McGill Undergraduate Mathematics Magazine

the δelta Ɛpsilon

Second Issue

Banach and Tarski paradoxically trying to feed themselves with math.

SUMS

McGill

McGill Undergraduate Mathematics Magazine

the δelta Ɛpsilon

WE WANT YOUR CONTRIBUTION!

The Delta-Epsilon is a math journal made for undergrads, by anyone with an interest in mathematics. We are looking for summer research papers, your takes on course material, book reviews, course reviews, puzzles, or any other piece of interesting mathematical miscellany.

You like the Delta-Epsilon? Join the team now!

You can e-mail us at

thedeltaepsilon@gmail.com

or visit us online at

http://sums.math.mcgill.ca/deltaepsilon

Contents

Letter From The Editors

Monday, November 26th, 2007

You have opened the second issue of the Delta-Epsilon. A slow, pleasant feeling of warmth is rising within you, as you glimpse the many hours of unbounded delight that will follow. However, reader, take note:

The Delta-Epsilon needs your help.

Its entire staff is leaving for graduate school next year, excepting Mr. Filip. We need replacements, or there will be no more issues. So send us an email or approach us in Burnside corridors if you're interested – it's a barrel of fun and an excellent way to contribute to departmental wellbeing.

This year's issue is much more research oriented than the last one. If you're unhappy with this, let us know. We make this journal for you, so we want to know you enjoy reading it, and believe it or not, we can adjust!

On a different note, keep sending your articles in – this year's issue has benefited from a flood of excellent contributions. And if you have any comments, suggestions or outrages to communicate, we do check our email from time to time.

The editors of the Delta-Epsilon
thedeltaepsilon@gmail.com
http://sums.math.mcgill.ca/delta-epsilon/

Letter From SUMS

Monday, November 26th, 2007

The Society of Undergraduate Mathematics Students (SUMS) would like to congratulate the δelta-ϵpsilon on the publication of its second issue. The δelta-ϵpsilon is a great achievement: undergraduates felt there was a need to showcase the incredible undergraduate research being performed by mathematics students at McGill, and so they created a journal to do just that.

SUMS is proud to support such a worthy cause, and we wish the δelta-ϵpsilon, and the undergraduate mathematics researchers at McGill, many more successful years. *Congratulations!*

Sincerely,

Nicholas Smith
SUMS President (for the SUMS Council)
http://sums.math.mcgill.ca/

Interview with Professor Eyal Goren

Michael McBreen

The Delta-Epsilon interviewed Professor Eyal Goren from the Department of Mathematics and Statistics early this spring, asking about his research as an arithmetic geometer, but also about what made him choose his path. This is what he had to say.

What research are you currently working on?

In the large, I'm an arithmetic geometer. My research combines number theory, which in essence studies integers and their various generalizations like algebraic integers (numbers that satisfy monic polynomial equations with integer coefficients), with algebraic geometry, which studies manifolds or *varieties* defined as the solutions to systems of polynomial equations in several variables.

Variety

Given a set of polynomial equations $P_i(x_1, x_2, x_3, \ldots, x_n) = 0$ in a field $\mathbb{K}$, the associated variety is the set of points $(x_1, x_2, \ldots, x_n) \in \mathbb{K}^n$ which satisfy the equations. Varieties generally have the structure of a manifold away from a smaller singular locus, where they may have jagged edges or self-intersections.

In arithmetic geometry, you might take a polynomial equation with integer coefficients, reduce the coefficients mod p and ask for solutions in characteristic p. This brings another dimension to the picture. The same equation gives a variety in characteristic p for every p, a complex variety when you look at complex solutions, and so on. Arithmetic geometry, in the large, makes use of this extra dimension to study problems that arise in number theory.

Can you picture varieties in characteristic p?

Yes, but it's not clear what the picture means. It gives an intuition or a way of organizing your thoughts, rather than any solid meaning. But still, if you have an equation for a line, you like to draw a line on the board because things behave rather similarly to usual geometry in many respects. Somehow, this whole geometric intuition makes arithmetic geometry work, and I enjoy very much translating questions about numbers into geometric questions.

My own research is deeply concerned with constructing units. Pick a polynomial, say a monic polynomial with integer coefficients, so that a root would be an algebraic integer. If its free coefficient is 1 or -1, the root would in fact be a unit. In other words, one can construct a ring whose elements are algebraic integers and that element would be invertible in that ring. It's not hard to see, because the free coefficient is the product of the roots of the polynomial, which are all algebraic integers. If it's 1, then the roots are invertible. That's not so hard, but the game is really played differently: you first pick the extension of $\mathbb{Q}$ in which you want the number to lie in, for instance you could pick $\mathbb{Q}[\sqrt{2}]$, and in this field there's a ring of integers, some of which are units. The question is how to find these units, and that turns out to be one of the major problems of this type of algebraic number theory. Some of the strongest tools we have come from arithmetic geometry.

Figure 1: Eyal Goren

The main idea in my research is that you take some variety over the complex numbers, which is defined by integer polynomials so that you can look at its reduction mod p for various primes p. When you do this you get all these (sometimes singular) varieties, and we think of them as a single geometric object. We take a function f on the variety which makes sense arithmetically, perhaps also defined with $\mathbb{Z}$ coefficients, and we evaluate it at some point x. Suppose $f(x) = \frac{a}{b}$ where a and b are algebraic integers, and $gcd(a, b) = 1$ (one can make this precise). We want to know if $f(x)$ is a unit.

There's an analogue of prime numbers called prime ideals, and for our ring of integers one can make sense of the statement "p appears in the denominator of $f(x)$". You don't have unique factorization into primes, but if you think of $f(x)$ as generating a principal ideal, there is a unique factorization into prime ideals. Hence, the ideal generated by $f(x)$ in the ring of algebraic integers of this field can be decomposed as a product of powers of prime ideals.

If an ideal p appears in the denominator of $f(x)$, there's another way to think about it, which is to say that $f(x) = \infty$ mod p. If it's in the numerator, then $f(x) = 0$ mod p. So the picture is that we have a variety given by

polynomial equations, a point x on this variety (that is, a solution to those equations), and a function f on the variety. And the function, the point and the variety can all be reduced mod p for most p. We define the 0-divisor of f roughly as the set of points where $f(x) = 0$, and the ∞-divisor as the points where $f(x) = \infty$, and the statement that p is in the denominator of $f(x)$ translates as saying that x mod p belongs to the ∞-divisor of f. All this works for general varieties and functions f, but the main idea here is to use a variety and a function where everything has an extra meaning. The varieties we are using are parameter spaces or "moduli spaces". The simplest example is the variety that classifies elliptic curves up to isomorphism, i.e. whose points correspond to isomorphism classes of elliptic curves.

Elliptic curves

An elliptic curve is the set of solutions (x, y) to the equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

where we require that the resulting curve be non singular: if we write the equation as $f(x, y) = 0$, there is no point (x_0, y_0) such that $\frac{d}{dx}f(x, y)|_{x_0,y_0}$ and $\frac{d}{dy}f(x, y)|_{x_0,y_0}$ are both zero. In $\mathbb{C}^2$, elliptic curves are tori.

You can look at a function on this space which vanishes exactly on the elliptic curves with some property - let's call it "spin". This is just a name, and there is no connection to spin in physics. Take a point x without this property, i.e. such that $f(x) \neq 0$. To say that p appears in $f(x)$ is to say that mod p, the elliptic curve has spin. That's now a question about elliptic curves. It sounds like a lot of machinery to solve a simple-minded question, but the truth is that we can almost never solve the simple question directly. By cleverly choosing spaces, functions and points, you can translate the original question of whether $f(x)$ is a unit into the question of whether the elliptic curve acquires a special property mod p, and these are questions we can say much more about.

A lot of my research is concerned with the behavior of elliptic curves and their higher dimensional generalizations when you reduce them mod p: seeing how various properties of these varieties behave when you reduce them. It turns out that those properties, when defined properly, are essentially geometric.

One interesting feature, and for many of us it's a very frustrating feature, is that we almost never have equations for our moduli spaces. We know that they exist and are algebraic varieties, but we can't really describe them with equations. Everything goes through these translations: every point in the space has an extra meaning.

We talked about spaces which classify elliptic curves, and among these there are those that have some special property. Similarly, you can have abelian varieties, which are higher dimensional analogues of elliptic curves, and they might have some interesting properties - part of the game is to define such good properties. And then you look at all abelian varieties with this property and you try to prove, for example, that they form a subvariety of the moduli space classifying all abelian varieties of this type. All this using pure thought, so to speak, never using equations: it would be horrible with equations, perhaps impossible.

The subvarieties that arise this way include those called Shimura varieties, and are very important in number theory and algebraic geometry. Therefore, one wants to study them further. For example, one may try to study the local nature of some property, in the sense that you have an abelian variety and you ask if I slightly deform it, will the deformation preserve this property. Usually the answer is no. You then ask yourself under what conditions will the deformation preserve the property. And if you can find such conditions, that tells you about the local structure of the varieties you are defining.

These are very roundabout techniques, and definitely when one is first exposed to all this one should be very suspicious as to whether the whole thing is worth the effort, but I think the answer is yes, we are proving stuff, and the spaces we obtain are important for physics and other applications.

Prof. Goren now tells us about a different aspect of his research, certain cryptographic tools called hash functions.

This part of my work is in collaboration with Kristin Lauter from Microsoft research. Amazingly, it relates to the units we discussed earlier, but it would be too long to explain the connection here.

Hash functions are critical tools in certain security protocols used over the internet, for example in the digital signature protocols used in online transactions. From a mathematical viewpoint, a hash function is a rather simple object; it takes a bit string of arbitrary length and produces a string of fixed length, say 32 bits. A very primitive example would be the following: if I wanted to check that nobody is tampering with my hard drive while I'm away from my office, I could use a function which takes the whole content of my hard drive and returns a 32 digit number. I would run this function before leaving the office and run it again when I come back, and if I get the same 32 digit number, it's very likely no-one messed with the hard drive while I was gone.

For this to work, you need functions which are very sensitive to small changes. If someone hacks into his bank account and adds a single digit to his savings, you want to detect that. You also want it to be very hard for a person to know which changes to make to modify the value of the hash function in any given way.

Many of the currently used hash functions are quite simple: you have this big potato, and you chop it up and fry it and so forth until it's unrecognizable. You try to do something very complicated and aggressive to the data and do it many times, and you hope it ends up properly

hashed. But it turns out these protocols aren't as secure as people thought, so there's a big search for good hash functions.

We propose to take this huge string of bits and use it as directions to walk on a graph. Imagine a graph where eight edges go out from every vertex. The first part of the string tells you which vertex to start on. Then you chop the rest of the string into 3 bit pieces, and each 3-bit sequence, which can encode 8 different possibilities, tells you which edge to move along next. Each vertex has a label, and the label of the end vertex is the output of the hash function.

You don't want someone to be able to modify a few bits of the input but still reach the same end vertex. You want a graph where if you change even a single bit of input, you could end up somewhere completely different. There are also other cryptographic requirements that put additional conditions on the process. So how do you construct a graph of this sort? It turns out that the best constructions that we know of come from number theory, and involve modular forms and elliptic curves, or abelian varieties, in characteristic p.

By using number theory to construct the graphs, we're able to translate the security requirements of the hash functions into questions about elliptic curves, for example. In other words, we translate the problem of cracking the hash function into a problem about elliptic curves. You can't really prove that a Hash function is secure: you can only show that the obvious attacks fail. In some sense, you play the devil's advocate by inventing methods of attack and showing that they fail. Since people have been thinking about the relevant number theoretic problems for more than a century now, we feel confident that the translated problem is truly hard, and this somehow justifies our faith in the security of the hash function.

You can also use these graphs to create pseudo-random number generators, or to sample data sets. One of the best ways to sample of data is to do it randomly, but there's a price to pay for randomness, in running time or otherwise: it's really very difficult to generate random numbers. So finding ways to mimic randomness is a big deal in computer science.

How did you get into mathematics?

It's really a series of events. When I was about to turn six, I became very sick and I had to spend the whole summer in bed. My dad got me books in math, because he was afraid I wouldn't be able to catch up in class, but I ended up studying all this math which was quite hard for a six year-old. I really enjoyed just staying in bed and doing those exercises.

Later, when I was ten, we learnt in school about divisibility properties: when is a number divisible by 3, by 5, by 11 and so on. I was totally obsessed with finding a rule for seven, which is very tricky. I can't really explain it, but the problem appealed to me. I spent that summer at my Grandparent's place in Haifa, which is a harbor city in the north of Israel, in this little villa. I remember spending the hours before falling asleep thinking about this problem, and eventually solving it, and that was a tremendous reward for me.

So what is the rule for divisibility by 7?

Actually, I've given it as an assignment in Algebra 1, so it's on my course webpage

When I was in High school, I remember buying those books of the Schaum series – because they were the cheapest, so I could afford them. I think I liked then to calculate a lot, and see what the answer is. I also had the good fortune to be in contact with a professor from the Hebrew University, who gave me real math books and helped me read them and understand them, so I was exposed to higher mathematics, but I was never sure I wanted to do math. My main interest in high school was biology. I was very interested in immunology, the workings of the immune system, which is a truly fascinating subject. Music was another possible career choice.

Basically, I got to studying math in university by elimination. Biology became a total mess at that time, because all the current theories about the immune system were discovered to be false, and there were too many theories coming out, so I thought "let these people figure out first what they want to say, and then we'll see." As for music, I played the piano for many years, and I realized the life of a performer was too difficult: very stressful, very competitive, and very few get to a position where they can actually play for an audience.

So I started university in math and physics, but after one semester I got very irritated with the physicists, because nothing was defined: what is mass? How do you know that those are the forces working on a ladder? And so forth. So I decided to transfer into math. Even then, I did a lot of other things during my studies. For instance, I had a break of two years working in agriculture. I did do some math during that period - I actually corresponded by letter with Ehud de Shalit from Hebrew University, who eventually became my thesis supervisor.

What got me back to mathematics, and what keeps me wanting to do research, is in some sense the same thing that had piqued my interest when I was five, or ten: I really wanted to understand why things were true, what is the structure there - *how do you tell*, what's the pattern. It comes from a place which is unmotivated by more sophisticated considerations.

How does it feel, when you suddenly go from a post-doc to a professor?

Actually, I think graduate school is the most exciting part. You have a lot of responsibilities as a professor, so in some sense graduate school is the time when you're the most carefree, and it's where your mind really expands. I re-

member, during my studies, encountering concepts that I had never thought of before, and it was very unsettling.

For instance, the first time I heard of the Banach-Tarski paradox. You take a solid unit ball in $\mathbb{R}^3$, you divide it into finitely many parts, and then using only rigid transformations - no bending or anything like that - you can reassemble those parts into two solid unit balls. For me this was very unsettling, I remember being deeply troubled by this phenomenon for weeks, because it shattered my intuition and the way I understood the relevance of mathematics.

Another example is the notion of different cardinalities of infinite sets. When I was fourteen, I was babysitting my neighbor's child. This neighbor had a degree in math, and one day he proved to me that there were the same number of integers as squares, and that was a revolution for me. I remember trying for weeks to check whether certain sets are the same size or not, and you could feel the mind physically rewiring itself to digest these new phenomena. As you progress, you get more professional and there are less and less instances like that, where you feel that a whole new universe is being opened to you. Graduate school is a great time for that. There are other discoveries later on, you discover your own theorems, but they're very rarely on the same fundamental level.

Any advice for undergraduates?

This is not just for undergraduates, this is universal advice. When people undertake a long term project, for instance getting a degree or a PhD, very often they tend to forget halfway why they've started it. They know they have to finish it, but they can't reconnect to the things that prompted them to undertake this project to begin with. You see this phenomenon when classes are cancelled and everyone is happy, which is pretty ironic, because you came to university to learn this stuff, you've made that choice. So it's good to try and reconnect to the reasons that got you to university, or made you go into the Ph.D. program and so on. This applies especially in math - I think people choose math for the same kind of reasons that I was describing earlier, and I think it's very important to reconnect to this desire to know, to learn more about these patterns, and to appreciate their beauty.

Jokes

A mathematician, a physicist, and an engineer are all given identical rubber balls and told to find the volume. They are given anything they want to measure it, and have all the time they need. The mathematician pulls out a measuring tape and records the circumference. He then divides by two times pi to get the radius, cubes that, multiplies by pi again, and then multiplies by four-thirds and thereby calculates the volume. The physicist gets a bucket of water, places 1.00000 gallons of water in the bucket, drops in the ball, and measures the displacement to six significant figures. And the engineer? He writes down the serial number of the ball, and looks it up. □

An engineer, a physist and a mathematician are sent to a desolated jail. The engineer is sent in first, alone, with nothing but a can containing his only potential source of food. After a few minutes, he looks into his pocket, finds some trash and make a can opener out of it. He then eats almost all the food and use the rest and the can to make a small bomb to break the wall of his cell. He escapes, retires at the age of 55 to go travel around the world on a boat, and lives happily ever after.
The physists is then sent in and, again, his only hope resides in his can of food. After a few hours, he takes a rock and writes on the ground. He then computes the exact angle and force at which he needs to throw his can to destroy both the can and the wall. He eats the food, escapes and starts a new ground-breaking theory in which everything is represented as tiny 24 dimensional cans.
The mathematician is then sent in the prison. After a week, the guardians come. They find him dead in his own blood, lying face down in the corner of the cell. In the exact middle, the can lay perfectly still, and closed. Around it, an elegant drawing accompagnies a beautiful proof of the sphere packing problem – written in blood. Next to the mathematician, the guardians find, and clean away in their ignorance, some text: "Theorem. If I don't open the can, I will die. Proof. Suppose not." □

On Primes in Arithmetic Progressions

Vincent Quenneville-Bélair

Dirichlet's theorem is proved using the Riemann Zeta Function and similar Dirichlet series based on characters. Indeed, the similar series satisfy an identity that will be used to derive an asymptote for the sum of the reciprocals of primes in some congruence class, under the condition that they are bounded away from zero.

Theorem (Dirichlet's Theorem)**.** *An arithmetic progression* $\{a + nm\}_{n=0}^{\infty}$ *where* $a, n, m \in \mathbb{Z}$ *contains infinitely many primes when* $\gcd(a, m) = 1$.

Introduction

Dirichlet proved the infinitude of primes in arithmetic progressions in 1837 using ideas from Euler's proof about the infinite number of primes – a task so great that it is claimed to be *the crowning achievement of the XIXth century*[1] in number theory. The theorem equivalently states that there exist infinitely many primes congruent to a mod m when a and m are coprime. The proof starts by noting that $\zeta(s)$, the Riemann Zeta Function, has a simple pole at $s = 1$ and continues with the definition of similar series with the key property that they are bounded away from zero. That is a major point: showing that these series are not zero as s approaches 1 from the right. A survey of characters, examples of periodic functions from the integers to the multiplicative group of the complex numbers, will be necessary in defining these Dirichlet L Series. Euler's genius comes into play when finding a factorization of all these series and deriving from them an asymptotic behaviour for sums of primes. It is worth noting before beginning the following simple result: an arithmetic progression contains at most one prime when $\gcd(a, m) > 1$.

Riemann Zeta Function

The Riemann Zeta Function has very special properties linked to extremely deep topics in mathematics – such as the Riemann Hypothesis, which claims that all the non-trivial zeros are on the $\Re(s) = 1/2$ line. The journey into the proofs of Dirichlet's Theorem starts with the study of the Riemann Zeta Function and similar series. It is important to know that the next definition only makes sense for $\Re(s) > 1$.

Definition 1. The Riemann Zeta Function, denoted $\zeta(s)$, is defined to be the following series for $\Re(s) > 1$: [2]

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

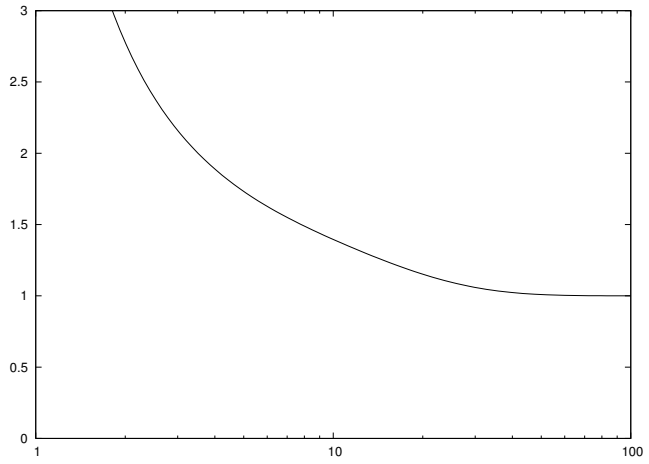


Figure 1: $\zeta(s)$ up to roughly 30 terms in the series on $1 < s < 100$.

Interestingly, $\zeta(s)$ can be extended uniquely to an analytic function with a simple pole at $s = 1$. The uniqueness follows from an important result in complex analysis stating that if two functions f and g are analytic on a domain D and that there exists a sequence $\{z_n\}$ of points in D accumulating at ω in D such that $f(z_n) = g(z_n)$ then $f = g$ everywhere in D [3].

Property 1. $\zeta(s)$ is absolutely convergent for $\Re(s) > 1$.

This will follow from the proof of property 2, but the adaptation of the proof is left to the reader.

A function has a pole of order N at an isolated point z_0 if it diverges to infinity if the limit of $(z - z_0)^N f(z)$ as z approaches the singular point z_0 is neither zero nor ∞.

Property 2. $\zeta(s)$ has a simple pole at $s = 1$

Proof. Using the integral test for series,

$$\int_1^{x+1} \frac{dt}{t^s} \leq \sum_{n=1}^{x} \frac{1}{n^s} \leq 1 + \int_1^x \frac{dt}{t^s}$$

$$\frac{-1}{s-1}\frac{1}{(x+1)^{s-1}} \leq \sum_{n=1}^{x} \frac{1}{n^s} - \frac{1}{s-1} \leq 1 - \frac{1}{s-1}\frac{1}{x^{s-1}}.$$

Fixing $s > 1$ and letting x going to infinity,

$$0 \leq \zeta(s) - \frac{1}{s-1} \leq 1.$$

The result follows from the fact that $\lim_{s\to 1^+} \zeta(s) = \infty$ but $\lim_{s\to 1^+} \zeta(s)(s-1) = 1$. □

[1] Citation from a number theory lecture on March 2nd 2007 by Professor Henri Darmon.

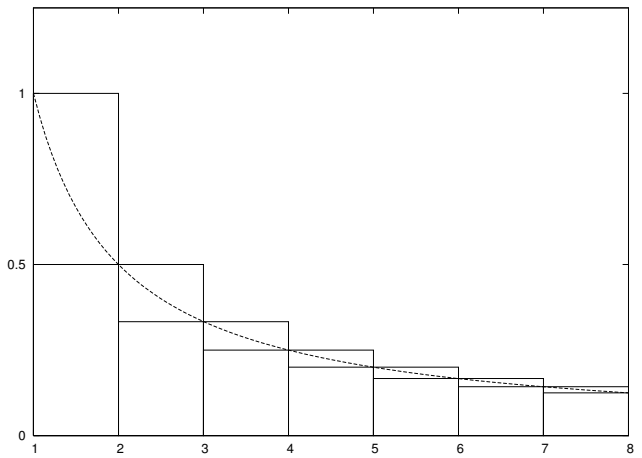


Figure 2: Approximating $1/x$ by sums.

Characters

Definition 2. A character χ modulo m (where $m \geq 1$) is a homomorphism from $(\mathbb{Z}/m\mathbb{Z})^*$ to $\mathbb{C}^*$. It is extended to all $\mathbb{Z}$ by setting $\chi(a) = 0$ when a is not coprime with m. [2]

From now on, $\alpha(n)$ is the order of n in $(\mathbb{Z}/m\mathbb{Z})^*$, i.e. the smallest integer strictly greater than 1 with $n^{\alpha(n)} \equiv 1 \pmod{m}$. Note that $\alpha(n)|\phi(m)$ by Euler's theorem, where $\phi(m)$ is the Euler-Phi function ($\phi(p) = p - 1$ when p is a prime).

These characters $\chi(n)$ modulo m can be viewed as multiplicative functions in the strict sense on $\mathbb{Z}$, which have period m, have image in $\mathbb{C}^*$, are zero when n is not coprime to m and take values among the $\alpha(n)^{th}$-roots of unity. Furthermore, it is an important fact that there are $\phi(m)$ distinct characters for a fixed modulus m. Indeed, the group formed by the characters is abstractly isomorphic to $(\mathbb{Z}/m\mathbb{Z})^*$ which has $\phi(m)$ elements. [1, 2] From now on, the modulus m is fixed.

Lemma 1. *If $\chi = 1$, $\sum_a \chi(a) = \phi(m)$, or otherwise zero. [2]*

Proof. For $\chi = 1$, the sum counts the number of elements in $(\mathbb{Z}/m\mathbb{Z})^*$ and is hence $\phi(m)$. If $\chi \neq 1$, consider multiplying by $\chi(b) \neq 1$, knowing that $b(\mathbb{Z}/m\mathbb{Z})^* = (\mathbb{Z}/m\mathbb{Z})^*$:

$$\chi(b)\sum_a \chi(a) = \sum_a \chi(ab) = \sum_a \chi(a).$$

Thus, $(\chi(b)-1)\sum_a \chi(a) = 0$ which implies that $\sum_a \chi(a) = 0$. □

The previous proof can be adapted to obtain the next lemma.

Lemma 2. *If $a \equiv 1 \pmod{m}$, $\sum_\chi \chi(a)$ is either $\phi(m)$ or 0. [2]*

Proof. If $a \equiv 1 \pmod{m}$, $\chi(a) = 1$ and thus the sum is counting the number of characters. If $a \not\equiv 1 \pmod{m}$, one takes $\chi'(a) \neq 1$ and

$$\sum_\chi \chi(a) = \sum_\chi \chi(a)\chi'(a) = \chi'(a)\sum_\chi \chi(a)$$

which implies as in lemma 1 that the sum is zero. The previous equation uses the fact that $\chi(a)$ takes values in the $\alpha(a)^{th}$-roots of unity and that multiplying by one of them simply permutes them: their sum is thus the same. □

Dirichlet L Functions

With characters in hand, one can define the Dirichlet L functions.

Definition 3. A Dirichlet L function [2] is a series

$$L(s,\chi) = \sum_{n=1}^{\infty} \chi(n)/n^s.$$

where $\chi(n)$ is a character modulo m.

Now, these series are absolutely convergent for $\Re(s) > 1$ from property 1 (their absolute value is bounded by the absolute value of $\zeta(s)$). Furthermore, one can use the next property to show that the series converges uniformly for $\Re(s) \geq \delta > 0$ if $\chi \neq 1$, which implies that $L(s,\chi)$ is continuous for $\Re(s) > 0$. The proof will be omitted: however, an interested reader can look for Abel's sum in references such as [2].

Property 3. For $\chi \neq 1$, $L(s,\chi)$ converges (maybe not absolutely) for $\Re(s) > 0$. [2]

Property 4. The Dirichlet L function can be factored in a manner similar to Euler's identity:

$$L(s,\chi) = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}.$$

Proof. With p denoting a prime from here onwards and p_r the greatest prime smaller than x,

$$\begin{aligned}\prod_{p<x}\left(1-\frac{\chi(p)}{p^s}\right)^{-1} &= \prod_{p<x}\left(1+\frac{\chi(p)}{p^s}+\frac{\chi(p)^2}{p^{2s}}+\ldots\right)\\ &= \lim_{e\to\infty}\sum_{0\leq e_1,e_2,\ldots,e_r\leq e}\frac{\chi(p_1^{e_1}p_2^{e_2}\ldots p_r^{e_r})}{(p_1^{e_1}p_2^{e_2}\ldots p_r^{e_r})^s}\\ &= \sum_{\text{n x-smooth}}\frac{\chi(n)}{n^s}\end{aligned}$$

where a number $n \in \mathbb{N}$ is x-smooth if all its prime factors are strictly smaller than x. The result is obtained by letting x going to infinity. □

Using exactly the same method, one obtains the factorization of $\zeta(s)$. Remark also that $L(s,1)$ is similar to $\zeta(s)$, in fact,

$$L(s,1) = \prod_{p\nmid m}\left(1-\frac{1}{p^s}\right)^{-1} = \zeta(s)\prod_{p|m}\left(1-\frac{1}{p^s}\right). \quad (1)$$

Asymptotes concerning Primes

By taking the logarithm of both sides in property 4, we get

$$\begin{aligned}\log L(s,\chi) &= -\log\prod_p\left(1-\frac{\chi(p)}{p^s}\right)\\ &= \sum_p\sum_{n\in\mathbb{Z}^+}\frac{\chi(p^n)}{np^{ns}}. \quad (2)\end{aligned}$$

Again, replacing χ by 1 yields the result for $\zeta(s)$:

$$\begin{aligned}\log\zeta(s) &= \sum_p p^{-s} + \sum_p\sum_{n=2}^{\infty}\frac{1}{np^{ns}}\\ &= \sum_p p^{-s} + O(1), \quad (3)\end{aligned}$$

with the left side diverging as s approaches 1. It follows that there are infinitely many primes and that the sum of their reciprocal diverges. Turning to equation 2 and dividing by $\chi(a)\neq 0$, summing over all characters and applying lemma 2,

$$\begin{aligned}&\sum_\chi \chi(a)^{-1}\log L(s,\chi)\\ &= \sum_p\sum_{\substack{n\in\mathbb{Z}^+\\ p^n\equiv a(m)}} n^{-1}p^{-ns}\phi(m)\\ &= \phi(m)\sum_{p\equiv a(m)} p^{-s} + \phi(m)\sum_p\sum_{\substack{n=2\\ p^n\equiv a(m)}}^{\infty} n^{-1}p^{-ns}\\ &= \phi(m)\sum_{p\equiv a(m)} p^{-s} + O(1). \quad (4)\end{aligned}$$

If we can show that for $\chi\neq 1$, $L(s,\chi)$ is non-zero as $s\to 1^+$, then the proof would be done: it would follow that the left hand side diverges, implying that there are infinitely many primes congruent to $a \pmod m$. [1]

Away from Zero

Naturally, it remains to prove that the Dirichlet L series are not zero as $s\to 1^+$.

Lemma 3. *Let $n\not\equiv 0 \pmod m$, $g(n)=\phi(m)/\alpha(n)$ and $T=p^{-s}$. Then*

$$\prod_\chi(1-\chi(n)T) = (1-T^{\alpha(n)})^{g(n)},$$

where $\alpha(n)$ is the order of n in $(\mathbb{Z}/m\mathbb{Z})^*$. [2]

Proof. First, consider W, the set of $\alpha(n)^{th}$-roots of unity. One has

$$\begin{aligned}\prod_{w\in W}(1-wT) &= 1-\sum_{w\in W}wT + \sum_{w_i\neq w_j\in W}w_iw_jT^2\\ &\quad -\ldots+(-1)^nT^{\alpha(n)}\\ &= 1-T^{\alpha(n)}.\end{aligned}$$

Recall that the sum of all the roots of unity yields zero. The result follows since there are $g(n)$ character modulo m such that $\chi(n)=w$. □

Having lemma 3 in hand (with $T=p^{-s}$), it is now possible to factorize $\zeta_m(s)$ using equation 4 for $\Re(s)>1$ where convergence is clear [2],

$$\begin{aligned}\zeta_m(s) &= \prod_\chi L(s,\chi) = \prod_\chi\prod_p\left(1-\frac{\chi(p)}{p^s}\right)^{-1}\\ &= \prod_{p\nmid m}(1-p^{-s\alpha(p)})^{-g(p)}. \quad (5)\end{aligned}$$

Property 5. $L(1,\chi)\neq 0$ when $\chi\neq 1$. [2]

Proof. Suppose $L(1,\chi)=0$ for some χ. It would imply that $\zeta_m(s)$ is convergent for $\Re(s)>0$ since, as mentioned before, the simple pole of $L(s,1)$ at $s=1$ would be removed by the zero of $L(s,\chi)$. Now, the right hand side of equation 5 is

$$\begin{aligned}\zeta_m(s) &= \prod_{p\nmid m}(1+p^{-\alpha(p)s}+p^{-2\alpha(p)s}+\ldots)^{g(p)}\\ &\geq \prod_{p\nmid m}(1+p^{-\phi(m)s}+p^{-2\phi(m)s}+\ldots)\\ &\geq \sum_{p\nmid m}\frac{1}{p^{\phi(m)s}}\end{aligned}$$

but this last series goes to infinity as $s\to 1/\phi(m)$ since the sum of the reciprocal of the primes diverges. This contradicts the convergence of $\zeta_m(s)$ and, hence, all $L(1,\chi)$ are non-zero. □

Conclusion

Now, the left side of equations 4 diverges since, for $\chi\neq 1$, $L(s,\chi)$ is bounded away from zero and so $\log(L(s,\chi))$ is bounded. Hence, the right side must diverge, because $L(s,\chi_1)$ diverges, and thus so does the sum of the reciprocal of the primes congruent to a modulo m. Done!

The author thanks professor Henri Darmon for his guidance during the writing of this article.

References

[1] Harold Davenport. *Multiplicative Number Theory.* Springer, third edition, 2000.

[2] Jean-Pierre Serre. *A Course in Arithmetic.* Springer, 1973.

[3] David Wunsch, A. *Complex Variables with Applications.* Pearson Education, third edition, 2005.

Object Detection Using Feature Selection and a Classifier Cascade

Rishi Rajalingham

To construct an object-detector, one must provide a classifier model with training data from which it will "learn" what distinguishes the object class. Commonly, the training data in question is a large set of labeled images of class and non-class objects, and the distinguishing features are edges extracted from these images. The set of all edges in the training data, or feature space, is large, and hence training a classifier is time-consuming. Furthermore, once trained, classification may be crude or slow in conventional methods. This paper will briefly describe the proposition to reduce the complete set of features, using François Fleuret's conditional mutual information maximization (CMIM) algorithm, to a few most informative features; hence, the training time is considerably reduced. Moreover, the classifier is trained as a cascade of weak classifiers, rejecting non-class images quickly, as per Viola and Jones [1] , thus reducing classification time as well.

Conditional Mutual Information Maximization

In [2], Fleuret introduces the probabilistic notion of mutual information, specifically conditional mutual information maximization (CMIM), to the field of object detection. The purpose of the CMIM algorithm is ultimately to select, from a given feature space, the small number of features that are deemed most informative and hence best represent a class of objects. It follows intuitively that classifying using this reduced set of features is far more efficient than using the complete set.

To understand CMIM, first recall from information theory the concept of entropy (H) of a variable: $H(U)$ represents the uncertainty of U. Moreover, the conditional entropy of a variable, $H(U|V)$, represents the uncertainty of U when V is known. (Thus it is trivial that, if U is a function of V alone, then $H(U|V) = 0$, and if U and V are independent, then $H(U|V) = H(U)$.) Using this, it is now possible to express conditional mutual information (I) as:

$$I(U;V|W) = H(U|W) - H(U|W,V).$$

The value $I(U;V|W)$ gives an idea of the information shared between U and V, given W. Within the object detection application, U must be understood as a class of objects, V as a feature about to be selected (or rejected), and W as the set of features already selected.

Thus, if the new feature V carries no or little new information on the class, given some pre-selected features, then both conditional entropy terms are equal, or similar, and the conditional mutual information is zero, or small. Likewise, if the new feature V brings forth much new information of the class U, which is what we seek, the conditional mutual information will be large. The reason for the term 'maximization' now becomes clear.

To further tie this to the present application, let $X_1, \ldots, X_N$ be the N features in the complete set. For common object detection problems, the variable N is in the order of tens of thousands. Likewise, let $X_{V(1)}, \ldots, X_{V(K)}$ be the K features in the reduced set, where K is in the order of tens. $\{X_{V(1)}, \ldots, X_{V(K)}\}$ can be obtained by iterating over the complete set: first selecting the most informative feature $X_{V(1)}$, and subsequently selecting, and adding to the reduced set, the feature $X_{V(i)}$ for which the conditional mutual information is largest. For the complete algorithm or implemented code, refer to [2].

The advantages of using this reduced set of features lie not only in efficiency in computation power and time, but also in theoretical performance. Indeed, by using fewer features per classifier, the phenomenon of "overfitting" is avoided. Overfitting occurs when too many parameters, or in this case, features, relative to the training data needlessly increases the complexity of the classifier model, possibly resulting in a very poor classifier.

Features and Filters

Due to its computational complexity, and consequently, large computation time, the CMIM algorithm necessitates the use of very crude features that take limited values. This is best done by using binary features, where a value of 1 indicates the presence of a specific edge, and 0, the absence. Features are obtained by running filters over images. The filters used in this approach are similar to the edge fragment detectors used in [3], and are called "crude edge detectors". Briefly, they return true (or binary 1) if the contrast between pixels across the supposed edge is greater than the contrast between pixels along the edge.

Running these filters of all eight orientations (see Figure 1), in neighborhoods of size varying between 1 and 7 pixels, at every pixel location of a training image of size 24×24, we obtain $20 \times 20 \times 8 \times 7 = 22400$ features in total.

Weak Classifier Models

The classifier models implemented are linear classifiers; given any image from which N input features are extracted, or alternatively given any feature vector $\vec{x} = (x_1, \ldots, x_N)$, the class is determined using:

$$f(\vec{x}) = sgn\left(\langle \vec{x}, \vec{\omega} \rangle + b\right),$$

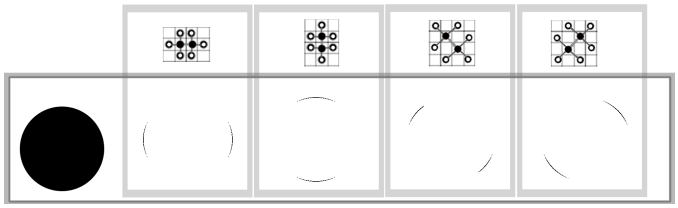


Figure 1: The crude edge detectors (top) return true if the contrast between the two pixels shown in solid dots is greater than the contrast of neighbouring pixels, shown with white circles. The detectors of four directions map the dark disk on the left to the corresponding edge maps (bottom).

for bias b and weight vector $\vec{\omega} = (\omega_1, \ldots, \omega_N)$ computed in the training phase, and standard inner product $\langle \vec{x}, \vec{\omega} \rangle$. The Signum (sgn) function returns the sign of its argument, and hence, the class of the image (positive or negative).

This concept may be understood geometrically by visualizing $\vec{x}$ as a point in N-space, while $\Pi(\vec{u}) = \vec{u} \cdot \vec{\omega} + b$ is the equation for an $(N-1)$-flat, or hyperplane, having normal vector $\vec{\omega}$ and constant term b. It should be clear that the hyperplane Π cuts the space in two, thus determining the class of any image $\vec{x}$ by its coordinates. What remains is to determine the particular Π for each classifier, or equivalently to determine its $\vec{\omega}$ and b. The following are methods for determining the weights and bias.

Perceptron

The classical Perceptron (see [4], [5]) provides an iterative method, the Perceptron learning algorithm, to compute the weight vector. The vector, $\vec{\omega}$, is initialized and iteratively corrected in the training process. If a training example is incorrectly classified, its feature vector is added or subtracted, depending on its true class, to the weight vector. This process is known to converge for linearly separable training sets.

Naive Bayesian Classifier

The naive Bayesian classifier classifies by comparing probabilities with a simple inequality. Let $\vec{x} = (x_1, x_2, \ldots, x_N)$ be the feature vector of an image, and Y (1 for positive, 0 for negative), its class label. Then

$$f(x) = \begin{cases} 1, & P(Y=1|X_1=x_1,\ldots,X_N=x_N) \\ & > P(Y=0|X_1=x_1,\ldots,X_N=x_N) \\ 0, & \text{else.} \end{cases};$$

or equivalently,

$$f(\vec{x}) = sgn\left\{\log \frac{P(Y=1|X_1=x_1,\ldots,X_N=x_N)}{P(Y=0|X_1=x_1,\ldots,X_N=x_N)}\right\}.$$

Now, recall from statistics that for events A, B, Bayes' Theorem states that

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}.$$

Hence, assuming the X_i's are conditionally independent (a naive assumption), and applying Bayes' Theorem, we have

$$\begin{aligned} & f(\vec{x}) \\ &= sgn\left\{\log \frac{\prod_{k=1}^N P(X_k=x_k|Y=1)}{\prod_{k=1}^N P(X_k=x_k|Y=0)} + \log\frac{P(Y=1)}{P(Y=0)}\right\} \\ &= sgn\left\{\sum_{k=1}^N \log \frac{P(X_k=x_k|Y=1)}{P(X_k=x_k|Y=0)} + \log\frac{P(Y=1)}{P(Y=0)}\right\} \end{aligned}$$

Thus, we have arrived at the linear form $f(\vec{x}) = sgn(\vec{x}\cdot\vec{\omega}+b)$, with

$$\omega_k = \log \frac{P(X_k=1|Y=1)P(X_k=0|Y=0)}{P(X_k=1|Y=0)P(X_k=0|Y=1)}.$$

The Bayesian weight computation required no iteration (and thus cannot fail to converge), while still outdoing the Perceptron in both speed and accuracy. Despite the naive assumption, experiments show that the Bayesian classifier results in lower error rates when dealing with 'real life' cases.

Bias

Assuming that the training set is well representative of the classes in question, one should see in the distribution of the weighted sum of inputs, two quasi-distinct regions representing the positive and negatives, respectively. It suffices then, for prescribed error rates, to estimate empirically a threshold value θ that best separates the classes. The bias b is then simply the negative of θ.

Strong Classifier: An Attentional Cascade

Given the classifier models, we may now string together several such weak classifiers to construct a strong one. The "attentional cascade" of Viola and Jones [1] does just this, and provides a way to achieve high detection rates while drastically decreasing detection time. The principal assumption is that, in any given image, the vast majority of objects will be negatives. In the specific case of face detection, this is known to be a well-founded assumption, as images seldom contain more than a dozen distinct faces, while the number of non-faces is generally on the order of tens of thousands.

Implementing an attentional cascade means constructing a decision tree. The cascading method discussed in this paper involves training each weak classifier sequentially using an increasing number of inputs until the error constraints are satisfied by that particular weak classifier. The constraints determining the characteristics of the strong classifier are on the true-positive and false-positive rates. The terms true-positive, false-positive, true-negative, and false-negative make reference to whether the classification was correct (true/correct or false/incorrect), and the classification return value (positive image or negative image). For example, in the case of a face detector, a misclassified face image is a false-negative.

Table 1: **Perceptron learning algorithm.**

- Given a training set of m labeled images: $D_m = \{(\vec{x_1}, y_1), (\vec{x_2}, y_2), \ldots, (\vec{x_m}, y_m)\}$, where the $\vec{x_i}$ is the feature vector for the i-th training image and y its corresponding label
- Given weights vector $\vec{\omega}$
- For each $(\vec{x_i}, y)$ pair in D_m
  - Initialize weights $\omega(j) \leftarrow 0$
  - Do until classifier converges or 5000 iterations:
    - Compute $\partial = \begin{cases} 1, & \text{if } \vec{x_i} \cdot \vec{w} \geq 0; \\ 0, & \text{else.} \end{cases}$
    - Update $\omega(j) \leftarrow \omega(j) + (\partial - y)x_i(j)$.

Table 2: **Cascade training algorithm.**

- Given a set of labeled training images
- Given an array of features for each image (output from Fleuret's CMIM algorithm [2])
- Initialize number of inputs to two ($n \leftarrow 2$)
- Do until all the features have been used
  - Train weak classifier i
  - Evaluate classification error on training set
  - If TP-rate > true positive constraint and FP-rate < false positive constraint
    - Move to next weak classifier ($i \leftarrow i + 1$)
    - Reset number of inputs ($n \leftarrow 2$)
  - Else
    - Increase number of inputs ($n \leftarrow n + 1$)

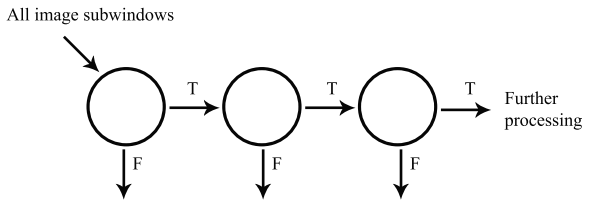


Figure 2: The attentional cascade structure. Early stages reject many negative sub-windows, thus increasing overall detection speed.

The advantage of the cascade structure lies in the time cost within the testing phase, where the object detector will be used on large images containing both positive and negative objects at a greatly skewed distribution. This requires performing a Raster scan of this test image, and looking at sub-windows of the image with the strong classifier. Since, as discussed above, most sub-windows are of negative images, classification will occur within the first few stages of the cascade for the majority of sub-windows, and hence the overall detection time is cut short. Naturally, a sub-window is classified as a positive only once it has reached the end of the cascade.

Results

Experiments done on this object detector, where a reduced set of features is used to train a cascade of classifiers, have shown that it compares and competes with more sophisticated, time-expensive models. Indeed, our classifier took minutes to train, compared to the days it took Viola and Jones, and resulted in smaller error rates than a single stage decision classifier, such as the one used by Fleuret.

Acknowledgments

The above article was a brief account of an NSERC project completed by ECSE undergraduate students Oliver Bates, Rishi Rajalingham, Meera Nair and Julien Cassis. This project could not have been completed without the guidance and support of Harkirat S. Sahambi and Dr. Martin Levine. Many thanks to Ioan Filip for his help with the formatting of this article.

References

[1] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, pp. 137-154, 2004.

[2] F. Fleuret, "Fast binary feature selection with conditional mutual information," *Journal of Machine Learning Research*, vol. 5, pp. 1531-1555, 2004.

[3] F. Fleuret and D. Geman, "Coarse-to-fine Face Detection," *International Journal of Computer Vision (IJCV)*, vol. 41, pp. 85-107, 2001.

[4] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," *Psychological Review*, vol. 65, pp. 386-408, 1958.

[5] A. B. J. Novikoff, "On Convergence Proofs on Perceptrons," presented at Symposium on the Mathematical Theory of Automata, 1962.

Optimizing Efficiency of a Geothermal Air Conditioner

Alexandra Ortan and Vincent Quenneville-Bélair

The underlying principle of a geothermal air-conditioning is to extract heat from the soil by running water through a series of pipes in the ground. However, since the installation costs of such a heat pump is very high, its configuration must be designed in such a way as to minimize them. The relationship between power output and controlable parameters such as pipe radius and length is investigated to this end. A derivation of the temperature profile of the soil is done in order to take advantage of the greatest temperature difference. Two models are used for the water running through the pipes: plug flow and Poisseuille flow, which were then used to predict the length of pipe necessary.

Problem Description

A geothermal heating system takes advantage of the fact that the temperature of the soil fluctuates slower than that of air, and in fact is almost stable at a certain depth. A series of pipes is buried in the ground following different configurations, and water is circulated through them. Thus the water either heats up or cools down depending on the season. A heat exchanger installed in the house then uses this water to either heat or cool the house and the water re-enters the cycle.

The configuration of the pipes through the ground can be either vertical or horizontal. As shown in the figures opposite, the pipes can be streched out or coiled together. A variant of the system is to put run the pipes through a pond of water, for better conductivity. A detailed analysis of each of them however could reveal the main differences and thus allow for better choices of the most appropriate configuration.

The efficiency of such a system relies on how much heat can be exchanged with the soil. Generally speaking, the longer the pipe carrying water through the ground, the better the heat exchange. However, other factors, such as flow rate, pipe radius and geometry of the pipe are also to be considered in calculating the heat transfer occurring between water and soil.

Soil Temperature Profile

The premise of the geothermal heating system is that the soil remains at almost constant temperature at a certain depth. In order to take the best advantage of that, one needs to know exactly how the soil responds to the seasonal temperature changes in the air and calculate that depth.

The variation of the temperature in function of the depth in the soil can be set up as a partial differential equation [1]. Indeed, it can be assumed that $\Theta(x,t)$, the temperature in function of the depth and of the time, respects the heat diffusion equation:

$$\Theta_t = \alpha \Theta_{xx},$$

where α is the thermal diffusivity of the soil. The seasonal variation of the soil's surface temperature yields a periodic boundary condition:

$$\Theta(0,t) = T_A + \Delta T e^{i\sigma t},$$

where T_A is the average temperature throughout the year and σ^{-1} is proportional to a month. Now, a trial function to transform the PDE in an ODE can be used: $\Theta(x,t) = T_A + Ae^{i\sigma t}W(x)$. It then follows that

$$W''(x) = \frac{i\sigma}{\alpha} W(x)$$

with $W(0) = 1$ and $\lim_{x\to\infty} W(x) = 0$. Trying then $W(x) = e^{mx}$ gives that $m = \pm\sqrt{\frac{\sigma}{2\alpha}}(1+i)$. Since $W(x)$ decays with x increasing, the negative value of m must be used. Finally, the result lies in the real part of $\Theta(x,t)$:

$$\begin{aligned} T(x,t) &= \Re\{\Theta(x,t)\} \\ &= T_A + e^{-\sqrt{\frac{\sigma}{2\alpha}}x} \cos\left(-\sqrt{\frac{\sigma}{2\alpha}}x + \sigma t\right). \end{aligned} \quad (1)$$

Using appropriate values for the constants gives that the ground temperature is almost uniformly 13°C below 2.5 m and that there is a temperature inversion at roughly 1.5 m.

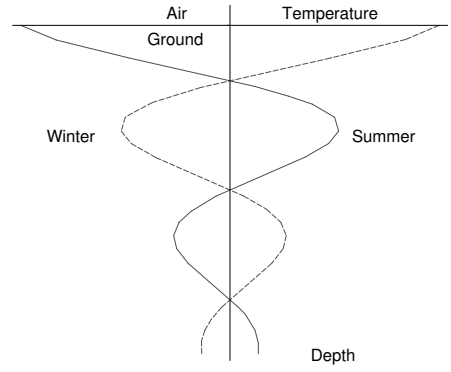


Figure 1: Temperature profile in °C of the ground for both summer and winter as a function of the depth.

Plug Flow

A typical annual household's energy consumption is about 75 MBTU or around 80 MJ. Assuming water enters the system at 3°C and heats up to the temperature of the soil,

that is 13°C, a volumetric flow of 30 L/min would be required to power the house. In a typical 1 cm pipes, that means a flow rate of roughly 2 m/s.

Under the assumption that the soil remains at a constant temperature T_a and that the pipe is straight, it is possible to find an equation for the power gained by a volume element. First, using the relationship between energy and heat capacity, one has

$$P = \rho_w \Delta V c_{pw} \frac{\Delta T}{\Delta t} \tag{2}$$

for a change of temperature ΔT in a time Δt of a volume element ΔV of water, and where ρ_w is the density of water and c_{pw} is its thermal capacity. Now, the heat input must be related to the heat ϕ_0 transferred from the soil to the pipe and the heat leaving the volume element by convection ϕ_1:

$$\Phi = \phi_0 - \phi_1,$$

where $\phi_0 = -hS(T - T_a)$ and $\phi_1 = \rho_w A u c_{pw} \Delta T$, with S being the surface area of a volume element, A the cross-sectional area of the pipe and u the velocity of water in the pipe and h the heat transfer coefficient.

Thus the governing equation for the temperature in the volume element is:

$$\begin{aligned} &\rho_w \Delta V c_{pw} \frac{\Delta T}{\Delta t} + \rho_w A u \Delta T c_{pw} \\ &= -hS(T - T_a). \end{aligned} \tag{3}$$

In order to avoid references to a unit system, the equation should be non-dimensionalized. Note that $\Delta V = A\Delta x = \pi R^2 \Delta x$ and that $S = 2\pi R \Delta x$, with R being the radius of the pipe. Taking the limit in which Δx and Δt go to 0,

$$\begin{aligned} &\rho_w R c_{pw} T_t(x,t) + \rho_w R u T_x(x,t) c_{pw} \\ &= -2h(T(x,t) - T_a) \end{aligned} \tag{4}$$

with boundary condition being $T(0,t) = T_1$ and the initial condition $T(x,0) = T_a$, on can define $\tilde{x} = \frac{x}{R}$, $\tilde{T} = \frac{T}{T_a}$, $\tilde{t} = \frac{2h}{\rho R c_{pw}} t$ and $\epsilon = \frac{\rho c_{pw}}{2h} u$. The equation now becomes

$$\tilde{T}_{\tilde{t}} + \epsilon \tilde{T}_{\tilde{x}} = 1 - \tilde{T} \tag{5}$$

with boundary conditions $\tilde{T}(0,t) = \frac{T_1}{T_a}$ and $\tilde{T}(x,0) = 1$.

Solving for the steady state of the previous equation, and dropping the tilas for convenience,

$$T(x) = (T_1 - T_a)\, e^{-x/\epsilon} + T_a.$$

Using this equation, it is possible to obtain the length of the pipe (as a function of radius, flow rate and initial temperature) needed by the water to reach a given temperature T_2, by solving for L in $T(L) = \frac{T_2}{T_a}$:

$$L = \frac{Q \rho_w c_{pw}}{2\pi R h} \ln\left(\frac{T_2 - T_a}{T_a - T_1}\right)$$

where $Q = \pi R^2 u$ is the volumetric flow rate. Note that the length is dependent on h which can vary by up to an order of magnitude, depending on the type of ground.

If $u(t) = 0$ in equation 4, it is possible to solve for $T(x,t)$ since

$$\frac{T_t(x,t)}{T(x,t) - T_a} = \frac{-2h}{R c_p \rho}.$$

Integrating with respect to t yields

$$T = C(x) \exp\left(\frac{-2h}{R c_p \rho} t\right) + T_a$$

where $C(x)$ is a function determined by the initial condition.

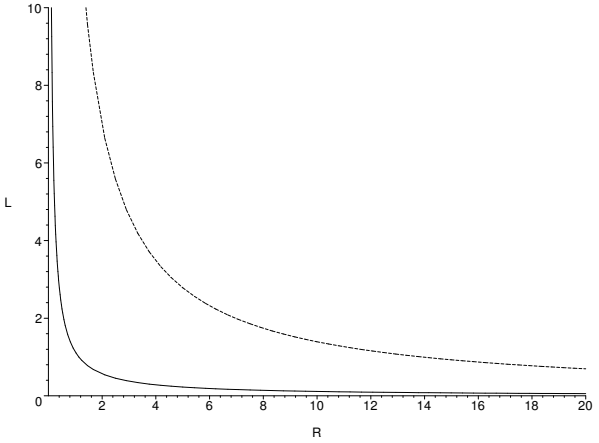


Figure 2: Expected pipe length in meters as a function of the radius of the pipe using plug flow. The lower values seem to be too large to be practical. The top curve uses $h = 55\text{W/Km}^2$, whereas the bottom curve uses $h = 675\text{W/Km}^2$.

Poiseuille Flow

A model refinement can be implemented by taking into account non-uniform velocity profile in the pipe. Assuming $u = u(r)$ with Poiseuille flow yields:

$$u = -\frac{\Delta P}{4\mu}\left(R^2 - r^2\right),$$

where ΔP is change of pressure (assumed to be constant and negative) of the fluid. Now, the flow rate is

$$Q = 2\pi \int_0^R u r dr$$

and the new energy equation

$$\begin{aligned} &\rho_w c_{pw} \overline{T}_t + \frac{Q \rho_w c_{pw}}{\pi R^2} \overline{T}_x \\ &= k_w \overline{T}_{xx} - \frac{2h}{R}\left(\overline{T} - T_a\right) \end{aligned} \tag{6}$$

where $T(x,t)$ was replaced by its average value over time, $\overline{T}(x,t)$.

Non-dimensionalizing gives

$$\frac{P_e}{B_i}\tilde{T}_{\tilde{t}} = \frac{1}{B_i}\tilde{T}_{\tilde{x}\tilde{x}} + 1 - \tilde{T},$$

from which the steady-state temperature distribution is

$$T(x) = (T_1 - T_a)\, e^{\frac{B_i}{P_e}x} + T_a.$$

Thus to heat the water to T_2, the length of the pipe must be

$$L = R\frac{\ln\left(\frac{T_2 - T_a}{T_1 - T_a}\right)}{P_e - \sqrt{P_e^2 + 2B_i}}$$

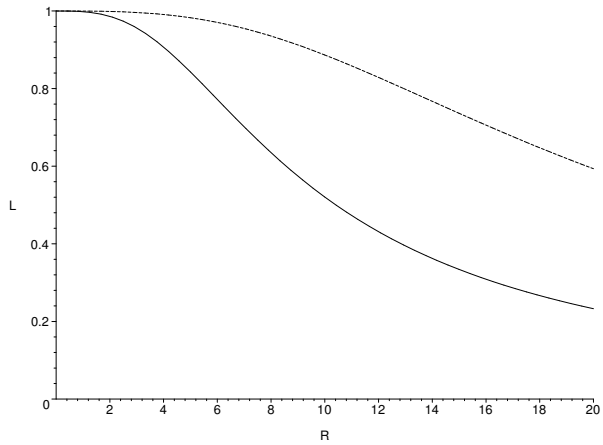


Figure 3: Ratio of the lengths of the pipe without and with Poiseuille flow taken into account as a function of the radius in meters. As long as the pipe is less than roughly 2m, the effects of Poiseuille flow are negligible. The top curve uses $h = 55\text{W/Km}^2$, whereas the bottom curve uses $h = 675\text{W/Km}^2$.

Conclusion

In order to optimize a geothermal air-conditing system, a model of the heat exchange between the pipe and the soil has been developped. It was observed that the temperature in the soil is not constant, which was confirmed by a derivation of the soil's temperature profile. In fact, an inversion occurs within just a few meters of the surface. This temperature difference can be taken advantage of to maximize the power output of the air-conditioning system. To understand the influence of the behaviour of water on the temperature profile in the pipe, the flow of water was modeled using both a plug flow or a Poisseuille flow. Both models have been used to predict the pipe length necessary for the extraction from the ground of enough heat to heat a house. The predicted lengths turned out to be very close in both cases for realistic pipe radii. The assumption that the soil remains at constant temperature along the pipe should give a lower bound on the pipe length as the soil is then able to give off more heat.

The authors would like to thank professor Burt Tilley for the support during the work on this project.

References

[1] Lin, C. Segel, L. A. Mathematics Applied to Deterministic Problems in the Natural Sciences, Classics in Applied Mathematics, 1998.

[2] US Department of Energy. [www.eia.doe.gov], 2007.

Jokes

An engineer, a physicist and a mathematician were asked to hammer a nail into a wall.
The engineer went to build a Universal Automatic Nailer – a device able to hammer every possible nail into every possible wall.
The physicist conducted series of experiments on strength of hammers, nails, and walls and developed a revolutionary technology of ultra-sonic nail hammering at super-low temperature.
The mathematician generalized the problem to a N dimensional problem of penetration of a knotted one dimensional nail into a N-1 dimensional hyper-wall. Several fundamental theorems are proved. Of course, the problem is too rich to suggest a possibility of a simple solution, even the existence of a solution is far from obvious. □

A mathematician was put in a room. The room contains a table and three metal spheres about the size of a softball. He was told to do whatever he wants with the balls and the table in one hour. After an hour, the balls are arranged in a triangle at the center of the table. The same test is given to a Physicist. After an hour, the balls are stacked one on top of the other in the center of the table. Finally, an Engineer was tested. After an hour, one of the balls is broken, one is missing, and he's carrying the third out in his lunchbox. □

CENTRE DE RECHERCHES MATHÉMATIQUES

Grâce à sa programmation de renommée internationale lancée par le CRM dans les années 80, ses ateliers scientifiques et ses activités de transfert et de formation, ses grandes conférences, ses 1500 chercheurs invités chaque année provenant des quatre coins du monde, grâce à ses neuf laboratoires impliquant 170 chercheurs de douze grandes universités du Québec et de l'Ontario, le Centre de recherches mathématiques (CRM) est une plaque tournante mondiale en sciences mathématiques.

With its world-renowned thematic programming introduced by the CRM in the 80's, its scientific workshops, outreach activities and grandes conférences, its 1,500 annual visiting scientists from around the world, and nine laboratories directly involving 170 researchers from twelve major universities in Quebec and Ontario, the Centre de recherches mathématiques (CRM) is a major hub for the mathematical sciences.

CRM propose:

• une programmation scientifique de niveau international
• des laboratoires de recherche de haute performance
• un point d'ancrage de plus d'une centaine de chercheurs membres qui se penchent sur la recherche fondamentale en mathématique et leurs applications, ainsi que de chercheurs invités
• des cours gradués auprès de jeunes chercheurs
• un lieu privilégié de rencontre où tous les membres bénéficient de nombreux échanges et collaborations scientifiques.

CRM proposes :

• world-renowned scientific programming
• high performance research laboratories
• a hub for over one hundred researchers members of CRM doing fundamental research in mathematics and their applications, as well as invited researchers
• advanced courses for young researchers
• a major meeting place where all members benefit from a large number of scientific exchanges and collaborative projects.

Les grandes conférences du CRM

Données par des scientifiques qui ont l'expérience de la communication, les Grandes conférences du Centre de recherches mathématiques s' adressent au public curieux de comprendre les développements récents les plus marquants en sciences mathématiques. De la cryptographie et de l'informatique quantiques au chaos des systèmes météorologiques ou financiers, en passant par l'imagerie cérébrale et les révolutions biotechnologiques, les conférences ont ceci en commun qu'elles cherchent à révéler la beauté et la puissance de la recherche mathématique de pointe dans un langage accessible à tous.

The CRM' s "Grandes conférences" series invites scientists with a gift for communicating to present the most exciting recent developments in mathematics to a curious general public. From cryptography and quantum computing to chaos in meteorological or financial systems, and brain imagery and revolutions in biotechnology, all of the conferences reveal the power and beauty of cutting-edge mathematical research in a language accessible to all.

Les Grands Conférenciers : Jean-Marie De Koninck, Bart de Smit, Ivar Ekeland, Francis Clarke et Jean-Paul Delahaye

www.crm.math.ca • CRM@CRM.UMontreal.CA

Table des caractères invariants de gl_2 sur un corps fini

Marc Desgroseilliers

Après une courte introduction concernant la théorie des représentations, nous calculons les classes de conjugaison de $gl_2(k)$ pour un corps fini k, puis sa table de caractères invariants par conjugaison. L'attrait de la technique utilisée vient du fait que les calculs effectués sont élémentaires et permettent de déduire des informations intéressantes sur le groupe associé.

La théorie des représentations

L'idée (très générale) derrière la théorie des représentations des groupes est d'étudier un groupe G à travers des homomorphismes $\rho : G \to GL(V)$ dans le groupe d'automorphismes d'un espace vectoriel judicieusement choisi. Habituellement, il est intéressant d'étudier les représentations – un vectoriel avec l'homomorphise ρ associé – dont les vectoriels sont dans un certain sens indécomposables. On dit alors qu'il s'agit d'une représentation irréductible. La théorie des caractères utilise la fonction de trace sur ces vectoriels pour déduire des propriétés intéressantes et utiles du groupe G. Cette théorie est entre autres une des pierres angulaires de la classification des groupes simples finis. Il est parfois fort difficile d'obtenir les caractères associés à une représentation irréductible pour un groupe, par exemple les groupes matriciels sur un corps fini. Dans l'article qui suit, nous nous proposons de calculer des caractères associés à l'algèbre de Lie, elle même liée au groupe matriciel en question. Nous nous bornons à dire qu'une fois que cette table de caractères associés est calculée, il est possible de déduire les caractères irréductibles du groupe matriciel, sans entrer dans les détails.

Énoncé du problème, notations, définitions

Soit $gl_2(\mathbb{F}_q)$ l'anneau des matrices de dimension 2 sur $\mathbb{F}_q$ le corps à q éléments, où $q = p^e$. Il y a une action naturelle de $GL_2(\mathbb{F}_q)$, le groupes des matrices inversibles, par conjugaison et nous notons $O\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ l'orbite de la matrice sous cette action. Soit Ψ un caractère additif non-trivial sur $\mathbb{F}_q$ (un homomorphisme du groupe additif de $\mathbb{F}_q$ dans le groupe multiplicatif de $\mathbb{C}$, par exemple $x \mapsto e^{\frac{2\pi i}{p} Tr_{\mathbb{F}_q/\mathbb{F}_p}(x)}$). Regardons l'homomorphisme $\Theta_X : (gl_2, +) \to \mathbb{C}^*; Y \mapsto \Psi(\mathrm{tr}(XY))$, qui est un caractère sur l'algèbre de Lie. En prenant $S_O := \sum_{X \in O(Y)} \Theta_X$ pour un Y donné, nous obtenons un caractère (puisque c'est une somme de caractères) qui est invariant par l'action de conjugaison de GL_2 définie par $g(\chi(X)) = \chi(gXg^{-1})$. L'intérêt de cette construction réside en le fait que ces caractères sont minimaux , en ce sens qu'ils ne peuvent pas être décomposés en somme de caractères invariants par l'action de GL_2. En effet, une telle décomposition partitionerait l'orbite et les caractères ne pourraient pas être GL_2-invariants. De plus, si l'orbite n'est pas triviale (elle ne contient pas l'identité), alors $\chi_{\text{triviale}} \notin S_O$. Nous pouvons alors utiliser le produit scalaire habituel $(\chi|\Phi) := \frac{1}{|G|}\sum_{g\in G} \chi(g)\overline{\Phi(g)}$ et les relations d'orthogonalité pour conclure que le produit scalaire entre le caractère trivial et Ψ est 0, et donc $\sum_{m\in\mathbb{F}_q} \Psi(m) = 0$. Nous souhaitons calculer S_O, c'est-à-dire

$$\begin{aligned}&\sum_{y\in O} \Psi(\mathrm{tr}(yx))\\ &= \sum_{m\in\mathbb{F}_q} |(y \in gl_2(\mathbb{F}_q) : y \in O, \mathrm{tr}(yx) = m)|\Psi(m) \qquad (1)\end{aligned}$$

pour un $x \in gl_2(\mathbb{F}_q)$ et une orbite O fixés.

Premièrement, observons que la somme est invariante par rapport au choix de deux x dans la même classe de conjugaison. Pour x et $hxh^{-1} = x' \in O'$ et $y \in O$

$$\begin{aligned}&\sum_{y\in O} \Psi(\mathrm{tr}(xy))\\ &= \frac{\sum_{g\in GL_2(\mathbb{F}_q)} \Psi\left(\mathrm{tr}\left(xgyg^{-1}\right)\right)}{|\mathrm{Stab}(y)|}\\ &= \frac{\sum_{g\in GL_2(\mathbb{F}_q)} \Psi\left(\mathrm{tr}\left(h^{-1}\left(hx'h^{-1}gyg^{-1}\right)h\right)\right)}{|\mathrm{Stab}(y)|}\\ &= \sum_{y\in O} \Psi(\mathrm{tr}(x'y))\end{aligned}$$

Classes de conjugaison

Le but de cette section est de classifier les classes de conjugaison de gl_2 et de compter le nombre d'éléments dans chaque classe. Nous utilisons sans distinction le vocabulaire de classe de conjugaison et d'orbite, en gardant en tête l'action de conjugaison de $GL_2(\mathbb{F}_q)$ sur $gl_2(\mathbb{F}_q)$. Nous observons que $|GL_2(\mathbb{F}_q)| = (q^2-1)(q^2-q)$. En effet, nous avons (q^2-1) choix pour la première ligne, et (q^2-q) pour la deuxième ligne (nous éliminons les multiples de la première ligne afin que le déterminant soit non-nul).

Cas 1: éléments centraux

Pour commencer, il est clair que les éléments centraux, qui commutent avec tous les autres éléments, sont seuls dans leur classe de conjugaison. Il y a q tels éléments.

Cas 2: éléments diagonalisables

Nous regardons maintenant les éléments diagonalisables avec valeurs propres distinctes. Soit $A := \left(\begin{smallmatrix} \alpha & 0 \\ 0 & \beta \end{smallmatrix}\right)$. Nous cherchons les éléments $g \in GL_2(\mathbb{F}_q)$ tels que gA=Ag. Si

$g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, les équations suivantes doivent être satisfaites:

$$b\beta = \alpha b \qquad\qquad \alpha c = \beta c$$

Comme $\alpha \neq \beta$, nous avons que $b = c = 0$ et donc $|\mathrm{Stab}(A)| = (q-1)^2$. Nous concluons que la taille de la classe de conjugaison d'un élément diagonalisable avec valeurs propres distinctes est $|GL_2(\mathbb{F}_q)|/|\mathrm{Stab}(A)| = q(q+1)$.

Cas 3: Une valeur propre, non diagonalisable

Nous considérons ici des matrices non diagonalisables avec une seule valeur propre. La forme normale de Jordan dans ce cas est $\left(\begin{smallmatrix} \alpha & 1 \\ 0 & \alpha \end{smallmatrix}\right)$. Comme précédemment, nous déterminons l'ordre du stabilisateur d'une telle matrice et arrivons aux équations suivantes:

$$c = 0 \qquad\qquad a = d$$

Nous concluons qu'il y a $q-1$ possibilités pour la valeur de a ($a \neq 0$ sinon le déterminant est nul) et q possibilités pour b. $|\text{Orbite}| = \frac{|GL_2(\mathbb{F}_q)|}{|\text{Stabilisateur}|} = (q-1)(q+1) = q^2 - 1$.

Cas 4: aucune valeur propre dans $\mathbb{F}_q$

Finalement, le polynôme caractéristique de la matrice peut être irréductible sur $\mathbb{F}_q$. Puisque le polynôme caractéristique est de degré 2, ses valeurs propres se situent dans une extension de $\mathbb{F}_q$ de degré 2, ou, plus simplement, $\mathbb{F}_{q^2}$. Soient τ et ω les deux valeurs propres de la matrice $\left(\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right)$ en question. Nous voulons trouver l'ordre du stabilisateur de cette matrice dans $GL_2(\mathbb{F}_q)$.

Soient X une matrice dont le polynôme caractéristique est irréductible sur $\mathbb{F}_q$, Y la matrice diagonale associée dans $\mathbb{F}_{q^2}$, h la matrice telle que $h^{-1}Xh = Y$ et F l'homomorphisme $F : gl_2(\mathbb{F}_{q^2}) \to gl_2(\mathbb{F}_{q^2}), (a_{ij}) \mapsto (a_{ij})^q$ dont les points fixes sont les matrices avec coefficients dans $\mathbb{F}_q$. Nous avons le diagramme suivant

$$\begin{array}{ccc} GL_2(\mathbb{F}_{q^2}) & \xrightarrow{\mathrm{Aut}_h} & GL_2(\mathbb{F}_{q^2}) \\ \downarrow F & & \downarrow F' \\ GL_2(\mathbb{F}_{q^2}) & \xrightarrow{\mathrm{Aut}_h} & GL_2(\mathbb{F}_{q^2}) \end{array}$$

où $\mathrm{Aut}_h(z) := h^{-1}zh$ et F' est définie de façon à rendre le diagramme commutatif. Dans ce cadre plus général, $|\mathrm{Stab}_{GL(\mathbb{F}_q)}(X)| = |\mathrm{Stab}_{GL(\mathbb{F}_{q^2})}(X)^F|$, où G^F dénote les points de G fixés par la fonction F.

Nous voudrions voir que $|\mathrm{Stab}_{GL(\mathbb{F}_{q^2})}(X)^F| = |\mathrm{Stab}_{GL(\mathbb{F}_{q^2})}(Y)^{F'}|$. Soit $g \in \mathrm{Stab}_{GL(\mathbb{F}_{q^2})}(X)^F = \mathrm{Stab}_{GL(\mathbb{F}_q)}(X)$. Alors $h^{-1}ghYh^{-1}g^{-1}h = h^{-1}gXg^{-1}h = h^{-1}Xh = Y$ d'où nous concluons $h^{-1}gh \in \mathrm{Stab}_{GL(\mathbb{F}_{q^2})}(Y)$ et $F'(h^{-1}gh) = h^{-1}gh$. De la même manière, on montre que pour $g' \in \mathrm{Stab}_{GL(F_{q^2})}(Y)^{F'}$, alors $hg'h^{-1} \in \mathrm{Stab}_{GL(\mathbb{F}_{q^2})}(X)^F$.

Soit Y comme ci-haut. Alors $F'(Y) = h^{-1}F(hYh^{-1})h = Y$ et puisque F est un homomorphisme, alors il s'agit en fait de la congugaison de $F(Y)$ par $h^{-1}F(h)$. Soit

$$T := \left\{ \left(\begin{array}{cc} \alpha & 0 \\ 0 & \beta \end{array} \right) \mid \alpha, \beta \in \mathbb{F}_{q^2} \right\}$$

le tore dans $\mathbb{F}_{q^2}$. On vérifie que le normalisateur du tore est le sous-groupe engendré par $\langle \sigma, T \rangle$, où $\sigma := \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Comme Y et $F(Y) \in T$, on conclut que $h^{-1}F(h) \in \mathrm{Normalisateur}_{GL(\mathbb{F}_{q^2})}(T)$ et donc qu'il peut s'écrire comme σt pour $t \in T$. Nous avons donc que $F'(Y) = Y$ ce qui entraîne $\sigma t F(Y) t^{-1}\sigma = Y$ ou $\sigma F(Y)\sigma = Y$ puisque deux éléments du tore commutent et que σ est son propre inverse. Nous concluons que Y est de la forme $\left(\begin{smallmatrix} \tau & 0 \\ 0 & \tau^q \end{smallmatrix}\right)$ pour $\tau \in \mathbb{F}_{q^2} \backslash \mathbb{F}_q$. De plus, $|\mathrm{Stab}_{GL(\mathbb{F}_{q^2})}(Y)^{F'}| = q^2 - 1$ puisque le choix d'un élément dans la case (1,1) de la matrice stabilisant Y spécifie complètement la matrice, et que le déterminant doit être non-nul. Nous concluons qu'il y a $q(q-1)$ éléments dans l'orbite d'un élément dont les valeurs propres ne sont pas dans $\mathbb{F}_q$.

De plus, supposons qu'on peut choisir $t = \left(\begin{smallmatrix} \alpha & 0 \\ 0 & \beta \end{smallmatrix}\right) \in T$ pour un X donné. En opérant un changement de base $\{e_1, e_2\} \mapsto \{e_1, \frac{\alpha}{\beta} e_2\}$, et en faisant un choix approprié de h, l'application F' se réduit à l'application de F, suivie de la conjugaison par $\sigma = h^{-1}F(h)$.

Table des caractères

Notre but est de remplir la table suivante avec la valeur de $\sum_{Y \in O} \Psi(\mathrm{tr}(XY))$ pour un X fixé dans chaque colonne. Nous avons la liberté de choisir le X qui nous convient le mieux (voir section 1).

Cette table possède une symétrie que nous utiliserons abondamment. En effet, nous avons:

$$\begin{aligned} &\sum_{y \in O} \Psi(\mathrm{tr}(xy)) \\ &= \frac{\sum_{g \in GL_2(\mathbb{F}_q)} \Psi\left(\mathrm{tr}\left(xgyg^{-1}\right)\right)}{|\mathrm{Stab}_{GL_2(\mathbb{F}_q}(y)|} \\ &= \frac{\sum_{g \in GL_2(\mathbb{F}_q)} \Psi\left(\mathrm{tr}\left(g^{-1}xgy\right)\right)}{|\mathrm{Stab}_{GL_2(\mathbb{F}_q)}(x)|} \frac{|\mathrm{Stab}_{GL_2(\mathbb{F}_q)}(x)|}{|\mathrm{Stab}_{GL_2(\mathbb{F}_q)}(y)|} \\ &= \frac{|O(y)|}{|O(x)|} \sum_{x \in O} \Psi(\mathrm{tr}(xy)) \end{aligned}$$

Autrement dit, la valeur dans la case (i,j) est un multiple de la valeur de la case (j,i), ce multiple dépendant uniquement de la taille des orbites en question.

Première ligne

La première ligne (et donc la première colonne, par l'observation précédente) est aisée puisqu'un élément cen-

	$\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}$	$\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}$	$\begin{pmatrix} \omega & 0 \\ 0 & \omega^q \end{pmatrix}$	$\begin{pmatrix} x & 1 \\ 0 & x \end{pmatrix}$
$O\begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix}$	$\Psi(2\alpha x)$	$\Psi(\alpha(x+y))$	$\Psi(\alpha(\omega+\omega^q))$	$\Psi(2\alpha x)$
$O\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$	$q(q+1)\Psi(x(\alpha+\beta))$	$q[\Psi(\alpha y+\beta x)+\Psi(\alpha x+\beta y)]$	0	$q\Psi(x(\alpha+\beta))+\Psi(\alpha x+\beta y)]$
$O\begin{pmatrix} \tau & 0 \\ 0 & \tau^q \end{pmatrix}$	$q(q-1)\Psi(x(\tau+\tau^q))$	0	$-q[\Psi(\omega\tau+\omega^q\tau^q)+\Psi(\omega\tau^q+\omega^q\tau)]$	$-q\Psi(x(\tau+\tau^q))$
$O\begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix}$	$(q^2-1)\Psi(2\alpha x)$	$(q-1)\Psi(\alpha(x+y))$	$-(q+1)\Psi(\alpha(\omega+\omega^q))$	$-\Psi(2x\alpha)$

tral est seul dans sa classe de conjugaison. Les valeurs sont donc, de gauche à droite, $\Psi(2\alpha x)$, $\Psi(\alpha(x+y))$, $\Psi(\alpha(\omega+\omega^q))$ et $\Psi(2\alpha x)$. Nous concluons que les valeurs de la première colonne sont, de haut en bas, $\Psi(2\alpha x)$, $q(q+1)\Psi(x(\alpha+\beta))$, $q(q-1)\Psi(x(\tau+\tau^q))$ et $(q^2-1)\Psi(2\alpha x)$.

Case (2,2)

Nous voulons calculer la cardinalité des $Y \in gl_2(\mathbb{F}_q)$ tels que

$$Y \in O\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \cap \operatorname{tr}\left(Y\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}\right) = m$$

pour ensuite faire la somme sur tous les $m \in \mathbb{F}_q$. Soit $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ une telle matrice. Nous avons, en comparant la trace et le déterminant:

$$a+d = \alpha+\beta \tag{2}$$
$$ad-bc = \alpha\beta \tag{3}$$
$$xa+yd = m \tag{4}$$

d'où $d = \frac{m-x(\alpha+\beta)}{y-x}$ et $a = \frac{-m+y(\alpha+\beta)}{y-x}$ en utilisant (1) et (3). Nous observons que 2 cas sont possibles: $ad = \alpha\beta$ ou $ad \neq \alpha\beta$. Dans le deuxième cas, pour $b \in \mathbb{F}_q*$ fixé, le choix de $c \in \mathbb{F}_q*$ est fixé et il y a donc $q-1$ matrices pour chaque m. Dans le cas où $ad = \alpha\beta$, nous déduisons que

$$m^2 - m(\alpha+\beta)(x+y) + xy(\alpha+\beta)^2 + \alpha\beta(y-x)^2 = 0.$$

L'équation est quadratique en m et donc

$$\begin{aligned} m &= \frac{(x+y)(\alpha+\beta) \pm \sqrt{\begin{matrix}(x+y)^2(\alpha+\beta)^2 \\ -4(xy)(\alpha+\beta)^2 \\ -4\alpha\beta(y-x)^2\end{matrix}}}{2} \\ &= \frac{(x+y)(\alpha+\beta) \pm \sqrt{(y-x)^2(\alpha-\beta)^2}}{2} \\ &= \alpha y+\beta x \quad ou \quad \alpha x+\beta y \end{aligned}$$

et nous pouvons calculer que ce résultat est toujours vrai en caractéristique 2. Si $m = \alpha y+\beta x$ ou $\alpha x+\beta y$, soit $b = 0$ et il y a q possibilités pour la valeur de c, ou $c = 0$ et il y a $q-1$ possibilités pour la valeur de b ($b = c = 0$ a déjà été compté). Nous concluons qu'il y a $2q-1$ possibilités pour $m = \alpha y+\beta x$ ou $\alpha x+\beta y$. Nous vérifions que tous les éléments de l'orbite ont été pris en considération puisque

$$2(2q-1) + (qk2)(q-1) = q(q+1) = \left|O\left(\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}\right)\right|.$$

En se rappelant que $\sum_{m\in\mathbb{F}_q}\Psi(m) = 0$, nous concluons que

$$\begin{aligned} &\sum_{Y\in O}\Psi\left(\operatorname{tr}\left(\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}Y\right)\right) \\ &= (q-1)\sum_{m\in\mathbb{F}_q\setminus\{\alpha y+\beta x,\alpha x+\beta y\}}\Psi(m) \\ &\quad + (2q-1)(\Psi(\alpha y+\beta x)+\Psi(\alpha x+\beta y)) \\ &= -(q-1)(\Psi(\alpha y+\beta x)+\Psi(\alpha x+\beta y)) \\ &\quad + (2q-1)(\Psi(\alpha y+\beta x)+\Psi(\alpha x+\beta y)) \\ &= q(\Psi(\alpha y+\beta x)+\Psi(\alpha x+\beta y)) \end{aligned}$$

Case (2,3)

Soit $Y \in O\left(\left(\begin{smallmatrix} \alpha & 0 \\ 0 & \beta \end{smallmatrix}\right)\right)$et $H\left(\begin{smallmatrix} \omega & 0 \\ 0 & \omega^q \end{smallmatrix}\right)H^{-1}$ un élément dont le polynôme caractéristique est irréductible sur $\mathbb{F}_q$. Nous avons que $F'(H^{-1}YH) = H^{-1}YH$ puisque $Y \in \mathbb{F}_q$ est F-stable. Soit $H^{-1}YH = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = Y'$. Nous cherchons alors, pour $m \in \mathbb{F}_q$, une solution

$$\begin{aligned} a\omega + d\omega^q &= m & ad-bc &= \alpha+\beta \\ a+d &= \alpha+\beta \end{aligned}$$

puisque la conjugaison par H n'affecte ni la trace, ni le déterminant de la matrice Y. De plus, $F'(Y') = \sigma F(Y')\sigma = Y'$ et donc $d = a^q$ et $c = b^q$. En remplaçant ceci dans les équations ci-haut, nous obtenons $d = \frac{m-(\alpha+\beta)\omega}{\omega^q-\omega}$. Nous devons maintenant résoudre $c^{q+1} = d^{q+1} - \alpha\beta$, pour $c \in \mathbb{F}_{q^2}$. Premièrement, $c \neq 0$, car sinon, la trace et le déterminant de Y et de $\left(\begin{smallmatrix} \omega & 0 \\ 0 & \omega^q \end{smallmatrix}\right)$sont les mêmes, une absurdité puisque une matrice est diagonalisable et l'autre pas. L'équation $x^{q+1} - d^{q+1} + \alpha\beta = 0$ ne peux avoir plus de $q+1$ solutions. Comme $(c^{q+1})^q = c^{q^2+q} = c^{q+1}$, nous concluons que $c^{q+1} \in \mathbb{F}_q$. Par le principe du pigeonnier, pour chaque valeur de c^{q+1} dans $\mathbb{F}_q*$, l'équation possède exactement $q+1$ racines. Ceci implique que

$$\sum_{Z\in O}\Psi\left(\operatorname{tr}\left(\begin{pmatrix} \omega & 0 \\ 0 & \omega^q \end{pmatrix}Z\right)\right) = 0$$

puisqu'il y le même nombre d'éléments pour chaque valeur de m.

Case (3,3)

Nous utilisons un argument similaire à la case (2,3). Soit $H^{-1}\left(\begin{smallmatrix}\omega & 0\\ 0 & \omega^q\end{smallmatrix}\right)H$ une matrice dont le polynôme caractéristique est irréductible sur F_q. On considère $Y=\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)= HXH^{-1}$, où $X \in O\left(\left(\begin{smallmatrix}\tau & 0\\ 0 & \tau^q\end{smallmatrix}\right)\right)$. Alors $F'(Y)=Y$ et nous avons

$$\begin{aligned} d &= a^q & c &= b^q \\ a^{q+1}-b^{q+1} &= \tau^{q+1} & a+a^q &= \tau+\tau^q \\ a\omega + a^q\omega^q &= m \end{aligned}$$

Si $a^{q+1}=\tau^{q+1}$, alors $b=c=0$ et $a \in \{\tau,\tau^q\}$ et donc $m=\omega\tau+\omega^q\tau^q$ ou $m=\omega\tau^q+\omega^q\tau$ (ces deux valeurs sont distinctes puisque τ et $\omega \in \mathbb{F}_{q^2}\setminus\mathbb{F}_q$). Pour un m fixé tel que $a^{q+1}\neq\tau^{q+1}$, nous avons $b\neq 0$. On cherche donc les solutions de l'équation $a^{q+1}-\tau^{q+1}=b^{q+1}$. Comme tous ces éléments sont dans $\mathbb{F}_q$, que $b\neq 0$ et qu'une équation de degré $q+1$ ne peut avoir plus de $q+1$ solutions (voir Case (2,3)), on en conclut qu'il y a exactement $q+1$ possibilités pour la valeur de b.

$$\begin{aligned} &\sum_{Y\in HOH^{-1}} \Psi\left(\mathrm{tr}\left(\begin{pmatrix}\omega & 0\\ 0 & \omega^q\end{pmatrix}Y\right)\right) \\ &= \Psi(\omega\tau+\omega^q\tau^q)+\Psi(\omega\tau^q+\omega^q\tau) \\ &\quad +(q+1)\sum_{m\in\mathbb{F}_q\setminus\{\omega\tau+\omega^q\tau^q,\omega\tau^q+\omega^q\tau\}}\Psi(m) \\ &= -q[\Psi(\omega\tau+\omega^q\tau^q)+\Psi(\omega\tau^q+\omega^q\tau)] \end{aligned}$$

Case (3,4)

Suivant la même approche que précédemment, nous cherchons les éléments $\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)\in O\left(\left(\begin{smallmatrix}\tau & 0\\ 0 & \tau^q\end{smallmatrix}\right)\right)$ tels que

$$\begin{aligned} a+d &= \tau+\tau^q & ad-bc &= \tau^{q+1} \\ x(a+d)+c &= x(\tau+\tau^q)+c = m \end{aligned}$$

$c\neq 0$ sinon $ad=\tau^{q+1}$ et donc $\{a,d\}=\{\tau,\tau^q\}$, une contradiction puisque la matrice désirée est dans $\mathbb{F}_q$. Pour $c\neq 0$ fixé, alors un choix pour la valeur a dans $\mathbb{F}_q$ détermine la valeur de d, ce qui assigne alors une valeur univoque à b. On voit donc que

$$\begin{aligned} \sum_{Y\in O}\Psi\left(\mathrm{tr}\left(\begin{pmatrix}x & 1\\ 0 & x\end{pmatrix}Y\right)\right) &= q\sum_{m\in\mathbb{F}_q\setminus\{x(\tau+\tau^q)\}}\Psi(m) \\ &= -q\Psi(x(\tau+\tau^q)) \end{aligned}$$

Case (4,2)

On cherche une matrice $\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$ telle que

$$\begin{aligned} a+d &= 2\alpha & ad-bc &= \alpha^2 \\ xa+yd &= m \end{aligned}$$

Similairement à la case (2,2), $a=\frac{2\alpha y-m}{y-x}$ et $d=\frac{m-2\alpha x}{y-x}$. Si $ad=\alpha^2$, alors $m^2-m(2\alpha)(x+y)+4\alpha^2xy+\alpha^2(y-x)^2$ et donc, en utilisant la formule quadratique, m=$2\alpha(x+y)$. Dans ce cas, $bc=0$ et il y a $2q-2$ possibilités (car $b=c=0$ est impossible puisque la matrice n'est pas diagonalisable). Si $ad\neq\alpha^2$, alors $bc\neq 0$ et il y a $q-1$ choix pour la valeur de b, ce qui détermine univoquement la valeur de c. On voit donc, en utilisant $\sum_{m\in\mathbb{F}_q}\Psi(m)=0$, que

$$\begin{aligned} &\sum_{Y\in O}\Psi\left(\mathrm{tr}\left(\begin{pmatrix}x & 0\\ 0 & y\end{pmatrix}Y\right)\right) \\ &= 2(q-1)\Psi(\alpha(x+y))-(q-1)\Psi(\alpha(x+y)) \\ &= (q-1)\Psi(\alpha(x+y)) \end{aligned}$$

Case (4,4)

On cherche les éléments $\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$ tels que

$$\begin{aligned} a+d &= 2\alpha & ad-bc &= \alpha^2 \\ x(a+d)+c &= x(2\alpha)+c = m \end{aligned}$$

On conclut que la valeur de c est complètement déterminée par le choix de m et vice-versa. Si $c=0$, alors $ad=\alpha^2$ et donc $a=d=\alpha$. Il y a donc $q-1$ choix pour b, car si $b=0$, la matrice est diagonale. Si $c\neq 0$, alors il y a q choix pour la valeur de a et les valeurs de b et d sont fixées. On conclut que

$$\begin{aligned} &\sum_{Y\in O}\Psi\left(\mathrm{tr}\left(\begin{pmatrix}x & 1\\ 0 & x\end{pmatrix}Y\right)\right) \\ &= (q-1)\Psi(2x\alpha)+q\sum_{m\in\mathbb{F}_q\setminus x(a+d)}\Psi(m) \\ &= (q-1)\Psi(2x\alpha)-q\Psi(2x\alpha) \\ &= -\Psi(2x\alpha) \end{aligned}$$

Ces recherches furent effectuées durant un stage d'été de l'Institut des Sciences Mathématiques. Je tiens à remercier mon superviseur Emmanuel Letellier pour son aide tout au long de mes explorations mathématiques.

References

[1] Serge Lang, *Algebra*, Revised Third Edition, New York, Springer, 2002.

[2] Rudolf Lidl, Harald Niederreiter, *Finite Fields*, Encyclopedia of Mathematics and its applications, vol 20, London, Addison-Wesley, 1983.

[3] Jean-Pierre Serre, *Représentations linéaires des groupes finis*, Paris, Methodes, 1998.

Interview with Benoit Charbonneau

Agnès F. Beaudry

This summer, the Delta-Epsilon interviewed prof. Benoit Charbonneau, at the time a postdoctorate student at McGill, now an assistant professor at Duke University. He talked to us about the mechanics of becomming a mathematician, and told us about his experience on the road.

How would you describe "being a postdoc"?

I usually describe it by saying that it is not a diploma. In Québec, we have a very special situation: we are considered students by the ministry of education. But neither McGill nor anybody else on the planet considers postdocs as students. It is a position that you have after your PhD, where you actually do research. You are not permanent: we want to see what you're made of.

How many postdocs does one usually do before getting a tenure track position? How does one make the transition?

It's a strange question, because these positions are not well-defined. We usually say you did two postdocs if you have been to different places. I've been at McGill for three years now, and technically, I have done two postdocs for the following reasons: the first year I was funded by Jacques Hurtubise, and the subsequent years I was funded by NSERC. So in funding terms, I did two different postdocs. However, I've been at the same university, so that's one postdoc. Now I'm going to Duke as a visiting assistant professor, and that's really considered to be a postdoc position: it is not a permanent position and is for young people. Some people find jobs right away, and others have to wait longer. It depends on various factors, like where you would like to be. I would like to end up in Québec, hopefully Montréal, but I might not necessarily want to go to France for example. So if there was a permanent position opening in France, I would not necessarily apply. Even then, if some position opens, they might not be in the right field. There has been some positions opening in Québec, but not in geometry. Hopefully, after five or six years of postdoc position, I'll be out of it, and in a regular faculty, tenure track position. We'll see!

What is the difference between a postdoc position and tenure track position?

Tenure track means you have three to five years to prove what you can do. This is the time when it's really publish or perish. The difference between tenure track and postdoc, is that you're much more left to yourself as a postdoc. The university just said, "we'll take you, for not that much money, for three years or two." Whereas, when you are a faculty member, there are saying, "we are taking you for two or three years, but we might also take you forever." That's the difference. With this also comes responsibility: they want to evaluate you not only on your publishing, but also on how much of an asset you are to the department, in helping with various committees, etc.

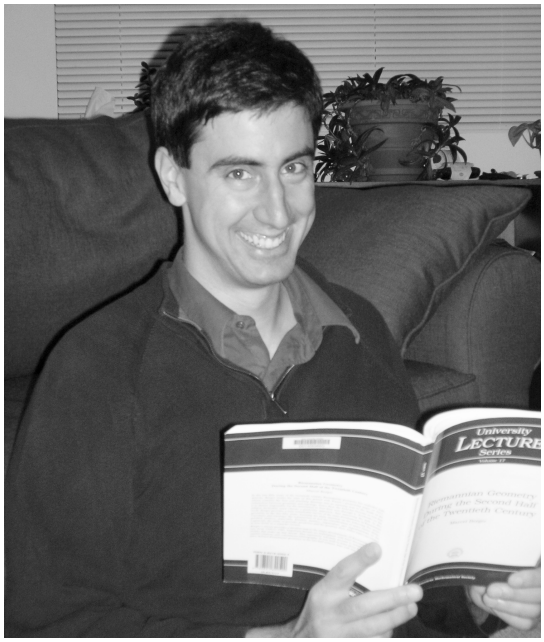


Figure 1: Benoit Charbonneau

How do you work as a postdoc, how does collaboration work?

Well, first of all, at most universities , the difference between being regular faculty and postodcs, is that as a postdoc you are going to work with someone. I have been working with Jacques Hurtubise for the past three years now, and I came to work with him. It does not mean that's the only thing I do. Collaborations develop in other ways as well. When the first paper that came out of my thesis was put on the archive, which is the repository of all new preprints that come out in math, a person in Brazil invited me to come to Brazil and I spent ten days there. It could have come to nothing, but after a few days, we realized that we had a question, and that we could work on that question. Most collaborations work that way: you have a question that you figure out you are both interested in, and you try to bring your tools, you try to think, you argue, you go back in your respective rooms and think about it.

How does the collaboration work with your professors?

You exit your PhD as an expert on a little something. What happened in my case is that Jacques is an expert in a different thing. He said: "Why don't we study what you have done and try to extend it." He is bringing different tools, different questions, and so am I. It's not the same type of relationship one has as an undergraduate when you are assigned a task to work on. In fact, if a project becomes boring, you go on to another subject.

What are you doing, why is it important, and how does it relate to other fields of mathematics and physics?

I'm doing differential geometry, but more specifically gauge theory, which is the study of vector bundles and connections. That's how mathematicians and differential geometers talk about it. For physicists, it is related to particles. They think of it in a very different way. What happens is that there is some equation that came out of physics, in my case Yang Mills' instanton equation, which is a differential equation for certain objects on certain spaces, and physicists are interested in that since it represents something that has to do with particles and with which they play. Mathematicians, on the other hand, realized that we can gain some information in topology by studying these equations. Although it's not the first thing that happened, it was very powerful. Also, these equations by themselves, their space of solutions, trying to understand what's going on became in itself a field of research. Physicist take what they want out of it and interpret it in their own way.

Did you expect to be working on these projects?

When I applied to MIT, I said that I wanted to work on Seiberg-Witten, which is another part of gauge theory. It's another concept coming from physics. In the simplest case, the Seiberg-Witten theory is an explanation of the mass gap. Seiberg-Witten was the correct formulation of physics to explain this phenomenon. My PhD supervisor Tom Mrowka used Seiberg-Witten to proved interesting results. I was at the right place at MIT and in the Boston environment to study in that field, but that's not what happened. My advisor offered me a problem which is not the problem I ended up solving. This problem required me to understand more of the Yang Mill theory.

Do you feel like a mathematician, after all this time?

Absolutely. I think of it this way. When you are an undergrad, you are *learning* to be a mathematician. Most people after their undergraduate degree are mathematicians in the sense that they know a lot of mathematics, that they are carrying this knowledge. That is the goal of the undergraduate degree. The goal of the masters degree, in my opinion, is to make you an autonomous reader. When you are an undergraduate, you probably are not able to go in books and figure things out by yourself, although perhaps some smart students can. You will never read a math book as you read a novel, but as a masters student, you learn patience. You are assigned something to read, to understand and to chew on, to bring back to life in your own words. Then, as a PhD student, you learn to be an autonomous researcher. If you define *mathematician* as somebody who does research, certainly, now I feel like a mathematician. I felt like a mathematician even before that, but I did not feel equipped to do research. Now I have a lot of projects and a lot of ideas.

Does it feel good to be a mathematician?

Oh yeah, absolutely: we are pushing back the frontiers of ignorance! Like discoverers. Although, maybe not everyday: sometimes it's depressing. You are always at the frontiers of your own ignorance. In fact, if you understand everything you do everyday, it is probably because you are not working hard enough. It's not comfortable: imagine spending two or three days on something when you have no clue what's going on. Then you feel like thinking "Ah! I'm a fraud, I don't know how to do it!" Then you have to go back to your victories of the past and see that you have been able to achieve some outstanding results - and this motivates you to keep trying until you succeed.

Jokes

A professor's enthusiasm for teaching precalculus varies inversely with the likelihood of his having to do it. □

The highest moments in the life of a mathematician are the first few moments after one has proved the result, but before one finds the mistake. □

The reason that every major university maintains a department of mathematics is that it is cheaper to do this than to institutionalize all those people. □

The Airplane Boarding Problem

Alexandra Ortan, Erin Prosk and Vincent Quenneville-Bélair

In order to increase the flying time of a plane, airplane companies try to minimize the boarding time, which is one of the most lengthy parts of a plane's turn time (the time gap between the moment it lands and the moment it takes off). Boarding time is increased by interferences between passengers: a passenger trying to attain his seat is blocked by either passengers in front of him who are stowing their luggage in the overhead (and thus blocking the aisle) or by seated passengers obstructing the access to his seat. To reduce boarding time, it is thus necessary to minimize the number of interferences between passengers by controlling the order in which they get onto the plane through a boarding policy.

As airline companies are looking to increase their profits, they are looking to maximize the flight time of planes. Delays can cost carriers around 22$US per additional minute spent on the ground [4]. If every plane is delayed for a few minutes at every flight, this can amount to considerable sums. Thus, airline companies have every advantage to minimize the turn time of a plane. While this time is also used for servicing and cargo handling, the determining factors are passenger deboarding, cabin and galley servicing and passenger boarding, with the latter taking up the biggest part [3]. The boarding time for an airplane can go from 30 to 60 minutes, of which deboarding takes 10-15, cleaning takes 15-20 and boarding takes up to 30 minutes [6]. While cleaning time can presumably not be much improved with the given resources, the boarding time can be improved by implementing more efficient strategies.

Most airlines assign seats prior to the boarding process, so this gives some control over the order in which the passengers get onto the plane by allowing the crew to call them according to a given method. While a lot of airlines use a back to front method, which boards passengers by blocks, starting from the back of the plane, some airlines have started experimenting new and more sophisticated strategies, like outside-in or rotating zones. However, finding the optimal boarding method seems to fall into the NP-hard complexity class (non-deterministic polynomial time) and hence announces itself not to be an easy task! [5]

Hypothesis

The models will only consider what happens in a plane with only one aisle, since even if bigger planes have two aisles, it is theoretically possible to avoid any interference between them by directing each passenger to the aisle closest to his seat. The same sort of argument can be applied to boarding using two doors or even to two levels; hence, using both doors, or floors, is equivalent to boarding two independent regions of the plane, each through a single door, because the aisle's saturation limit can be easily attained through only one door and thus further people coming in the way would not be able to advance faster.

It is assumed that a passenger will always walk (at constant speed) toward his assigned seat unless his passage through the aisle is blocked (aisle interference) or the access to his seat is obstructed by a passenger seated in the same row in a seat closer to the aisle (seat interference) since passengers are seen to have a desire to reach their place rapidly. It is further assumed that every passenger will seat only at his assigned seat and will put his luggage (in a fixed time [4]) into the overhead bin directly above his seat (which will be assumed to always accommodate as much luggage as is needed) since these are events that are commonly expected.

Variables

Independent variables. The total number K of passenger on the plane – each one moving at speed v in the aisle and causing a time delay t_A for an aisle interference and t_S for a seat interference – are assigned seats according to a predefined sequence (called Boarding Policy) of seating regions in the airplane – having a seat configuration with Δx as the width of one of the X rows of Y seats.
Dependent variable. The total time, considering the number N_A of different aisle interferences and the number N_S of seat interferences, taken to seat everyone is T.

Model A

Counting interferences

A seat in the plane is denoted by (x, y), where $x \in \{0, ..., X\}$ is the row number and $y \in \{-y_{\min}, ..., -1, 1, ..., y_{\max}\}$ is the seat number, while pairs $(x, 0)$ representing the aisle. Every passenger is assigned a unique seat, and the order in which the passengers board onto the plane is given by the sequence $((x_1, y_1), (x_2, y_2), ..., (x_K, y_K))$.

The boarding sequence can be partitioned into subsequences

$$(x_1, ..., x_{m_1}), (x_{m_1+1}, ..., x_{m_2}), ..., (x_{m_{n-1}+1}, ..., x_{m_n})$$

such that $x_{m_i+1} > ... > x_{m_{i+1}}$ and $x_{m_i} \leq x_{m_i+1} \forall i$, in other words a minimal decomposition into strictly decreasing subsequences by row number [1].

Passengers belonging to a same subsequence will never interfere with each other, but consecutive subsequences will always interfere, and so the number of such subsequences $N_{\leq}$ 1 will be the number of aisle interferences. It may however happen that two aisle interferences occur simultaneously, in which case they should not be counted twice. Given a boarding sequence, two aisle interferences will occur simultaneously if $|x_{m_i} - x_{m_j}| = m_i - m_j$, meaning that although x_{m_i} and x_{m_j} do not belong to the same subsequence, they can still be seating at the same time. The number $N_=$ must be subtracted form the number of aisle interferences.

Another source of aisle interference is if the whole aisle is full, which occurs if there exists a strictly decreasing subsequence of $(x_1, ..., x_K)$ of length X. Let the number of distinct such subsequences be N_X. Thus $N_A = N_{\leq} - 1 - N_= + N_X$ is the total number of aisle interferences.

A seat interference occurs if for a given row, the passenger closest to the aisle gets seated before another further from the aisle. That is, if for a given (x, y), the rank of $(x, y + 1)$ in the boarding sequence is greater than that of (x, y) if $y > 0$ and smaller than that of (x, y) if $y < 0$. Total number of times this occurs is N_S.

Bachmat et al. use a very similar method to count interferences [1], with the difference that they assume passengers to be infinitely thin, while this method attempts to account for that factor.

Calculating boarding time

The best boarding time is achieved by calling the passengers furthest from the aisle on one side in descending order, then the passengers furthest from the aisle on the other side and then similarly for columns closer and closer to the aisle [2]. In that case, the total boarding time will be merely the time it takes Y groups of passengers to walk the distance $X\Delta x$ to their seats at speed v (so $YXv\Delta x$) plus X times the time t_A it takes for a passenger to put luggage away and sit down. If however the passengers are called in a different order, there will be seat interferences and N_A aisle interferences instead of just X, and these will cause time delay of $t_A(N_A + 1) + t_s N_S$. Thus the total boarding time will be:

$$T = YXv\Delta x + (N_A + 1)t_A + N_S t_S$$

Algorithm

Given a boarding sequence, the numbers N_A and N_S can be computed as described above by a computer program. First, $N_A = N_S = 0$. The list is checked in order for strictly decreasing sequences of consecutive x-component (add to N_A), for strictly decreasing list (of length greater than Y) of consecutive x-component (add to N_A) and also for simultaneous aisle interferences. If it finds simultaneous aisle interferences, it removes the double counting by subtracting one to N_A. Further, both people interfering are checked for seat interferences - if it is happening, the minimum of the two number of seat interferences is removed to N_S. Hence, N_A and N_S have been calculated.

Model B

This model is based on a computer algorithm made in MatLab. (1) At the beginning seats are grouped in region (A, B, ...) listed according to the Boarding Policy . (2) Then, each person (numbered p_1, p_2, ..., p_n according to their position in the line) receives a random seat in the first group still having available places. Now, p_1 is loaded on the starting point x_0. (3) p_1 will now check if the row of his assigned seat corresponds to its current row. If it is, he will receive delay points (the number of iteration corresponding to t_A and $m \cdot t_S$, where m is the number of people seating on the same side of the row and that are closer to the aisle than the seat of p_1) and will stay at his current location. If it is not, he will move to $x_0 + 1$ and p_2 will now appear on the now free starting point x_0. (4) Owners of non-zero delay points lose one. Those for whom their delay points just fall back to zero will seat down to their seat. (5) Step 3 is repeated for p_i up to p_n, but will stay at their position if their number of delay points are non-zero. Step 4 is executed. (6) One is added to the time and steps 3 through 5 are repeated. The algorithm terminates when everyone is seated.

Boarding Policies

Random Policy. Although passengers have an assigned seat, the order in which they enter the plane is completely unregulated by the airline.

Back-to-Front Policy. The passengers are called onto the plane by blocks, starting from the back and proceeding to the front.

Rotating Zones Policy. As in the back-to-front policy, the plane is divided in zones, but here a zone closer to the front of the plane boards before the one at the very back has completely finished, so that the usage time of the aisle is not wasted. For longer planes, more than two zones could be boarded at the same time.

3	3	3	3	3	2	2	2	2	2	1	1	1	1	1	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
3	3	3	3	3	2	2	2	2	2	1	1	1	1	1	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
3	3	3	3	3	2	2	2	2	2	1	1	1	1	1	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
3	3	3	3	3	2	2	2	2	2	1	1	1	1	1	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
3	3	3	3	3	2	2	2	2	2	1	1	1	1	1	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
3	3	3	3	3	2	2	2	2	2	1	1	1	1	1	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1

Outside-in Policy. With this method, passengers with seats furthest from the aisle are called in first, followed by the passengers with seats closer to the aisle, regardless of the row.

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	3	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Outside-in and Back-to-Front Policy. The idea is very similar to the outside-in policy, but for a given column, passengers are called in by blocks.

4	4	4	4	4	4	4	4	4	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	5	5	5	5	5	5	5	5	5	5	5	5	5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
6	6	6	6	6	6	6	6	6	6	6	6	6	6	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
6	6	6	6	6	6	6	6	6	6	6	6	6	6	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
5	5	5	5	5	5	5	5	5	5	5	5	5	5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
4	4	4	4	4	4	4	4	4	4	4	4	4	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Reverse Pyramid. The method is similar to the previous one, except that the passengers in a given block furthest from the aisle board right after the passengers in the previous block, in a column closer to the aisle.

3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1
4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2
5	5	5	5	5	5	5	5	5	5	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3
5	5	5	5	5	5	5	5	5	5	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	1	1	1	1	1	1	1	1	1	1	1

Results

Model A for Back-to-Front Policy over 5 tryouts for 120 people in a plane of 30 rows with 2 seats on each side predicts an average boarding time of 817 seconds, with a standard deviation of 15%.

For the same plane configuration and policy, Model B predicts only 520 seconds, which suggests that at least one of the models is not perfect. In fact, if they are compared to actual data [6], they both underestimate boarding time. For comparison purposes however, this is not very important.

Model B has been used to predict boarding times for all the above-mentionned strategies on three different plane configurations. Taking the average over 10 trials suggests that the best boarding policies in all cases are the outside-in and back-to-front and the reverse-pyramid policies, with an insignificant difference between them.

Conclusion

The proposed models provide a reasonable representation of plane boarding scenarios. The results obtained allow a strong analysis of various boarding procedures, however much further investigation is recommended. The algorithms are easily adaptable to incorporate human random and chaotic behaviour. Though more structured loading policies, such as reverse-pyramid and block boarding obtain shorter boarding times under the proposed conditions, it is hypothesized that these times will be significantly affected if human behaviour is modeled more accurately.

The authors would like to thank professor Nilima Nigam for her unconditional encouragement throughout the Modeling Contest in Mathematics and her help in the preparation thereof.

References

[1] E. Bachmat, D. Berend, L. Sapir, and S. Skiena. Optimal boarding policies for thin passengers. 2006. Working paper.

[2] E. Bachmat, D. Berend, L. Sapir, S. Skiena, and N. Stolyarov. Analysis of airplane boarding times. 2006. Working paper.

[3] S. Marelli, G. Mattocks, and R. Merry. The role of computer simulation in reducing airplane turn time. *Aero Magazine*. Issue 01, Boeing.

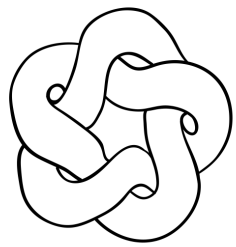
[4] M. Pan. Efficient boarding procedures for midsized passenger aircraft. 2004.

[5] M.H.L. Van Den Briel, J.R. Villalobos, and G.L. Hogg. The aircraft boarding problem. 2003. In Proceedings of the 12th Industrial Engineering Research Conference (IERC-2003), Nr 2153, CD-ROM.

[6] H. van Landeghem and A. Beuselinck. Reducing passenger boarding times in airplanes: A simulation based approach. *European Journal of Operations Research*, 142, 2002. No 2.

Jokes

Prove the four colors theorem in this particular example of 12 regions.



□

Spectrum and Expansion of Biregular Graphs

Rosalie Bélanger-Rioux and Ioan Filip

Graphs with a strong expansion property are extremely useful in many areas of mathematics and computer science, particularly in the design of efficient algorithms. It is, however, very difficult to explicitly construct infinite families of good expanders. In this paper, we study the spectrum of biregular graphs and show how it is related to their expansion coefficient. We also describe a construction of biregular expanders from elliptic curves. Finally, we present some experimental results on the second largest eigenvalues of biregular graphs with degrees 2 and 7.

Definition and preliminaries

A *graph* $G = (V, E)$ consists of a set V of vertices and a set $E \subseteq V \times V$ of edges between the elements in V. We say that a graph G is *bipartite* and we write $G = (L, R, E)$ if V can be partitioned into two subsets L and R (called "left" and "right" vertices) such that the edges in E are from elements of L to elements of R. Finally, we call a bipartite graph a *biregular graph* if every left vertex has degree d_L and every right vertex has degree d_R (by degree of $v \in V$ we mean the number of edges in E incident to v). This implies that $d_L |L| = d_R |R|$.

An important property of graphs is the notion of *expansion*: how much larger than $S \subset V$ can its set of neighbors be. Graphs with a high expansion coefficient are extremely useful in computer science for the design of efficient algorithms, for combinatorial optimization and for constructing error-correcting codes, but they are also important in other fields like statistical physics.

Using a probabilistic approach, it is not difficult to prove the existence of infinitely many graphs with good expanding properties. However, explicit constructions of families of good expanders are hard to find and only appeared in the late 80's (see [6]).

Our main interest is constructing biregular graphs with high expansion. The motivation for this particular class of graphs comes from applications where the structures involved consist of two types of fundamentally different nodes. For instance, when analyzing complex networks such as the Internet, one can distinguish between users and servers, viewed as vertices on a large graph. In this paper, we make rigorous the notion of expansion for a biregular graph G and present some theoretical as well as experimental results relating expansion to the spectrum of G.

Definition 1. An $(L, R, d_L, d_R, c) - expander$ is a bipartite graph on the sets of left vertices L and right vertices R, where the maximal degree of a left vertex is d_L and the maximal degree of a right vertex is d_R, such that $d_L \leq d_R$ and for every $X \subset L$,

$$|\partial(X)| \geq \left(\frac{d_L}{d_R} + c\left(1 - \frac{|X|}{|L|}\right)\right)|X|. \tag{1}$$

Here, $\partial(X) = \{r \in R \mid \exists l \in X \text{ with } (l, r) \in E\}$, denotes the set of neighbors of X.

Let $G = (L, R, E)$ be a connected bipartite graph with the sets $L = \{l_1, \ldots, l_{|L|}\}$ of left vertices and $R = \{r_1, \ldots, r_{|R|}\}$ of right vertices. (By connected we mean that for all x, y in G, there exists a path in G from x to y.) We define the incidence matrix M_G of G as follows

$$m^G_{i,j} = \begin{cases} 1 & (l_i, r_j) \in E \\ 0 & \text{otherwise} \end{cases}.$$

Next, we define the adjacency matrix A_G (note it is real and symmetric, so diagonalisable) such that

$$a^G_{i,j} = \begin{cases} 1 & (v_i, v_j) \in E \\ 0 & otherwise \end{cases}.$$

where $v_k = l_k$ for $1 \leq k \leq |L|$ and $v_{k'+|L|} = r_{k'}$ for $1 \leq k' \leq |R|$. In fact,

$$A_G = \begin{bmatrix} 0 & M \\ M^t & 0 \end{bmatrix}.$$

The eigenvalues of A_G are called the spectrum of G. It is easy to prove that the spectrum of G, a connected biregular graph, is

$$\sqrt{d_L d_R} = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n = -\sqrt{d_L d_R},$$

with $\lambda \neq 0$ being an eigenvalue if and only if $-\lambda$ is an eigenvalue, if and only if λ is an eigenvalue of both M and M^T. Note also that

$$A^2_G = \begin{bmatrix} MM^T & 0 \\ 0 & M^T M \end{bmatrix}, \tag{2}$$

so the ij^{th} entry of A^2_G is the total number of paths of length 2 between v_i and v_j, with any row or column sum equal to $d_L d_R$. It then follows that $\pm\lambda$ is an eigenvalue of A_G if and only if λ^2 is an eigenvalue of A^2_G. Moreover, $\lambda \neq 0$ is an eigenvalue of A^2_G if and only if it is an eigenvalue of both MM^T and $M^T M$. Finally, we know that the eigenvector associated both with the largest eigenvalue of A_G and A^2_G is $e_1 = \frac{(1,\ldots,1)}{\sqrt{|L|+|R|}}$. The other vectors of the orthonormal eigenbasis of A^2_G will be called $e_2, \ldots, e_{|L|+|R|}$. We are now ready to prove the following theorem:

Theorem 1. *Let $G = (L, R, E)$ be a connected biregular graph as defined above. Let $\lambda_2 = \lambda_2(A_G)$. Then G is an (L, R, d_L, d_R, c)-expander, with $c = \frac{d_L d_R - \lambda_2^2}{d_R^2}$.*

Proof. (See [2]) Let $A = A(G)$, and let X be a subset of the left vertices of G such that $|X| \leq \alpha|L|$, with x its characteristic vector (so the i^{th} coordinate of x's is 1 if $v_i \in X$, else it is 0). We have that $xx^T = \|x\|^2 = |X|$. We shall bound $\frac{|\partial(X)|}{|X|}$ from below. Let $C = xA$. We know that $f(x) = x^2$ is a convex function, and so for any real numbers a_i, $i = 1, \ldots, k$ we have

$$\left(\sum_{i=1}^{k} a_i^2\right) \geq \frac{1}{k}\left(\sum_{i=1}^{k} a_i\right)^2.$$

Further, since there are precisely $|\partial(X)|$ coordinates C_j of C that are not zero, we have:

$$\|C\|^2 = \sum_{j=1}^{|L|+|R|} C_j^2 \geq \frac{\left(\sum_{j=1}^{|L|+|R|} C_j\right)^2}{|\partial(X)|} = \frac{(|X|\, d_L)^2}{|\partial(X)|},$$

and so

$$\frac{|\partial(X)|}{|X|} \geq \frac{|X|\, d_L^2}{\|C\|^2}. \tag{3}$$

Remains to be found the required upper bound for $\|C\|^2$. First, we expand x in terms of our orthonormal eigenbasis so that

$$AA^T x^T = \sum_{i=1}^{|L|+|R|} \lambda_i^2 \gamma_i e_i.$$

Then,

$$\|C\|^2 = \|xA\|^2 = \langle xA, xA\rangle = xAA^T x^T = \sum_{i=1}^{|L|+|R|} \lambda_i^2 \gamma_i^2.$$

Now, we know that,

$$\gamma_1 = \langle x, e_1\rangle = xe_1^T = |X|/\sqrt{|L|+|R|},$$

so, since $\lambda_2 \geq \ldots \geq \lambda_n$,

$$\begin{aligned}\|C\|^2 &\leq \lambda_1^2\gamma_1^2 + \lambda_2^2 \sum_{i=2}^{|L|+|R|} \gamma_i^2 \\ &= \gamma_1^2\left(\lambda_1^2 - \lambda_2^2\right) + \lambda_2^2 \|x\|^2 \\ &= \frac{|X|^2}{|L|+|R|}\left(d_L d_R - \lambda_2^2\right) + \lambda_2^2 |X|.\end{aligned}$$

Thus from 3 we obtain the following:

$$\frac{|\partial(X)|}{|X|} \geq \frac{d_L^2}{\frac{|X|}{|L|+|R|}\left(d_L d_R - \lambda_2^2\right) + \lambda_2^2}$$

or

$$\partial(X) \geq |X| \frac{d_L^2}{\alpha\left(d_L d_R - \lambda_2^2\right) + \lambda_2^2}$$

for all $X \subseteq L, |X| \leq \alpha|L|$. Thus we have found a lower bound for the expansion of subsets of the left vertices of any connected biregular graph. Putting $\alpha = 1$ and rearranging (see [1] for the case $|L| = |R|$) to obtain the form required by Definition 1, we get:

$$|\partial(X)| \geq \left(\frac{d_L}{d_R} + \left(\frac{d_L d_R - \lambda_2^2}{d_R^2}\right)\left(1 - \frac{|X|}{|L|}\right)\right) \cdot |X|.$$

□

Ramanujan graphs : definition and examples

We remind the reader that we are in fact interested in explicit constructions of biregular graphs that are good expanders for any left size. Since a smaller second eigenvalue implies a larger expansion, we would like to study the asymptotic behavior of the second largest eigenvalue of a bipartite graph's adjacency matrix as the left size of the graph grows. The following theorem is a generalisation of the Alon-Boppana bound for the eigenvalues of regular graphs.

Theorem. *Among the biregular graphs $G = (L, R, E)$ we have*

$$\liminf_{|L|\to\infty} \lambda_2(G) \geq \sqrt{d_L - 1} + \sqrt{d_R - 1}.$$

This motivates in part the definition of *Ramanujan* biregular graphs, graphs with their second largest eigenvalue as small as possible in the asymptotic sense.

Definition 2. A biregular graph $G = (L, R, E)$ is *Ramanujan* if any non-trivial eigenvalue λ of A_G satisfies

$$|\sqrt{d_L - 1} - \sqrt{d_R - 1}| \leq |\lambda| \leq \sqrt{d_L - 1} + \sqrt{d_R - 1}.$$

We continue with two interesting examples of biregular graphs. The first example is a specific infinite class of biregular graphs of left degree 2 and the second is based on a construction using projective curves.

Example 1: Ramanujan graphs of left degree 2

Let $n, k \in \mathbb{N}$ with $n \geq k \geq 1$. Construct the graph $G = (L, R, E; n, k)$ as follows.

Let $L = \{1, 2, \ldots, n\}$ and R be the set whose elements are the subsets of L of size k. Thus $|L| = n$ and $|R| = \binom{n}{k}$. An element $x \in L$ is connected to an element $S \in R$ if and only if $x \in S$. The graph we obtain is biregular with $d_L = \binom{n-1}{k-1}$ and $d_R = k$.

Let M be the incidence matrix of G. Given $i, j \in L$, there are precisely

$$\#\{S \in R : i \in S, j \in S\} = \binom{n-2}{k-2}$$

subsets of L of size k containing both i and j, so for $i \neq j$, this is also the number of walks of length 2 from i to j.

Therefore, we have that

$$MM^t = \begin{bmatrix} \binom{n-1}{k-1} & \binom{n-2}{k-2} & \cdots & \binom{n-2}{k-2} \\ \binom{n-2}{k-2} & \binom{n-1}{k-1} & \cdots & \binom{n-2}{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \binom{n-2}{k-2} & \binom{n-2}{k-2} & \cdots & \binom{n-1}{k-1} \end{bmatrix}$$
$$= \binom{n-2}{k-1} I_n + \binom{n-2}{k-2} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix}.$$

It is easy to see that the second eigenvalue of MM^t is $\binom{n-2}{k-1}$ and that its spectrum is

$$\mathrm{Spec}(MM^t) = \left\{ k\binom{n-1}{k-1}, \binom{n-2}{k-1} \right\}.$$

Consequently, by definition 2, our graph $G = (L, R, E; n, k)$ is Ramanujan if and only if

$$\sqrt{\binom{n-1}{k-1} - 1} - \sqrt{k-1} \leq \sqrt{\binom{n-2}{k-1}} \leq \sqrt{\binom{n-1}{k-1} - 1} + \sqrt{k-1}.$$

Observe that for $k = 2$, these inequalities are trivial and thus the graphs $G = (L, R, E; n, 2)$ are Ramanujan. Moreover, note that the right inequality holds for any n, k with $n \geq k \geq 1$ and so by theorem 1, these graphs are relatively good expanders.

Example 2: Biregular graphs from projective curves

Let $\mathbb{F}$ be the field with q elements and let $\mathbb{P}^2(\mathbb{F})$ be the projective plane over $\mathbb{F}$. The general linear group $GL_{n+1}(\mathbb{F})$ acts on $\mathbb{A}^3 \backslash \{0\}$ by multiplication by a matrix $M \in GL_{n+1}(\mathbb{F})$ and $PGL_3(\mathbb{F}) = GL_3(\mathbb{F})/\mathbb{F}^\times$ acts similarly on $\mathbb{P}^2$.

Let $f(x, y, z)$ be a homogeneous polynomial of degree $d = 3$ and observe that the equation $f(x : y : z) = 0$ is well defined. Denote by $Z_f(\mathbb{F})$ be the hypersurface of degree 3 and dimension 1 given by

$$Z_f(\mathbb{F}) = \{x \in \mathbb{P}^2(\mathbb{F}) \mid f(x) = 0\}.$$

We further assume that $q > 16$ and that $f(x, y, z)$ is irreducible and non-singular on $Z_f(\mathbb{F})$, that is, on Z_f,

$$\left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right) \neq (0, 0, 0).$$

Now for every $M \in PGL_3(\mathbb{F})$, define ${}^M f(x) = f(M^{-1}x)$. Note that ${}^M f$ is still a homogeneous polynomial of degree d and $f(x) = 0$ if and only if ${}^M f(Mx) = 0$. Therefore $Z_{{}^M f}(\mathbb{F}) = M \cdot Z_f(\mathbb{F})$. For simplicity, we write $S(M)$ for $Z_{{}^M f}(\mathbb{F})$. Clearly, $S(M)$ depends only on an $M \in PGL_3(\mathbb{F})$:

$$S(M) = M \cdot S(I) = \{Mv \mid v \in S(I)\}.$$

From our assumptions and by the theorems of Hasse-Weil and Bézout, we can conclude that $S(M) = S(I)$ if and only if ${}^M f = \text{constant} \cdot f$. We write $g \sim f$ for $g = \text{constant} \cdot f$.

We can now construct the bipartite graph $G = (L, R, E)$ as follows. The left vertices are in fact the sets $S(M)$,

$$L = \{S(M) \mid M \in PGL_3(\mathbb{F})\}.$$

Note that $S(M) = S(N)$ if and only if $N^{-1}M \cdot S(I) = S(I)$, if and only if ${}^{N^{-1}M} f \sim f$, and that PGL_3 acts on L by $M_1 \cdot S(M_2) = S(M_1 M_2)$. The right vertices are the points in the projective plane, so $R = \mathbb{P}^2(\mathbb{F})$. We put an edge $(r, l) \in E$ where $r \in R$ and $l = S(M) \in L$ if $r \in S(M)$, which is equivalent to the condition that $M^{-1}r \in S(I)$.

Clearly we have that $|R| = q^2 + q + 1$. A computation shows that the graph is biregular, with $d_L = \#Z_f(\mathbb{F})$. Also, we must have that

$$|L| = \frac{\#PGL_3(\mathbb{F})}{\#\mathrm{Stab}(S(I))},$$

so in order to find d_R we can compute the size of the stabilizer of $S(I)$, namely $\#\mathrm{Stab}(S(I))$. Because f is not linear (as $d > 1$), we have

$$\{M \in PGL_3(\mathbb{F}) \mid^M f \sim f\} \hookrightarrow Aut(Z_f(\overline{\mathbb{F}})),$$

where $\overline{\mathbb{F}}$ is an algebraic closure of $\mathbb{F}$. Applying once again Bézout's theorem we obtain

$$Mv = v, \forall v \in Z_f(\mathbb{F}) \Rightarrow Mv = v, \forall v \in Z_f(\overline{\mathbb{F}}),$$

such that if $v \mapsto Mv$ is an automorphism of $Z_f(\mathbb{F})$, then M must be a scalar matrix.

Let $E[3](\mathbb{F})$ be the set of points of order 3 on E and assume the j-invariant for E, $j(E)$, is neither 0 nor 1728. Observing that if M preserves E, then M must permute the points of order 3 on E, we can conclude that

$$\{M \in PGL_3 \mid^M f \sim f\} \hookrightarrow \langle Id, i \rangle \cdot E[3](\mathbb{F}),$$

where i is the hyperelliptic involution and Id is the trivial automorphism. It is well known that $|E[3](\mathbb{F})| = 1, 3$ or 9. So if $3 \nmid N = q + 1 + err$, and taking $err = 0$, $q \equiv 1 \bmod 3$, we have a simple characterization of $\{M \in PGL_3(\mathbb{F}) : {}^M f \sim f\}$ because $E[3](\mathbb{F})$ is trivial.

Now for $r_1 \neq r_2 \in R$ and given $r_3 \neq r_4$, there exists a matrix M with $Mr_1 = r_3$ and $Mr_2 = r_4$. Consequently, we have that the number of common neighbors of r_1 and r_2, which is

$$u = |\{S(M) \mid M^{-1}r_1 \in S(I)\} \cap \{S(M) \mid M^{-1}r_2 \in S(I)\}|,$$

is independent of r_1, r_2.

Consider the matrix $D = MM^t$ with entries $(d)_{r_i r_j}$ = the number of walks of length 2 in the graph $G = (L, R, E)$ between vertices r_i and r_j. Observe that

$$D = \begin{bmatrix} d_R & u & \dots & u \\ u & d_R & & \vdots \\ \vdots & & \ddots & u \\ u & \dots & u & d_R \end{bmatrix}$$
$$= (d_R - u)I + u \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix},$$

has spectrum

$$\text{Spec}(D) = d_R - u + u \begin{cases} |R|, & \text{with multiplicity 1;} \\ 0, & \text{with multip. } |R| - 1. \end{cases} \quad (4)$$

Hence, the second largest eigenvalue of A_G is $\lambda_2(A_G) = \sqrt{d_R - u}$ and according to the definition of biregular Ramanujan graphs (see definition 2), $G = (L, R, E)$ is Ramanujan if and only if

$$\sqrt{d_R - 1} - \sqrt{d_L - 1} \le \sqrt{d_R - u} \le \sqrt{d_R - 1} + \sqrt{d_L - 1},$$

which is equivalent to

$$d_R + d_L - 2 - 2\sqrt{(d_R - 1)(d_L - 1)} \le \frac{d_R(|R| - d_L)}{|R| - 1} \le d_R + d_L - 2 + 2\sqrt{(d_R - 1)(d_L - 1)}.$$

The second inequality always holds; the first does not for large values of q because

$$|L| = \frac{|PGL_3(\mathbb{F})|}{\text{Stab}(S(I))} = \frac{q^2(q^3 - 1)(q^3 - q)}{\text{Stab}(S(I))} = \Theta(q^8),$$

and so

$$d_R = \frac{|L| d_L}{|R|} = \Omega(q^6).$$

It can be verified in fact that asymptotically (as $q \to \infty$), the reverse inequality holds. Hence the graphs $G = (L, R, E)$ described above are not Ramanujan as $|\mathbb{F}| \to \infty$. Expansion, however, is not reduced with the graphs not being Ramanujan (cf. theorem 1).

Experimental results

To conclude our article, let us present the computational results from our experiments on the distribution of the eigenvalues of biregular graphs. We constructed random simple connected biregular graphs, using the following idea (see [4]). First, build an array A of size $d_L|L|$, with its first d_R cells containing r_1, the next d_R cells containing r_2, etc. Then, permute the cells of A by a random permutation in $S_{d_L|L|}$ to get the array A', which defines a bipartite (multi)graph G, where the neighbors of l_1 are the first d_L entries of A', and so on. If G is a multigraph, return to array A, take another permutation and check if it defines a simple connected bipartite graph.

We found that, among 159 random simple connected biregular graphs of degrees 2 and 7, with $L = 1001$, about 79 of those were Ramanujan, while among 518 graphs of left size 39998, about 71 were Ramanujan. In fact, as is expected, this percentage tends to slowly decrease as L grows larger (see Fig. 1).

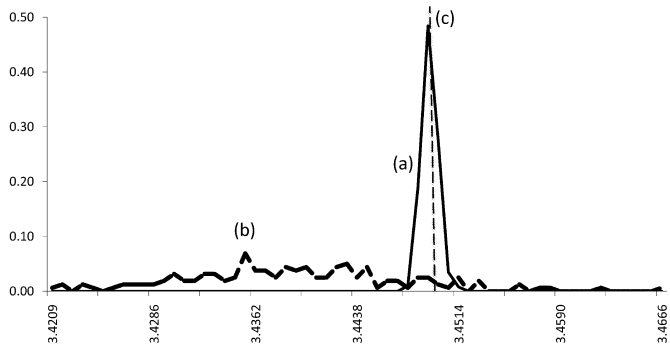


Figure 1: The distribution of the second largest eigenvalue in random biregular graphs of left degree 2, right degree 7 with left size 39998 (curve (a)) and left size 1001 (curve (b)). The vertical dotted line (c) indicates the Ramanujan bound.

Acknowledgments

We thank Professor Eyal Goren for his guidance and support, which were crucial to our progress on this project. We are grateful to him for introducing us to this topic. We also thank Professor Nilima Nigam for her help with Matlab.

References

[1] Alon, Noga. "Eigenvalues and Expanders", *Combinatorica* , Vol. 6, 1986, p. 83-96.

[2] Tanner, R. Michael. "Explicit concentrators from generalized N-gons", *SIAM J. Alg. Disc. Meth.*, Vol. 5, No. 3, September 1984, p. 287-293.

[3] Hoory, S. Linial, N. Wigderson, A. "Expander graphs and their applications." *Bulletin of the American Mathematical Society*, Vol. 43, No. 4, 2006, pp. 438-562.

[4] Jakobson, D. Miller, S. D. Rivin, I. Rudnick, Z. "Eigenvalue Spacings For Regular Graphs", *Emerging Applications of Number Theory*, Minneapolis, Vol. 109, 1996.

[5] Solé, P. "Ramanujan Hypergraphs and Ramanujan Geometries." *Institute for Mathematics and Its Applications*, Vol. 109, 1999, pp. 583-590.

[6] Lubotzky, A. Phillips, R. and Sarnak, P. "Ramanujan Graphs", *Combinatorica*, Vol. 8, 1988, pp. 261-277.

Partially Observable Markov Decision Processes

Yang Li

In the field of artificial intelligence, many people are interested in finding new algorithms that enable an agent to act intelligently in a world. Planning how to act in a stochastic world is a major problem in the field. An intelligent agent must usually rely on an imperfect model of the world to plan its actions. To improve the model used, the agent can learn a better model through experience; this makes learning another important problem in the field. Markov Decision Processes (MDPs) and Partially Observable Markov Decision Processes (POMDPs) are widely studied mathematical models for these problems, originating from operations research [SS73]. POMDPs are more expressive than MDPs because they model both partial observability (which can result, for example, from having imperfect sensors) and probabilistic transitions (which are the result of the environment being stochastic). Unfortunately POMDPs are much harder to learn and to use for planning. In this paper we present a new algorithm for learning POMDPs from data. The main idea is to work with histories of actions and observations. We show that the new representation can be used successfully to plan a good behavior.

Background

MDPs are formally represented as a 4-tuple $(A, S, P(\cdot), R(\cdot))$, where A is the action space, S is the state space, P is the transition function $P : S \times A \times S \to [0, 1]$ that gives the probability that an action a will take the agent from state s to state s' and R is the reward function $R : S \to \mathbb{R}$ which gives the agent's immediate reward after moving to state s [SB98]. On every time step t, the agent will observe the state of its environment, s_t and pick an action a_t. This will cause the environment to transition to a new state s_{t+1} (determined by P) and the agent will receive a reward r_{t+1} determined by R. The behavior of an agent can be described by a policy $\pi : S \to A$ that specifies what action should be taken in each state. The value function $V^\pi : S \to \mathbb{R}$ expresses how good it is to be in a certain state s if policy π will be used to choose actions. It is defined as the expected value of the sum of rewards that will be obtained when starting in s. In an MDP with a finite state and action set there is a unique optimal value function, V^*, and a corresponding optimal policy π^*.

For MDPs with continuous or large discrete state space, it is often handy to use state aggregation as it can reduce the state space to a constant number of aggregates G [SB98]. An aggregate G is a set of states combined together to have a single value, and thus optimal action. Clearly, given an optimal policy π^* over a non-aggregated state space and an optimal policy π'^* over an aggregated state space, the expected value $V^{\pi^*}(s) \geq V^{\pi'^*}(s)\ \forall s \in S$. This is because, if an aggregate contains states with a different optimal action under π^*, at least one of these states will be forced to take a non-optimal action under π'^*.

This introduces the idea of splitting the state aggregates into smaller state aggregates. Since splitting will never decrease the performance of the optimal policy, by splitting aggregates we can improve the optimal policy. However, we do not want to split aggregates in which the optimal policy is the same after splitting. In other words, we hope to split aggregates in such a way that the expected return is significantly improved.

A POMDP is simply an MDP, but instead of being able to observe the states, the agent is only able to observe some features related to the state, according to some probability distribution. Since the agent may observe certain features of a state, but not enough to distinguish all states from one another [KLC98], one set of features may match several different states. This induces state aliasing over observation features. Thus, we would like to represent each state as an unique history of observations and actions, or at least, separate the aliased states to maximize the value function.

For this to be possible, we assume a weaker version of the Markovian property. Instead of assuming the Markovian property that states that future observations and expected reward depend only on the present observation and not on any past observations [SB98], we assume that the future observations and expected reward depend only on the past n observations (and actions) and not on any anterior ones.

Under this assumption, which is much more reasonable than the original one, we built an algorithm that enhances a history by one unit of experience each time it thinks it is useful for maximizing the expected return.

Problem and Approach

For many problems, we would like our agents to be able to improve their policy and internal world representation in an online fashion, as they observe new data. We expect the agent to be able to observe features in the past that can help it guess in which state it is at the present time, and thus determine which action it should undertake next.

Updating the Variable Length Histories Table (VLHT)

A history is a sequence of action-observation pairs. Our algorithm will maintain a table of histories, such that each entry corresponds to a distinct state as perceived by the agent. The agent will decide what histories should be included in the table based on the data it observes. Each entry in the VLHT will have an associated value (which is the estimate of expected return if the agent acts according to policy π from that point on, having that history) and a corresponding action. In our approach, the agent will gather data in episodes, where each episode consists of at most m steps. After m steps or after reaching a goal state, the episode ends and the return information for all the histories in the VLHT which have been experienced during the episode are updated.

Enhancing histories

Enhancing histories is done by adding more action-observation pairs (thus making it longer). Deciding which history to enhance is the trickiest part of the problem. We set a threshold on the variance of the returns following a history to determine whether the history is representative of a single state (or a group of states which have the same optimal action), or if it may correspond to different states. This technique is simple, but it is very efficient and works well enough for our purpose. To do this, the agent remembers the different observations after each of its histories and keeps the sum and squared sum of the returns that have been obtained so far for each history, in order to compute the variance quickly and whenever necessary.

Our agent is expected to be able to learn the length of the history necessary to distinguish aliased states which impair the performance of the agent. The length of the history will vary depending on the world and the features observable by the agent. The agent should stop adding experience units to any history after the optimal policy can be represented.

Algorithm 1 Variable Length History Algorithm

```
VLHT ← new table, {}
while conditions C not met do
  Learn(randomAction, m, parameters)
end while
```

The VLH-algorithm

As seen in Algorithm 1 and 2, the VLH-algorithm is recursive and tells the agent to stop when the number of steps remaining decreases to 0 or when the goal is reached.

In the recursion step of the algorithm, the agent, when at state s, uses its sensors to gather observation o and undertakes an action a After ending in state s', it will, again, use its sensors to observe o' and decide which action to take next. Actions are chosen in an ϵ-greedy fashion according to the VLH table; that is, the best action as predicted by the table is chosen with probability $(1-\epsilon)$ and a uniformly random action is chosen with probability ϵ.

Each entry of the VLH table is defined recursively: (o, a) is a valid entry and $e(o, a)$ is a valid entry if e is a valid entry. Thus, each entry of the table will be a sequence of experience units, (o, a), of different lengths. Since the history of the actions of an agent is also a sequence of experience units, by comparing the value of each sequence $e(o_t, a_1), e(o_t, a_2), \cdots, e(o_t, a_n)$, we know which action a_i to choose when we observe o_t and have history e.

Algorithm 2 Recursive Function of the VLH-algorithm

```
Learn(action, m, parameters)
  o, a, s, r ← observation, action, state, reward
  s' ← state after undertaking action a in s
  a' ← action selected in a randomized way.
  o', r' ← observation at state s', reward gained
  h ← longest history in VLHT matching agent's
  if agent in goal state then
    s ← random state
    return reward
  else if m = 0 then
    return reward
  end if
  MCreturn ← r + γLearn(a', m − 1, parameters)
  if (o, a) in VLHT then
    error ← r + γMCreturn − VLHT[h]
    Update VLHT[h] (∑MCret, ∑MCret²)
    for all sub-histories(*) sh of h do
      VLHT[sh] ← VLHT[sh] +α error
    end for
  else
    VLHT[(o, a)] ← MCreturn
  end if
  if h has unusual outcomes then
    h' ← h+unit < o, a > followed by h
    VLHT[h'], ET[h'] ← MCreturn, 0
  end if
```

(*) sub-histories of h which ends with the same (o, a) tuple as h.

An important property to notice is that the shorter length histories will never be made obsolete by longer ones, and are quite useful, due to the fact that the agent does not always have a history which matches an entry of the table. Furthermore, whenever the agent starts fresh, or thinks that its history is inaccurate due to noise, the agent may purposefully delete its history and start collecting experience units from scratch; this will require knowledge about the short histories.

When an episode is over, the agent will update its table and representation of the world. Updating the table is straightforward as we update the history and all sub-histories equally by $r+\gamma$MCreturn $-$ VLHT$[h]$, a common error term. We also update the sum of the returns and the

sum of squared returns for each observation following for the corresponding VLHT table entry.

The last and key part of the algorithm remains. In order to determine which history to enhance, we use the sum and sum squared of the Monte Carlo returns and calculate a variance. Clearly, aliased states with different optimal policy or very different optimal returns, will have variance that is very high, and conversely, non-aliased states should have very similar returns, thus low variance.

Results and Analysis

To test the VLH-algorithm, we use a simple π shaped grid-world with 9 states (Figure 1). The agent is equipped with a compass and can detect in which direction it can move, thus {North, South, East, West} form the whole action space A and $o \subset A$ is an observation.

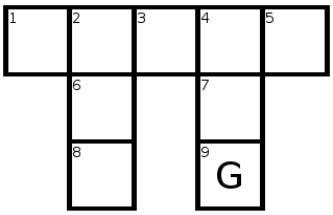


Figure 1: Grid-world (π shaped) with 3 pairs of aliased states (2 & 4, 6 & 7, 8 & 9)

Figure 2 shows the performance of the true, state-based optimal policy (horizontal top line) compared to the performance of policies found by our algorithm (averaged over 30 independent runs). As we can see, the average performance of the near-optimal policy was not significantly better than the one of the policies found by the VLH-algorithm. This is really encouraging, given that VLH constructs all its representation from observed data, without having any prior knowledge about the environment.

Discussion

The first version of the VLH-algorithm seems promising as it is simple and yet able to learn in environments with uncertainty. McCallum proposed a similar algorithm in his PhD thesis; however, the VLH- algorithm has a different approach for estimating which histories should be enhanced.

There are many adjustments which can improve the performance of the algorithm. The most important one is merging similar histories together. This would lower both the space complexity and the time complexity of the algorithm.

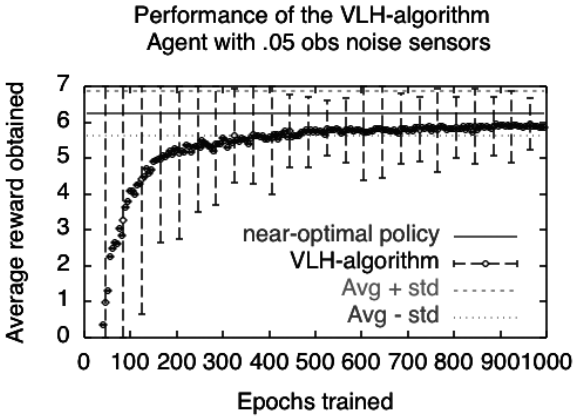


Figure 2: Performance graph

The project was realized during Summer 2007 and funded by a NSERC USRA supervised by Dr. Doina Precup.

References

[SS73] R. D. Smallwood and E. J. Sondik. The optimal control of partially observable Markov processes over a finite horizon. Operation research, 1973.

[M05] K. P. Murphy. A Survey of POMDP Solution Techniques. 2005.

[SB98] R. Sutton and A. Barto. Reinforcment Learning: An Introduction. MIT Press, 1998.

[McC95] A. McCallum. Reinforcement Learning with Selective Perception and Hidden State. PhD thesis, Univ. Rochester, 1995.

[KLC98] L. P. Kaelbling, M. Littman, and A. Cassandra. Planning and acting in partially observable stochastic domains. Artificial Intelligence, 101, 1998.

Jokes

Relations between pure and applied mathematicians are based on trust and understanding. Namely, pure mathematicians do not trust applied mathematicians, and applied mathematicians do not understand pure mathematicians. □

Fun Results in Algebraic Topology

Agnès F. Beaudry

Algebraic topology provides very elegant tools to prove fun results such as Brouwer's Fixed Point Theorem, which says that every continuous function from the disk to itself has a fixed point; or the Borsuk-Ulam theorem, which states that every continuous map from the sphere to $\mathbb{R}^2$ sends a pair of antipodal points to the same point in $\mathbb{R}^2$. I want to give here a taste of the machinery. To do this, I will provide an intuitive proof of Brouwer's Fixed Point theorem, which illustrates the basic ideas which are at the root of algebraic topology.

Introduction

Throughout, S^1 denotes a circle, D^2 a closed disk and $I = [0, 1]$ the closed unit interval. You can picture S^1 in $\mathbb{R}^2$ as the curve $x^2 + y^2 = 1$ and D^2 as the set $\{(x, y)|x^2 + y^2 \leq 1\}$.

My goal is to give an intuitive idea of how one proves the following theorem using algebraic topology. You might have seen it proved using more analytic tools in one of the undergraduate analysis classes. The point here is not so much the result (although it's a great result), but the tools used to prove it.

Theorem. Brouwer's Fixed Point Theorem *Every continuous map $\varphi : D^2 \to D^2$ has a fixed point, i.e., there is an $x \in D^2$ such that $\varphi(x) = x$.*

I will prove Brouwer's fixed point theorem by contradiction, showing that, if there is a map $\varphi : D^2 \to D^2$ which has no fix point, then we can continuously deform the disk D^2 onto the circle S^1. Some key topological properties of S^1 make this impossible: this is what we will be looking at. The proofs will not be rigorous, although everything we say can be made rigorous with more machinery.

Paths

We start with X, a topological space. Formally, a path from x to y on X is a continuous map $g : I \to X$ such that $g(0) = x$ and $g(1) = y$. You can picture this as a parametrized curve joining two points. An example is a straight line in $\mathbb{R}^2$ joining $\mathbf{x}$ and $\mathbf{y}$, which we can write as $l : I \to \mathbb{R}^2$

$$l(t) = t\mathbf{y} + (1 - t)\mathbf{x}.$$

At $t = 0$, the line is at $\mathbf{x}$, and at $t = 1$, it is at $\mathbf{y}$.

Another example is the arc of a circle. Say $\mathbf{x} = (1, 0)$ and $\mathbf{y} = (-1, 0)$. Then we can join $\mathbf{x}$ and $\mathbf{y}$ by an arc of the unit circle. We can parameterize this as $g : I \to \mathbb{R}^2$

$$g(t) = (\cos(\pi t), \sin(\pi t)).$$

Similarly we can join $\mathbf{x}$ and $\mathbf{y}$ by $h : I \to \mathbb{R}^2$

$$h(t) = (\cos(-\pi t), \sin(-\pi t)).$$

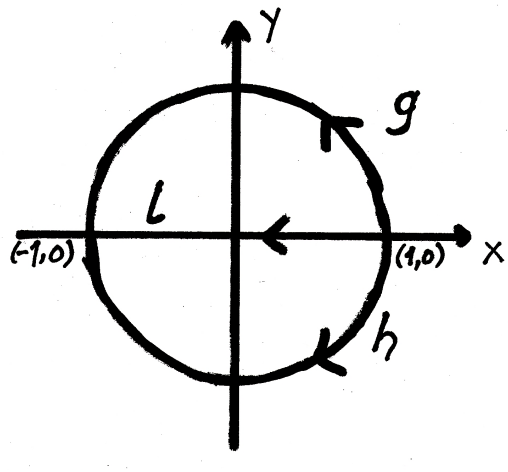


Figure 1: Some paths in $\mathbb{R}^2$.

With the line $l(t) = t\mathbf{y} + (1 - t)\mathbf{x}$, we now have three paths joining the points $(\pm 1, 0)$. In fact, there are uncountably many such paths. How different are they? Suppose we take a rope and pin one end to $(1, 0)$ and pin the other end to $(-1, 0)$. We would like to say that it does not matter where we set down the rope on the plane, how much we stretch it, or how we move it around (without lifting it): as long as we do not cut it, it remains the same rope joining the same two points.

We do the same thing for paths and say that two paths are equivalent, or *homotopic*, if we can continuously deform one path into the other, without cutting our curves. More precisely, two paths are homotopic if there is a family of intermediate curves continuously transforming one path into the other.

For a general topological space X we say the following: let $g_0 : I \to X$ and $g_1 : I \to X$ be continuous maps joining x and y, i.e., $g_i(0) = x$ and $g_i(1) = y$ for $i = 1, 2$. We say that g_0 and g_1 are *homotopic* if there exists a family of continuous maps $F(s, t) = f_s(t) : I \times I \to X$ such that $F(0, t) = f_0(t) = g_0(t)$, $F(1, t) = f_1(t) = g_1(t)$ with $F(s, 0) = x$ and $F(s, 1) = y$ for all $s \in I$. This means that the first curve is g_0, the last curve g_1, and each intermediate curve is a path joining x and y.

For example, letting $l(t)$ and $g(t)$ be the paths in $\mathbb{R}^2$ described above, then

$$F(s, t) = f_s(t) = (\cos(\pi t), s \cdot \sin(\pi t))$$

is a homotopy from $l(t)$ to $g(t)$.

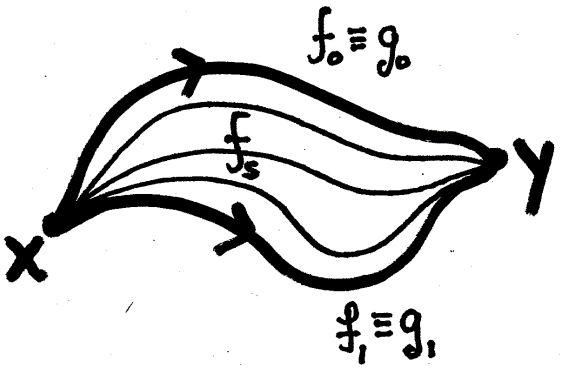


Figure 2: A homotopy between g_0 and g_1.

The idea is this: pick any two points on your surface X, say a torus, draw two curves or paths lying on your surface joining the points. If you can stretch and twist and shrink the first curve into the other, without cutting it, then the two paths are homotopic. The left-hand picture below shows two paths on the torus which are homotopic, while the right-hand picture shows two non-homotopic paths.

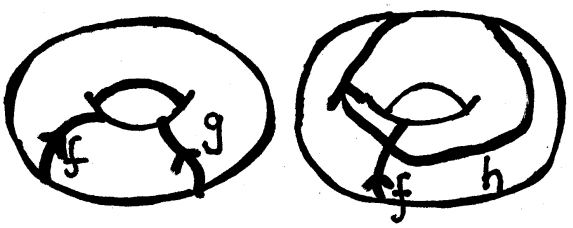


Figure 3: Homotopic and non-homotopic loops on the torus.

Loops

Now fix a point x_0 on your surface X. Paths which start and end at x_0 are called *loops* with base point x_0.

Definition 1. A *loop* is a path $f : I \to X$ such that $f(0) = f(1)$.

We want to know which loops are homotopic. Let's do the exercise on D^2. Let x_0 be the center of this disk. The first loop is the 'do-nothing' loop, i.e., $c : I \to D^2$ with

$$c(t) = x_0, \ \forall t \in I.$$

Take another loop based at x_0, say $f : I \to X$. Can we deform f into the constant loop, i.e., can we shrink f to the center? Intuitively, this seems possible. Indeed, since D^2 is convex we can join any point on the loop f to the point x_0 by a straight line, and then shrink f to x_0 along these lines. The homotopy is given by $F(s,t) = f_s(t) : I \times I \to D^2$

$$f_s(t) = sx_0 + (1-s)f(t).$$

What this means is that any loop based at x_0 on the disk D^2 is homotopic to the constant loop at x_0. The only thing we used to prove this is that, for any point $x \in D^2$, there is a line joining x and x_0. This holds for any point on the disk D^2, not just the center. Hence the loops based at any point are homotopic to the constant loop at that point. We state this result as a lemma, for it will be useful later:

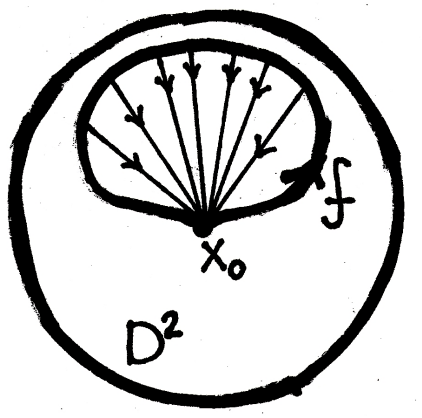


Figure 4: Homotopy from f to the constant loop at x_0.

Lemma 1. *Any loop $f : I \to D^2$ based at x is homotopic to the constant loop at x.*

Another example is $\mathbb{R}^2$. Pick any point $x_0 \in \mathbb{R}^2$. As for the disk, any point can be joined to x_0 by a straight line. Hence 1 gives a homotopy from any loop at x_0 to the constant loop at x_0.

Though it seems intuitively obvious, it is harder to show that any loop $f : I \to S^2$ based at x_0 is homotopic to the constant loop at x_0. Here S^2 denote a sphere, which you can visualize in $\mathbb{R}^3$ as the surface $x^2 + y^2 + z^2 = 1$.

In all these examples, we could shrink the loops to a point. This is not always the case. The loops on a torus T^1 are not all trivial. Let $x_0 \in T^1$. Let f be the loop based at x_0 which wraps around the torus in the vertical direction and h the loop which wraps around T^1 in the horizontal direction (see the figure of the torus above).

The loop h is not homotopic to the constant loop, since we cannot pull it across the hole back to the point x_0. It "contains" the hole while the constant loop does not. This is a fundamental difference between the two loops. On the other hand, the loop f does not surround the hole, but it encloses the inside of the torus. To shrink it to the constant loop, we would need to cut through the torus. Hence f and h are both essentially different from the constant loop at x_0. Are they homotopic to each other? The answer is no, since h contains the hole of the torus and f does not. To deform h into f, we would need to pull it through the hole. Similarly, we cannot deform f into h. Therefore, we have at least two non-homotopic loops which are not homotopic to the constant loop. Of course, we can make these intuitive arguments rigorous.

The Fundamental Group

Now comes the machinery. Let X be a path-connected topological space, i.e., for any two points x, y on X, there exists a path joining x and y. To every such space we attached a group as follows: we fix a point on X, say x_0, and we consider all the loops with base point at x_0. We put an equivalence relation on this set, and say that two loops f and g are equivalent, denoted $f \sim g$, if they are homotopic (check that this is in fact an equivalence relation!) Given a loop f based at x_0, we denote by $[f]$ the set of homotopic paths,

$$[f] := \{g | g \sim f\}.$$

Given two loops f and g based at x_0. It does not make sense to compose f and g as functions, because the range of f is X, while the domain of g is I. But we can certainly follow the path f, and then the path g, and still get a loop based at x_0. Hence we define $g \circ f$ as

$$g \circ f := \begin{cases} f(2t) & \text{if } t \in [0, 1/2]; \\ g(2t-1) & \text{if } t \in [1/2, 1]. \end{cases}$$

This is well-defined since $f(2 \cdot 1/2) = g(2 \cdot 1/2 - 1) = x_0$.

Let c be the constant loop at x_0, i.e.,

$$c : I \to X,\ c(t) = x_0\ \forall t \in I.$$

Given any loop f at x_0, it is easy to see that $f \circ c \sim c \circ f \sim f$. Also, given f, we get another loop for free, the loop $-f : I \to X$,

$$-f(t) = f(1-t).$$

We can deform the composition $f \circ -f$, into the constant loop. I leave this as an exercise.

This implies that the loop $f \circ -f \sim c$. Similarly, $-f \circ f \sim c$. Furthermore, it does not matter in what order we compose three loops f, g and h. The resulting loop $f \circ (g \circ h) \sim (f \circ g) \circ h$, that is, composition is associative. You can also check that if $[f_1] = [f_2]$ and $[g_1] = [g_2]$, then

$$[f_1 \circ g_1] = [f_2 \circ g_2].$$

Therefore, we can define

$$[f] \circ [g] := [f \circ g].$$

This gives the set of loops modulo homotopy equivalence a *group* structure. We call this group *the fundamental group* and denote it by $\pi_1(X, x_0)$.

The good thing is that this definition does not depend on x_0. Namely, if y_0 is another point of X, the groups $\pi_1(X, x_0)$ and $\pi(X, y_0)$ are isomorphic. Indeed, since X is path-connected, there is a path joining x_0 and y_0,

$$t : I \to X,\ t(0) = x_0 \text{ and } t(1) = y_0.$$

For any loop f based at y_0, the path $t^{-1} \circ f \circ t$ is a loop based at x_0.

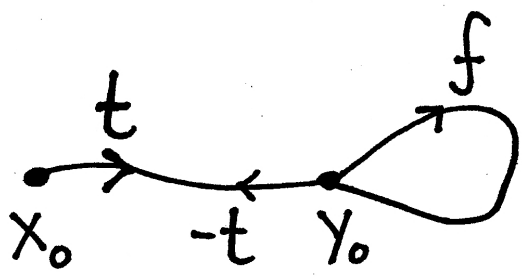


Figure 5: An isomorphism between $\pi_1(X, x_0)$ and $\pi_2(X, y_0)$.

The isomorphism $\phi : \pi_1(X, y_0) \to \pi_1(X, x_0)$ is given by

$$[f] \mapsto [t^{-1} \circ f \circ t].$$

Once can check that ϕ is indeed a homomorphism. It is clearly invertible, since $\phi^{-1} : \pi_1(X, x_0) \to \pi_1(X, y_0)$ defined as

$$[g] \mapsto [t \circ g \circ t^{-1}]$$

is an inverse of ϕ. Therefore we can write $\pi_1(X, x_0) := \pi_1(X)$ without ambiguity.

To go back to our example, let x_0 be a point on D^2. Lemma 1 tells us that $f : I \to D^2$ can be deformed to the constant loop at x_0. This means that $\pi_1(D^2) = 0$. It contains only the equivalence class of the trivial loop. Similarly, our above discussion shows that $\pi_1(\mathbb{R}^2) = 0$.

The Fundamental Group of the circle S^1

Our next goal is to compute $\pi_1(S^1)$. This is the key topological property we need to prove our theorem. Everything we will do in the next section can be made completely rigorous, but I'll just give a general overview of what's going on.

Fix a point $x_0 = 0$ on S^1 and view the points of S^1 as angles $0 \leq \theta < 2\pi$. What are the loops based at 0? There is always the trivial loop $c(I) = 0$. Let f be the loop given by a half counterclockwise rotation around the circle (up to π) followed by a half rotation in the clockwise direction, ending at 0. We can deform l into the constant loop at x_0, using the same argument as in the proof that $-f$ is the inverse of f in $\pi_1(X)$.

Now suppose we consider a full counterclockwise loop l_1 around the circle. We cannot deform this into the constant loop, because both end points of l_1 are pinned down at 0, and we cannot pull the loop out of the circle to bring it back to 0. So here, we have a non-trivial element of $\pi_1(S^1)$! Now suppose we take the loop which goes twice around the circle in the counterclockwise direction, i.e. the composition of l_1 with itself; call this loop l_2. Intuitively, it appears that l_1 and l_2 are not homotopic. This is harder to prove, so we will take it on fate. In fact, if we let l_n where $n \in \mathbb{Z}$ be the loop "go around the circle n times" or "l_1^n" (if n is negative, we mean compose $-l_1$ with itself $|n|$ times, and $l_0 = c$), then one can show that $l_n \sim l_m$ if

and only if $m = n$. Also, these are the only equivalence classes, i.e., if $f : I \to S^1$ is a loop based at 0, then $f \sim l_n$ for some $n \in \mathbb{Z}$.

Let me give a more precise idea of what's going on. We look at the map $f : \mathbb{R} \to S^1$ define by $\varphi(x) = e^{2\pi ix}$. This map sends all the integers to 1. Also, if we take a small enough neighborhood of a point x in $\mathbb{R}$, say $I_x = (x - 1/2, x + 1/2)$, the restriction of φ to I_x is a homeomorphism onto its image. Hence φ is a local homeomorphism. We say that φ is a covering map of S_1. Now it's a general fact that any loop γ on X can be lifted to a unique path in $\mathbb{R}$ starting at 0. That is, there exists a unique path $g(t)$ in $\mathbb{R}$ such that $g(0) = 0$ and $\varphi \circ g = \gamma$.

Letting γ_0 be loop in S^1, and let g_0 be this unique path. Since $\gamma_0(1) = 1$, it must be that $g_0(1) = n \in \mathbb{Z}$. If γ_1 is homotopic to γ_0, the homotopy $\gamma_s(t)$ lifts to a family $g_s(t)$ of continuous maps from g_0 to g_1 where g_1 is the unique lift of γ_1 starting at 0. But since $\gamma_s(1) = 1$ for all s, we have $g_s(1) \in \mathbb{Z}$ for all s. We know that $g_0(1) = n \in \mathbb{Z}$, hence by continuity, it must be that $g_1(1) = n$.

An isomorphism from $\pi_1(S_1)$ to $\mathbb{Z}$ is given by associating to each equivalence class of loops this integer. We state this as a theorem:

Theorem. *The fundamental group of the circle $\pi_1(S^1)$ is isomorphic to $\mathbb{Z}$.*

Retracts

Suppose we have a topological space X, and a subset A of X (to which we give the subspace topology.) It might be possible to *retract* X onto A, i.e., to shrink X to A, as in the following picture.

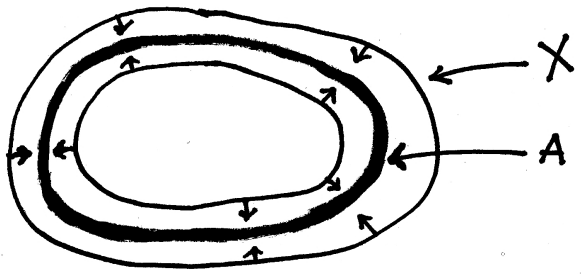


Figure 6: A retract of X onto a subspace A.

To make this idea concrete, we define a *retraction* of a topological space X onto a subspace A to be a continuous function $r : X \to A$, such that r restricts to the identity map on A: $r|_A = id_A$.

Why do we care about retractions in the context of the fundamental group? Suppose A is a subspace of X, and g and h are paths in A. It might be ambiguous whether or not they are homotopic by looking at A only. However, if X retracts onto A, and as paths of X, g and h are homotopic, we can conclude that g and h are homotopic in A also. This is what the following lemma says.

Lemma 2. *Let X and A be as above, and let $g : I \to A$ and $h : I \to A$ be paths in X with image in A. Suppose g and h are homotopic when viewed as paths in X, i.e., there is a homotopy $f_t : I \to X$, from g to h. If there is a retraction $r : X \to A$ of X onto A, then the family of maps $r \circ f_t : I \to A$ is a homotopy of g and h in the space A.*

The next proposition is the key to the proof of theorem , and this is here all the work pays off. The proof of theorem is just the usual gymnastic of mathematics with this proposition.

Proposition 1. *There is no retraction of D^2 onto S^1.*

Proof. For the sake of contradiction, suppose that such a retraction r existed. Consider the loops l_1 and l_0 based at 0 on S^1. We argued above that they represent distinct homotopy classes of $\pi_1(S^1)$. As paths on D^2, lemma 1 shows that there is a homotopy $f_t : I \to D^2$ such that $f_0 = l_1$ and $f_1 = l_0$. Consider, $r \circ f_t : I \to S^1$, $t \in I$. By proposition 2, this is a homotopy of l_1 and l_0 *in* S^1! But l_0 and l_1 are not homotopic in S^1, a contradiction. Therefore, there is no such retraction r. □

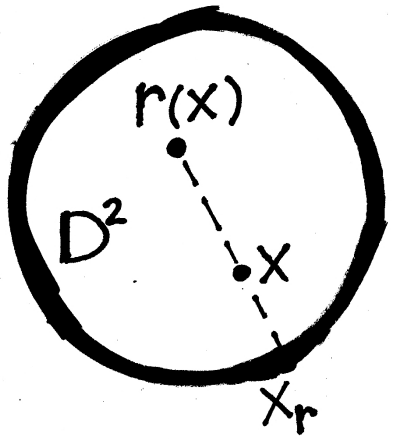


Figure 7: A would-be retract of D^2 onto S^1.

As promised, here's a proof of theorem

Proof. **Brouwer's Fixed point Theorem**

Suppose that $\varphi : D^2 \to D^2$ is a continuous map which has no fixed point. This means that $x \neq \varphi(x)$ for all $x \in D^2$, so we can draw a ray starting at $\varphi(x)$ going through x. This ray must meet the boundary of D^2, namely S^1, at one and only one point x_r. Let $r : D^2 \to S^1$ be defined by $r(x) = x_r$ (see figure below). This map is clearly continuous, because $\varphi(x)$ is continuous. What does it do to S^1? If $x \in S^1$, draw the ray starting from $\varphi(x)$ passing through x. This ray meets S^1 in one point x_r. Clearly $x_r = x$. Therefore $r(x) = x$. Hence $r|_{S^1} = id_{S^1}$. This means that r is a retraction of D^2 onto S^1, contradicting proposition 1. Therefore, all maps $\varphi : D^2 \to D^2$ have a fix point. □

Other Interesting Results

Now that was just to give you a taste of what's going on. The essential element was that $\pi_1(S^1) \simeq \mathbb{Z}$. As a conclusion, I state here are a few other interesting theorems which are consequences of that fact.

Theorem. The Borsuk-Ulam Theorem *For every continuous map $\psi : S^2 \to \mathbb{R}^2$, there are antipodal points x and $-x$ on S^2 such that $\psi(x) = \psi(-x)$.*

Corollary 1. *Whenever three closed sets A_1, A_2 and A_3 have the property that $S^2 = A_1 \cup A_2 \cup A_3$, then at least one of the sets A_i contains a pair of antipodal points, x and $-x$.*

Theorem. Hamburger Theorem *If A_1, A_2 and A_3 are closed and bounded sets in $\mathbb{R}^3$, there exists a plane which cuts each A_i into two parts of equal measure.*

Theorem. Fundamental Theorem of Algebra *Every non-constant polynomial $f(x) \in \mathbb{C}[x]$ has a root. That is, there is an $\alpha \in \mathbb{C}$ such that $f(\alpha) = 0$.*

This article comes from notes written for a mini-lecture I gave at the Program for Young Scientists (PROMYS) at Boston University. I would like to thank prof. Glenn Stevens and everyone involved in PROMYS for hosting such a great program and making these mini-lectures possible.

References

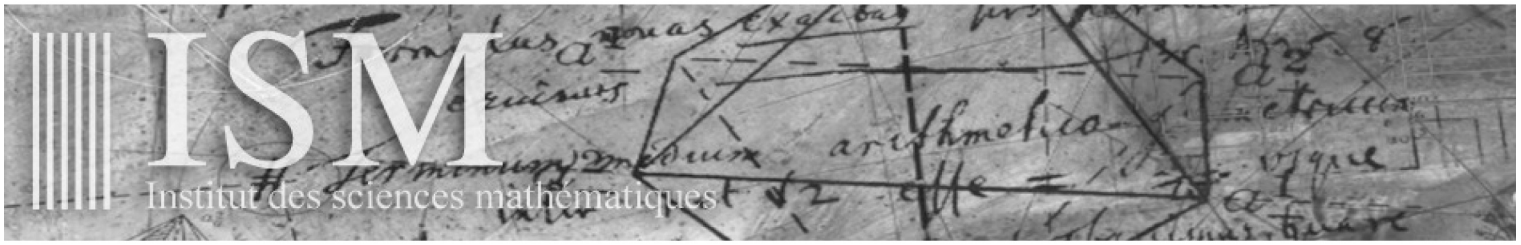
1. Allen Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, 2001

Jokes

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. □

A mathematician organizes a lottery promising an infinite amount of money. Naturally, a lot of people buy tickets and the mahematician earns a lot of money from the sales. When the lucky winner is finally announced, he goes to see the mathematician, eager to get his prize. The mathematician looks at him calmly and says: "Here's one dollar. Come back tomorrow and I'll give you half a dollar. Then the day after I'll give you $\frac{1}{3}$ of a dollar..." □

A lecturer tells the students to learn the phone-book by heart.
The mathematicians are baffled: 'By heart? Are you kidding?'
The physics students ask: 'Why?'
The engineers sigh: 'Do we have to?'
The chemistry students ask: 'Until next Monday?'
The accounting-students (scribbling): 'Until tomorrow?'
The medicine students ask: 'Should we start on the Yellow Pages?'
The laws students answer: 'We already have.' □



What is the ISM? The Institut des sciences mathématiques (ISM) is a consortium of seven Québec universities (Concordia, Laval, McGill, Université de Montréal, Université du Québec à Montréal, Université du Québec à Trois-Rivières and Université de Sherbrooke) that coordinates the graduate programs in mathematics and statistics of the participating universities. If you are a registered student at one of these universities, you can participate in the ISM's activities. The ISM allows students to take courses at any of the member universities, to participate in the weekly student seminar and the Quebec graduate student conference, and to take advantage of several scholarship programs.

ISM courses. Currently there are eleven scientific program groups in the ISM, each one offering a wide selection of advanced courses. The complete list of ISM courses can be found on the ISM webpage.

ISM Seminars. Most of the scientific programs organize weekly seminars in which students can participate. The ISM also sponsors the weekly graduate student seminar, organized by and for students from the four Montreal universities.

The Quebec Graduate Student Conference. This conference was born out of the weekly graduate student seminar. Each year the conference attracts approximately ninety participants from Quebec and elsewhere.

ISM Scholarships. The ISM offers two funding programs for graduate students: scholarships and travel bursaries for students who wish to present a paper at a conference. In the summer, undergraduate students can apply for Summer 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@uqam.ca
www.math.uqam.ca/ism

Mathematical Digest

Nan Yang

Modern mathematics can be very abstract. A layman flipping through a mathematics textbook may be so overwhelmed by the formalism that she loses sight of the very ideas which the formalism is trying to capture, even though the ideas themselves may be very intuitive and simple. In this issue of Mathematical Digest we will take a look at two of them, isomorphisms and homomorphisms, from a non-mathematical perspective.

Isomorphism

The word 'isomorphic' appears in many courses describing many different things. You may hear it in set theory describing ordered sets, or linear algebra describing vector spaces, or abstract algebra describing groups, rings, fields, etc. You probably already know what it means: two things are isomorphic if they are, *in the context of interest*, the same thing. Take, for example, chess (perhaps the best example available). Observe the following picture:

Figure 1: Chess

The picture on the left is a standard chessboard, made of wood; the picture on the right is a chessboard made out of, well, sidewalk. Yet we say that they are the 'isomorphic', that they are the same game. Intuitively, this is because if you were to place them side by side, each move on one corresponds uniquely to a move on the other. A game to which chess is not isomorphic is, for instance, tic-tac-toe: for one thing, tic-tac-toe has no pieces to move!

It is possible to make the notion of isomorphism of games more rigorous as follows. The chessboard can be in one of finitely many states. At every stage the chessboard is in a given state. It is possible to move from one state to another given that there is a single piece on the first state that can be moved to a corresponding piece on the second state by a legal move. We then connect these two states. If we represent the states by nodes, and the possibility of 'moving' from one state to another by an edge, then we can think of Chess as a very large tree (of graph theory), where repeated moves such as a knight moving back and forth between two positions is described by an infinite branch. Then, given any game which has the above properties (states, transition of states, etc), we can assign a unique tree to that game, up to isomorphism (this isomorphism is well defined, for more information see a textbook on graph theory). Two games are isomorphic, then, when the trees assigned to them are isomorphic.

But I digress. What else can be a good, concrete analogy for 'isomorphism'? Let's look at something a little more esoteric, and perhaps a bit far-fetched.

Beethoven's 6th symphony is written in F major. Its first few notes are:

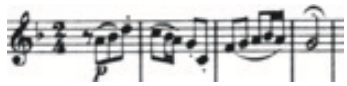


Figure 2: Pastoral

But suppose it were written in some other key, say C major. Of course, it certainly would *sound* different if it were written in C major. And any music student would argue with me that if it were performed in C major, it is not the same piece of music. But I can dare say that those two pieces are 'isomorphic'. How so? Because you can literally 'translate' one into the other by 'shifting' notes up or down (the technical details will not be discussed here) and thus information is not lost in translation. If for some reason the only surviving copy of the score for the symphony has been 'shifted down' so that it's in C major, but we know that it should be in F major, we can easily translate it back to F major without loss of information. In this sense, the two pieces of music are 'isomorphic'.

I mentioned the loss of information, because this is important for the next concept that we will be discussing, homomorphism.

Homomorphism

You may have encountered homomorphisms before isomorphisms because isomorphisms are, usually, 'bijective homomorphisms'. Therefore homomorphisms are more general. There is a very simple and concrete illustration of the idea that two things are homomorphic: acronyms.

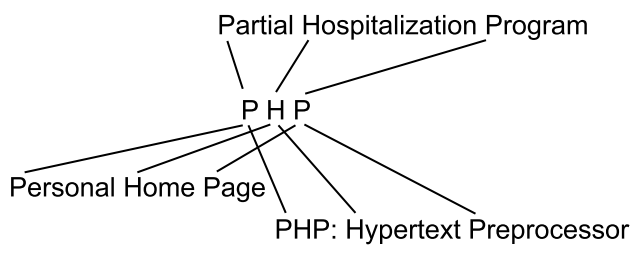


Figure 3: PHP

See? Think of the set of finite sequences of letters without spaces (i.e., ASDF, WADS, AGEFAWE) as having a 'structure'. Then we can think of the set of finite sequences of words as having an even *finer* structure. When we map from sequences of words to sequences of letters by using acronyms, some information is lost, but not all. This is the idea behind a 'non-bijective' isomorphism, or homomorphism. By analogy of the example of a surviving copy of Beethoven's 6th symphony being in the wrong key, suppose an error correction algorithm works by taking the first letter from every word of a text file and concatenating them to form a sequence of letters, and transmitting both parts. Suppose for some reason only the 'error-correcting' part is received, then it would *not* be possible to translate it back to the original text, since any translation would not be unique; but it is also possible to *rule out* texts that cannot have been sent. For example, if you received error-correction code 'N.A.N.' but not the text, you could not say for certain whether the text sent had been 'Not a Number', 'National Academy of Neuropsychology' or any other possibilities. But you could say that it had *not* been 'Society of Undergraduate Math Students'.

Once Upon a Time in a p-adic Approximation Lattice

Vincent Quenneville-Bélair

The type of a p-adic number z will be helpful in the study of the growth of the rational approximation made by the lattice $L_n(z)$. An interesting result about the type is that it is able to detect rational numbers since, for these numbers, it exists and takes on a specific value. Another key point is that it has a maximal value.

Introduction

Once upon a time in a p-adic approximation lattice, the norm of the shortest non-zero vector was expected to yield a good rational approximation. Lenstra, Lenstra and Lovásv conceived a very efficient and now widely used algorithm for computing an almost orthogonal basis for a lattice. The result obtained from it yields a very good approximation of the smallest vector of a lattice which, in turn, is a good rational approximation of a number. Indeed, a special kind of lattice will possess such an approximation as its smallest vector – hence, the need for a such a fast such algorithm. p-adic approximation lattices, as they are called, can be viewed as a p-adic analogue of the continued fraction expansion of a real number which can give very good approximation to a real number. The growth of this good approximation, the smallest vector, is expected to be exponential and this is what the type of a lattice attempts to capture. Interestingly, the type will be able to detect rational numbers. In general however, it will be shown that the type might not exist.

Definition 1 ([4])**.** A lattice Λ is a set $\mathbb{Z}a_1 + ... + \mathbb{Z}a_d$ with $a_1, ..., a_d$ linearly independent real vectors in $\mathbb{R}^d$.

Lattices are essentially a vector space over $\mathbb{Z}$. No big deal! The fundamental parallelepiped P for the lattice Λ corresponds to

$$\{t_1a_1 + ... + t_da_d \text{ s.t. } 0 \leq t_1, ..., t_d < 1\}$$

which can be identified with $\mathbb{R}^d/\Lambda$. The volume of a lattice is defined as the volume of its fundamental parallelepiped.

$$\det(\Lambda) = |\det(a_1, ..., a_d)| = \text{Vol}\left(\mathbb{R}^d/\Lambda\right).$$

Type of a p-adic number

One of the motivation behing the study of the type is the problem of rational recognition. Given a rational number presented as a real number, how can it be recognise as a rational number? The naive way of doing would be to look whether the decimal expansion is periodic: if it is than the number is rational! In practice however, this is not a very convenient way and the type might be able to give better results as it will be seen with lemma 1.

Before considering rational detection, one will look at rational approximation. A special kind of lattices can be used in order to obtain good rational approximation to a p-adic number. Their real equivalent could be seen as the well-known continued fraction expansion. Continued fractions also give better algorithm of rational detection over the naive method mentioned above. To attempt to make a good rational approximation, the linear form

$$a_1 + a_2z$$

where $a_1, a_2 \in \mathbb{Z}$ and $z \in \mathbb{Q}_p$ could be made small. In the p-adic sense, it means to make it divisible by a high power of p with p a prime as usual. Indeed, one can consider a lattice of the following form:

$$L_n(z) := \{(a_1, a_2) \in \mathbb{Z}^2 | a_1 + a_2z = 0 \pmod{p^n}\} \quad (1)$$

where $z \in \mathbb{P}^1(\mathbb{Q}_p) = \mathbb{P}^1(\mathbb{Z}_p)$ and $n \in \mathbb{Z}^+$. It is a sublattice of $\mathbb{Z}^2$, but $p^n\mathbb{Z}^2$ is contained in it. There is a natural homomorphism from $\mathbb{Z}^2$ to $\mathbb{Z}/p^n\mathbb{Z}$ with kernel corresponding to $L_n(z)$ from equation 1. This implies that

$$\#\left(\mathbb{Z}^2/L_n(z)\right) = \text{Vol}\left(L_n(z)\right) = p^n$$

since the volume of the lattice has integer height and base and is thus known to be an integer.

Remark. $\mathbb{Q}_p$ corresponds to the field of p-adic numbers and $\mathbb{Z}_p$ are the p-adic integers. A p-adic number is small when it is divisible by a high power of p. Actually, the p-adic norm, $|z|_p$, is the highest power of p that divides z. In order to look at the projective line of $\mathbb{Q}_p$, one needs to homogenize equation 1: any denominators should be cleared. If $z = X/Y$ with $X, Y \in \mathbb{Z}_p$,

$$a_1 + a_2z = a_1Y + a_2X = 0.$$

Now, the case when $Y = 1$ is taken care of by equation 1. If $Y \neq 0 \pmod{p}$ then diving through by Y reduces the equation to equation 1. If $Y = 0$, $z = \infty$: it is a point "at infinity". Similarly, if $Y = 0 \pmod{p^{n_0}}$ for some $n_0 > 0$ then, for the first few n at least, z will behave like a point "at infinity".

Going back now to equation 1. For a fixed n, $z = -a_2'/a_1' \pmod{p^n}$ where the shortest vector is (a_1', a_2'). As n goes to infinity, the smallest vector gives a better and better rational approximation to the p-adic number z. The expected behaviour is that m_n, the norm of the smallest non-zero vector, should grow exponentially with respect to large values of n. That is what the type of a p-adic number tries to capture.

Definition 2 ([6])**.** $z \in \mathbb{P}^1(\mathbb{Q}_p)$ is said to be of type $\alpha \in \mathbb{R}$ if the lattice $L_n(z)$ satisfies

$$\log(m_n) = \alpha n + O_n(1)$$

where

$$m_n = \min_{\substack{\lambda \in L_n(z) \\ \lambda \neq 0}} \|\lambda\|.$$

This definition of the type is readily extended to higher dimensions. However, a focus on the two dimensional case will be maintained for simplicity. It is important to mention that the existence of the type for all p-adics is not guaranteed. Indeed, there exists numbers for which m_n does not converge. However, if z is a rational number, the type exists. Even more, the type is an indicator of rationality: heuristically, if z is a rational number, the shortest vector should after a while remain the same and thus the type should be zero.

Lemma 1 ([6])**.** type(z) = 0 *if and only if* $z = (y : x) \in \mathbb{P}^1(\mathbb{Q})$.

Proof. Suppose that $z \in \mathbb{P}^1(\mathbb{Q})$. Then $z = (y : x)$, so that, if $a_1 = -x$ and $a_2 = y$, $a_1 + a_2 z = 0 \pmod{p^n}$. Thus, $m_n \leq \|(-x, y)\|$ which implies that $\log m_n$ is bounded and thus that the type is zero.

Suppose now that type$(z) = 0$. It implies that $m_n \leq e^C$ for some constant $C \in \mathbb{R}$. However, there are only finitely many points of the lattice with norm smaller than e^C since the lattice is discrete. It implies that there is such an element for infinitely many n in the sequence of smallest non-zero vectors. Call it (x, y). Hence, $z = -\frac{x}{y}$, a rational number. □

Even though the type of rational numbers is zero, it can be shown that the average type is $\log(p)/2$ [6]. Actually, this value of the type is maximal. Thus most numbers have it.

Lemma 2. *The maximal value of the type is* $\log(p)/d$ *where* d *is the dimension of the lattice.*

A reader wishing to prove this lemma could look at Hermite's theorem or at Minkowski's Convex Body theorem. It is not by accident that Minkowski's theorem appears here: it may be considered as a fundamental element of the geometry of numbers which relates to many elements in this article. [1]

Theorem (Hermite's theorem [5, 3])**.** *There exists a constant* $\mu_d \in \mathbb{R}^{>0}$ *dependent only on* d *such that*

$$m_n^d(z) \leq \mu_d \mathrm{Vol}(\Lambda)^2$$

where Λ *is a lattice of dimension* d.

Hermite's theorem will be stated without proof, but an interested reader can look into [3].

Proof of lemma 2. Using theorem from Hermite,

$$m_n \leq M_1^{1/2} \leq (\mu_d \mathrm{Vol}(\Lambda)^2)^{\frac{1}{2d}}.$$

Since $\mathrm{Vol}(L_n(z)) = p^n$,

$$\frac{\log(m_n)}{n} \leq \frac{2n\log(p)}{2dn} + \frac{log(\mu_d)}{2dn}$$

which goes to $log(p)/d$ as n goes to infinity. □

Existential Problem

As mentioned earlier, not all p-adic numbers have a type. For instance, the Liouville number $z_1 = \sum_{n=1}^{\infty} p^{n!}$ does not have one since its coefficients are too scarce.

Before jumping to proving that the type does not necessarily exist in general, it is important to explain how this approximation was obtained. A fast algorithm to compute the shortest vector of a lattice is needed. This is where the Lenstra-Lenstra-Lovàsv algorithm comes in. This algorithm is able to find a vector very close to the shortest non-zero one in the lattice. The approximation of the type is made using $\log(m_n)/n$ which converge, by definition 2, towards the type as n becomes large. Figures in this articles show $\log(m_n)/n$ as n grows. In particular, z_1 in figure 1 can be seen to diverge.

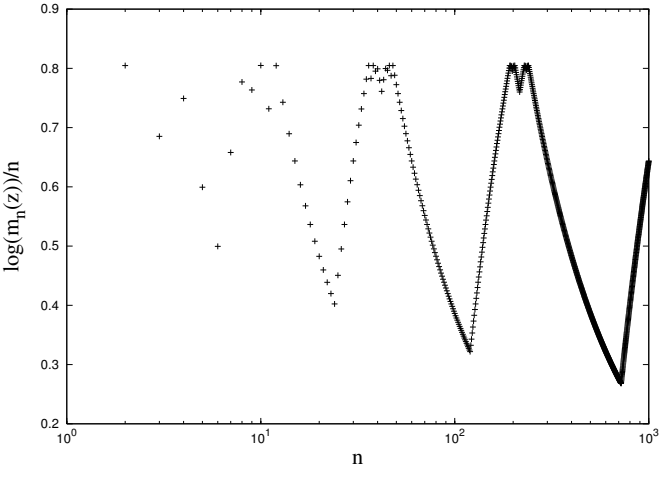


Figure 1: This graph shows $\log(m_n(z_1))/n$ in function of n for $z_1 = \sum_{i=1}^{\infty} 5^{i!}$. As it can be obversed from these numerical considerations, the type does not exist in this case.

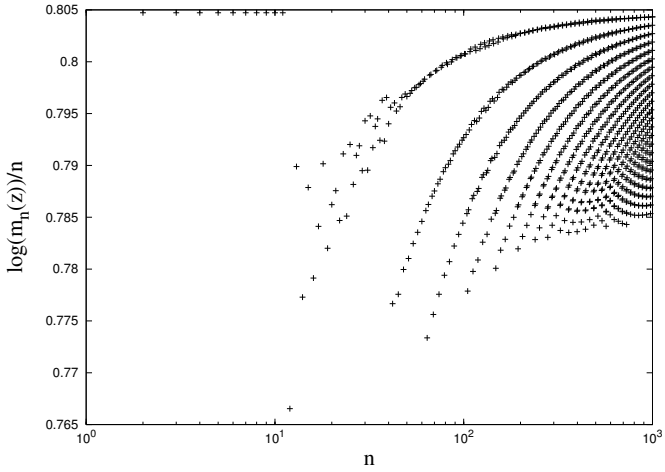


Figure 2: Same axis as figure 1. The 5-adic number giving this graph was generated in a way that would make it reach type $= \pi/4$. This lovely pattern rather seems to converge to the maximal type.

Lemma 3. *type(z) does not exist for* $z = z_1 = \sum_{n=1}^{\infty} p^{n!}$ *and* $z = z_2 = \sum_{n=j}^{\infty} \mathrm{tower_k(j)}$ *for* $k \in \mathbb{Z}^{>2}$ *where* $\mathrm{tower_k(j)} = \mathrm{p}^{\mathrm{tower_{k-1}(j)}}$ *and* $\mathrm{tower_0(j)} = \mathrm{j}$.

Proof. First, one considers $z_1' = z_1 \pmod{p^{(N+1)!}}$ for N large. The norm of the smallest vector of $L_n(z_1)$ is smaller than the norm of $(-z_1', 1)$ for all n such that $N! < n \leq (N+1)!$. Also, $(z_1')^2 + 1 \leq 4(z_1')^2 \leq 4N^2p^{2N!}$. Hence,

$$\frac{\log(m_{(N+1)!})}{(N+1)!} \leq \frac{\log(1+(z_1')^2)}{2(N+1)!} \leq \frac{\log(4N^2p^{2N!})}{2(N+1)!} \leq \frac{1}{N+1}\log(p) + \frac{\log(2N)}{(N+1)!}$$

which goes to zero as $N \to \infty$. Since z_1 is not a rational number, it cannot have type zero and so it does not have one. It is possible to adapt the previous steps to $z = z_2$ easily.

Second, one considers $z_2' = z_2 \pmod{pM}$ where $M = \mathrm{tower_{k-1}(N)}$ for N large. One notes that $(z_2')^2 + 1 \leq 4N^2p^{2\mathrm{tower_{k-1}(N)}} = (2Np^M)^2$. For all n such that $M = \mathrm{tower_{k-1}(N)} < \mathrm{n} \leq \mathrm{tower_{k-1}(N+1)} = \mathrm{M'}$,

$$m_n < \sqrt{1+(z_2')^2}$$

and so

$$\frac{\log(m_{M'})}{M'} \leq \frac{\log(1+\mathrm{tower_k(N)}^2)}{2M'} \leq \frac{M}{M'}\log(p) + \frac{\log(2N)}{M'}$$

which goes to zero as N goes to infinity. However, z_2 is not a rational since its base p expansion is not periodic: the type is not zero if it has one. [2] Thus, the type(z_2) does not exist. □

The algorithm mentioned above for approximating the type can also be adapted to seek for new values of type. Indeed, up to now, the only values of the type observed are zero and the maximal. The general idea is that the series expansion is developed term by term while attempting to keep $\log(m_n)/n$ close to the desired type. Figure 2 shows an attempt to generate a number with type $\pi/4$. An interesting pattern can be observed!

Conclusion

To wrap up, the type of a p-adic number corresponds to the exponential growth rate of p-adic approximation lattices. Some interesting properties can be observed. A p-adic number is rational if and only if its type is zero. Further, a maximal value was shown to exist. However, the existence of the type is not guaranteed. For instance, the Liouville number $z_1 = \sum_{n=0}^{\infty} p^{n!}$ and $z_2 = \sum_{n=0}^{\infty} \mathrm{tower_k(n)}$ with $k \in \mathbb{Z}^{>2}$ do not have one. Up to now, if a p-adic number has a type, it only has been observed to be either zero or maximal. Some numerical results showed attempts to find numbers with different values. They were obtained using the Lenstra-Lenstra-Lovásv algorithm that is able to find a vector that is close to the smallest non-zero one. And thus, the smallest vector lived happily ever after, for it was now well approximated by the algorithm...

The author would like to thank professor Henri Darmon and Dr. Christian Wüthrich for their support and advices throughout the elaboration of this article.

References

[1] J.W.S. Cassels. *An Introduction to the Geometry of Numbers*. Springer-Verglag, 1959.

[2] Wiliam Leveque. *Fundamentals of Number Theory*. Dover, New York, 1996.

[3] M. Pohst and H. Zassenhaus. *Algorithmic algebraic number theory*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, 1990.

[4] H. Silverman, Joseph and John Tate. *Rational Points on Elliptic Curves*. Undergraduate Texts in Mathematics. Springer, New York, 1992.

[5] Nigel Smart. *The Algorithm Resolution of Diophantine Equation*. London Mathematical Society, Cambridge, 1998.

[6] Christian Wüthrich. On p-adic elliptic logarithms and p-adic approximation lattices. 2006.

On Nodes and Knots on S^3

Tayeb Aissiou and Sergei Dyda

We show how simple links and torus knots are generated by the restriction of nodal sets of complex-valued spherical harmonics to S^3. We will detail the construction of two classes of knots using some types of polynomials.

Introduction

A knot is generically a closed, 1-dimensional curve in $\Re^3$. By considering the nodal set of a *complex-valued* function on a compact 3-dimensional manifold (the set of points where the function vanishes) it is therefore possible to generate a knot. In particular, one can consider these sets for complex-valued eigenfunctions of the Laplacian or of the Schrödinger operator. M. Berry studied this problem for the hydrogen atom in [1].

In the current paper, we study nodal sets of complex-valued eigenfunctions of the Laplacian for the round metric on S^3. We realize S^3 as a set $\{(z,w) \in \mathbf{C}^2 : |z|^2 + |w|^2 = 1\}$. Spherical harmonics on S^3 are homogeneous polynomials $P(z,w,\bar{z},\bar{w})$ satisfying

$$\Delta P = (\partial_z \partial_{\bar{z}} + \partial_w \partial_{\bar{w}})P = 0,$$

restricted to S^3, [4]. In particular, we note homogeneous polynomials in z, w only (or in $\bar{z}, \bar{w}$ only) give rise to spherical harmonics. We call the harmonics arising from degree n polynomials *eigenfunctions of degree* n, where $n = 1, 2 \ldots$.

We show how two classes of knots can be generated using these types of polynomials and detail their construction.

Homogeneous polynomials in z, w

Consider a homogeneous polynomial $P(z,w) = a_n z^n + a_{n-1} z^{n-1} w + \ldots + a_0 w^n$ and look at its nodal set $N := \{(z,w) : P(z,w) = 0\}$. Dividing through by w^n, we find that $P(z/w) = a_n (z/w)^n + a_{n-1}(z/w)^{n-1} + \ldots + a_0 = 0$. Suppose for simplicity that P has n distinct complex roots, $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, and that $\alpha_j \neq 0$ for all j (this can be achieved by requiring that $a_0 \neq 0$). It follows that the nodal set $N = \{(z,w) : z = \alpha_j w, 1 \leq j \leq n\}$. In particular, $|z| = |\alpha_j| \cdot |w|$. Substituting into the equation $|z|^2 + |w|^2 = 1$, we find that

$$|w|^2 = \frac{1}{1+|\alpha_j|^2},\ |z|^2 = \frac{|\alpha_j|^2}{1+|\alpha_j|^2}, \qquad 1 \leq j \leq n.$$

It follows that the set N can be parametrized by

$$\left\{ \left(\frac{e^{it}}{\sqrt{1+|\alpha_j|^2}}, \frac{\alpha_j e^{it}}{\sqrt{1+|\alpha_j|^2}} \right) : \qquad t \in [0, 2\pi], 1 \leq j \leq n \right\}.$$

We shall call the j-th component $C(\alpha_j)$. The set N is thus a link with $\leq n$ components C_j, each diffeomorphic to S^1.

Proposition 1. The components $C(\alpha_j)$ are disjoint.

Proof of Proposition 1. To understand when $C(\alpha_i) \cap C(\alpha_j) \neq \emptyset$, we assume that $i = 1, j = 2$ (say), and that $(z_0, w_0) \in C(\alpha_1) \cap C(\alpha_2)$. Assume that $\alpha_1 \alpha_2 \neq 0$. Then $z_0 \neq 0$ and $w_0 \neq 0$, since $|z_0|^2 + |w_0|^2 = 1$. Therefore, we can divide through in the system of equations

$$z_0 = \alpha_1 w_0,$$
$$z_0 = \alpha_2 w_0,$$

to get $\alpha_1 = \alpha_2$. So the circles are disjoint unless $\alpha_1 = \alpha_2$, assuming $\alpha_j \neq 0$. □

We next remark that $C(\alpha)$ is a *leaf* in the *Hopf fibration*, which is a map $H : S^3 \to (\mathbf{C} \cup \infty) \cong S^2$ given by

$$H(z_1, z_2) = z_1/z_2,$$

where $|z_1|^2 + |z_2|^2 = 1$.

Proposition 2. The fibres in the Hopf fibration are linked.

Proof of Proposition 2.

Let s: $S^3/(0,0,0,1) \to \Re^3$ be the usual stereographic projection. We wish to show that any two fibres $s \circ H^{-1}(Q), s \circ H^{-1}(P)$ are linked. We begin by noting that $s \circ H^{-1}(-1,0,0)$ is a circle in the y-z plane. Let $s \circ H^{-1}(P)$ be another fibre. Applying the maps defined above, we realize that this fibre intersects the y-z plane both inside and outside the unit circle. This shows the two fibres are linked. To complete the proof it suffices to show that $s \circ H^{-1}(-1,0,0)$ can be deformed continuously into any other fibre of the Hopf fibration. Since each knot is determined solely by the complex roots α_i, a continuous deformation of one fibre into another is equivalent to a continuous curve through $\mathbf{C}^1$. Hence to continuously deform the y-z unit circle into any other fibre without intersecting some fixed fibre defined by the root α_i is equivalent to finding a continuous path linking two points in $\mathbf{C}^1/\alpha_i$. Since this is always possible, we conclude the Hopf fibration is linked. □

>From the above argument we can also conclude that n distinct fibers in the Hopf fibration are isotopic to each other, and hence determine the same link.

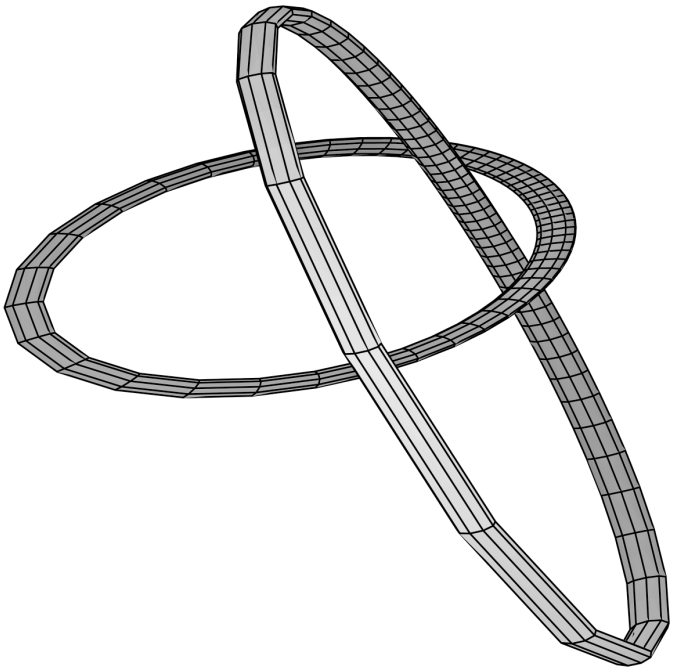


Figure 1: Two leafs of the Hopf fibration (links) plotted using the above construction.

Torus knots

Let $m > 1, n > 1, (m, n) = 1$ be two relatively prime integers.

Proposition 3. There exist two (real) spherical harmonics P, Q on S^3 such that the nodal set of $P + iQ$ has connected components isomorphic to the (m, n) torus knot.

Proof of Proposition 3. The idea is to choose a real-valued polynomial $P(z, \bar{z}, w, \bar{w})$ that will be "responsible" for rotation around the torus, and real-valued polynomial $Q(z, \bar{z}, w, \bar{w})$ that will be "responsible" for localization on the torus. Then the components of the nodal set $\mathcal{N}(P + iQ) = \mathcal{N}(P) \cap \mathcal{N}(Q)$ will define the torus knot.

We proceed to choose

$$P(z, \bar{z}, w, \bar{w}) = z^m \bar{w}^n + w^n \bar{z}^m. \tag{4}$$

This polynomial is harmonic (since every term is harmonic), and its nodal set is given by

$$(z/\bar{z})^m = -(w/\bar{w})^n.$$

Letting $z = |z|e^{it}, w = |w|e^{is}$, we find that

$$2mt = -\pi + 2ns \quad (\mod 2\pi),$$

so t and s are related by

$$t = \frac{\pi(2j-1)}{2m} + \frac{ns}{m},$$

thus defining the rotation around the torus.

We next describe the polynomial $Q(z, \bar{z}, w, \bar{w}) = Q(|z|, |w|)$. According to [3, Thm. 4.1], its restriction to $S^3 = \{|z|^2 + |w|^2 = 1\}$ is given by the Jacobi polynomial $P_n^{0,0} = P_n$ (this is the only spherical harmonic invariant under $\mathrm{SO}(2) \times \mathrm{SO}(2) \subset \mathrm{SO}(4)$). The zeros of P_n give rise to the two-dimensional tori in S^3, while the polynomial P defines the rotation around these tori.

□

Explicit construction for $m + n$ even

We construct the polynomial Q explicitly in the case when $m + n$ is even, say $m + n = 2k$. We look for a polynomial Q of the form

$$\sum_{j=0}^{k} a_{k-j} |z|^{2(k-j)} |w|^{2j}.$$

Setting a_k=1 and demanding that Q be harmonic imposses that

$$a_j = (-1)^j \binom{k}{j}^2.$$

Accordingly, if we let $|z|^2/|w|^2 = x$, we find that

$$Q_{2k} = |w|^{2k} \left(\sum_{j=0}^{k} (-1)^j \binom{k}{j}^2 x^j \right) := |w|^{2k} q_k(x),$$

where $q_k(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j}^2 x^j$. Form a related generating function

$$\tilde{q}_k(x) := \sum_{j=0}^{k} \binom{k}{j}^2 x^j.$$

Claim 1.

$$\tilde{q}_k(x) = \sum_{j=0}^{k} \binom{k}{j}^2 x^j = (1-x)^k P_k((1+x)/(1-x)),$$

where P_k is the k-th Legendre polynomial.

Proof of the Claim: This can be done by induction on k by showing the Legendre polynomial recursion relation

$$(k+1)\, P_{k+1}(x) - (2k+1)\, x\, P_k(x) + k\, P_{k-1}(x) = 0.$$

is valid. The details are left as an exercise to the reader. □

It follows from Claim 1 that

$$q_k(x) = (1+x)^k P_k((1-x)/(1+x)).$$

Accordingly, if we let $-1 \le \alpha_k < \alpha_{k-1} < \ldots < \alpha_1 < \alpha_0 \le 1$ denote the roots of $P_k(x)$, then the roots β_j of $q_k(x)$ satisfy $(1-\beta_j)/(1+\beta_j) = \alpha_j$, or

$$\beta_j = \frac{1-\alpha_j}{1+\alpha_j}.$$

It follows that all β_j are real and positive.

The roots of $Q_k(|z|, |w|)$ lying on the sphere $|z|^2 + |w|^2 = 1$ have the form

$$|z|^2 = \frac{\beta_j}{1+\beta_j}, \quad |w|^2 = \frac{1}{1+\beta_j}.$$

This defines tori in S^3, and together with P a torus knot. □

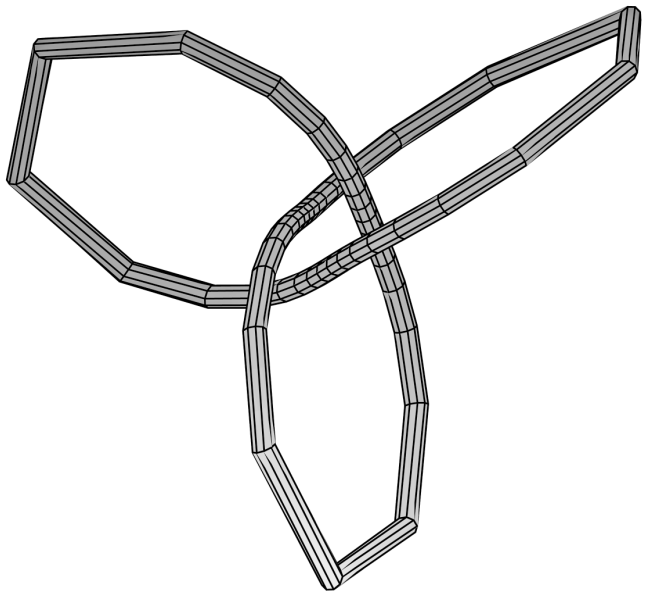


Figure 2: A trefoil (3,2) torus knot plotted using the above construction.

Conclusion and Future Work

We have discovered some interesting examples of knots generated by nodal sets of harmonic polynomials, notably links which are the fibres of the Hopf fibration and torus knots. Future work will attempt to probe the relationships, if any, between the polynomial properties and its associated knot. In particular, we are interested in the follwing questions.

(i) Which links and knots appear as nodal sets of complex-valued spherical harmonics on S^3, and what is the minimal degree of the corresponding harmonic.

(ii) For links and knots arising in (i), express the corresponding link and knot invariants through quantities related to eigenfunctions.

Acknowledgements

Thank you to Prof. Dima Jakobson for his patience and guidance. We would also like to thank NSERC and McGill University for funding this work.

References

[1] M. Berry. Knotted Zeros in the Quantum States of Hydrogen. Foundations of Physics, Vol. 31 (2001), No. 4, 659–667.

[2] A. Eremenko, D. Jakobson and N. Nadirashvili. On nodal sets and nodal domains on S^2 and $\mathbf{R}^2$. math.SP/0611627, to appear in Annales de l'Institut Fourier.

[3] T. Koornwinder. The addition formula for Jacobi Polynomials and Spherical Harmonics. Siam J. Appl. Math., Vol. 25, No. 2 (1973), 236–246.

[4] P. Kramer. An invariant operator due to F Klein quantizes H Poincarés dodecahedral 3-manifold. J. Phys. A: Math. Gen. 38 (2005) 3517-3540.

[5] L. Rudolph. Braided surfaces and Seifert ribbons for closed braids. Comment. Math. Helv. 58 (1983), no. 1, 1–37.

[6] L. Rudolph. Algebraic functions and closed braids. Topology 22 (1983), no. 2, 191–202.

Jokes

"Do you love your math more than me?"
"Of course not, dear - I love you much more."
"Then prove it!"
"OK... Let R be the set of all lovable objects..." □

A Few Problems in Analytic Number Theory

Maksym Radziwill

The four problems below belong to a branch of Number theory called Analytic Number Theory. My hope is that they give a glimpse of this beautiful subject.

1. What is the probability that two integers are coprime?

Suppose your friend picks a very large integer x (say, bigger than the number of atoms in the universe). You *randomly* choose two integers m, n in the interval $[1, x]$. What is the probability that $(m, n) = 1$? As your friend takes x bigger and bigger this probability will tend to $6/\pi^2$. In other words:

$$\lim_{x\to\infty} \frac{1}{x^2} \sum_{\substack{1 \leq m,\, n \leq x \\ (m,\, n)\, =\, 1}} 1 = \frac{6}{\pi^2} \tag{1}$$

There is another more leisurely way to state problem 1. Suppose you are placed at the origin of the plane $\mathbb{R}^2$. Suppose further that at each lattice point of $\mathbb{R}^2$ (each except the origin) there is a beautiful girl studying math. Each is beautiful and *different* from the others. If you draw a square with end vertices at $(\pm R, \pm R)$ how many distinct girls inside the square can you see without moving? Clearly not all of them! If you see the blonde with coordinate $(1, 1)$ you cannot see the brunette with coordinate $(2, 2)$. However, if someone lets $R \longrightarrow \infty$ the percentage of girls in the square that *you* (or as a matter of fact anybody else at the origin) will be able to see, will tend to $6/\pi^2$. Now, such a useful thing clearly needs a proof. Here it is.

Proof. We are going to prove (1). The Moebius function μ is the unique function $\mathbb{N} \longrightarrow \mathbb{C}$ such that for all integers $n \in \mathbb{N}$ the relation

$$\sum_{d|n} \mu(d) = \begin{cases} 1 \text{ if } n = 1 \\ 0 \text{ otherwise} \end{cases}$$

holds. (Dear reader : if you bother, find an explicit form for $\mu(n)$ and note that $|\mu(n)| \leq 1$ for all $n \in \mathbb{N}$. We'll use that last relation.) In particular,

$$\sum_{\substack{d|m \\ d|n}} \mu(d) = \begin{cases} 1 \text{ if } (m, n) = 1 \\ 0 \text{ otherwise} \end{cases}$$

for all $m, n \in \mathbb{N}$. The above observations gives that

$$\begin{aligned} \sum_{\substack{m,\, n \leq x \\ (m,\, n)\, =\, 1}} 1 &= \sum_{m,n \leq x} \sum_{\substack{d|m \\ d|n}} \mu(d) \\ &= \sum_{d \geq 1} \mu(d) \sum_{\substack{m,\, n \leq x \\ d|m \\ d|n}} 1 \\ &= \sum_{d \leq x} \mu(d) \left(\sum_{\substack{m \leq x \\ d|m}} 1 \right)^2 \end{aligned} \tag{2}$$

Exactly $\lfloor x/d \rfloor = x/d + O(1)$ integers $m \leq x$ are divisible by d. Therefore (2) becomes

$$\sum_{\substack{m,\, n \leq x \\ (m,\, n)\, =\, 1}} 1 = x^2 \sum_{d \leq x} \frac{\mu(d)}{d^2} + O\left(x \sum_{d \leq x} \frac{1}{d} \right) \tag{3}$$

Since $1/t$ is stricly decreasing and continuous, the bound

$$\sum_{d \leq x} \frac{1}{d} = O\left(\int_1^x \frac{\mathrm{dt}}{t} \right) = O\left(\ln x \right)$$

holds. We're almost finished. We want to replace

$$\sum_{d \leq x} \frac{\mu(d)}{d^2} \quad \text{by} \quad \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}$$

By doing so we over/under estimate by

$$\sum_{d > x} \frac{\mu(d)}{d^2} = O\left(\sum_{d > x} \frac{1}{d^2} \right) = O\left(\int_x^{\infty} \frac{\mathrm{dt}}{t^2} \right) = O\left(\frac{1}{x} \right)$$

It follows that (3) is equal to $6x^2/\pi^2 + O(x \ln x)$. □

2. How many squarefree integers $\leq x$?

In some (naive) sense this problem is the "inverse" of problem 3. How would we approach it? Any integer n can be written as $n = ab^2$ with a squarefree. Now $n = ab^2$ is squarefree if and only if $b = 1$. Is the Moebius function of any help here? Yes. Just remark that, for $n = ab^2$ with a squarefree,

$$\sum_{d^2|n} \mu(d) = \sum_{d|b} \mu(d) = \begin{cases} 1 \text{ if } b = 1 \text{ (i.e. } n \in \mathcal{S}) \\ 0 \text{ otherwise} \end{cases}$$

where $\mathcal{S}$ is the set of squarefree integers. Therefore, the number of squarefree integers $\leq x$ is,

$$\begin{aligned}\sum_{n\leq x}\sum_{d^2|n}\mu(d) &= \sum_{d\leq\sqrt{x}}\mu(d)\sum_{\substack{n\leq x\\ d^2|n}}1\\ &= \sum_{d\leq\sqrt{x}}\mu(d)\left(\frac{x}{d^2}+O(1)\right)\\ &= x\sum_{d\leq\sqrt{x}}\frac{\mu(d)}{d^2}+O\left(x^{1/2}\right)\end{aligned}$$

Again, we want to replace,

$$\sum_{d\leq\sqrt{x}}\frac{\mu(d)}{d^2} \quad \text{by} \quad \sum_{d=1}^{\infty}\frac{\mu(d)}{d^2}=\frac{1}{\zeta(2)}$$

Doing so, we make an error of

$$O\left(\sum_{d>\sqrt{x}}\frac{1}{d^2}\right)=O\left(\int_{\sqrt{x}}^{\infty}\frac{\mathrm{d}t}{t^2}\right)=O\left(\frac{1}{\sqrt{x}}\right)$$

It follows that the number of squarefree integers less than x is equal to $x/\zeta(2)+O\left(x^{1/2}\right)$.

3. How many powerful integers $\leq x$?

We say that an integer $n\in\mathbb{N}$ is powerful if all the exponents in its prime factorization are greater than 2. All squares are powerful. Are there a lot more of powerful numbers than squares? No. Not even three times more. This glorious fact certainly deserves a proof. To get started: is there an explicit way to write down a powerful number? Note that any integer $\alpha\geq 2$ can be written as $2n+3m$ with $n\geq 0$ and $m\in\{0,1\}$. Thus, any powerful number can be written as n^2m^3 with m squarefree. Hence, the number of powerful integers $\leq x$ is,

$$\begin{aligned}\sum_{\substack{n^2m^3\leq x\\ m\in\mathcal{S}}}1 &= \sum_{\substack{m\leq x^{1/3}\\ m\in\mathcal{S}}}\sum_{n\leq\sqrt{x/m^3}}1 \qquad (4)\\ &= \sqrt{x}\sum_{\substack{m\leq x^{1/3}\\ m\in\mathcal{S}}}\left(\frac{\sqrt{x}}{m^{3/2}}+O(1)\right)\\ &= \sqrt{x}\sum_{\substack{m\leq x^{1/3}\\ m\in\mathcal{S}}}m^{-3/2}+O\left(x^{1/3}\right)\end{aligned}$$

We use our old trick and replace the finite sum above by the infinite sum

$$\sum_{m\in\mathcal{S}}m^{-3/2}=\frac{\zeta(3/2)}{\zeta(3)}$$

This operation costs

$$O\left(\sum_{m>x^{1/3}}\frac{1}{m^{3/2}}\right)=O\left(\int_{x^{1/3}}^{\infty}\frac{\mathrm{d}t}{t^{3/2}}\right)=O\left(x^{-1/6}\right)$$

We conclude that (4) is equal to $\left(\frac{\zeta(3/2)}{\zeta(3)}\right)\cdot\sqrt{x}+O\left(x^{1/3}\right)$.

4. How many abelian groups of order $\leq x$?

We will need some heavy machinery for this one! Let $a(n)$ denote the number of abelian groups of order n. By the fundamental theorem (for finite abelian groups) we know that $a(\cdot)$ is (weakly) multiplicative. Moreover for any prime p and exponent $\alpha\geq 0$ we have $a(p^\alpha)=P(\alpha)$, where $P(\alpha)$ is the number of partitions of α. The number of partitions of α corresponds to the number of solutions in $\mathbb{N}$ to $\alpha=x_1+2x_2+3x_3+\ldots$.

For complex $|z|<1$,

$$\begin{aligned}\sum_{n=1}^{\infty}P(n)z^n &= \sum_{x_1,x_2,\ldots\in\mathbb{N}}z^{x_1+2x_2+3x_3+\ldots}\\ &= \left(\sum_{x_1\in\mathbb{N}}z^{x_1}\right)\left(\sum_{x_2\in\mathbb{N}}z^{2x_2}\right)\ldots\\ &= \prod_{j=1}^{\infty}\frac{1}{1-z^j}\end{aligned}$$

Now, for complex $\mathrm{Re}\,s>1$, (the product below is taken over the primes),

$$\begin{aligned}\sum_{n=1}^{\infty}\frac{a(n)}{n^s} &= \prod_p\left(\sum_{j=0}^{\infty}\frac{a\left(p^j\right)}{p^{js}}\right)\\ &= \prod_p\left(\sum_{j=0}^{\infty}P(j)\left(\frac{1}{p^s}\right)^j\right)\\ &= \prod_p\prod_{j=1}^{\infty}\frac{1}{1-p^{-sj}}=\prod_{j=1}^{\infty}\zeta\left(sj\right)\end{aligned}$$

where in the last line we interchanged the two products. (The motivated reader can check that this is legitimate). Let's define $b(n)$ such that $\sum_{n=1}^{\infty}b(n)n^{-s}=\zeta(2s)\zeta(3s)\ldots$. Thus $\sum_{d|n}b(d)=a(n)$ and

$$\sum_{n\leq x}|b(n)| \quad =O_\varepsilon\left(x^{1/2+\varepsilon}\right)$$

for any fixed $\varepsilon>0$ (the implicit constant in the big $-O$ depends on ε). The last estimate is a consequence of the (absolute) convergence of $\sum_{n=1}^{\infty}b(n)n^{-s}$ for all $\mathrm{Re}\,s>1/2$.

Now, we are ready to *charge*. The number of (non-isomorphic, of course) abelian groups of order $\leq x$ is

$$\begin{aligned}\sum_{n\leq x}\sum_{d|n}b(d) &= \sum_{d\leq x}b(d)\sum_{\substack{n\leq x\\ d|n}}1\\ &= \sum_{d\leq x}b(d)\left(\frac{x}{d}+O(1)\right)\\ &= x\sum_{d\leq x}\frac{b(d)}{d}+O_\varepsilon\left(x^{1/2+\varepsilon}\right)\end{aligned}$$

We replace the finite sum

$$\sum_{d\leq x}\frac{b(d)}{d} \quad \text{by} \quad \sum_{d=1}^{\infty}\frac{b(d)}{d}=\prod_{j=2}^{\infty}\zeta(j)$$

making an error of

$$\begin{aligned} O\left(\sum_{d>x}\frac{|b(d)|}{d}\right) &= O\left(\frac{1}{x^{1/2-\varepsilon}}\cdot\sum_{d>x}\frac{|b(d)|}{d^{1/2+\varepsilon}}\right) \\ &= O\left(x^{-1/2+\varepsilon}\right) \end{aligned}$$

where the last estimate follows from the convergence of $\sum_{d=1}^{\infty}|b(d)|d^{-1/2-\varepsilon}$. We conclude that the number of abelian groups $\leqslant x$ is

$$\left(\prod_{j=2}^{\infty}\zeta(j)\right)\cdot x \quad + \quad O_{\varepsilon}\left(x^{1/2+\varepsilon}\right)$$

With a little more work Erdos and Sarkozy showed that

$$\sum_{1\leq n\leq x}a(n)=\left(\prod_{j=2}^{\infty}\zeta(j)\right)\cdot x+O\left(x^{1/2}\right).$$

For an more precise estimate, see Ivic's book.

Concluding remarks

In the end, what is Analytic number theory? The answer, if such a thing exists, is approximated by the following quote of Henryk Iwaniec (one of the great Analytic number theorists of this century): *"Analytic number theory pursues hard classical problems of an arithmetical nature by means of best available technologies from any branch of mathematics, and that is its beauty and strength. Analytic number theory is not driven by one concept; consequently it has no unique identity."*

Let us note that the problems given above, altough not hard (if we are satisfied with the remainder terms as given) and folklore, contains the germs of extremely deep problems. Improving the error term in problem 2, requires the use of analytical techniques. In fact, establishing a very sharp error term is a problem at the depth of the Riemann Hypothesis. The estimates in problems 2, 3 and 4 can all be improved by the use of complex analysis. (see [2])

It is my hope that the reader perceive a certain coherence in the demarche used to solve the above problems, so that Iwaniec's remark gains in strength. It took me a while to realize how true and wonderful it is!

References

[1] Edmund Hlawka. *Geometric and analytic number theory* Springer-Verlag, c1991.

[2] Harold Davenport. *Multiplicative Number Theory.* Springer, third edition, 2000.

[3] Aleksandar Ivic. *Lectures on mean values of the Riemann Zeta function* Springer-Vlg, c1991.

Jokes

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 lecturer: "Now we'll prove the theorem. In fact I'll prove it all by myself." □

Q: What does an analytic number theorist say when he's drowning?
A: Log-log, log-log, log... □

Q: Why can't you grow wheat in $\mathbb{Z}/6\mathbb{Z}$?
A: It's not a field. □

Q: Why is the integral along the contour of Western Europe zero?
A: Because all the Poles are in the Eastern part. □

What is the difference between an argument and a proof? An argument will convince a reasonable man, but a proof is needed to convince an unreasonable one. □

Graduate Studies: Applications and Beyond

Leonid Chindelevitch

Those of you who are entering your last year of undergraduate studies are probably considering different options for what lies ahead. Those of you who still have a year or two to go may not be so worried yet, but nevertheless the question of where to go next invariably comes to mind from time to time. And when you are dealing with existential questions of this kind, it always helps to know that you are not alone, and that other people have gone through the same ordeal.

Since enrolling in graduate school is certainly one of the most popular options, In this short article I am going to answer some questions that you will undoubtedly be asking yourselves when the time comes to apply to graduate school. My hope is that it will help you avoid some of the difficulties that I had to face at this time two years ago.

But first, some background questions to give you an idea of where I am coming from.

What did you do in undergrad? I was a McGill undergraduate student in the Joint Honours program in Mathematics and Computer Science from 2003 to 2006. What convinced me to go on to graduate school were my three successful NSERC-USRA summer research experiences.

What university are you currently enrolled in? I am starting the second year of a PhD program in Applied Mathematics at MIT (Massachusetts Institute of Technology - if you apply there, make sure you get the name right). That's in the U.S., but only a 5-6 hour drive away from Montreal.

What are you specializing in? My areas of application are biology and linguistics. This is a little bit unusual (the "standard" areas of application for mathematics being physics, computer science, and combinatorics), but that is what matches my research interests.

What are the good schools in your field of studies? Generally speaking, there are a lot of excellent universities in Canada and the United States, in both pure and applied mathematics. In Canada, McGill, University of Toronto and University of British Columbia are particularly strong in mathematics (both pure and applied), while Waterloo and Simon Fraser University are very strong in applied mathematics. In the US, the top universities for pure math would be Harvard, MIT, Princeton, Berkeley and Stanford; for applied math: MIT, New York University, Berkeley, Brown, Johns Hopkins, and I am certainly forgetting some others. You can consult the latest annual ratings available at the Careers and Placement Center (CAPS), located in the Brown building. I also used their help to improve my personal statement (it's a free service, so take advantage of it)!

When did you first think about where you wanted to go? I believe that I first heard about MIT in my first year of undergrad. The context was that it's a well-reputed school which is very difficult to get in. It was at that time that I decided to make it a challenge for myself to get into graduate school there. I like setting high goals, although I didn't believe I could achieve this one until I actually got my acceptance letter.

When should we start thinking about it? It's never too early to start thinking about where you would like to go. However, you should definitely start planning your application process in September of your last year. If you wait longer, things may get difficult (especially if you are applying to universities in the US). Also, make sure to apply for funding before the deadlines in mid-October!

And now, the questions which are directly related to the application process. I would like to mention that my friend Marco Carone, whose story is similar to mine, has written an excellent and very detailed overview of the application process, which is available online[2].

When should we start applying? Different universities have different deadlines. If you are going to be applying to US universities, their deadlines are earlier (typically around mid-December). Canadian universities have later deadlines (usually February or March), but if you can get all of your applications out of the way before Christmas break, you will save yourself a lot of time in your last semester. I did everything just before the deadlines, but in retrospective, I wish I hadn't.

How much time does it take to apply? This depends on the number of universities you apply to. I applied to six (and I wouldn't recommend applying to more than 10). The first application is always the hardest. The statement of purpose, the biographical information, the reference letters, the GRE scores, the textbook information - believe it or not, some schools actually ask you to list all the textbooks you have used during your undergraduate career - all these take a long time to put together. Each subsequent

[2] www.math.mcgill.ca/students/undergrad/gradschoolapp

one gets easier, but it is still a considerable time investment. On average, I would estimate about 5-10 hours per university (extended over a period of several days, because of all the different components).

Finally, some questions related to graduate studies in general, and PhD studies in the US in particular.

How does one pay for one's graduate studies? Usually, the university will mention the conditions of your admission in your acceptance letter; in particular, you will be notified about the source of funding for your studies. Typically, your tuition will be free (covered by the department); you will also receive some money for living expenses. In exchange, the university will usually impose some teaching duties on you, such as grading papers or teaching a section in one of the undergraduate courses. Since grading is not as much fun as teaching, make sure to find out what requirements you need to fulfill to get to teach.

Also, you may or may not be able to take the funding that you receive from NSERC abroad (FQRNT is usually more flexible in that respect); you may also have to negotiate with your university. Some universities will allow you to combine external funding with departmental funding, some will not (in order to ensure everyone is at the same level); still others (like MIT) will settle on a compromise (e.g. a lower departmental funding plus your external funding).

What's more important in considerering a grad school: good researchers, good teaching, good grad students? One of my friends (a professor) once told me that, although you learn a lot from your teachers, you learn even more from your peers. Of course, in the ideal world, you would have everything - great students, great professors and exciting research. In reality, there may be a slight trade-off. My suggestion is to visit all the universities where you have been accepted and that you are considering (they will usually pay for your travel expenses, even the US ones) and talk to both the professors that you are interested in and to the students. Most importantly, you have to be in an environment which stimulates you, and you are the only one who can decide that. Don't take other peoples' word for it!

What are the pros and cons of doing a PhD without doing a Masters? If you are sure that a PhD is what you ultimately want, then going into a PhD program directly (which is what usually happens in the US) is a good option. Despite the common misconception, a PhD does not mean you have to continue in academia afterwards - it actually gives you a lot of options. The disadvantage is that you have to go all the way, as you may not get any diploma otherwise (in a way, you can think of a Masters degree as a safety net for that case). Many universities (including MIT) have so-called "qualifying exams", and if you fail them twice, you leave without a degree; my qualifying exams are coming up in two weeks, so I really shouldn't be writing this right now... oh well.

On the other hand, if your research background is not as strong as you would like it to be, or if you are not sure that you would like to go all the way to a PhD, then doing a Masters degree first is preferable. This is what typically happens in Canadian universities (although you can get a Masters degree in some US universities too). The disadvantage is that overall, it may take a year longer to get your PhD if you go through the Masters program first.

What kind of undergraduate students are fit for a PhD in mathematics? That's a provocative question! As the goal of most PhD programs is to produce original research (whether in the form of a thesis or otherwise), the most important "fitness" criterion is an interest and an ability to do research. The other qualities you will need are perseverance, patience and a willingness to work long hours when necessary; if you do not mind spending several hours on a challenging assignment problem, then you are an excellent PhD candidate.

Are extracurricular activities important to get accepted in mathematics? Unfortunately, I don't know. I am still not sure what it was in my resume that got me in, but I am certain that my involvement in extracurricular activities did not hinder my cause (I was actively involved in the SUMS council, was the president of the 2006 CUMC, and also ran or helped run several student organizations). From what I understand, the universities look most of all at your research ability, but it also helps if you show you are a "well-rounded" person.

Compare MIT students with McGill students. I haven't talked to enough graduate students at McGill to get a representative sample. The main difference between the undergraduates at McGill and those at MIT is that those at McGill actually sleep...

Are you as involved in the MIT student life as you were here? Unfortunately, not as much, but it's not for lack of opportunity. I am involved in a lot of extracurricular activities and clubs, but mostly as an active member rather than an executive (with the exception of the MIT Esperanto Club).

How was the transition from the Canadian to the American education system? Relatively smooth - the one thing that really gets on my nerves is the use of the imperial system in the US. From the experience of my friends who are doing graduate studies in Canada, I would say that the difference between undergrad and graduate school is more significant than the difference between the Canadian and the American education systems.

What's the meaning of life? The meaning of life is to find the meaning of life! Sorry about the infinite recursion...

Credits

The Delta-Epsilon Editing Team

In alphabetical order

- Agnès F. Beaudry
- Ioan Filip
- Michael McBreen
- Alexandra Ortan
- Vincent Quenneville-Bélair
- Nan Yang

Cover Art and Design

- Michael McBreen
- Mathieu Ménard

Acknowledgements

Now has come the time to say some *thanks*! We would like to thank professor Eyal Goren, professor Benoit Charbonneau and Leonid Chindelevitch for taking the time to speak with us and for all the great advice they sent out to the undergraduates. Thank you SUMS for your multi-dimensional support; if it would be as elegant, we truly believe that you would give us infinitely many parameters to work with. We are grateful to the Science Undergraduate Society, the Arts Undergraduate Society, the Faculty of Science, the Department of Mathematics and Statistics, the Institut des sciences mathématiques and the Centre de recherche mathématiques for providing the funding. We also take this occasion to thank the Science Undegraduate Society for awarding us the Best Student Publication Award for last year's issue. We also would like to thank Alex Dergachev for having freely let us use his server without which editing a journal is a mess. Finally, thank you, reader, for the thing you are defined to do! Without you, needless to say, we would be useless.