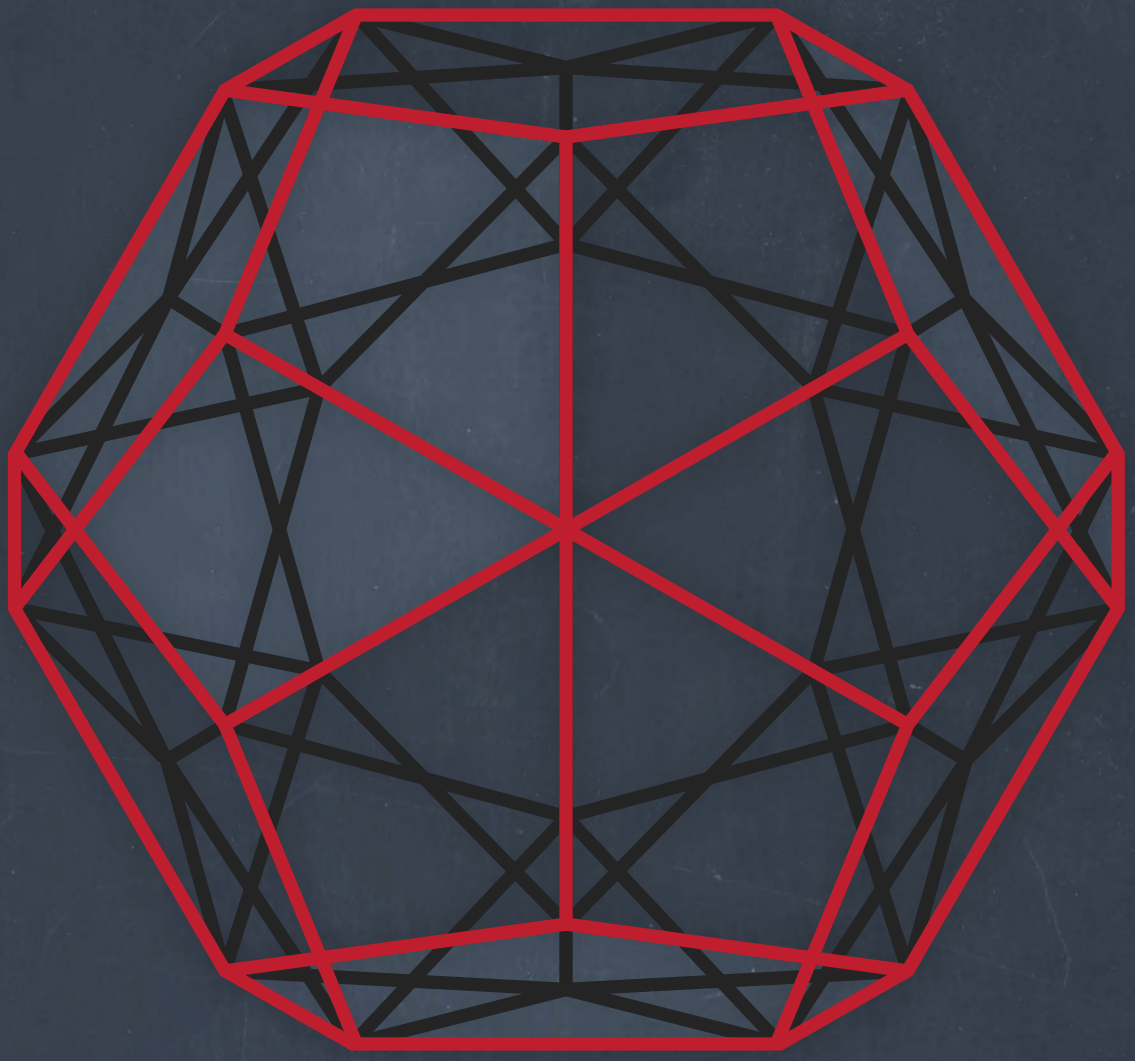
McGill



McGill Undergraduate Mathematics Journal

the δelta ε psilon

Spring 2014

Seventh Issue

Issue 7, published March 28th, 2014
ISSN: 1911-8996

THE δELTA-εPSILON

MCGILL UNDERGRADUATE MATHEMATICS JOURNAL

CONTENTS

LETTER FROM THE EDITORS

The Delta-Epsilon team has faced difficult times in the past two years, with a delayed publication and a sharp decline in the number of submissions. Thanks to the hard work of our volunteers, we are nevertheless able to perpetuate the tradition and present you the 7^{th} edition. There is hope that continuation will be possible, with more enthusiasm from the rapidly increasing cohort of mathematics students at McGill. We believe this issue will provide interesting readings and allow incoming students to discover new research fields. The Delta-Epsilon takes great pride in offering undergraduate students the possibility to get a first access to the article submission process, graduate peers reviewers a chance to exploit their skills and the editing team experience in publishing.

Particular efforts have been put into improving graphical readability and the fluency of the reviewing, despite lack of submissions and many missed deadlines. We are as always looking for volunteers interested in helping the undergraduate mathematical community; please contact us if you are eager to participate or give a hand. I invite U1 and U2 students to look at the past issues online. Everyone is invited to submit their work, whether research, neat explanation of small piece of mathematics, interview or expository material. Share your passion with others: submit to the Delta-Epsilon.

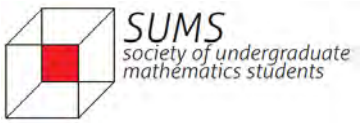
Léo Belzile, Editor
(on behalf of the δelta-εpsilon editing team)
thedeltaepsilon@gmail.com

LETTER FROM SUMS

If mathematics happened in a concrete building would it exist?
The act of doing math is rather difficult to define, but the current state of math would never be possible without the ability to read and write. It follows that math is founded on the idea of sharing, and we are taught to share usually no later than at the undergraduate level.

With the seventh issue the δelta-εpsilon continues to be a testimony to the high quality research being done by McGill's undergraduate math student body and the passion they have for sharing their love of the subject with others. The Society of Undergraduate Mathematics Students (SUMS) is proud to see the efforts of both the contributors and editors come together into this inspiring journal.
On behalf of SUMS congratulations to the δelta-εpsilon team!

Thomas Ng, SUMS President
(On behalf of SUMS council)
`http://sums.math.mcgill.ca/`



SOLVING PELL'S EQUATION USING ORTHOGONAL TRANSFORMATIONS

Nikita Lvov and Xi Sisi Shen

We devise a method to solve a generalized version of Pell's equation using the theory of orthogonal matrices. It will be shown that infinitely many solutions can be found provided that at least one nontrivial solution exists and is known.

Consider the problem of finding integers (x, y) that solve the equation:

$$x^2 - 2y^2 = 1 \tag{1}$$

The solutions $(x, y) = (3, 2)$ and $(1, 0)$ immediately jump out. After some thought, one might realize that $(17, 12)$ is also a solution and that $(99, 70)$ is another one.

A natural question one might then ask is if there exist infinitely many solutions. Equation 1 is known as a specific instance of Pell's equation, $x^2 - ny^2 = 1$, and is indeed well-known[1] to admit infinitely many solutions, provided that n is not a perfect square.

In this article, we shall demonstrate, by means of a simple method, that eq. 1 does have infinitely many solutions and that, much more generally, a relatively wide class of equations satisfy the following theorem:

Theorem 1. *Suppose an equation of the form*

$$x_1^2 - \sum_{k=2}^{n} a_k x_k^2 = 1 \tag{2}$$

where $\{a_k\}$ are strictly positive integers and $n > 1$, admits one solution in the strictly positive integers. Then, for any $N > 0$, it also admits an integer solution $\{x_i\}_{i=1}^n$ that satisfies $x_i > N\ \forall i$. Consequently this equation must have infinitely many integer solutions.

We will prove this statement by first finding a transformation that takes the set P of positive integer solutions (defined as the set of all n-tuplets of positive integers that solve eq. 2) into itself. We will then show that applying this transformation reiteratively to the given solution generates infinitely many solutions that meet the condition $x_i > N\ \forall i$.

In order to find a way to derive such a transformation, let us first consider the simpler and seemingly unrelated problem of finding linear transformations $M : \mathbb{R}^n \to \mathbb{R}^n$ that preserve the Euclidean vector norm, *i.e.* those for which $\|Mv\| = \|v\|, \forall v \in \mathbb{R}^n$.

Such an M should satisfy, $\forall v \in \mathbb{R}^n$:

$$\begin{aligned} (Mv)^\top I(Mv) &= v^\top Iv \\ v^\top(M^\top IM - I)v = 0 \quad &\Rightarrow \quad M^\top IM = I \end{aligned}$$

The last relation for M is easily recognized as the one describing the set of orthogonal matrices, denoted by $O(n)$. More generally, we shall denote by $O_s(n)$ the set of all $n \times n$ matrices M such that:

$$M^\top SM = S$$

for a given symmetric matrix S. It is easy to see that if $M \in O_s(n)$, then $(Mv)^\top S(Mv) = v^\top Sv$ for any $v \in \mathbb{R}^n$. It is also useful here to observe that the composition of two elements of $O_s(n)$, say, M and N, is another element of $O_s(n)$ since

$$(MN)^\top S(MN) = N^\top(M^\top SM)N = N^\top SN = S$$

We can now consider the question of how to generate elements of $O_s(n)$. To answer this, we return to the Euclidean case that was first considered and we note that reflection about any hyperplane (that passes through the origin) is a linear transformation that preserves the Euclidean norm of any vector in $\mathbb{R}^n$. Indeed, taking e to be the unit normal to the hyperplane of reflection, the action T of the reflection upon $v \in \mathbb{R}^n$ can be described by:

$$T(v) = v - 2e\langle e, v\rangle$$

where $\langle e, v\rangle = e^\top Iv$ is the Euclidean inner product of e and v. It is now straightforward to verify that

$$\begin{aligned} \langle T(v), T(v)\rangle &= \langle v - 2e\langle e, v\rangle, v - 2e\langle e, v\rangle\rangle \\ &= \langle v, v\rangle \end{aligned}$$

for any vector $v \in \mathbb{R}^n$. Thus, T gives rise to an orthogonal matrix.

In fact, it is easy to see that exactly the same holds if the inner product $\langle\cdot,\cdot\rangle$ is replaced by the inner product $\langle\cdot,\cdot\rangle_s$, defined as $\langle u, w\rangle_s = u^\top Sw$,

[1] see for example Section 7.3 of Chapter 2 of Borevich and Shafarevich (1966)

provided that e now satisfies $\langle e, e\rangle_s = e^\top Se = 1$, *i.e.* we have

$$[v - 2e(e^\top Sv)]^\top S[v - 2e(e^\top Sv)] = v^\top Sv$$

$\forall v \in \mathbb{R}^n$, for any symmetric matrix S, and any e such that $e^\top Se = 1$. This gives us the following lemma:

Lemma 2. *If $e \in \mathbb{R}^n$ satisfies $e^\top Se = 1$ and S is a symmetric matrix, then the matrix $M_e = I - 2ee^\top S$ belongs to $O_s(n)$.*

Let us momentarily pause to see how this works for the case of $\mathbb{R}^2$ and $S = I$. If $e^\top Ie = 1$ for a vector $e \in \mathbb{R}^2$, then e has the general form $e = (\cos\theta, \sin\theta)$. The resulting orthogonal matrix associated with e is hence

$$\begin{aligned} M_e &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - 2\begin{pmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{pmatrix} \\ &= \begin{pmatrix} -\cos 2\theta & -\sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{pmatrix} \\ &= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos 2\theta & \sin 2\theta \\ -\sin 2\theta & \cos 2\theta \end{pmatrix} \end{aligned}$$

The matrix M_e therefore describes a rotation composed with a reflection, and is indeed orthogonal. Slightly more interesting is the case when

$$S = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The general form of a vector $e \in \mathbb{R}^2$ for which $e^\top Se = 1$ is now $e = (\cosh\beta, \sinh\beta)$. A computation analogous to the one above shows that the matrix M_e associated with this vector has the following form:

$$\begin{aligned} M_e &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \\ &\quad 2\begin{pmatrix} \cosh^2\beta & \cosh\beta\sinh\beta \\ \sinh\beta\cosh\beta & \sinh^2\beta \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} \cosh 2\beta & \sinh 2\beta \\ \sinh 2\beta & \cosh 2\beta \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

This describes a reflection composed with a transformation that is known in special relativity as a Lorentz transformation.

Let us now apply Lemma 2 to find the desired transformation that takes positive integer solutions of eq. 2 to positive integer solutions. To make notation more concise, we shall understand the phrase "a positive matrix" from here on to refer to a matrix with all entries positive. Furthermore, a vector in $\mathbb{R}^n$ will be termed "positive solution" if its coordinates are positive integers that solve eq. 2. A solution will thus be denoted either as a vector, say x, or as an n-tuplet of integers, say $\{x_i\}_{i=1}^n$.

First of all, we can observe that $x = \{x_i\}_{i=1}^n$ is a solution of eq. 2 if and only if $x^\top Sx = 1$ where

$$S = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & -a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -a_n \end{pmatrix}$$

Hence, in order for a transformation $\mathbb{R}^n \to \mathbb{R}^n$ which will take positive solutions of eq. 2 to positive solutions, it is sufficient for its associated matrix M to belong to $O_s(n)$ and to have only positive integer entries.

Now, suppose that eq. 2 admits at least one solution in the strictly positive integers. Therefore, there exists a vector of strictly positive integers e_1 that satisfies $e_1^\top Se_1 = 1$. Furthermore, e_1 is distinct from the vector $e_0 \coloneqq (1, 0, \ldots, 0)$ satisfying $e_0^\top Se_0 = 1$. Denoting the i^{th} entry of e_1 by $(e_1)_i$, we have that

$$[(e_1)_1]^2 = 1 + \sum_{i=2}^{n} a_i[(e_1)_i]^2$$

and therefore, since (e_1) is positive and integer-valued, $(e_1)_1 \geq 2$. Of course, we also have that $(e_1)_i \geq 1$ for all i.

We know from Lemma 2 that $M_{e_1} = I - 2e_1e_1^\top S$ belongs to $O_s(n)$. It is also clear that M_{e_1} has only integer entries. Thus, it satisfies all the requirements we desire except for positivity. The latter can be obtained using e_0. As $e_0^\top Se_0 = 1$, e_0 also gives rise to an element $M_{e_0} = 1 - 2e_0e_0^\top S$ of $O_s(n)$. Since the composition of two elements of $O_s(n)$ is another element of $O_s(n)$,

$$M = M_{e_1}M_{e_0} = (I - 2e_1e_1^\top S)(I - 2e_0e_0^\top S) \in O_s(n)$$

Let us now show that M is positive. It is easy to see that $I - 2e_0e_0^\top S$ is simply the matrix

$$\begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

which we can denote by I'. Substituting this gives:

$$\begin{aligned} M = M_{e_1}M_{e_0} &= (I - 2e_1e_1^\top S)I' \\ &= I' + 2e_1e_1^\top(-SI'). \end{aligned}$$

The matrix $-SI'$ is a diagonal matrix with positive entries on the diagonal, whose first entry is equal to 1. Hence, $2e_1e_1^\top(-SI')$ is a positive matrix whose first term is $2[(e_1)_1]^2 \geq 8$ and whose i^{th} entry on the diagonal is bounded below by $2[(e_i)_i]^2 \geq 2$. This implies that M itself is a positive matrix with $M_{11} \geq 8-1 = 7$ and $M_{ii} \geq 2+1 = 3$. Also, M belongs to $O_s(n)$ and has integer entries. Thus, takes positive integer solutions of eq. 2 to positive integer solutions.

Therefore, $M^k e_1$ is a positive integer solution to eq. 2 for any k, and the i^{th} coordinate of this solution satisfies

$$\begin{aligned}(M^k e_1)_i &\geq (M^k)_{ii}(e_1)_i \geq (M_{ii})^k (e_1)_i \\ &\geq (M_{ii})^k \geq 3^k\end{aligned}$$

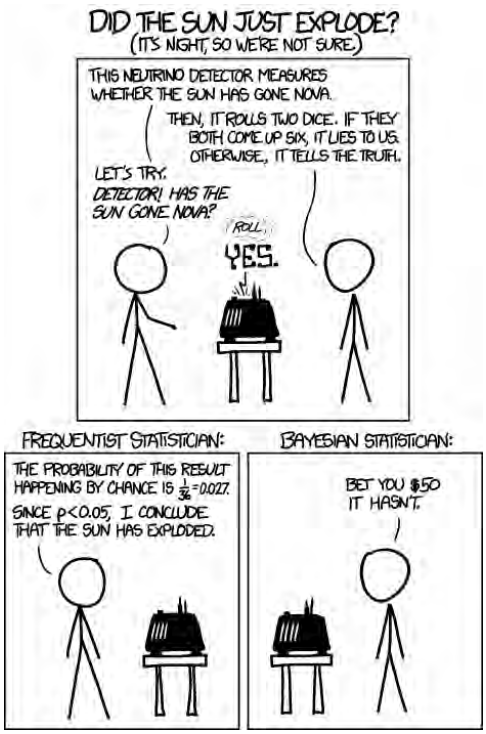
This implies that $\{M^k e_1\}_{k=1}^{\infty}$ is indeed an infinite family of solutions to eq. 2. Furthermore, for any $N > 0$, we can choose k sufficiently large that $(M^k e_1)_i > N \; \forall i$, thus proving Theorem 1.

It is worth noting that the above inequalities hold because the matrices in question are non-negative. Should this requirement be relaxed, we could still obtain non-negative integer solutions (just by changing the sign of the negative entries in the solution vectors). However, it would be possible that $M^N e_1 = M^K e_1$ for some finite distinct $N > 0$ and $K > 0$ and we would not be able to generate infinitely many solutions and conclude the theorem.

REFERENCES

Borevich, A. I. and Shafarevich, I. R. (1966). *Number theory*. Translated from the Russian by Newcomb Greenleaf. Pure and Applied Mathematics, Vol. 20. Academic Press, New York.

JOKES AND COMICS



xkcd 1133: *Frequentists vs Bayesians*
'Detector! What would the Bayesian statistician say if I asked him whether the–' [roll] 'I AM A NEUTRINO DETECTOR, NOT A LABYRINTH GUARD. SERIOUSLY, DID YOUR BRAIN FALL OUT ? ' [roll] '... yes.'

An Introduction to Belyi Surfaces

Matthew Stevenson

We outline the basic theory of Belyi surfaces, up to Belyi's theorem (Belyĭ (1979)), which characterizes these spaces as precisely those Riemann surfaces that are defined over $\overline{\mathbb{Q}}$. We then detail the Brooks-Makover construction of a Belyi surface from an oriented cubic graph. Consequently, we can apply the model of Bollobás in order to randomly pick a Belyi surface. Finally, we briefly explain the relationship between Belyi surfaces and Grothendieck's *dessins d'enfants*.

1 Belyi Surfaces

A Riemann surface is a 1-dimensional complex manifold *i.e.* a connected orientable surface such that every point has a neighborhood homeomorphic to $\mathbb{C}$. Some common examples are the Riemann sphere $\mathbb{P}^1(\mathbb{C})$, a surface of genus 2 (pictured below), or the analytic continuation of a holomorphic function (e.g. $f(z) = \log z$).

Figure 1: A surface of genus 2

Let $f\colon S_1 \to S_2$ be a nonconstant holomorphism between two Riemann surfaces. We say that f is an n-sheeted branched covering if there exists a finite subset $C(f) \subset S_2$ such that:

- if $x \in S_2 - C(f)$, then $|f^{-1}(y)| = n$
- if $x \in C(f)$, then $1 \leq |f^{-1}(y)| < n$

We call $C(f)$ the set of *critical values* of f, and say that f is *ramified* or *branched* over $x \in C(f)$. Note that outside of these degenerate points, f is a degree n covering map.

Suppose S is a Riemann surface. Consider such a map $\beta\colon S \to \mathbb{P}^1(\mathbb{C})$. If $C(\beta) \subseteq \{0, 1, \infty\}$ then β is said to be a *Belyi function*, and in this case S is called a Belyi surface.

Consider the map $\beta\colon \mathbb{P}^1(\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})$ given by $z \mapsto z^n$ for any integer $n \geq 1$. Clearly β is nonconstant and holomorphic. Moreover, for any fixed $w \in \mathbb{P}^1(\mathbb{C}) - \{0, \infty\}$, the equation $z^n = w$ has n distinct solutions by the Fundamental Theorem of Algebra. However, $\beta^{-1}(0) = \{0\}$ and $\beta^{-1}(\infty) = \{\infty\}$, so β is ramified at $\{0, \infty\}$. In particular, β is a Belyi function and the Riemann sphere is a Belyi surface.

Now that we understand the most basic Belyi surface, let us describe Belyi's incredible characterization of a general Belyi surface. First, note that certain Riemann surfaces arise as the zero set of a two-variable complex polynomial. For a subfield $K \subset \mathbb{C}$, a *Riemann surface over K* is an irreducible polynomial in $K[x, y]$. For example, the elliptic curve

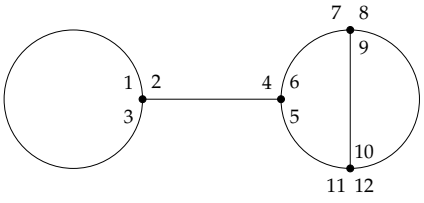
$$y^2 = x^3 + ax + b\colon a, b \in \mathbb{Q}$$

defines a torus over $\mathbb{Q}$, since all of the coefficients are rational.

In 1979, Belyĭ proved the remarkable result that a Riemann surface S is defined over $\overline{\mathbb{Q}}$ iff S is a Belyi surface. This implies that Belyi surfaces are dense in the space of all Riemann surfaces.

2 The Brooks-Makover Construction

While studying the first eigenvalue of the Laplacian on hyperbolic Riemann surfaces, Brooks and Makover developed a model that takes as input an oriented cubic graph and outputs a compact, finite-area, Belyi surface. Indeed, beginning with a finite cubic graph Γ (that is say Γ is a 3-regular graph on n vertices), define an orientation $\mathcal{O}$ on Γ by adding a cyclic orientation of the edges incident to each vertex. For example, consider the orientation on the graph below:



Notice that the orientation is represented at a vertex by a 3-cycle, read counterclockwise. The orientation $\mathcal{O}$ of the graph is then completely determined by a product α of disjoint 3-cycles. In

the example above, α is given by

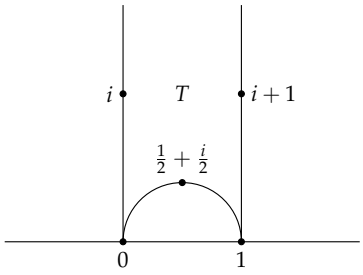
$$\alpha = (1,3,2)(4,5,6)(7,9,8)(10,11,12).$$

The purpose of representing the orientation by such a permutation α is to demonstrate that the edges incident to a given vertex can be cyclically reordered, just as the cycles $(1,3,2)$ and $(2,1,3)$ are equivalent as elements of a symmetric group.

Now, take the triangle T in the hyperbolic plane $\mathbb{H}^2$ with vertices at $\{0,1,\infty\}$. Its hyperbolic area is,

$$\text{area}(T) = \int_0^1 \int_{\sqrt{\frac{1}{4}-(x-\frac{1}{2})^2}}^{\infty} \frac{dy\,dx}{y^2} = \pi.$$

In particular, the area of T is finite with respect to the hyperbolic metric. On each of the bounding geodesics of T, pick a "midpoint" of that side, namely $\{i, i+1, \frac{i+1}{2}\}$. The term "midpoint" is a misnomer, since each bounding geodesic is of infinite length, but we use these points as a reference. The scenario is thus the following:



Given an oriented cubic graph $(\Gamma, \mathcal{O})$ on n vertices, we associate each vertex v_i to such a hyperbolic triangle T_i. For every edge between v_i and v_j, identify one side of T_i to one side of T_j; this 'gluing' procedure is uniquely determined by the following two conditions:

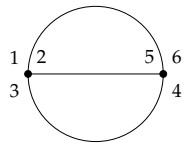
1. the "midpoints" of the two sides are glued together
2. the gluing preserves the orientations of T_i and T_j

The result of all of the gluing is a noncompact surface $S^O(\Gamma, \mathcal{O})$, which will have a bunch of cusps (there will be exactly one cusp for each LHT path[1]). Also note that $S^O(\Gamma, \mathcal{O})$ will have finite hyperbolic area equal to πn, as we have glued together n hyperbolic triangles T each of area equal to π.

Finally, we want to consider the conformal compactification $S^C(\Gamma, \mathcal{O})$ of the surface $S^O(\Gamma, \mathcal{O})$. Each cusp neighbourhood of $S^O(\Gamma, \mathcal{O})$ is conformally equivalent (isometric, even) to a punctured disk. If each cusp neighbourhood of $S^O(\Gamma, \mathcal{O})$ is replaced with a punctured disk and then all of the punctures are filled in; the resulting compact surface is $S^C(\Gamma, \mathcal{O})$. Moreover, adding finitely many points does not change the area, so $S^C(\Gamma, \mathcal{O})$ also has finite area.

A natural question to ask now is why are the surfaces constructed in this manner significant. One can show, as in Brooks and Makover (2004), that S is a Belyi surface iff there exists an oriented cubic graph $(\Gamma, \mathcal{O})$ such that $S = S^C(\Gamma, \mathcal{O})$.[2] Therein lies the key idea of the Brooks-Makover construction: one can understand a Belyi surface by understanding a finite graph.

In the first section, we showed that the Riemann sphere $\mathbb{P}^1(\mathbb{C})$ is a Belyi surface, so there must exist some oriented cubic graph $(\Gamma, \mathcal{O})$ such that $\mathbb{P}^1(\mathbb{C}) = S^C(\Gamma, \mathcal{O})$. While there is no way of reversing this process in general, let us apply the Brooks-Makover process to the following oriented graph $(\Gamma, \mathcal{O})$ and see what happens.



Gluing together two hyperbolic triangles T_1 and T_2 according to the above recipe, we obtain a surface $S^O(\Gamma, \mathcal{O})$ that appears spherical, but has 3 cusps. Replacing each cusp neighborhood with a punctured disk, results is a sphere with 3 punctures. Finally, to realize the conformal compactification $S^C(\Gamma, \mathcal{O})$, artificially fill in each puncture to get a sphere *i.e.* $S^C(\Gamma, \mathcal{O}) = \mathbb{P}^1(\mathbb{C})$.

3 Random Belyi Surfaces

Without delving too deeply into the theory of random graphs, the procedure to randomly choose a Belyi surface is given below, and re-

[1] A left-hand-turn (LHT) path on the oriented graph $(\Gamma, \mathcal{O})$ is a path where we always turn "left" in the orientation $\mathcal{O}$ at each vertex. One can take "left" turn at a vertex to denote, for example, the smallest number associated to an edge incident to the vertex by $\mathcal{O}$.

[2] Note that the conformal compactification $S^C(\Gamma, \mathcal{O})$ is indeed a compact surface. Belyi's Theorem implies that Belyi surfaces are dense in the space of all Riemann surfaces, so we are faced with the (perhaps counter-intuitive) fact that any Riemann surface can be approximated arbitrarily-well by a compact surface.

lies crucially on the construction of Brooks and Makover.

The random choice of an oriented cubic graph $(\Gamma, \mathcal{O})$ on n vertices gives a random Belyi surface $S^C(\Gamma, \mathcal{O})$, by applying the Brooks-Makover construction to $(\Gamma, \mathcal{O})$. However, there are exactly 2^n possible orientations on any cubic graph with n vertices. The problem of randomly choosing an oriented cubic graph is then reduced to randomly choosing a cubic graph on n vertices, as one can then just pick (uniformly, say) one of the orientations.

Bollobás's configuration model Bollobás (1985) provides an algorithm for randomly choosing a k-regular graph on n vertices in a "sensible" manner. In the case of a cubic graph, Bollobás's strategy can be explained as follows:

To choose a random cubic graph on n vertices, let $\{v_1, \cdots, v_n\}$ denote its vertex set. Now, imagine filling a bag with $3n$ balls, where each ball is labelled with a vertex v_i. Each v_i will appear on exactly 3 balls in the bag. Randomly pick a pair (v_i, v_j) of balls from the bag and add an edge between vertices v_i and v_j in our graph. Repeat this process until we have picked every ball from the bag, and the result is a 3-regular graph on n vertices. (This is indeed possible since finite 3-regular graphs have an even number of vertices.)

Recall that Belyi's Theorem implies that Belyi surfaces are dense in the space of all Riemann surfaces. Thus, by extension we can pick a random Riemann surface.

Primarily using the ideas presented herein, Brooks and Makover were able to conjecture that the expected genus of a random Riemann surface obeys a particular distribution. In Gamburd (2006), Gamburd proved their conjecture, by translating the problem into estimating the limiting behaviour of certain probability measures on the alternating group. For the interested reader, these first three sections provide sufficient detail and background to tackle Gamburd's proof.

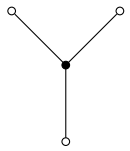
4 Grothendieck's *Dessins d'Enfants*

Given a Belyi surface S and a Belyi function $\beta\colon S \to \mathbb{P}^1(\mathbb{C})$, a *dessin d'enfant* is a bipartite graph whose vertex set is determined by:

1. A black vertex for each $x \in \beta^{-1}(0)$.
2. A white vertex for each $x \in \beta^{-1}(1)$.

The edges are given by the connected components of the preimage of $(0, 1)$ by β. The bipartition of the vertex set is given by those vertices that are colored black and white.

For example, the Belyi function $\beta\colon \mathbb{P}^1(\mathbb{C}) \to \mathbb{P}^1(\mathbb{C})$ given by $z \mapsto z^3$. As shown in the first section, $\beta^{-1}(1)$ will have three components but $\beta^{-1}(0)$ will only be a point. The set $\beta^{-1}((0, 1))$ consists of 3 components, connecting each component of $\beta^{-1}(1)$ to $\beta^{-1}(0)$. Thus, the *dessin d'enfant* associated to this Belyi function is the following:



Note that for Belyi surfaces that are more complicated than $\mathbb{P}^1(\mathbb{C})$, the Belyi functions are generally more complicated as well. Consequently, the corresponding dessin d'enfant can become quite difficult to compute (although still much simpler than working with the Belyi surface itself).

Therefore, a Belyi surface and a Belyi function give rise to a dessin d'enfant. Grothendieck remarked, as in Schneps (1994), that a dessin determines a covering of $\mathbb{P}^1(\mathbb{C})$ that is ramified at most at 0, 1, and the point added at ∞. Thus, the dessin also determines a Belyi function, and hence a Belyi surface. The absolute Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on the Belyi surface[3], and so by extension on the dessin. In particular, the dessins d'enfants provide certain useful invariants under the action of this Galois group.

Indeed, these dessins d'enfants, though simple in appearance, have had a profound impact on the study of Belyi surfaces and Belyi functions. For a more complete and quite beautiful description of these multifaceted objects, please see Schneps (1994).

References

Belyĭ, G. V. (1979). Galois extensions of a maximal cyclotomic field. *Izv. Akad. Nauk SSSR Ser. Mat.*, 43(2):267–276, 479.

Bollobás, B. (1985). *Random graphs*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London.

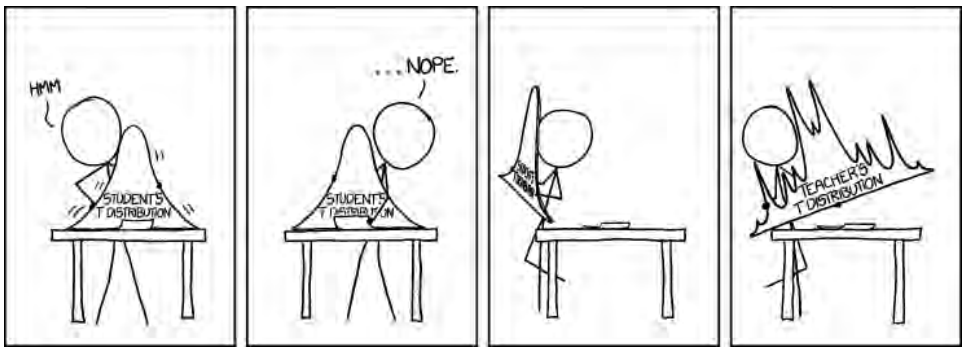
[3]By Belyi's Theorem, the Belyi surface is determined by a polynomial in 2 complex variables with coefficients in $\overline{\mathbb{Q}}$; thus, $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on the coefficients of this polynomial in the natural way, which induces the action on the surface.

Brooks, R. and Makover, E. (2004). Random construction of riemann surfaces. *Journal of Differential Geometry*, 68(1):121–157.

Gamburd, A. (2006). Poisson-Dirichlet distribution for random Belyi surfaces. *Ann. Probab.*, 34(5):1827–1848.

Schneps, L., editor (1994). *The Grothendieck theory of dessins d'enfants*, volume 200 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge. Papers from the Conference on Dessins d'Enfant held in Luminy, April 19–24, 1993.

JOKES AND COMICS

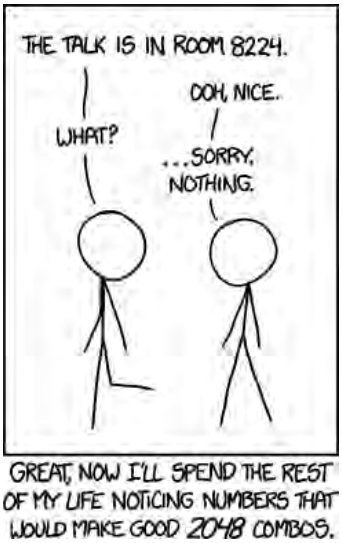


xkcd 1348: *t distribution*

If data fails the Teacher's t test, you can just force it to take the test again until it passes.

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.

One day a farmer called up an engineer, a physicist, and a mathematician and asked them to fence in the largest possible area with the least amount of fence. The engineer made the fence in a circle and proclaimed that he had the most efficient design. The physicist made a long, straight line and proclaimed "We can assume the length is infinite..." and pointed out that fencing off half of the Earth was certainly a more efficient way to do it. The mathematician just laughed at them. He built a tiny fence around himself and said, "I declare myself to be on the outside."



xkcd 1345: *Digits*

It's taken me 20 years to get over skyline tetris.

Interview with Dr. Linan Chen

Meng Zhao

Biography in a box

Nationality	Chinese
Speciality	Probability
Alma mater	MIT
Coming from	MIT, McGill (postdoc)
Teaching lately	Calculus 1, Ad. Probability 2

$\delta\varepsilon$: Tell me a little about your personal and academic background.

I grew up in the north east of China, in Liaoning province. After high school, I went to Tsinghua University in Beijing for my undergrad studies and there I got my B.Sci. After that, I was admitted to the math PhD program at MIT. During my graduate studies, I got interested in probability theory and its connections with various other fields. In 2011, I completed my PhD thesis on this topic and after that I joined McGill University as a postdoc, where I have spent almost 3 years. Starting in September 2014, I will become an assistant professor at McGill and I am very much looking forward to it.

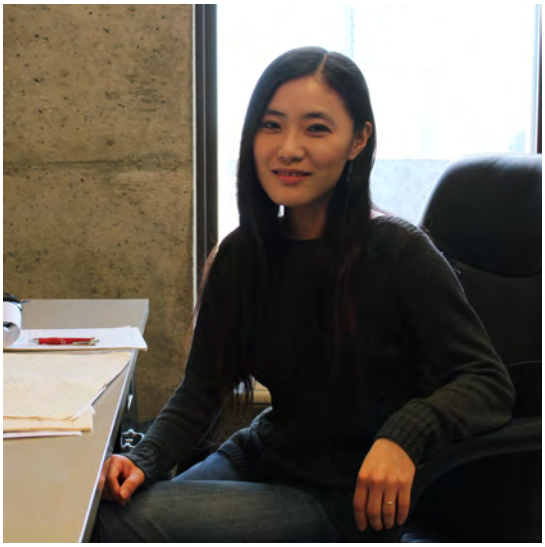
$\delta\varepsilon$: How did you become interested in probability?

Probability theory has a long history; many people become interested in this field because of its combinatorial side which is discrete probability, indeed the original root of probability theory. Nowadays, it has given rise to many meaningful problems. However, that's not how I got started. I actually started with the analytical side. In particular, I was shown by my PhD advisor, Professor Daniel Stroock, the connections between probability theory and partial differential equations. I learned from him how a deterministic phenomenon, for example heat diffusion, can be described and studied through the underlying random process; and then how this random process is driven by deterministic partial differential equations known as Kolmogorov's equations. I got immediately interested in this and other profound connections between the random world and the deterministic world.

$\delta\varepsilon$: What are your current research interests?

After PhD, I continued working on analytical probability theory. I studied in particular its connection with infinite dimension analysis. Right now, I'm working on problems related to Gaussian free fields. Heuristically speaking, a Gaussian free field is a Gaussian process parametrized by multiple parameters. You can think of it as a Brownian motion but driven by more than one time parameter. Imagine I give you two clocks, and you have to time your process using two clocks at the same time, or even more clocks.

Gaussian free fields have various applications in physics and analysis and can also be used to model random geometric objects – for example, a random manifold or a random surface. I have collaborated with Professor Dmitry Jakobson on various projects related to random metrics and random measures. We build these random geometric objects using Gaussian free fields and we study their properties.



Picture 1: Dr. Linan Chen, who is joining the Department as Assistant Professor

$\delta\varepsilon$: What is a random surface?

Think of a plane. A plane is not random, right? It's completely flat. Now repeatedly select a point at random and make it into either a bump

or a valley. In some sense you are randomly reshaping the landscape of this plane. You can think of this as one way of getting a random surface. If you and I both do a simulation, we would get different surfaces. The space of all these random surfaces has an underlying probability distribution. You may favour a certain kind of configuration and penalize other kinds.

$\delta\varepsilon$: Compare your experiences with undergrad in China, PhD at MIT, and postdoc at McGill? For example, did each step prepare you well for the next?

[LAUGHS] One thing I can say for sure is the winters got colder and colder from one place to the next. I was in a different stage in my studies at each place – there was no overlap – so it's hard to compare. For my undergrad studies, I basically followed a well-planned and very strict curriculum in China. In hindsight I think it's important to lay down a solid foundation at the early stage of mathematical study. Even after I had an idea of what I wanted to do, this was very helpful; I was working really hard and doing a *lot* of practice.

By contrast, at MIT the PhD curriculum has a lot more flexibility: a lot less required coursework and a lot more research training. The transition is that you spend less time doing homework problems, which are designed to be solved efficiently in a short amount of time, and far more time tackling "real" research problems that need to be broken down into small steps. The researcher must plan a strategy to achieve a goal which may not even happen. As a postdoc, besides pushing forward with the work that I started in my PhD, I also tried to broaden my research field. For example, I've been trying to make more connections between subjects which are familiar to me with ones that are new to me. At every stage of my study, the guidance and the support of my advisors have been incredibly important. Meanwhile, I have had to become increasingly independent; it's like when an eagle must eventually leave the nest and learn to fly by herself.

$\delta\varepsilon$: What is the collaborative research process like?

A collaborative research process is a combination of learning and teaching. You will sometimes learn more than you teach, and other times vice versa. It's a lot of exchange of knowledge, and bringing different skill sets to one problem. In my experience it can be very efficient to work with someone. On the one hand, your collaborator can bring a whole new perspective to the problem; on the other hand, you are motivated – in some sense you're forced – to keep track and to be super organized with your thoughts because you need to make someone else understand them. And same for your collaborator.

Oftentimes, it's helpful to talk to someone when you get stuck. I've gotten inspired about my research problem by talking with peer graduate students that are not strictly my collaborators and who are doing completely different things.

$\delta\varepsilon$: How does it feel to be a postdoc on the verge of being an assistant professor?

I am very excited and a little bit stressed about this transition. Being a professor is like starting a new era which certainly brings me more resources and opportunities to further my research interests but which also comes with a lot of responsibilities, bigger teaching loads, and higher expectations – as you can imagine. But after all, being a faculty member, I look forward to doing more things for the department, serving the community, and doing more things for the students. As for teaching, you know, I love teaching. It is really a very good experience for me to have taught two very different courses: graduate level advanced probability and freshmen-level calculus. I learned a lot. In the future I definitely look forward to teaching more courses.

Relating Deterministic Finite Automata and Boolean Circuits

Leah Weiner

To help us gain some understanding, we begin by looking at two of the simplest computational models. A natural computational model for performing sequential computations is the finite automaton. For parallel computations however, it is natural to use boolean circuits. These two models of computation have surprising relationships. Below, we will define these two models and demonstrate the relationships between them. It turns out that the languages which can be recognized by a boolean circuit of constant depth and a polynomial number of gates (with respect to the number of input nodes) are exactly the languages which are p-reducible to a language which can be recognized by a specific subset of finite automata.

1 Introduction to regular languages

An *alphabet* Σ is a non-empty, finite set of symbols. For example, $\{a, b, c\}$ and $\{0, 1\}$ are two distinct alphabets. Any finite string of symbols over the alphabet Σ is called a *word*. The empty word, denoted ϵ, is the word of length zero. Further examples of words over the alphabet $\{a, b, c\}$ include a, acb, and $caaabbccca$.

In computer science, we are generally only interested in subsets of all possible words over an alphabet Σ. We denote the set of words of length n over the alphabet Σ by Σ^n, and the set of all possible words over the alphabet Σ by Σ^*. A subset of words $L \subset \Sigma^*$ is called a *language* over the alphabet Σ. Languages can contain either a finite or an infinite number of words. One example of a (infinite) language over the alphabet $\{a, b, c\}$ is $L = \{ab, aabb, aaabbb, ...\} = \{a^n b^n : n > 0\}$.

Let us now define operations on languages. Given any two languages L_1 and L_2 over the same alphabet Σ, we define the following operations:

1. Concatenation: $L_1 L_2 = \{w_1 w_2 : w_1 \in L_1, w_2 \in L_2\}$
2. Union: $L_1 \cup L_2 = \{w : w \in L_1 \text{ or } w \in L_2\}$
3. Intersection: $L_1 \cap L_2 = \{w : w \in L_1 \text{ and } w \in L_2\}$
4. Kleene's star: $L_1^* = \{w = w_1 w_2 \dots w_n : n \geq 0, w_i \in L_1\}$

Note that the empy word $\epsilon = \emptyset^*$, and that the language consisting of every word over Σ is denoted by Σ^* (coinciding with our earlier definition). When the meaning is clear, we identify a word $w \in \Sigma$ with the singleton $\{w\}$, so we may write, for example, $\{w\}^* = w^* = \{\epsilon, w, ww, ...\}$. Examples of languages that can be constructed using the operations defined above include $ab^* = \{a, ab, abb, abbb, ...\}$ and $(ab)^* = \{\epsilon, ab, abab, ababab, ...\}$.

We are interested in looking at a specific family of languages. The class of *regular languages* occurs naturally in computer science and their properties are therefore studied in great depth. The class of regular languages over the alphabet Σ is the smallest class of languages closed under union, concatenation and Kleene's star that contains every singleton $\{a\}$, for each $a \in \Sigma$, and the empty set $\emptyset$.

An example of a regular language over the alphabet $\Sigma = \{a, b, c\}$ is $L = \{a, ab, abb, abbb, abbbb, ...\} = \{ab^n : n \geq 0\} = ab^*$, the set of words that begin with an a and end with any number of b's. Every finite language is regular. Not every language is regular, however. A common first example of a language that is not regular is the language described by $\{a^n b^n : n > 0\}$, which we will investigate later.

2 Introduction to deterministic finite automata

A deterministic finite automaton (DFA) operating over an alphabet Σ is a computational model which takes as input any word in Σ^* and has a binary output (i.e. 'yes' or 'no'). Each DFA operates over only one alphabet. A DFA is formally represented by a 5-tuple $(Q, \Sigma, \delta, q_0, F)$ where Q is a set of states, Σ is an alphabet, δ is a transition function $\delta : Q \times \Sigma \to Q$, q_0 is the start state, and F is the set of accepting states.

DFAs can be represented on paper as directed graphs (See Figure 1). Each state of the DFA corresponds to a node in the graph, where the start node is specified by the upper left-most arrow, and accepted states are represented by a double circle. The transition function is represented

as labeled arrows connecting nodes, where the arrow from state q_1 to q_2 is labeled $a \in \Sigma$ if and only if $\delta(q_1, a) = q_2$. Generally, the graphical representation of DFAs are not drawn completely. In what follows, if we omit any arrows or nodes, it is assumed that these arrows lead only to rejection states. For example, refer to Figure 1 and Figure 2. Both DFAs accept exactly the same words. We can see, however, that the DFA in Figure 2 omits the implicit rejection arrows and nodes.

When given a word in its input alphabet, the DFA will process the word as specified by its transition function. It is easiest to see this by picturing the on-paper representation of a DFA. The DFA begins by default at the start state. The sequence of states is determined by following the labeled arrows and reading one symbol of the word at a time. When the DFA has finally processed the last symbol of the word, it observes its current state. If the DFA is in an accepting state, then it will output "yes", and the word is in the language recognized by the DFA. However, if at the end the DFA finds itself in a rejecting state, then it will output "no", corresponding to the words that are not in the language recognized by the DFA.

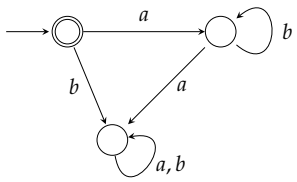


Figure 1: This DFA recognizes the language ab^*, and includes all possible transitions.

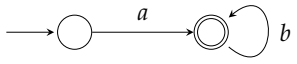


Figure 2: This DFA recognizes the language ab^*, leaving transition arrows which lead to rejection implied.

Now that we have described DFAs, let us look at the relationship between DFAs and regular languages. The famous theorem of Kleene's states:

Theorem 1 (Kleene's Theorem)**.** *A language L is regular iff L can be recognized by some DFA.*

Kleene's theorem allows us to prove our earlier claim about the regularity of the language $\{a^n b^n : n > 0\}$.

Kleene's theorem tells us that given any language L, if we can construct a DFA that accepts *only* the words in L, then L is a regular language. Moreover, if L is a regular language, we will be able to construct a DFA that accepts exactly the words in L. Although we will not prove this theorem here, we will use the result.

With Kleene's theorem it is now easy to see that the language $\{a^n b^n : n > 0\} = \{ab, aabb, aaabbb, ...\}$ is not a regular language, as it cannot be recognized by a DFA.

Proposition 2. *The language* $\{a^n b^n : n > 0\} = \{ab, aabb, ...\}$ *is not regular.*

Proof. Suppose the above language were regular so that it is then recognized by some DFA. For any $n \in N$, let q_n be the final state of the DFA when the word a^n is inputted. Since the DFA accepts only words of the form $a^n b^n$, it follows that $q_i \neq q_j$ for any $i \neq j$. For if they were the same state, then the DFA would accept both $a^i b^i$ and $a^i b^j$. Since a DFA has, by definition, only finitely many states, this yields a contradiction. □

An easy corollary of the above theorem regards the complements of regular languages.

Corollary 3. *Regular languages are closed under complementation*

Proof. If L is a regular language, then L is recognized by some DFA $D = (Q, \Sigma, \delta, q_0, F)$. Consider the DFA $D' = (Q, \Sigma, \delta, q_0, Q - F)$. Clearly, D' accepts a word iff D rejects that word. Therefore, D' recognizes the complement of L. □

The class of *star-free language* is defined as the smallest class of languages closed under unions, concatenations and complements, that contains every singleton $\{a\}$, $a \in \Sigma$, as well as the empty set $\emptyset$. It follows from Corollary 3 that every star-free language is regular. Intuitively, a star free language is a language that can be completely described without using the * operation.

One easy example of a star-free language is Σ^*, as Σ^* can be completely described by the star-free expression $\overline{\emptyset}$. Another example of a star-free language is the language introduced above, ab^* (over the alphabet $\Sigma = \{a.b\}$). Although the expression ab^* is certainly the cleanest description of this language, it can be completely describe without the use of Kleene's star. In fact, $ab^* = a\overline{\Sigma^* a \Sigma^*} = a\overline{\overline{\emptyset} a \overline{\emptyset}}$. Here, the claim

is really that $b^* = \overline{\Sigma^* a \Sigma^*} = \overline{\overline{\emptyset} a \overline{\emptyset}}$, with the latter expression begin obviously star-free. To see this equality, note that $\overline{\Sigma^* a \Sigma^*}$ is precisely the set of words that do not contain the letter a. Since $\Sigma = \{a, b\}$, this is precisely the set of words that contain only the letter b (together with ϵ), which is exactly b^*. Over the alphabet $\{a, b\}$, it is also clear that ϵ is star-free. In fact, $\epsilon = \overline{\overline{a\overline{\emptyset}} \cup \overline{b\overline{\emptyset}}}$, and a similar expression for ϵ holds for any alphabet Σ, yielding that ϵ is always star-free regardless of the underlying alphabet.

A more complicated example of a star-free language over the alphabet $\Sigma = \{a, b\}$ is the language that is completely described by $(ab)^* = \{\epsilon, ab, abab, ababab, ...\}$. To see this, we will show that $(ab)^*$ can be equivalently described by the language $K = \{\epsilon\} \cup a\overline{\emptyset} \cap \overline{\emptyset} b \cap \overline{\overline{\emptyset} aa \overline{\emptyset}} \cap \overline{\overline{\emptyset} bb \overline{\emptyset}}$. In order to gain an understanding about which words lie in K, it is useful to replace $\overline{\emptyset}$ with Σ^* in the expression, and obtain $K = \{\epsilon\} \cup a\Sigma^* \cap \Sigma^* b \cap \overline{\Sigma^* aa \Sigma^*} \cap \overline{\Sigma^* bb \Sigma^*}$. In fact, the claim is made almost immediately obvious using this expression.

Proposition 4. *The language $(ab)^*$ is star-free and is given by the expression $K = \{\epsilon\} \cup a\overline{\emptyset} \cap \overline{\emptyset} b \cap \overline{\overline{\emptyset} aa \overline{\emptyset}} \cap \overline{\overline{\emptyset} bb \overline{\emptyset}}$*

Proof. We only have to prove that $(ab)^* = K$. By the above expression, all the words in K begin with a, end with b, and contain no consecutive a's or b's (together with ϵ, of course). These are exactly the words in $(ab)^*$.

□

It is natural now to ask if there is a methodological way of determining whether or not a given language is star-free. In fact, a characterization of star-free languages comes in the form of DFAs.

We say that a DFA is *periodic* if there exists a word that, when processed by the DFA, induces a cycle in the directed graph representation of the DFA. To be more precise, if the sequence of states that the DFA passes through while processing the word in question are $q_0, ..., q_n$, then for this DFA to be periodic, we must have $q_k = q_i$ for some $k \neq i$, where δ is the transition function of the DFA and q_0 is the start state. Recall that if the word inducing the cycle is $w = w_1 ... w_n$, the states are given by $q_i = \delta(q_{i-1}, w_i)$A DFA is *aperiodic* if it is not periodic.

Theorem 5 (Schützenberger's Theorem)**.** *A regular language is star-free iff the DFA that recognizes the language is aperiodic.*

It can be shown that if an n-state DFA is periodic, then there exists a cycle induced by a word w and such a word can be found of length at most n^n. Thus, it is decidable if a given language is star-free.

3 Introduction to circuits

An n-input boolean function is a function $f : \{0,1\}^n \to \{0,1\}$, i.e. f takes a 0-1 string of length n as input and outputs a 0 or a 1. We are interested in building complicated boolean functions (circuits) out of a certain subset of simple boolean functions, which are called *gates*. We will restrict ourselves to the AND gate $\wedge$, which is given by

$$\wedge(x_1, ..., x_n) = \begin{cases} 1 & \text{if } x_1 = x_2 = ... = x_n = 1 \\ 0 & \text{otherwise} \end{cases}$$

and the OR gate $\vee$, which is given by

$$\vee(x_1, ..., x_n) = \begin{cases} 0 & \text{if } x_1 = x_2 = ... = x_n = 0 \\ 1 & \text{otherwise} \end{cases}$$

While there is really a different AND (resp., OR) gate for every finite number of boolean inputs, we will abuse notation slightly and use the symbol $\wedge$ (resp., $\vee$) to refer to every AND (resp., OR) gate, regardless of the number of inputs.

Given n boolean inputs $X = \{x_1, x_2, ..., x_n\}$, we define the set of literals to be $\overline{X} = \{x_1, x_2, ..., x_n, \overline{x_1}, \overline{x_2}, ..., \overline{x_n}\}$ where $\overline{x}$ denotes the complement of x. Our convention is that the boolean statement x is true iff $x = 1$, and $\overline{x}$ is true iff $\overline{x} = 1$, so that $x = 0$.

A circuit is a directed acyclic graph whose nodes are $\wedge$ or $\vee$ gates, with boolean inputs feeding into the n base nodes of the tree. A circuit which takes in input strings of length n can be thought of as a function $C_n : \{0,1\}^n \to \{0,1\}$. Each of these nodes is a gate which outputs either 0 or 1 (based on the boolean inputs it receives) to other nodes (gates) in the graph according to the layout of the directed edges which connect the nodes. A circuit has a distinguished output node; upon input of the n boolean variables, the output of C_n is just the output of that distinguished node.

The *depth* of a circuit is the number of layers of gates in a circuit; that is the height of the graph representing this circuit on paper. The height of

a circuit is well-defined.[1] The *size* of a circuit is the total number of gates in the circuit, the total number of nodes of the tree representing this circuit on paper. The *fan-in* of a node in an acyclic graph refers to the number of input edges to that node. We say a node has *bounded fan-in* if this number of input edges is finite.

Notice that each individual circuit C_n operates over an input of some fixed length n. A boolean language, however, generally has words of arbitrary length. Thus, when we talk about a circuit operating over some boolean language, we are actually referring to a *family* of circuits. Given a language $L \subset \{0,1\}^*$, we say the family of circuits C recognizes L, where $C = \cup_n C_n$, and where C_n recognizes $L \cap \{0,1\}^n$ for each natural number $n \geq 1$ (that is, C_n outputs 1 if and only if the input word is an element of $L \cap \{0,1\}^n$).

Families of circuits are what we call *non-uniform* models of computation because inputs of different length can perform different computations and thus requiring an infinite description. In practice, it is not feasible to have a model of computation that requires an infinite description; if the circuit which solves the system requires an unreasonable amount of time to be constructed, then the problem isn't truly solvable. In order to address this issue, we consider *uniform* family of circuits. For a uniform family of circuits there is a rule to construct the next circuit, based on the previous ones, and this construction is guaranteed to be done in some given amount of time. More formally, we say a family of circuits C_n is *polynomial-time uniform* if there exists a polytime Turing Machine M such that M runs in polynomial time and $\forall n \in \mathbb{N}$, whenever M receives 1^n as input, M outputs a description of C_n.

In what follows, we will abuse the language and refer to the uniform family of circuits which recognizes the language L simply as a circuit. Similarly, we say a (family of) circuits has depth (resp., size) equal to the maximum depth (resp., size) over each circuit of the sequence. It is then natural to inspect the complexity of circuits. Here, we introduce two classes of circuits.

Definition 1. *A boolean language L is in NC^i if there exists a boolean circuit of depth $O(\log^i n)$ and is of polynomial number of gates, where each gate has a fan-in of at most two which recognizes exactly the words in L (i.e. the circuit outputs 1 if the circuit's input is a word in L, and outputs 0 otherwise). We may say that a circuit C is in NC^i if C recognizes a language in NC^i. We denote $NC = \cup_{i \geq 1} NC^i$ We say a sequence of circuits is in NC^i if each circuit of the sequence is in NC^i.*

Definition 2. *A boolean language L is in AC^i if there exists a boolean circuit of depth $O(\log^i n)$, and is of polynomial number of gates, and unbounded fan-in which recognizes exactly the words in L (i.e. the circuit outputs 1 if the circuit's input is a word in L, and outputs 0 otherwise). We may say that a circuit C is in AC^i if C recognizes a language in AC^i. We denote $AC = \cup_{i \geq 1} AC^i$ We say a sequence of circuits is in AC^i if each circuit of the sequence is in AC^i.*

Theorem 6. *$NC = AC$ with bounded fan-in and $NC^i \subset AC^i \subset NC^{i+1}$ $\forall i \geq 1$.*

Proof. The first inclusion, that $NC^i \subset AC^i$ for each i follows immediately from the definitions.

The second inclusion $AC^i \subset NC^{i+1}$ is also relatively straight-forward. To see this, consider some language that can be recognized by a circuit C (with bounded fan-in) in AC^i. From C, we will construct a circuit that is in NC^{i+1} and is computationally equivalent to C, so that it recognizes the same language as C.

To do this, consider replacing each gate of C with fan-in more than two by a binary tree with gates of fan-in at most two and of depth $O(\log^1 n)$. Since each gate of C has bounded fan-in by assumption, this iterative construction will terminate. Furthermore, we have increased the depth only by a factor of $\log n$, giving this modified circuit a depth of $O(\log^{i+1} n)$ and fan-in at most 2.

In this way we have created a computationally equivalent circuit in NC^{i+1} which recognizes the same language. Thus, $AC^i \subset NC^{i+1}$.

□

We will be predominantly interested in looking at languages in AC^0, the class of circuits of polynomial size and constant depth. Given an integer constant k, let AC^0_k be the subclass of AC^0 consisting of circuits of polynomial size and depth exactly k. Notice that $AC^0 = \cup_{k \geq 1} AC^0_k$

4 Relating Automata and Circuits

With this understanding of DFAs and boolean circuits, the question of how we can relate these

[1] To see this, recall that a circuit is an acyclic graph. The nodes with no incoming edges are at height 0. Each node is at the height given by the shortest directed path from any root node to this node. Generally, the roots children are at height 1, their children at height 2, etc.

two natural models of computation arises.

Notice that DFAs can recognize languages over any alphabet and can be finitely described while all circuits take in inputs over $\{0,1\}$ and are described as an infinite sequence of (finite) circuits. Clearly, the structures and the input languages of the two computational models we wish to compare differ, and this difference must be resolved before further analysis.

Given any language L over the alphabet $\{0,1\}$ that is recognized by some circuit, it is easy to construct a DFA which recognizes L, given that L is a regular language. Creating a circuit which recognizes the same language as a given DFA is trickier. To do so, we must convert the DFA's (arbitrary) language to binary strings. This conversion is not so obvious and is described below.

Given a word over an arbitrary alphabet Σ, we want to convert this word to a binary string so that it can be processed by a circuit. This conversion can be done as follows. For each word of length n in Σ, we create a circuit with $|\Sigma|n$ input gates. Each input gate is assigned a unique pairing (i,σ) for $i = 1,2,...,n$ and for each $\sigma \in \Sigma$. Suppose we would like to input the word $w = w_1w_2...w_n \in \Sigma^*$. At the input gate corresponding to (i,σ) we ask "does $w_i = \sigma$", where an affirmative response corresponds to that node receiving input 1 and a negative response corresponds to that node receiving input 0.

With this conversion from words over an arbitrary alphabet to the alphabet $\{0,1\}$ in mind, we will abuse language. We may now say that an language L over an arbitrary alphabet is recognizable by a circuit, in which case we mean that the words of L are first converted into binary strings in the way described above, and then inputted into a circuit. (It is important to note that this conversion adds only a polynomial number of gates and only one layer of gates to the circuit).

Next, we are interested in comparing DFAs with one another. DFAs operate over arbitrary alphabets and we thus require some method for comparing two arbitrary alphabets. The notion of *p-reducibility* allows us to "switch" between any two languages in the following sense.

Let Σ and Γ be two arbitrary alphabets, and consider the languages $L \subset \Sigma^*$ and $K \subset \Gamma^*$. Then the language L is *p-reducible* to K, denoted $L \leq_p K$, if for each n, there is some function $\phi_n : \Sigma^n \to \Gamma^{s_n}$, where s_n is polynomial in the size of n, and such that:

1. $a = a_1...a_n \in L$ iff $\phi_n(a) = \phi_n(a_1...a_n) = b_1, b_2, ..., b_{s_n} \in K$.

2. For each $j \in \{1,...,s_n\}$, b_j depends only on some a_{i_j}. That is, there exists some function $f_j : \Sigma \to \Gamma$ so that $b_j = f_j(a_{i_j})$ whenever $b_1...b_{s_n} = \phi_n(a_1...a_n)$.

Point 1 states that the function ϕ_n is an injection between words of length n over L and words of length s_n over K.

Point 2 states that each letter b_j of the word $b \in K$ depends only on the i_jth letter of the word a, where $\phi_n(a) = b$.

We say that $L \subset \Sigma^*$ is *p-recognizable* if there is some regular language $K \subset \Gamma^*$ such that $L \leq_p K$.

Theorem 7. *L is p-recognizable iff there exists a circuit in NC^1 which recognizes the language L.*

Here, when we say a circuit is recognizing the p-recognizable language L, we are assuming the conversion of words in L to binary strings as described previously.

Recall that NC^1 is the set of circuits of logarithmic depth, of polynomial size and whose gates have binary inputs. To show the first direction, suppose that L is a p-recognizable language. Our goal then is to find a circuit $C \in NC^1$ which recognizes L.

We will first show the case for when L is a regular language.

Let $D = (Q, \Sigma, \delta, q_0, q_f)$ be the DFA that recognizes this regular language L. (Without loss of generality we assume D has only one accepting state.)

Now, let $w = a_1a_2...a_n$ be any word in L, where each a_i is a symbol of L's alphabet. Then $\delta(q_0, a_1, ..., a_n) = q_f$ if and only if there exists some state q such that $\delta(q_0, a_1, ..., a_{\lfloor n/2 \rfloor}) = q$ and $\delta(q, a_{\lfloor n/2 \rfloor + 1}, ..., a_n) = q_f$. That is, there must exist some intermediate state q of the DFA D which is "touched" when D is processing the word $w = a_1a_2...a_n$.

With this intuition, we can construct a circuit which recognizes L. (Note that the circuit we are constructing is actually a sequence of circuits, as each word in L corresponds to one circuit we are constructing.)

The construction of the circuit in NC^1 which recognizes the word $w \in L$ can be done as follows.

Recall that every DFA has a finite number of states. Thus, there is a finite number of possible "intermediate states" which could be the state where $\delta(q_0, a_1, ..., a_{\lfloor n/2 \rfloor}) = q$ and

$\delta(q, a_{\lfloor n/2 \rfloor+1}, ..., a_n) = q_f$. In fact there are $|Q|$ possible such states.

The first level of the circuit will consist of $|Q|$ AND gates, with each gate corresponding to a unique $q_i \in Q$. Each of these nodes will have fan-in of two. The first input to the node corresponding to state q_i will ask "does $\delta(q_0, a_1, ..., a_{\lfloor n/2 \rfloor}) = q_i$" while the second input will ask "does $\delta(q_i, a_{\lfloor n/2 \rfloor+1}, ..., a_n) = q_f$". The answer will be yes to both of these question if and only if the the node corresponds to the appropriate intermediate state q. That is, the output of the node corresponding to state q_i in this first layer of the circuit we are constructing will be 1 iff $\delta(q_0, a_1, ..., a_{\lfloor n/2 \rfloor}) = q_i$ or $\delta(q_i, a_{\lfloor n/2 \rfloor+1}, ..., a_n) = q_f$.

The construction of the second (and subsequent) levels of this circuit is similarly done by adopting a system of divide and conquer, checking for intermediate gates in the sub-word $a_1 ... a_{\lfloor n/2 \rfloor}$ and $a_{\lfloor n/2 \rfloor+1}, ..., a_n$. In this way, we can construct a circuit with depth $O(\log n)$.

Each node in the last layer of the circuit outputs into an OR gate with fan-in $|Q|$ (one input for each $q \in Q$). This final OR gate outputs 1 iff the original word was recognized by the DFA D, and thus is a word in L.

Now, manipulating the gates in the same fashion as we did in the proof of $AC = NC$ gives a circuit of depth $\log n$ with fan-in at most two.

We have thus shown this direction of the theorem holds for regular languages. We can now extend the argument to the case where L is p-recognizable simply by incorporating a so-called pre-processing phase. This phase consists of converting L into K, where $L \leq_p K$ and K is a regular language. (Recall this is possible by the definition of p-recognizability.) The pre-processing takes no time and each b_j depends on only one a_{ij}. Thus, $L \in NC^1$ as desired.

For the proof of the reverse implication, we refer the reader to Barrington's PhD thesis at MIT in 1986 Barrington (1989).

We are now interested in the size of a circuit necessary to recognize a given language.

5 The universal DFA

The universal DFA D_k, for $k > 1$, is defined as $D_k = (Q, \Sigma, \delta, q_1, q_1)$, where $Q = \{q_1, q_2, ...q_{k+1}\}$, $\Sigma = \{a, b, c\}$, and

$\delta(q_i, a) = q_{i+1}, i \neq k, k+1$

$\delta(q_i, a) = q_i\ , i = k, k+1$

$\delta(q_i, b) = q_{i-1}, i \neq 1, k+1$

$\delta(q_1, b) = q_{k+1}$

$\delta(q_{k+1}, b) = q_{k+1}$

$\delta(q_i, c) = q_{i+1}, i \neq k, k+1$

$\delta(q_k, c) = q_k$

$\delta(q_{k+1}, c) = q_{k+1}$

Figure 3 depicts the universal DFA D_2. We will mostly be interested in the universal DFA.

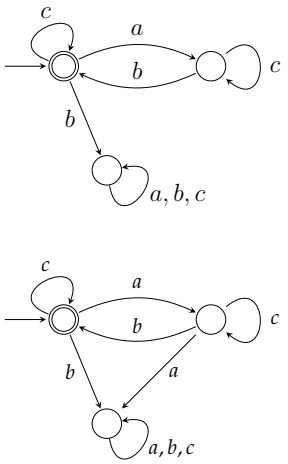


Figure 3: The Universal DFA D_2 (above) and another, similar DFA

The following theorems, although not be proven here, are important to note.

Theorem 8. *L is p-recognizable by an aperiodic DFA iff there exists a circuit in AC^0 which recognizes L.*

Theorem 9. *L is p-reducible to a star-free language iff there exists a circuit in AC^0 which recognizes L.*

Theorem 10. *L is p-reducible to a language in D_k iff there exists a circuit in AC^0_k which recognizes L.*

We will show part of this last theorem for the special case $k = 2$.

Theorem 11. *L is p-reducible to the language recognized by the DFA D_2 iff there exists a circuit in AC^0_2 which recognizes L.*

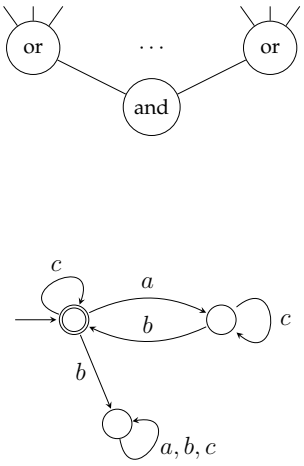
Proof idea:

We now demonstrate one example to give some intuition towards understanding the implication if $C \in AC^0_2$ recognizes the language L then L is p-reducible to the language recognized by the DFA D_2.

Let $C \in AC^0_2$ which recognizes the boolean language K (the "conversion" of the language L

to a language of boolean strings). We want to show that K is p-reducible to some regular language.

Consider the circuit on the left in the figure below.



Now, consider the map $\Phi : \{0, 1, \text{“gate change”}\} \to \{a, b, c\}$ where $\Phi(0) = c$, $\Phi(1) = a$ and $\Phi(\text{“gate change”}) = b$.

If the word w is recognized by the depicted circuit, then the output is 1. This means that each OR gate of the circuit must be outputting a 1 (so that the final AND gate correctly outputs 1). A bit of observation shows that the mapping Φ described above correctly maps the circuit to the shown DFA D_2; a input w is accepted by the circuit iff the word $\Phi(w)$ is accepted by the DFA D_2.

For example, never will a word over $\{0, 1\}$ from the circuit map to a word over $\{a, b, c\}$ which has two consecutive b's. This is clear since we never have a “gate change” twice in a row in the circuit.

6 Future Work

The two DFAs shown in Figure 3 look quite similar. We would like to conjecture that D_2 cannot be reduced to the DFA shown to the left in Figure 3. That is, we would like to show that it is not the case that languages in AC_2^0 are p-reducible to this second DFA. Although this conjecture has not yet been proven, an idea of how to prove this statement is as follows. Let K be the language recognized by the universal DFA D_1 and let L be the language recognized by the second DFA. Essentially, we would like to show that $K \not\leq_p L$. To do this, we suppose that $K \leq_p L$. If this were true, we'd have a program which maps the words in K to the words in L.

In order to arrive at the desired contradiction, we need to find two words, say w_1 and w_2 where $w_1 \in K$ and $w_2 \notin K$ and such that the program maps w_1 and w_2 to two words which are either both accepted or both rejected by the second DFA. If this situation were to arise, we would have a problem, as K accepts only one of these two words, but L recognizes neither, thus demonstrating that the two DFAs are not equivalent.

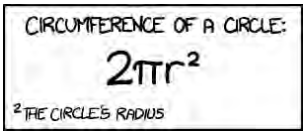
7 Conclusion

We have described two natural models of computation and have demonstrated some relationships between these two models. Future work includes rigorously proving the conjecture stated just above.

References

Barrington, D. A. (1989). Bounded-width polynomial-size branching programs recognize exactly those languages in nc[1]. *Journal of Computer and System Sciences*, 38(1):150 – 164.

Jokes and Comics



xkcd 1185: *Circumference formula*
Assume r' refers to the radius of Earth Prime, and r'' means radius in inches.

LIE POINT SYMMETRIES AND MAGNETIC MONOPOLES

Maxence Mayrand

We find all Lie point symmetries of the system of nonlinear second-order ordinary differential equations that govern the classical motion of a charged particle in the field of a magnetic monopole and use Noether's Theorem to derive the complete solution.

1 INTRODUCTION

This paper illustrates the usefulness of the method of Lie point symmetries for solving difficult systems of nonlinear ordinary differential equations. We will give a brief introduction to the theory and then apply it to a problem that arises from an interesting physical system. But first, to put this in context, recall that a well known experimental fact in physics is that magnetic poles always come in pairs of two different types: a "north" pole and a "south" pole. But in 1931, the famous physicist Dirac showed that the mere existence of one isolated magnetic pole (monopole) in the Universe necessarily implies that the electric charge of any particle must be an integral multiple of some fixed universal quantity. Although no monopole had ever been observed, it was the first time in history that a theoretical argument led to the intriguing fact that electric charge is quantized—a phenomenon that was (and still is) well established by many rigorous experiments. This then generated a great enthusiasm in the physics community which put magnetic monopoles among one of the major subject of interest (see Rajantie (2012) for a good historical and theoretical overview of monopoles.)

In this paper, we ask for the classical motion of an electrically charged particle in the field of a magnetic monopole. This gives a system of three nonlinear second-order differential equations involving first order derivatives:

$$\begin{aligned} \ddot{x} &= \frac{\dot{y}z - y\dot{z}}{(x^2+y^2+z^2)^{3/2}} \\ \ddot{y} &= \frac{\dot{z}x - z\dot{x}}{(x^2+y^2+z^2)^{3/2}} \\ \ddot{z} &= \frac{\dot{x}y - x\dot{y}}{(x^2+y^2+z^2)^{3/2}} \end{aligned} \tag{1}$$

where the dots denote derivative with respect to an independent variable t (time in this case). Very few such systems are integrable, but we will see that the method of Lie point symmetries enables us to solve it completely.

In §2, we will explain how to arrive at (1). In §3, we will give a short introduction to the method of Lie point symmetries. In §4, we will show how to obtain the symmetries of (1) and in §5 and §6, we will use them to derive the complete solution.

2 THE EQUATIONS OF MOTION

The magnetic field produced by a monopole at the origin is

$$\mathbf{B} = g\frac{\mathbf{r}}{r^3}, \tag{2}$$

where $\mathbf{r} = (x, y, z)$ is the radial vector in $\mathbb{R}^3$, $r = |\mathbf{r}|$ and g is a constant called the *magnetic charge*. Notice the close analogy to the electric field $\mathbf{E} = e\frac{\mathbf{r}}{r^3}$ of an electric charge, or to $\mathbf{F} = GMm\frac{\mathbf{r}}{r^3}$ of Newtonian gravity—these are all "inverse square laws". In the case of monopoles, however, it is a very different problem because of the special way in which the magnetic field acts on charged particles. Indeed, particles in an electromagnetic field are subject to the Lorentz force

$$\mathbf{F} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right), \tag{3}$$

where $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields respectively, $\mathbf{v} = \dot{\mathbf{r}} = \frac{d}{dt}\mathbf{r}(t)$, $\times$ is the usual vector cross-product and q is the electric charge of the particle. Thus, as opposed to the two other cases, it is not a central-force problem. Indeed, by Newton's second law and (3), the equation of motion of a particle of mass m in the field (2) is

$$m\ddot{\mathbf{r}} = gq\frac{\dot{\mathbf{r}} \times \mathbf{r}}{r^3} \tag{4}$$

By scaling the variables appropriately so that we get rid of the constants and by expanding with respect to each Cartesian coordinate, we get (1).

3 LIE POINT SYMMETRIES

The goal of this section is to briefly introduce the method of Lie point symmetries to solve differential equations and to give the main results that will be used in this paper. For more details see Stephani and MacCallum (1989) or Olver (1993) from which this section is largely inspired.

3.1 Overview

Roughly speaking, a point symmetry of an ordinary differential equation (ODE) is a smooth invertible transformation of the dependent and independent variables that leaves the ODE invariant; that is, it sends every solution of the ODE to another solution. We will be interested more particularly in Lie point symmetries which are, briefly, those that vary continuously with a parameter. For example, the ODE $y'' + y = 0$ has the point symmetries $(x, y) \mapsto (-x, y)$ and $(x, y) \mapsto (x, ay)\ \forall a \in \mathbb{R}$ where only the later is a Lie point symmetry.

In the later part of the nineteenth century, the mathematician Sophus Lie made an astonishing discovery: practically all the known methods for solving ODEs (separation of variables, integrating factor, variation of parameter, Wronskian solutions, etc.) rely, in fact, on the symmetries of the ODE. Furthermore, he developed a systematic way of finding the symmetries and using them to get an integration procedure.

Moreover, for systems with a Lagrangian (defined below), there is a deep connection between symmetries and conservation laws (i.e. functions of the dependent variables, their time derivatives and time that are constant along the solutions). This interdependence is encapsulated in a beautiful theorem due to Noether (1918). It basically shows how to construct the conservation laws from certain kind of symmetries. With enough of these, the system can usually be solved. In physics, for instance, this theorem tells how time translation invariance leads to conservation of energy and how the spherical symmetry of space leads to conservation of angular momentum.

3.2 The Symmetry Condition

For a smooth invertible map

$$\begin{array}{ccc} \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ (x,y) & \longmapsto & (\tilde{x}(x,y), \tilde{y}(x,y)), \end{array} \tag{5}$$

we denote

$$\tilde{y}' = \frac{d\tilde{y}}{d\tilde{x}}, \quad \ldots, \quad \tilde{y}^{(n)} = \frac{d\tilde{y}^{(n-1)}}{d\tilde{x}}$$

By definition, (5) is a point symmetry of the ODE

$$H(x, y, y', \ldots, y^{(n)}) = 0 \tag{6}$$

if and only if

$$H(\tilde{x}, \tilde{y}, \tilde{y}', \ldots, \tilde{y}^{(n)}) = 0 \tag{7}$$

whenever (6) is true. This is the general symmetry condition for point symmetries.

In principle, we can find the symmetries of (6) by using the derivative rule

$$\frac{d\tilde{y}}{d\tilde{x}} = \frac{\tilde{y}_x + y'\tilde{y}_y}{\tilde{x}_x + y'\tilde{x}_y} \tag{8}$$

(where we denote $y_x := \frac{\partial y}{\partial x}$, etc.) with (7) to get a PDE for $\tilde{x}(x, y)$ and $\tilde{y}(x, y)$. However, this PDE is in general too hard to solve, so we need another way of seeing symmetries. To do so, we restrict to a special type of point symmetries called *Lie point symmetries*, or more precisely *one-parameter local Lie groups of point symmetries*. Roughly speaking, these are the point symmetries that vary smoothly with a parameter ε and such that setting $\varepsilon = 0$ gives the identity transformation. They can thus be written

$$\begin{aligned} \tilde{x}(x, y; \varepsilon) &= x + \xi(x, y)\varepsilon + \mathrm{O}(\varepsilon^2) \\ \tilde{y}(x, y; \varepsilon) &= y + \eta(x, y)\varepsilon + \mathrm{O}(\varepsilon^2) \end{aligned} \tag{9}$$

for some functions ξ, η. Moreover, since the transformation is smooth, there are also functions $\eta'(x, y, y'), \ldots, \eta^{(n)}(x, y, y', \ldots, y^{(n)})$ such that

$$\begin{aligned} \tilde{y}' &= y' + \eta'\varepsilon + \mathrm{O}(\varepsilon^2) \\ &\vdots \\ \tilde{y}^{(n)} &= y^{(n)} + \eta^{(n)}\varepsilon + \mathrm{O}(\varepsilon^2) \end{aligned}$$

We define the *infinitesimal generator* of the transformation (9) to be

$$\mathbf{X} = \xi\frac{\partial}{\partial x} + \eta\frac{\partial}{\partial y} + \eta'\frac{\partial}{\partial y'} + \cdots + \eta^{(n)}\frac{\partial}{\partial y^{(n)}} \tag{10}$$

Notice that $\mathbf{X}$ is completely equivalent to the finite transformations $\tilde{x}(x, y; \varepsilon)$ and $\tilde{y}(x, y; \varepsilon)$ since these functions can be recovered by solving

$$\frac{d\tilde{x}}{d\varepsilon} = \xi(\tilde{x}, \tilde{y}), \quad \frac{d\tilde{y}}{d\varepsilon} = \eta(\tilde{x}, \tilde{y})$$

with initial conditions $\tilde{x}(x, y; 0) = x$, $\tilde{y}(x, y; 0) = y$. In fact, we will see that the whole method of Lie point symmetries makes use only of the form (10) of the transformations—we don't even need to know the corresponding finite transformation. Since the functions $\eta', \ldots, \eta^{(n)}$ are then fully determined by ξ and η, we sometimes only write $\mathbf{X} = \xi\frac{\partial}{\partial x} + \eta\frac{\partial}{\partial y}$ for the infinitesimal generator of a transformation and refers to (10) as the *extended* infinitesimal generator.

By expanding (8) in powers of ε, we can show that in general

$$\eta^{(k)} = \frac{d}{dx}\eta^{(k-1)} - y^{(k)}\frac{d\xi}{dx} \tag{11}$$

We put another restriction by assuming thereafter that the ODE $H = 0$ satisfies the requirement that not all first derivatives of H vanish on $H = 0$. Equivalently, this means that it can be put in the form

$$y^{(n)} = \omega(x, y, y', \ldots, y^{(n-1)}) \tag{12}$$

Now, it is interesting to observe that a generator as in given in (10) corresponds to a Lie point symmetry of $H = 0$ exactly if

$$\mathbf{X}H \equiv 0 \pmod{H = 0}$$

where $\equiv$ means that the equation must be identically true for all variables $x, y, y', \ldots, y^{(n-1)}$ and "mod $H = 0$" means that $y^{(n)}$ must be eliminated by means of $H = 0$.

To get an efficient way of computing symmetries, we will need to associate to the ODE (12) the partial differential operator

$$\mathbf{A} = \frac{\partial}{\partial x} + y'\frac{\partial}{\partial y} + y''\frac{\partial}{\partial y'} + \cdots + \omega\frac{\partial}{\partial y^{(n-1)}} \tag{13}$$

Now, it can be shown that (10) is the infinitesimal generator of a Lie point symmetry of (12) if and only if

$$\mathbf{X}\mathbf{A} - \mathbf{A}\mathbf{X} = \lambda\mathbf{A} \tag{14}$$

for some function λ. This is what will enable us to actually find the symmetries. Indeed, (14) gives rise to a PDE for $\xi(x, y)$ and $\eta(x, y)$, and since these functions are independent of $y', \ldots, y^{(n)}$, it splits into an overdetermined system of PDEs that is usually possible to solve if symmetries exist.

3.3 Systems of ODEs

We only talked about single ODEs, but all the results of this section apply as well to systems of ODEs. Since in this paper we are dealing with a system of second-order ODE, we will simply restate the results for this particular case. But first, we need some notation. We will use coordinates $x^1, \ldots, x^N$ (the indeces in the superscript are not to be confused with exponents) and use Einstein summation convention, which says that when an index appears both as an upper and a lower index in a single term, summation over all possible values of that index is assumed. For example, in Euclidean space we would have $(x^1, x^2, x^3) = (x, y, z)$ and

$$x^a\frac{d}{dx^a} = x\frac{d}{dx} + y\frac{d}{dy} + z\frac{d}{dz}$$

Using that notation (which is standard for this field of study), a general system of N second-order ODE reads

$$\ddot{x}^k = \omega^k(x^i, \dot{x}^i; t), \quad i, k = 1, \ldots, N \tag{15}$$

and its corresponding partial differential operator is

$$\mathbf{A} = \frac{\partial}{\partial t} + \dot{x}^k\frac{\partial}{\partial x^k} + \omega^k\frac{\partial}{\partial \dot{x}^k}$$

A general extended infinitesimal generator of a transformation now takes the form

$$\mathbf{X} = \xi(x^i, t)\frac{\partial}{\partial t} + \eta^k(x^i, t)\frac{\partial}{\partial x^k} + \dot{\eta}^k(x^i, \dot{x}^i, t)\frac{\partial}{\partial \dot{x}^k}$$

With these new operators $\mathbf{X}$ and $\mathbf{A}$, the symmetry condition (14) still holds. By comparing the coefficients of $\frac{\partial}{\partial t}$ in (14) we get $\lambda = -\mathbf{A}\xi$, and so the coefficients of $\frac{\partial}{\partial x^k}$ yields

$$\dot{\eta}^k = \mathbf{A}\eta^k - \dot{x}^k\mathbf{A}\xi \tag{16}$$

which is the formula analogous to (11). For the coefficients of $\frac{\partial}{\partial \dot{x}^k}$, we have

$$\mathbf{X}\omega^k = \mathbf{A}\dot{\eta}^k - \omega^k\mathbf{A}\xi \tag{17}$$

By inserting (16) into (17) and expanding the result we get (for $\omega^a_b := \frac{\partial \omega^a}{\partial x^b}$, etc.)

$$\begin{aligned}
&\xi\omega^a_t + \eta^b\omega^a_b + \left(\eta^b_t + \dot{x}^c\eta^j_c - \dot{x}^b\xi_t - \dot{x}^b\dot{x}^c\xi_c\right)\frac{\partial\omega^a}{\partial\dot{x}^b} \\
&+ 2\omega^a\left(\xi_t + \dot{x}^b\xi_b\right) + \omega^b\left(\dot{x}^a\xi_b - \eta^a_b\right) + \dot{x}^a\dot{x}^b\dot{x}^c\xi_{bc} \\
&+ 2\dot{x}^a\dot{x}^c\xi_{tc} - \dot{x}^c\dot{x}^b\eta^a_{bc} + \dot{x}^a\xi_{tt} - 2\dot{x}^b\eta^i_{tb} - \eta^a_{tt} \\
&= 0, \quad \text{for } a = 1, \ldots, N
\end{aligned} \tag{18}$$

Solving this equation will give the Lie point symmetries of (15). But don't be scared by its length; since ξ and η^k do not depend on any $\dot{x}^i$, it will split into many small partial differential equations.

3.4 Noether's Theorem

We are now ready to state in more details the close connection between symmetries and conservation laws.

Theorem 1 (Noether). *Suppose* $\mathbf{X} = \xi(x^i,t)\frac{\partial}{\partial t} + \eta^k(x^i,t)\frac{\partial}{\partial x^k} + \dot{\eta}^k(x^i,\dot{x}^i,t)\frac{\partial}{\partial \dot{x}^k}$ *is the generator of a Lie point symmetry of a system of ordinary differential equations derived from a Lagrangian* $\mathcal{L}$*. That is, the equations of the system are obtained by substituting a function* $\mathcal{L}(x^i,\dot{x}^i,t)$ *into the Euler-Lagrange equations*

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{x}^k} = \frac{\partial \mathcal{L}}{\partial x^k} \tag{19}$$

If there exists a function $V(x^i,t)$ *such that*

$$\mathbf{X}\mathcal{L} + \frac{d\xi}{dt}\mathcal{L} = \frac{dV}{dt}, \tag{20}$$

then

$$\varphi = \xi\left(\dot{x}^k\frac{\partial \mathcal{L}}{\partial \dot{x}^k} - \mathcal{L}\right) - \eta^k\frac{\partial \mathcal{L}}{\partial \dot{x}^k} + V$$

is constant along the solutions of the system.

Symmetries satisfying condition (20) are usually called *Noether symmetries* and the corresponding functions φ are called *conservation laws*.

4 Lie Point Symmetries of the Charge-Monopole System

4.1 Solving the Symmetry Condition

As mentioned above, to find the Lie point symmetries of (1) we need to solve (18). But first, we will write the system (1) in a more compact form by using Einstein summation. To this end, we introduce the Levi-Civita symbol ε_{ijk}, which is defined to be completely antisymmetric with $\varepsilon_{123} = 1$. That is, $\varepsilon_{ijk} = 1$ if (ijk) is an even permutation of (123), $\varepsilon_{ijk} = -1$ if (ijk) is an odd permutation of (123) and $\varepsilon_{ijk} = 0$ otherwise. We will often raise an index (e.g. $\varepsilon^k{}_{ij}$) in order to be consistent with Einstein summation convention, but the indices are always to be read from left to right. Moreover, we denote $r = \sqrt{(x^1)^2+(x^2)^2+(x^3)^2}$. With that notation, (1) takes the simple form

$$\dot{x}^k = r^{-3}\varepsilon^k{}_{ij}\dot{x}^i x^j, \quad k = 1,2,3$$

Substituting into the symmetry condition (18), we get

$$\begin{aligned}
&-\eta^b r^{-5}\left(3x^b\varepsilon^a{}_{ij}\dot{x}^i x^j + r^2\varepsilon^a{}_{bi}\dot{x}^i\right)\\
&\quad + \left(\eta^b_t + \dot{x}^c\eta^b_c - \dot{x}^b\xi_t - \dot{x}^b\dot{x}^c\xi_c\right)r^{-3}\varepsilon^a{}_{bi}x^i\\
&\quad + 2r^{-3}\varepsilon^a{}_{ij}\dot{x}^i x^j\left(\xi_t + \dot{x}^b\xi_b\right)\\
&\quad + r^{-3}\varepsilon^b{}_{ij}\dot{x}^i x^j\left(\dot{x}^a\xi_b - \eta^a_b\right) + \dot{x}^a\dot{x}^b\dot{x}^c\xi_{bc}\\
&\quad + 2\dot{x}^a\dot{x}^c\xi_{tc} - \dot{x}^c\dot{x}^b\eta^a_{bc} + \dot{x}^a\xi_{tt} - 2\dot{x}^b\eta^a_{tb}\\
&\quad - \eta^a_{tt} = 0, \qquad \text{for } a = 1,2,3
\end{aligned} \tag{21}$$

Since ξ and η^k depend only on x^i and t, this equation is a polynomial in the three variables $\dot{x}^1,\dot{x}^2,\dot{x}^3$, so we can equate to zero each coefficient to get a system of equations. We can then solve each equation starting from those of the higher order terms to the lower. However, this procedure is significantly lengthy and the details are intricate, so we will only show the first few steps and state the final result.

For the coefficient of the term cubic in $\dot{x}^a$, we get

$$\xi_{aa} = 0 \tag{22}$$

There are no term cubic in $\dot{x}^i$ for $i \neq a$. For the terms quadratic in $\dot{x}^a$, we find

$$r^{-3}\varepsilon^b{}_{aj}x^j\xi_b + 2\dot{x}^b\xi_{ab} + 2\xi_{ta} - \eta^a_{aa} = 0, \tag{23}$$
$$b\neq a$$

whence $\xi_{ab} = 0$ for $b \neq a$. Thus, by (22) we have

$$\xi(x,t) = C_k(t)x^k + B(t)$$

for some functions $C_k(t)$, $B(t)$. Putting this back into (23), we get

$$\eta^a_{aa} = 2\dot{C}_a + r^{-3}\varepsilon^j{}_{ai}x^i C_j$$

Now, because x^a does not appear in $\varepsilon^j{}_{ai}x^i C_j$ and because

$$\frac{\partial^2}{(\partial x^a)^2}\left(\frac{r}{r^2-(x^a)^2}\right) = r^{-3},$$

we have

$$\begin{aligned}
\eta^a(x,t) = \dot{C}_a(x^a)^2 + \frac{r}{r^2-(x^a)^2}\varepsilon^j{}_{ai}x^i C_j\\
+ \alpha^a(x,t)x^a + \beta^a(x,t),
\end{aligned}$$

for some functions α^a and β^a that are independent of x^a.

By continuing in this way, we obtain the complete solution of (21):

$$\begin{aligned}\xi(x,t) &= a_0 + 2a_1 t + a_2 t^2\\ \eta^1(x,t) &= (a_1 + a_2 t)x^1 + b_2^1 x^2 + b_3^1 x^3\\ \eta^2(x,t) &= -b_2^1 x^1 + (a_1 + a_2 t)x^2 + b_3^2 x^3\\ \eta^3(x,t) &= -b_3^1 x^1 - b_3^2 x^2 + (a_1 + a_2 t)x^3\end{aligned}$$

for constants $a_i, b_k^j \in \mathbb{R}$. Hence, we conclude that the charge-monopole system has exactly six linearly independent Lie point symmetries:

$$\begin{aligned}\mathbf{X}_1 &= x^2\frac{\partial}{\partial x^1} - x^1\frac{\partial}{\partial x^2}\\ \mathbf{X}_2 &= x^3\frac{\partial}{\partial x^1} - x^1\frac{\partial}{\partial x^3}\\ \mathbf{X}_3 &= x^3\frac{\partial}{\partial x^2} - x^2\frac{\partial}{\partial x^3}\\ \mathbf{X}_4 &= \frac{\partial}{\partial t}\\ \mathbf{X}_5 &= 2t\frac{\partial}{\partial t} + x^1\frac{\partial}{\partial x^1} + x^2\frac{\partial}{\partial x^2} + x^3\frac{\partial}{\partial x^3}\\ \mathbf{X}_6 &= t^2\frac{\partial}{\partial t} + tx^1\frac{\partial}{\partial x^1} + tx^2\frac{\partial}{\partial x^2} + tx^3\frac{\partial}{\partial x^3}\end{aligned}$$

By using (16), we get the corresponding extended infinitesimal generators:

$$\begin{aligned}\mathbf{X}_n &= \varepsilon_{nij}\left(x^j\frac{\partial}{\partial x^i} + \dot{x}^j\frac{\partial}{\partial \dot{x}^i}\right), \quad n = 1,2,3\\ \mathbf{X}_4 &= \frac{\partial}{\partial t}\\ \mathbf{X}_5 &= 2t\frac{\partial}{\partial t} + x^k\frac{\partial}{\partial x^k} - \dot{x}^k\frac{\partial}{\partial \dot{x}^k}\\ \mathbf{X}_6 &= t^2\frac{\partial}{\partial t} + tx^k\frac{\partial}{\partial x^k} + (x^k - t\dot{x}^k)\frac{\partial}{\partial \dot{x}^k}\end{aligned}$$

4.2 Interpretation of the Symmetries

The first three symmetries $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ correspond to the fact that the system is invariant under three-dimensional rotation through the origin. Indeed, rotation of the (x,y)-plane by an angle ε is

$$\begin{aligned}\tilde{x}(x,y;\varepsilon) &= x\cos\varepsilon - y\sin\varepsilon\\ \tilde{y}(x,y;\varepsilon) &= x\sin\varepsilon + y\cos\varepsilon,\end{aligned}$$

and we can check that the infinitesimal generator corresponding to this transformation is

$$\mathbf{X} = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial x}$$

Thus, $\mathbf{X}_1$ is a rotation about the x^3-axis, $\mathbf{X}_2$ about the x^2-axis and $\mathbf{X}_3$ about the x^1-axis. That is, $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ together give an arbitrary three-dimensional rotation. This could have been anticipated since the charge-monopole system has a spherical symmetry—there is no preferred direction of space.

Now, $\mathbf{X}_4$ correspond to the fact that the equations of motion are independent of the time t. Any solution can be translated in time $t \mapsto t + \varepsilon$ to give another solution.

To explain $\mathbf{X}_5$ and $\mathbf{X}_6$, we first express them in spherical coordinates:

$$\begin{aligned}\mathbf{X}_5 &= 2t\frac{\partial}{\partial t} + r\frac{\partial}{\partial r}\\ \mathbf{X}_6 &= t^2\frac{\partial}{\partial t} + tr\frac{\partial}{\partial r}\end{aligned}$$

For $\mathbf{X}_5$, it gives the transformation

$$\tilde{t}(t,r;\varepsilon) = e^{2\varepsilon}t, \quad \tilde{r}(t,r;\varepsilon) = e^{\varepsilon}r$$

and for $\mathbf{X}_6$ we have

$$\tilde{t}(t,r;\varepsilon) = \frac{t}{1-\varepsilon t}, \quad \tilde{r}(t,r,\varepsilon) = \frac{r}{1-\varepsilon t}$$

5 Conservation Laws

We will now use Noether's Theorem to obtain conservation laws from the symmetries just found. But first, we need to find a Lagrangian, i.e. a function $\mathcal{L}$ that gives the system of equations (1) when substituted into the Euler-Lagrange equations (19).

5.1 The Lagrangian

It can be shown (see Taylor (2005)) that in general, the motion of a particle of charge q and mass m under the Lorentz force (3) has—at least locally—the Lagrangian

$$\mathcal{L}(\mathbf{r}, \dot{\mathbf{r}}, t) = \frac{1}{2}m\dot{\mathbf{r}}^2 - q\left(V - \dot{\mathbf{r}}\cdot\mathbf{A}\right) \qquad (24)$$

where V and $\mathbf{A}$ (called the electric and magnetic potentials) are defined such that the electric and magnetic fields are respectively

$$\mathbf{E} = -\nabla V - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla\times\mathbf{A}$$

In the case of the charge-monopole system, i.e. when $\mathbf{E} = 0$ and $\mathbf{B} = g\frac{\mathbf{r}}{r^3}$, a solution is to take $V = 0$ and $\mathbf{A}$ in spherical coordinates (r,θ,φ) with

$$\mathbf{A}_r = 0, \quad \mathbf{A}_\theta = 0, \quad \mathbf{A}_\varphi = \frac{g(1-\cos\theta)}{r\sin\theta} = \frac{g}{r}\tan\frac{\theta}{2}$$

Indeed, by using the formula for $\nabla \times \mathbf{A}$ in spherical coordinates, we get

$$\begin{aligned}\nabla \times \mathbf{A} &= \mathbf{e}_r \frac{1}{r\sin\theta}\left(\frac{\partial}{\partial\theta}(\sin\theta A_\varphi) - \frac{\partial A_\theta}{\partial\varphi}\right) \\ &\quad + \mathbf{e}_\theta\left(\frac{1}{r\sin\theta}\frac{\partial A_r}{\partial\varphi} - \frac{1}{r}\frac{\partial}{\partial r}(rA_\varphi)\right) \\ &\quad + \mathbf{e}_\varphi\frac{1}{r}\left(\frac{\partial}{\partial r}(rA_\theta) - \frac{\partial A_r}{\partial\theta}\right) \\ &= \mathbf{e}_r\frac{1}{r\sin\theta}\left(\frac{g}{r}\sin\theta\right) = g\frac{\mathbf{r}}{r^3}\end{aligned}$$

Now, inserting in (24) and switching to the dimensionless variables of (1), we conclude that the system has the Lagrangian

$$\mathcal{L}(\mathbf{r},\dot{\mathbf{r}},t) = \frac{\dot{x}^2+\dot{y}^2+\dot{z}^2}{2} + \frac{x\dot{y}-\dot{x}y}{x^2+y^2}\left(1-\frac{z}{r}\right) \tag{25}$$

on $\mathbb{R}^3 \setminus \{z\text{-axis}\}$.

However, notice that the vector potential $\mathbf{A}$ is not defined for $\theta = \pi$. Such a singularity is unavoidable since we can show that no global magnetic vector potential is possible. Indeed, from basic vector analysis we know that $\nabla\cdot(\nabla\times\mathbf{A}) = 0$ for any $\mathbf{A}$, but we will show that $\iiint_E \nabla\cdot\mathbf{B}dV = 4\pi g$ for any volume E that contains the origin which then contradicts that $\mathbf{B} = \nabla\times\mathbf{A}$. Let S_a be a sphere, contained in E, of radius a centred at the origin, and let E_0 be the volume enclosed by $S_a \cup \partial E$. By the Divergence Theorem,

$$\begin{aligned}\iiint_{E_0}\nabla\cdot\mathbf{B}dV &= \iint_{S_a\cup\partial E}\mathbf{B}\cdot d\mathbf{S} \\ &= \iint_{\partial E}\mathbf{B}\cdot d\mathbf{S} - \iint_{S_a}\mathbf{B}\cdot d\mathbf{S}\end{aligned}$$

But when $\mathbf{r} \neq (0,0,0)$, it is a straightforward computation to show that $\nabla\cdot\mathbf{B} = 0$ which implies that the left hand side is zero, and that the surface integral on ∂E is equal to that on the sphere of radius a. Thus, using the Divergence Theorem again,

$$\begin{aligned}\iiint_E\nabla\cdot\mathbf{B}dV &= \iint_{\partial E}\mathbf{B}\cdot d\mathbf{S} = \iint_{S_a}\mathbf{B}\cdot d\mathbf{S} \\ &= \iint_{S_a} g\frac{\mathbf{r}}{r^3}\cdot\frac{\mathbf{r}}{r}dS = \frac{g}{a^2}\iint_{S_a}dS \\ &= \frac{g}{a^2}(4\pi a^2) = 4\pi g\end{aligned}$$

This problem was first thought as a strong indication that no magnetic monopole could exist, but Dirac showed, that we can nevertheless build a consistent theory of magnetic monopoles by allowing the potential $\mathbf{A}$ to be discontinuous on an infinitesimal "string"—as he called it—which start at the charge and extend to infinity (Dirac (1931)). This string carries the magnetic flux necessary to explain that $\iint_S \mathbf{B}\cdot d\mathbf{S} = 0$ everywhere, but since it is infinitesimal, particles don't "see" it and thus behave exactly as if $\mathbf{B} = g\frac{\mathbf{r}}{r^3}$.

5.2 *Application of Noether's Theorem*

For $\mathbf{X}_1$ with the Lagrangian (25), we get

$$\mathbf{X}_1\mathcal{L} + \frac{d\xi_1}{dt}\mathcal{L} = 0$$

Hence, by taking $V = 0$ in (20), we see that $\mathbf{X}_1$ is a Noether symmetry. Noether's Theorem then tells us that

$$\varphi_1 = x\dot{y} - \dot{x}y + 1 - \frac{z}{r}$$

or equivalently,

$$\ell_z = \dot{x}y - x\dot{y} + \frac{z}{r}$$

is constant. Now, since the system is invariant under three dimensional rotation, rotating this conservation law must also give conservation laws. Hence, we have that

$$\begin{aligned}\ell_x &= \dot{y}z - y\dot{z} + \frac{x}{r} \\ \ell_y &= \dot{z}x - z\dot{x} + \frac{y}{r}\end{aligned}$$

are also conserved quantity. More compactly, the vector

$$\mathbf{L} = \dot{\mathbf{r}}\times\mathbf{r} + \frac{\mathbf{r}}{r}$$

is conserved along the solutions.

For $\mathbf{X}_4$, the condition (20) is trivially satisfied and we get the conservation law

$$\varphi_4 = \frac{1}{2}\dot{\mathbf{r}}^2$$

That is, the speed of the particle is constant.

For $\mathbf{X}_5$, we find

$$\mathbf{X}_5\mathcal{L} + \frac{d\xi_5}{dt}\mathcal{L} = 2\mathcal{L}$$

but $2\mathcal{L} \neq \frac{dV}{dt}$ for any function V, so it is not a Noether symmetry.

For $\mathbf{X}_6$, we get

$$\mathbf{X}_6\mathcal{L} + \frac{d\xi_6}{dt}\mathcal{L} = x\dot{x} + y\dot{y} + z\dot{z} = \frac{d}{dt}\left(\frac{r^2}{2}\right)$$

from which we find, by Noether's Theorem, the conservation law

$$\varphi_6 = \frac{1}{2}(t\dot{\mathbf{r}} - \mathbf{r})^2$$

We note that all these conservation laws can be verified by simply differentiating with respect to time and inserting the equation of motion whenever $\ddot{\mathbf{r}}$ appears.

Before deriving the complete solution, we highlight two interesting facts that arise from these conservation laws. First, the magnitude of the angular momentum, i.e. $|\dot{\mathbf{r}} \times \mathbf{r}|$, is conserved. Indeed, we find

$$(\dot{\mathbf{r}} \times \mathbf{r})^2 = \mathbf{L}^2 - 1 \tag{26}$$

Second, if we calculate the angle θ between the position $\mathbf{r}$ and the fixed vector $\mathbf{L}$ we get

$$\cos\theta = \frac{\mathbf{r}\cdot\mathbf{L}}{|\mathbf{r}||\mathbf{L}|} = \frac{1}{|\mathbf{r}||\mathbf{L}|}\mathbf{r}\cdot\left(\dot{\mathbf{r}}\times\mathbf{r} + \frac{\mathbf{r}}{|\mathbf{r}|}\right) = \frac{1}{|\mathbf{L}|} \tag{27}$$

which is constant. Hence, the symmetries of the system already tell us that the particle is restricted to move at constant speed on the surface of a cone (with axis is in the direction of $\mathbf{L}$). Notice that from (26), we know that $0 < \frac{1}{|\mathbf{L}|} \leq 1$, so (27) is well-defined.

6 Complete Solution

To sum up what we have found so far, the system

$$\ddot{\mathbf{r}} = \frac{\dot{\mathbf{r}}\times\mathbf{r}}{r^3} \tag{28}$$

has the conservation laws

$$\begin{aligned} \mathbf{L} &= \dot{\mathbf{r}}\times\mathbf{r} + \frac{\mathbf{r}}{r} \\ \dot{\mathbf{r}}_0^2 &= \dot{\mathbf{r}}^2 \\ \mathbf{r}_0^2 &= (\mathbf{r} - t\dot{\mathbf{r}})^2 \end{aligned} \tag{29}$$

where we introduced the initial conditions $\mathbf{r}_0 = \mathbf{r}(0)$, $\dot{\mathbf{r}}_0 = \dot{\mathbf{r}}(0)$ of the system. Note that the constant vector $\mathbf{L}$ can also be expressed completely in terms of the initial conditions: $\mathbf{L} = \dot{\mathbf{r}}_0 \times \mathbf{r}_0 + \frac{\mathbf{r}_0}{|\mathbf{r}_0|}$.

To derive the complete solution, we will first get an exact expression for $r^2 = \mathbf{r}(t)^2$. By (29), we have

$$\mathbf{r}_0^2 = (\mathbf{r} - t\dot{\mathbf{r}})^2 = \mathbf{r}^2 - 2t\mathbf{r}\cdot\dot{\mathbf{r}} + t^2\dot{\mathbf{r}}_0^2,$$

whence

$$\frac{dr^2}{dt} = 2r\frac{dr}{dt} = 2r\frac{\mathbf{r}\cdot\dot{\mathbf{r}}}{r} = \frac{r^2 + t^2\dot{\mathbf{r}}_0^2 - \mathbf{r}_0^2}{t}$$

This is a simple first-order linear equation whose solution is

$$r(t)^2 = \mathbf{r}_0^2 + Ct + \dot{\mathbf{r}}_0^2 t^2 \tag{30}$$

for some constant C. By differentiating (30) with respect to t and by evaluating at $t = 0$, we get $C = 2\mathbf{r}_0\cdot\dot{\mathbf{r}}_0$, whence

$$r(t)^2 = (\mathbf{r}_0 + \dot{\mathbf{r}}_0 t)^2 \tag{31}$$

Now, by substituting $\dot{\mathbf{r}}\times\mathbf{r} = \mathbf{L} - \frac{\mathbf{r}}{r}$ in the system (28) and using (31), we find

$$\ddot{\mathbf{r}} + \frac{\mathbf{r}}{|\mathbf{r}_0 + \dot{\mathbf{r}}_0 t|^4} = \frac{\dot{\mathbf{r}}_0\times\mathbf{r}_0 + \frac{\mathbf{r}_0}{|\mathbf{r}_0|}}{|\mathbf{r}_0 + \dot{\mathbf{r}}_0 t|^3},$$

which is three separated second-order linear inhomogeneous ordinary differential equations—one for each Cartesian coordinate. We can thus solve for each coordinate separately by using the standard technique of the Wronskian. If $\dot{\mathbf{r}}_0 \times \mathbf{r}_0 \neq 0$, we find

$$\begin{aligned} \mathbf{r}(t) = |\mathbf{r}_0 + \dot{\mathbf{r}}_0 t| \Bigg(& \frac{\frac{\mathbf{r}_0}{|\mathbf{r}_0|} + \dot{\mathbf{r}}_0\times\mathbf{r}_0}{1 + |\dot{\mathbf{r}}_0\times\mathbf{r}_0|^2} \\ & + \mathbf{C}_1 \sin\left(\sqrt{1 + \frac{1}{|\dot{\mathbf{r}}_0\times\mathbf{r}_0|^2}}\arctan\frac{\dot{\mathbf{r}}_0\cdot(\mathbf{r}_0 + \dot{\mathbf{r}}_0 t)}{|\dot{\mathbf{r}}_0\times\mathbf{r}_0|}\right) \\ & + \mathbf{C}_2 \cos\left(\sqrt{1 + \frac{1}{|\dot{\mathbf{r}}_0\times\mathbf{r}_0|^2}}\arctan\frac{\dot{\mathbf{r}}_0\cdot(\mathbf{r}_0 + \dot{\mathbf{r}}_0 t)}{|\dot{\mathbf{r}}_0\times\mathbf{r}_0|}\right)\Bigg) \end{aligned}$$

where the constant vectors $\mathbf{C}_1$ and $\mathbf{C}_2$ can be found by solving the system of linear equations

$$\mathbf{r}(0) = \mathbf{r}_0, \quad \dot{\mathbf{r}}(0) = \dot{\mathbf{r}}_0$$

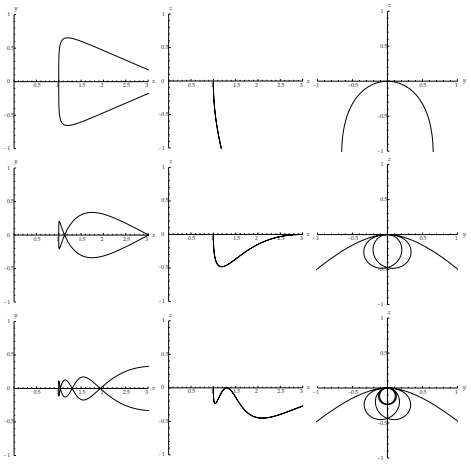
If $\dot{\mathbf{r}}_0 \times \mathbf{r}_0 = 0$, then, due to the conservation of the magnitude of the angular momentum (26), we have

$$|\ddot{\mathbf{r}}| = \frac{1}{r^3}|\dot{\mathbf{r}}\times\mathbf{r}| = \frac{1}{r^3}|\dot{\mathbf{r}}_0\times\mathbf{r}_0| = 0,$$

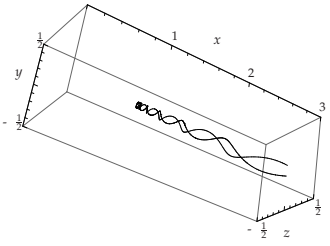
whence

$$\mathbf{r}(t) = \mathbf{r}_0 + \dot{\mathbf{r}}_0 t$$

To illustrate graphically some solutions, we take $\mathbf{r}_0 = (1,0,0)$ and $\dot{\mathbf{r}}_0 = (0,1/n,0)$ for some $n > 0$. For $n = 2,5,9$, we get the following graphs, where the columns correspond to the x-y, x-z and y-z planes respectively and the rows correspond to $n = 1,5,9$ from top to bottom.



Also, if we set $n = 16$, we get the following three-dimensional graph:



7 Conclusion

We found that the method of Lie point symmetries can be used to solve even something as complex as a system of nonlinear second-order ordinary differential equations with first order derivatives like (1). This particular system has already been solved by other authors—notably by Henri Poincaré in 1896: Poincaré (1896)—but what we have learned here is that we can solve such difficult systems by only a systematic use of the theory of Lie point symmetries. Although all the conservation laws obtained can be easily verified, arriving at them without prior knowledge of such a powerful method would have required great insights and experience with that kind of equation.

8 Acknowledgement

I would like to thank Professor Niky Kamran for his supervision and for introducing me to the fascinating topic of Lie groups applied to differential equations and in particular to this interesting physical problem. His guidance throughout this summer research experience was invaluable.

References

Dirac, P. A. M. (1931). Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society of London. Series A*, 133(821):60–72.

Noether, E. (1918). Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257.

Olver, P. J. (1993). *Applications of Lie groups to differential equations*, volume 107 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.

Poincaré, H. (1896). Remarques sur une expérience de m. birkeland. *Comptes rendus hebdomadaires de l'Académie des sciences*, 123:530–533.

Rajantie, A. (2012). Introduction to magnetic monopoles. *Contemporary Physics*, 53(3):195–211.

Stephani, H. and MacCallum, M. A. H. (1989). *Differential equations : their solution using symmetries*. Cambridge [England] ; New York : Cambridge University Press.

Taylor, J. R. (2005). *Classical Mechanics*. University Science Books.

Jokes and Comics

These days, even the most pure and abstract mathematics is in danger to be applied.

INTERVIEW WITH PR. MIKAËL PICHOT

Mathilde Gerbelli-Gauthier

Mathilde met with Prof. Mikaël Pichot to chat about his research, the process of doing mathematics, and family life. The following is a summary of their conversation.

Biography in a box

Nationality	French
Speciality	Group Theory
Alma mater	École Normale Supérieure de Lyon
Coming from	University of Tokyo
Teaching lately	Higher Algebra 1 and 2

$\delta\varepsilon$: How would you describe your research?

My research is principally in algebra, but I am interested by connections between algebra and other branches of mathematics. For example, I use a lot of functional analysis. I try to prove results about infinite groups by using tools from geometry and analysis. There are different approaches to mathematics, but for me it is easier to solve problems by combining tools from different areas. I like to explore the links between these different areas.

An example of something that has been coming up a lot in my research is the idea of Gromov that one can study an infinite group by looking at how this group is represented inside a finite group. For example, one can understand the free group on k generators $\mathbb{F}_k$ by looking at all the homomorphisms from $\mathbb{F}_k$ to the symmetric group S_n, for all n. A way to obtain an invariant of the free group is to count them. In the case of $\mathbb{F}_k$, it is not too hard since a homomorphism is determined by the image of each of the generators of $\mathbb{F}_k$. So you have approximately $n!^k$ homomorphisms, which is of the same order of magnitude as n^{nk}. To study the asymptotics of this as n goes to infinity, you take the logarithm of this number of homomorphisms, which is $nk \log n$. Then by dividing by a factor of $n \log n$ you recover k which is the rank of the free group. Then you try to obtain similar invariants for other infinite groups, but since they are no longer free, you relax the notion of a homomorphism, and instead count maps called quasi-homomorphisms.

An interesting aspect of this way of looking at groups is that they behave like large systems of particles, and so the techniques are very similar to the ones used when you are studying entropy of gases, for example. Your infinite group corresponds to a gas in a certain state. Counting all the homomorphisms to finite groups corresponds to counting all possible microstates to understand the entropy of the gas. So the methods I use have a lot if similarities with Boltzmann's entropy formula.

Picture 1: Pr. Mikaël Pichot

$\delta\varepsilon$: Do you have a favorite group?

…I don't know if I have a favorite group, but there are groups with which I interact daily. These are lattices in the p-adic analogues of Lie Groups.

$\delta\varepsilon$: How do you go about solving problems?

For me it's a very local process. You start with a white page of paper, and a problem for which you have an idea which is generally quite precise. This could be a statement you want to prove or a sequence of explicit steps that form an argument. At this point of the work, it's not even very important what the big picture is. In fact, the general idea seems to evolve on its own while you go through the steps towards the statement you want to prove, to the extent that it sometimes seems that the proof is dictated by the statement you want to prove. This is because a proof has to be "locally coherent": given a starting point, there are very few different ways to

reach your goal. This process leads you to writing pages and pages of mathematics. Sometimes it works and you end up proving something and you write an article from it. If it doesn't then you just wrote pages and pages of mathematics that you end up throwing in the garbage. Sometimes (rarely!), you land on a gold mine.

For example, one of my collaborators and I have this family of groups that we started looking at ten years ago while trying to solve a problem. We never solved the problem, but this family of groups turned out to have extremely rich properties.

$\delta\varepsilon$: You are talking about your collaborator. How does collaboration work? How do you find collaborators in mathematics?

It depends on the context. For example, I met this collaborator when I was a graduate student. I had proven a result about buildings, a branch of mathematics that was not very familiar to me. The proof was horrible, with dozen of cases and sub-cases. I spent an afternoon explaining it to him. He knew much more about buildings than I did, as this was his field of research. So we reworked the proof together, and with his input it became much more beautiful and conceptual. We've kept collaborating ever since!

$\delta\varepsilon$: How did you start doing mathematics?

I was always interested in mathematics. When I was a kid, I could spend hours trying to factor polynomials of degree 2 and 3. This was at the time when I knew nothing about polynomials, not even the quadratic formula, so I generally failed miserably. As I advanced in my studies the density of people that I met who had these sorts of stories got higher and higher. When you talk to mathematicians, it seems like everyone has these examples of very simple problems that they did for fun when they were young.

$\delta\varepsilon$: You have a 3 year-old son. How is it being both a dad and a mathematician?

That's a hard question. It's also a complex situation. It's wonderful to have a kid, but the amount of mathematics you can do decreases significantly. A thing I loved to do before I was a dad was to go for long walks on my own when I was trying to solve problems. Now when I go out for walks, it's with my son. Of course, I can't be thinking about mathematics, because I want to talk to him, teach him words, show him things... But sometimes I end up thinking about mathematics anyways.

AN INTRODUCTION TO GROUPS

Michael Snarski

The purpose of this article is to provide a hands-on introduction to abstract algebra. The exposition attempts to develop intuition rather than build formalism as there is already a vast literature which does the latter. We follow the principle of examples first, definitions later. The reader need have no familiarity with algebra and very little mathematical experience period.

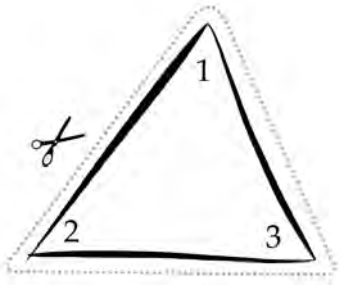
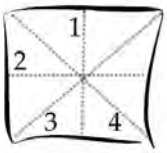


Figure 1: Your own triangle

SO WHAT DO GROUPS *do*?

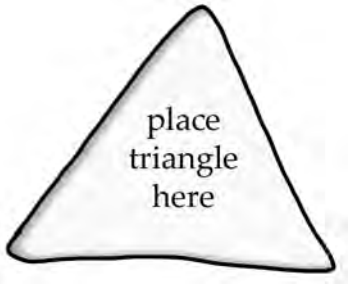
Groups give glimpses of symmetry. For instance, if one wanted to fold a rectangular piece of paper once in two equal halves, they would probably do it along the dashed line 1 or 2 in the figure below.



On the other hand, a square offers two additional lines of symmetry along which one can fold the paper such that the resulting layers coincide. One could argue there is more symmetry in the square; groups provide a way of capturing and measuring this symmetry. Consider the triangle on the right column; its yours, we're giving it to you.

The first thing a mathematician wanting to study the "symmetries" of this triangle might do is label the vertices – this has already been done for you. People usually label things because it gives them a feeling of control: you may not have a clue what's going on, but at least you can blame someone.

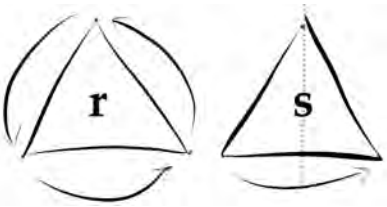
Here's a silly question: how many *distinct* ways can one place this triangle into the following triangle box?



By distinct, we mean distinguishable order and position of the numbers on the vertices of the triangle you got. If you like, you can cut it out and physically check the possibilities.

By doing so, you will expose the *group structure* of these triangles. By group structure, we mean roughly that there is a collection of 'objects' which have rules for interacting with each other. We'll make this precise shortly.

For now, let's say that that **r** denotes rotation by 120° counter-clockwise and that **s** denotes flipping the triangle along the vertical axis:



Rather tautologically, **r** and **s** *act* on this triangle. Moreover, each of these actions can be reversed (one can always rotate 120° clockwise and flip the triangle back). We can also *compose* these actions: we can flip and rotate twice (apply **s**, *then* **r** twice, denoted as $\mathbf{r}^2\mathbf{s}$) or rotate twice, then flip (denoted as $\mathbf{s}\mathbf{r}^2$). Finally, we could rotate,

then rotate the opposite way, effectively changing nothing. This is known as the *identity* operation, which we denote by **e**.

The following diagram summarizes these actions.

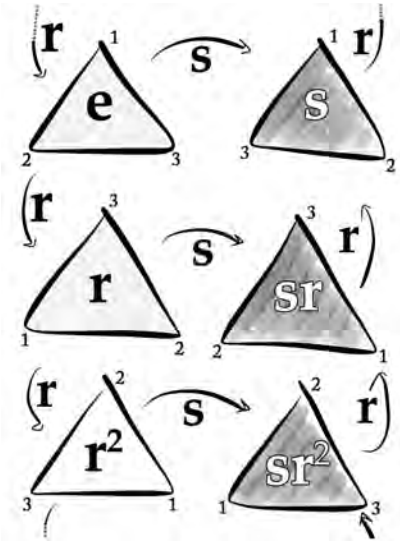


Figure 2: The triangle map

The above map makes sense only once we've agreed upon the identity operation, **e**. Let us declare the triangle you got in its original orientation to be the identity. It is the top-left triangle in the above diagram.

With this reference point, we can determine the composition of any sequence of **r**'s and **s**'s using Figure 2. To illustrate, suppose we want to determine the position of the vertices of $\mathbf{s} \circ \mathbf{r} \circ \mathbf{r} = \mathbf{srr}$. We compute the composition $\mathbf{s} \circ \mathbf{r} \circ \mathbf{r}$ starting from the right-most element (in this case, **r**). We start at the identity. We then follow the arrow **r** to the triangle in the middle left of Figure 1. Next, we follow the arrow **r** which lands in the bottom left at $\mathbf{r}^2$, and finally we follow the arrow **s** to the right, landing in the bottom right at $\mathbf{sr}^2$.

Notice that in the figure we formally distinguish between the group element (the letter inside the triangle) and the actions of **r** and **s** (the arrows between the triangles). However, each element is uniquely defined by its action on the identity element and so we interchange words like *element* and *action* or *operation* freely. This last fact is one of the reasons we must specify the element **e**. The element we choose to be the identity is simply the starting point from which to start following the arrows **r** and **s**.

Repeating this process for various combinations of **r** and **s**, we can make a table for the various relationships between the six triangles, thought of as "acting" on each other. That is, for every two elements in the list $G = \{\mathbf{e}, \mathbf{s}, \mathbf{r}, \mathbf{sr}, \mathbf{r}^2, \mathbf{sr}^2\}$, we can compose them to form a third element *in the same list*. Note that $\mathbf{r}^3 = \mathbf{r} \circ \mathbf{r} \circ \mathbf{r} = \mathbf{e}$ and $\mathbf{s}^2 = \mathbf{s} \circ \mathbf{s} = \mathbf{e}$.

$\nearrow$	**e**	**s**	**r**	**sr**	$\mathbf{r}^2$	$\mathbf{sr}^2$
e	**e**	**s**	**r**	**sr**	$\mathbf{r}^2$	$\mathbf{sr}^2$
s	**s**	**e**	**sr**	**r**	$\mathbf{sr}^2$	$\mathbf{r}^2$
r	**r**	$\mathbf{sr}^2$	$\mathbf{r}^2$	**s**	**e**	**sr**
sr	□	$\mathbf{r}^2$	□	**e**	**s**	**r**
$\mathbf{r}^2$	$\mathbf{r}^2$	**sr**	**e**	□	**r**	□
$\mathbf{sr}^2$	□	□	**s**	□	**sr**	□

Table 1: The composition $\mathbf{a} \circ \mathbf{b}$ is what would have happened had we applied the operation **b** *first*, and *then* **a**, where **a** is in G. For simplicity, we denote this composition by **ab**. [NOTE: THE ORDER IS IMPORTANT!]

The above table characterizes the group entirely and it is sometimes called the *Cayley table* of a group. Every element of the group appears exactly once in each row and column, like in a Sudoku puzzle. We left some of the entries blank in case you'd like to practice composing the operations of this group – a "grouppoku" if you will.

Of course, making these kind of tables is too tedious in practice. Nobody has time for that. Let's be minimalists. We only really used three relationships to create the table:

- $\mathbf{r}^3 = \mathbf{e}$.
- $\mathbf{s}^2 = \mathbf{e}$.
- $\mathbf{rs} = \mathbf{sr}^2 = \mathbf{sr}^{-1}$, or equivalently, $\mathbf{rsrs} = 1$.

For instance, if we want to know the composition $\mathbf{sr}^2 \circ \mathbf{r}^2 \circ \mathbf{s}$, we could simply compute:

$$\mathbf{sr}^2 \circ \mathbf{r}^2 \circ \mathbf{s} = \mathbf{s} \circ (\mathbf{r} \circ \mathbf{r} \circ \mathbf{r}) \circ \mathbf{r} \circ \mathbf{s} = \mathbf{s} \circ \mathbf{e} \circ \mathbf{r} \circ \mathbf{s}$$
$$= \mathbf{srs} = \mathbf{r}^{-1} = \mathbf{r}^2.$$

It is a fact – and a surprising one – that our group is the only one with six elements which satisfies these three relationships. We summarize this as

$$G = \langle r, s \mid r^3 = s^2 = rsrs = e \rangle.$$

This is known as the *presentation* of a group and r, s are its *generators*. The generators play the roles of the arrows **r** and **s** in Figure 2.[1]

You might ask, 'But, what about *rssrrrssrsrsrssr*?' It turns out that any concatenation of r's and s's can be simplified to one of the six elements of the group. Indeed,

$$\begin{aligned} rssrrrssrsrsrssr &= r(ss)(rrr)(ss)(rsrs)r(ss)r \\ &= rrr = e. \end{aligned}$$

The fact that we can place those parentheses anywhere and strike out every instance of ss, rrr or $rsrs$ is a consequence of *associativity*. It's the same reason $2 \times 3 \times 5 = 6 \times 5 = 2 \times 15$, but it is *not* the same reason $2 \times 3 = 3 \times 2$. That second property is called *commutativity*. Note that our group $G = \{\mathbf{e}, \mathbf{s}, \mathbf{r}, \mathbf{sr}, \mathbf{sr}^2\}$ is *not* commutative. We make this precise in the definition at the end of the section.

There is another, more general way of talking about the "triangle" group (a more technical name is "dihedral group of order six"). Stare at 2 a bit more, and ask yourself what essential information the picture contains.

All we really need to know is how **r** and **s** permute the ordered numbers [123].

For instance, **r** takes the number 1 to the number 2, 2 to 3 and 3 to 1, whereas **s** fixes 1 and swaps 2 and 3. We can write this down as

$$\mathbf{r} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \mathbf{s} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix},$$

with the precise meaning of the " = " sign yet to be explained.

Once again, it is important to note that each array of numbers defines an operation, and that we can *compose* these operations.

Say we wanted to apply **s** first, then **r** (so $\mathbf{r} \circ \mathbf{s} = \mathbf{rs}$ – note the order!). We should be able to write down a corresponding array of numbers:

$$\mathbf{rs} = \begin{pmatrix} 1 & 2 & 3 \\ \square & \square & \square \end{pmatrix}.$$

To figure out where **rs** takes 2, you can think: "*Two goes to three and three goes to one, so two goes to one.*" Repeat for each entry to obtain the corresponding array. Doing this for each triangle in our "map" (2),

$$\mathbf{e} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad \mathbf{s} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$
$$\mathbf{r} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad \mathbf{sr} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$
$$\mathbf{r}^2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad \mathbf{sr}^2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

Observe that every array in the right column has a "fixed point". Each of these arrays corresponds to flipping the triangle along the line passing through the fixed vertex and the midpoint of the opposing side. Refer to Figure 1 to see this. Probably the most important thing to note is that the *relative positions* of the numbers make the compositions work out. For instance, applying **r** simply cycles the list $1, 2, 3$, essentially shifting it right.

Before moving on, we pause to briefly reflect on what we've learned so far and to make some of the ideas more precise.

The first thing to understand is that each of the ways of displaying this "dihedral group of order six" has its distinct advantages and disadvantages. Having multiple ways of expressing the same idea is fundamental to mathematics, because some difficult problems become routine calculations when viewed the right way. What is common in all these perspectives is that we could compose operations, and that each of these operations is reversible. We use this to motivate our first official definition.

Definition 3 (Group)**.** *A finite group G is a set of elements $\{g_1, g_2, \ldots, g_n\}$ with a binary operation $\circ$ such that:*

- *There is an element $\boldsymbol{e}$ such that $\boldsymbol{e} \circ g = g \circ \boldsymbol{e} = g$ for any element g of the set G;*
- *Any element g has an inverse g^{-1} such that $g \circ g^{-1} = g^{-1} \circ g = \boldsymbol{e}$;*
- *The operation $\circ$ is* associative*: for any elements g_1, g_2, g_3 in G, $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$.*

[1] A common way of representing the relationships in Figure 2 is by means of a "Cayley graph." These graphs provide a wonderful avenue for exploring the interplay between groups and geometry.

SUBGROUPS AND QUOTIENTS

We introduce some **terminology and notation** for this section. The *order* of a (finite) group G is the number of elements it contains and is denoted by $|G|$. Let's assume from now on that $|G| = n$ (i.e., $G = \{g_1, \dots, g_n\}$). The statement 'H is a subgroup of G' will be denoted by $H < G$, and the statement $g \in G$ means 'g is an element of G'. We often adopt the convention of dropping the '$\circ$' when writing group elements, so $g \circ h_1 \circ h^{-1}$ is the exact same thing as gh_1h^{-1}.

Now, if $G = \{g_1, \dots, g_n\}$ is a group, a subgroup is presumably a subset[2] H of G such that H is itself a group.

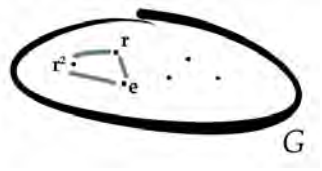
To make this clearer, let's go back to our original example with triangles.

$\nearrow$	**e**	**s**	**r**	**sr**	$\mathbf{r}^2$	$\mathbf{sr}^2$
e	**e**	**s**	**r**	**sr**	$\mathbf{r}^2$	$\mathbf{sr}^2$
s	**s**	**e**	**sr**	**r**	$\mathbf{sr}^2$	$\mathbf{r}^2$
r	**r**	$\mathbf{sr}^2$	$\mathbf{r}^2$	**s**	**e**	**sr**
sr	□	$\mathbf{r}^2$	□	**e**	**s**	**r**
$\mathbf{r}^2$	$\mathbf{r}^2$	**sr**	**e**	□	**r**	□
$\mathbf{sr}^2$	□	□	**s**	□	**sr**	□

Consider first the cells shaded black, including the **e** in the top left. The elements in black cells are **e** and **s**. Note that $\mathbf{se} = \mathbf{es} = \mathbf{e}$ and $\mathbf{ss} = \mathbf{e}$. The list $S = \{\mathbf{e}, \mathbf{s}\}$ is *closed under composition* and contains the inverse of every element in the list as well as the identity element **e**. The subset S is in fact a *subgroup* of G. The same holds for $R = \{\mathbf{e}, \mathbf{r}, \mathbf{r}^2\}$.

It's natural to ask how the structure of a group relates with its subgroups. This is a surprisingly deep question. Probably the most fundamental way of studying this relationship is by *quotienting*, and it comes up everywhere in mathematics. The rest of this section is dedicated to motivating and describing this 'quotient' structure.

To begin, let's recall that individual elements *act* on each other. We can generalise this to whole subgroups acting on an element. For instance, the left column of Figure 2 is the result of the action of the subgroup R on the identity **e**. It is the list of elements $R \cdot \mathbf{e} = \{\mathbf{e}, \mathbf{re}, \mathbf{r}^2\}$ obtained by composing each element of $r \in R$ with **e**.[3] This list, which we denote $R \cdot \mathbf{e}$, is known as the *orbit* of **e** under the action of R.

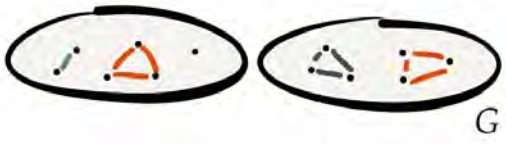


In the same way, we can act a subgroup H on the *whole group* G by acting H on eacth element g of G to obtain different orbits $H \cdot g$. If $H = \{h_1, h_2, \dots, h_k\}$, then $H \cdot g = \{h_1 g, h_2 g, \dots, h_k g\}$, and the action of H on G gives you a collection of orbits, $\{H \cdot g_1, H \cdot g_2, \dots, H \cdot g_n\}$, one for each of the n elements of G.

Ask yourself, (i) what can you say about the orbits? (ii) Can they intersect? (iii) How many distinct orbits are there? We will answer all these questions.

The first thing to notice that is that every element g will be in *some* orbit, because H is a (sub)group and therefore contains the identity element, **e**. Therefore, the orbit $H \cdot g$ contains the element $\mathbf{e}g = g$.

Now, consider the following diagram of orbits resulting from the action of some subgroup H on G:



Which is more likely: the situation on the left, or the one on the right?

It turns out that the picture on the left *can never happen*, and the situation on the right *must always happen.* In other words, the orbits of two different elements are either totally distinct or identical.

To make this more plausible, first consider that the orbit $H \cdot g$ has the exact same number of elements as H does (say k). To see this, suppose that $H = \{h_1, h_2, \dots, h_k\}$ has k *distinct* elements (so $h_i \neq h_j$ for any $i \neq j$), and recall that $H \cdot g = \{h_1 g, h_2 g, \dots, h_k g\}$. It's clear there are *at most* k elements. If there were less, then two elements of the list would be equal, i.e.

$$h_i g = h_j g.$$

But we are in a group, so we can multiply by g^{-1} on the right on both sides to obtain $h_i = h_i g g^{-1} = h_j g g^{-1} = h_j$, so $h_i = h_j$ and it must be that $i = j$.

Let's make this into an official proposition.

[2] By subset I mean that every element in H is also in G. This is written $H \subset G$.
[3] In this case, we are acting R from the left ($R \cdot \mathbf{e}$), so this is a *left* action.

Proposition 1. *The orbits $H \cdot g$ partition the group G into distinct sets of order k, where $k = |H|$.*

Proof. From the above discussion, each orbit $H \cdot g$ has $|H| = k$ elements. We wish to show that any two orbits are either completely distinct or identical.

Suppose that $H \cdot g_1$ and $H \cdot g_2$ are two orbits which share an element in common. We will show that $H \cdot g_1 = H \cdot g_2$. NOTE CAREFULLY THIS DOES NOT MEAN $hg_1 = hg_2$ for any h, but that if you give me some $h_1 g_1$, then I can find $h_2 \in H$ such that $h_2 g_2 = h_1 g_1$.

So, if $H \cdot g_1$ and $H \cdot g_2$ share an element in common, we have $h_1 g_1 = h_2 g_2$ for some $h_1, h_2 \in H$. Multiply by h_1^{-1} on the left on both sides to obtain

$$g_1 = h_1^{-1} h_2 g_2.$$

Note that $h_1^{-1} h_2$ is an element of H, since H is a subgroup.

Now, let $h_1' g_1$ be any other element of $H \cdot g_1$. Then setting $h_2' = h_1' h_1^{-1} h_2$, we get

$$h_1' g_1 = h_1'(h_1^{-1} h_2 g_2) = (h_1' h_1^{-1} h_2) g_2 = h_2' g_2,$$

so we have expressed $h_1' g_1$ as

$$(\textit{some element of } H) \cdot g_2.$$

Therefore, $h_2' g_2 \in H \cdot g_2$. □

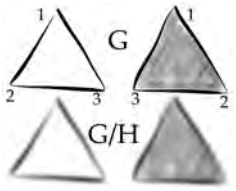
Corollary 2 (Lagrange)**.** *The order (the number of elements) $|H| = k$ of any subgroup divides the order $|G| = n$ of the group.*

Proof. The action partitions the group G into disjoint orbits of size k (orbits are represented by dotted lines and $k = 4$ in the diagram), and the union of these orbits makes up the whole group G. If m denotes the number of orbits, then $mk = n$, so k divides n.

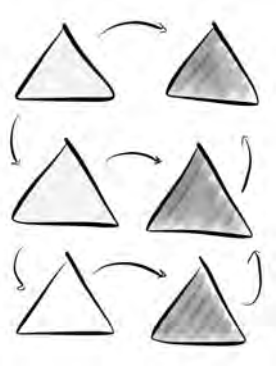
□

At this point we've answered two of the three questions we asked earlier: there are precisely $|G|/|H| = n/k$ distinct orbits, and they cannot intersect. However, we can say a lot more about these orbits by looking at the quotient structure they induce.

To motivate the idea of 'quotienting', let's consider a realistic example. Suppose we're playing, 'put the triangle in the box'. The rules of the game are that you are told to put the triangle in the box a certain way and then you do it – it's oddly satisfying. What if you are farsighted and don't have your glasses, so that you can't see the numbers on the triangle?



In other words, Figure 2 becomes



The "r" arrows do nothing – we are blind to rotations. All we see are the *orbits* of G under the action of R, the subgroup of rotations. This is a quotient structure.

However, we shouldn't forget the group G completely. We can still flip the triangle over – i.e., apply **s** – and if we flip twice then we recover the original triangle.

To make this precise, we have to define the composition of two orbits. That is, you need to know how to 'multiply' *orbits* as though they were elements in a group, like for instance $R \cdot \mathbf{r}$ and $R \cdot \mathbf{rs}$. Do it in the most natural way:

$$R \cdot \mathbf{r} \circ R \cdot \mathbf{rs} = R \cdot (\mathbf{r} \circ \mathbf{rs}) = R \cdot \mathbf{sr}.$$

This construction is called the *quotient* of G by R and is denoted G/R. In general, if you are given a subgroup $H = \{h_1, \ldots, h_k\}$, the composition of two elements $H \cdot g_1$ and $H \cdot g_2$ in G/H is defined by

$$(H \cdot g_1) \circ (H \cdot g_2) = H \cdot (g_1 g_2),$$

where $g_1 g_2$ is computed according to the group law of G.

So we have a respectable candidate for a group law on G/H, but we're not out of the woods yet. Could anything go wrong?

The problem is that we have different *representatives* for the same element: $\mathbf{e}$, $\mathbf{r}$ and $\mathbf{r}^2$ all fall into the same orbit under the action of R. More generally, if

$$H \cdot g_1 = H \cdot g_2, \quad H \cdot g_3 = H \cdot g_4,$$

(i.e., g_1, g_2 fall in one orbit, g_3, g_4 fall in another), how can you ensure that

$$(H \cdot g_1) \circ (H \cdot g_3) = (H \cdot g_2) \circ (H \cdot g_4)? \quad (1)$$

There is an elegant condition which specifies precisely when (1) holds. We require that H be a *normal* subgroup, and we relegate the explanation and proof to three exercises.

Problem 3. *Show that $H \cdot g = g \cdot H$ (i.e., $\{h_1 g, \dots, h_k g\} = \{g h_1, \dots, g h_k\}$) if and only if $ghg^{-1} \in G$ for any h in H. If this holds for some subgroup $H < G$, we say H is* normal *in G.*

Problem 4. *Show that the quotient G/H is well-defined and that G/H is a group if ghg^{-1} is in H for every h in H. To do this, let g_1, g_2, g_3, g_4 be as given above. You must show that every element in $H \cdot g_1 g_3$ is contained in $H \cdot g_2 g_4$, i.e. for any h_1 there is an h_2 such that $h_1 g_1 g_3 = h_2 g_2 g_4$.*

Problem 5. *Convince yourself that this assumption cannot be dropped by considering the subgroup $S = \{\boldsymbol{e}, \boldsymbol{s}\}$ of flips in the triangle group. In particular, $S \cdot \boldsymbol{sr} = S \cdot \boldsymbol{r}$ and $S \cdot \boldsymbol{sr}^2 = S \cdot \boldsymbol{r}^2$, but the composition does not agree. Generalize this to any group G and subgroup H.*

THE FIRST ISOMORPHISM THEOREM

In the previous section, we asked ourselves how a group related to its subgroups. We saw that by acting a subgroup $H < G$ on G, we partition G into disjoint sets of equal size called orbits. If (AND ONLY IF!) it turns out that $ghg^{-1} \in H$ for every $g \in G$, then the operation $H \cdot g_1 \circ H \cdot g_2$ is well-defined (independent of the choice of representatives g_1 and g_2) and we can talk of the quotient group G/H, which retains some – but not all – of the structure of G. G/H is, in a sense, a 'coarse' description of G.

The purpose of this section is to formulate *the exact same idea* in terms of mappings. In other words, we want to define a 'structure-preserving' function ϕ such that $\phi(g) = H \cdot g$. In order to go through with this construction, we have to define the notion of a *morphism*.

Definition 4. -morphism, *suffix. From the Greek μορφeta (morphe, 'form, shape') + -ism. 1. the state of having a specified form or shape.*

Okay, maybe that wasn't super helpful. Try this one.

Definition 5. *A homomorphism from a group G with operation $\circ_G$ to a group H with operation $\circ_H$ is a function (a 'map') $\phi : G \to H$ such that*

$$\phi(g_1 \circ_G g_2) = \phi(g_1) \circ_H \phi(g_2) \quad (2)$$

for every $g_1, g_2 \in G$. If e_G denotes the identity element of G, and e_H that of H, then we must have $\phi(e_G) = \phi(e_H)$.

Notice that we distinguish between the operation in G and the operation in H. For instance, when we took $G/H = \{H \cdot \mathbf{e}, H \cdot \mathbf{s}\}$ we *knew* what $H \cdot \mathbf{e} \circ H \cdot \mathbf{s}$ 'meant', but in reality we cheated: the operation $\circ$ doesn't even make sense for $H \cdot \mathbf{e}$! See the second example below if you are confused about this distinction.

Moreover, in this definition H need not be a subgroup of G. This definition is completely general, and it only refers to the underlying structure of the group. 2 is saying: it doesn't matter if we operate in G *first* and *then* send it to H, or if we send the elements to H first and then operate on them.

Example. (A silly one.) Consider ϕ to be the function which translates numbers from English to French:

$$\phi(\text{three} +_E \text{seven}) = \phi(\text{ten}) = \text{dix}$$
$$\phi(\text{three} +_E \text{seven}) = \text{trois} +_F \text{sept} = \text{dix}.$$

Hopefully, you'll believe that this works for any $g_1, g_2 \in$ integers. In some sense, ϕ is a homomorphism.

Example. (A non-trivial one.) First, convince yourself that $\mathbb{R}$, the real numbers, is a group under the operation $+$:

- For any $x, y, z \in \mathbb{R}$, we have $x + y \in \mathbb{R}$ and $(x + y) + z = x + (y + z)$.
- the inverse of x is $-x$,
- the group identity e is 0.

Compare with Definition 3. Take the time to internalize the fact that $+$ is *nothing more than a symbol*. We could've said $3 \circ 6 = 9$ and $0 \circ x = x$. The symbol we use to denote the operation has nothing to do with the structure.

Remark: This is also our first example of an *infinite* group.

Perhaps surprisingly, there is a homomorphism

$$\phi : \mathbb{R} \to \mathbb{R}_{x>0},, \qquad (\phi \text{ sends } \mathbb{R} \text{ to } \mathbb{R}_{x>0})$$

where $\mathbb{R}_{x>0}$ is the group of *strictly positive* real numbers with the operation being multiplication $\times$ (another stupid symbol). [Convince yourself that $\mathbb{R}_{x>0}$ is a group: what is the identity? what is the inverse of some $x \in \mathbb{R}_{x>0}$? Is the operation associative?]

The homomorphism ϕ which relates $\mathbb{R}$ to $\mathbb{R}_{x>0}$ is the most important function in mathematics, $\phi(x) = e^x$:

$$\phi(x+y) = e^{x+y} = e^x \times e^y = \phi(x) \times \phi(y)$$
$$\phi(0) = 1.$$

So in other words, it doesn't matter if we add and then exponentiate, or exponentiate and then add. Add and translate to French, or translate into French and then add. Notice also that $\phi(0) = 1$, which is the identity of multiplication.

There's also something special about this example: the function e^x also has an *inverse* defined on the positive real numbers: the logarithm, $\phi^{-1}(y) = \log(y)$. One can check that it satisfies the properties of a homomorphism.

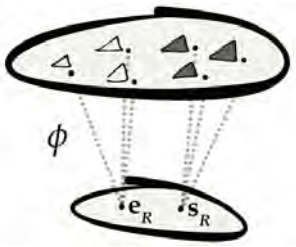
This example provides an example of an *invertible* homomorphism – an *isomorphism*. Going back to the etymological definition, we see that the words make a lot of sense, even though they might sound intimidating at first: homomorphism – similar structure; iso-morphism – identical structure[4].

The notion of isomorphism abstracts the idea that there may be many representations of the same structure. There's no difference between permutations of the numbers [123] and rigid motions of a triangle, $\langle r, s : r^3 = s^2 = rsrs \rangle$; the two groups are isomorphic. We will use this example to demonstrate the so-called 'First isomorphism theorem.'

Our first step is to construct a homomorphism $\phi : G \to G/R$. Let's call the identity of G/R $\mathbf{e}_R$, and let's denote the only other element $\mathbf{s}_R$. The map ϕ should take anything in the orbit $R \cdot \mathbf{e}$ to $\mathbf{e}_R$ and anything in the orbit $R \cdot \mathbf{s}$ to $\mathbf{s}_R$. Thus,

$$\phi(g) = \begin{cases} \mathbf{e}_R & \text{if } g = \mathbf{e}, \mathbf{r} \text{ or } \mathbf{r}^2 \\ \mathbf{s}_R & \text{if } g = \mathbf{s}, \mathbf{sr} \text{ or } \mathbf{sr}^2 \end{cases}$$

In other words, ϕ is not wearing its glasses. It treats every element in the same orbit equally.



Naturally, we can ask about the inverse of a homomorphism ϕ, but in this case there are multiple elements in G mapping to the same elements in G/R. There is a distinguished name for the pre-image of the identity: it is called the *kernel*, and is denoted $\ker\phi$. We have the following happy fact.

Proposition 6. *The kernel* $\ker\phi$ *of a group homomorphism* $\phi : G \to H$ *is a subgroup of* G.

Proof. It suffices to show that $\ker\phi$ is *closed under composition* and contains its inverses.[5] If $g_1, g_2 \in \ker\phi$, then

$$\begin{aligned} \phi(g_1 \circ_G g_2) &= \phi(g_1) \circ_H \phi(g_2) \\ &\quad \text{(since } \phi \text{ is a homomorphism)} \\ &= \mathbf{e}_H \circ_H \mathbf{e}_H \\ &\quad \text{(since both } g_1, g_2 \in \ker\phi) \\ &= \mathbf{e}_H \\ &\quad \text{(since } \mathbf{e}_H \text{ is the identity in } H) \end{aligned}$$

The fact that $\ker\phi$ contains its inverses follows from the fact that $\phi(g^{-1}) = (\phi(g))^{-1}$ (why, and why is the fact true?). □

Proposition 7. *The kernel* $\ker\phi$ *of a group homomorphism* $\phi : G \to H$ *is* normal.

Proof. It is required to show that $ghg^{-1} \in \ker\phi$ for every $h \in \ker\phi$. From the defining property of a homomorphism,

$$\begin{aligned} \phi(g \circ h \circ g^{-1}) &= \phi(g) \circ \phi(h) \circ \phi(g^{-1}) \\ &= \phi(g) \circ \mathbf{e}_H \circ (\phi(g))^{-1} \\ &= \phi(g) \circ (\phi(g))^{-1} = \mathbf{e}_H. \end{aligned}$$

In the above, we avoided specifying where the operation $\circ$ is taking place, but you should figure it out if it's not clear. □

[4] Actually, this might be a bastardisation of Greek. The prefixes homo- and iso- seem to mean the same thing.

[5] Why does associativity follow, and why don't we require that $\mathbf{e} \in \ker\phi$?

So we have a subgroup, $\ker\phi < G$. We know that acting subgroups on the group gives us information about structure; let's do that. What can you say about the orbits $\ker\phi \cdot g$? Any $k \in \ker\phi$ satisfies $\phi(k) = \mathbf{e}_H$, so for any $kg \in \ker\phi \cdot g$,

$$\phi(k \circ g) = \phi(k) \circ \phi(g) = \mathbf{e}_H \circ \phi(g) = \phi(g).$$

In other words,

every element in the orbit of $\ker\phi \cdot g$ maps to the same place as g under ϕ!

Moreover, since $\ker\phi$ is a normal subgroup, $G/\ker\phi$ is itself a group with the operation defined by

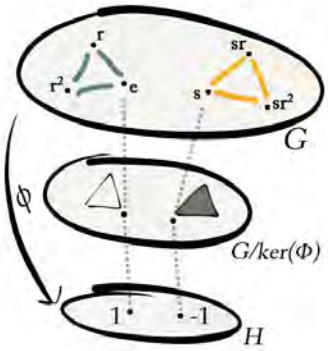
$$\ker\phi \cdot g_1 \circ \ker\phi \cdot g_2 = \ker\phi \cdot (g_1 \circ g_2),$$

as before. Thus we have the following two observations:

- ϕ sends *each* element in $\ker\phi \cdot g$ to the same h in H, and
- Each $\ker\phi \cdot g$ is a *single* element in the quotient $G/\ker\phi$.

These observations require one to hold many definitions in mind at the same time which can be difficult and confusing, especially if this is one's first encounter with an isomorphism theorem. The essence of these two facts is that the elements of $G/\ker\phi$ and H are in one-to-one correspondence. From this we might guess that the groups H and $G/\ker\phi$ are isomorphic. The first isomorphism theorem says that this is indeed the case.

An alternative perspective is that quotienting out by $\ker\phi$ strips away all the information which just gets killed[6] by the homomorphism ϕ, analogous to the way in which quotienting by R strips away the information contained by rotations.

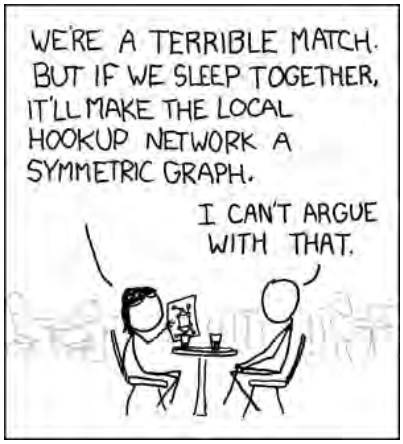


This theorem is so remarkable, so wonderful, that we will end our discussion with it.

Theorem 8 (First isomorphism theorem)**.** *Suppose G and H are groups and $\phi : G \to H$ a homomorphism. Then* $\ker\phi < G$ *(is a subgroup),* $\ker\phi \triangleleft G$ *(is a* normal *subgroup), and* $G/\ker\phi \cong H$ *(is* isomorphic *to H).*

Moreover, given any normal subgroup $K \triangleleft G$, there is a homomorphism $\pi : G \to G/K$. This is called the natural projection *of G onto G/K.*

JOKES AND COMICS



xkcd 404: *Convincing Pickup Line*

Check it out; I've had sex with someone who's had sex with someone who's written a paper with Paul Erdös!.

[6] Technical term for 'sent to the identity' – no joke.

VIEWING WEIGHTED GRAPHS AS A DIFFERENTIABLE MANIFOLD: A DISCRETE ANALOG OF COURANT'S THEOREM

Thomas Ng

Eigenvalues and eigenvectors of operators have been studied extensively due to their mathematical elegance and direct applications in many other disciplines such as general relativity. This has led to intense study of differential operators in the form of Spectral Geometry. Matrices, however, can be seen as operators on finite dimensional space, and surface in various seemingly unrelated fields such as Graph Theory. We provide a brief introduction to Spectral Graph Theory and illustrate the relationship with Spectral Theory of Manifolds including a reformulation of the proof of a discrete version of Courant's theorem on the interlacing of nodal sets of eigenfunctions by Chapon.

1 INTRODUCTION

Consider a finite simple graph $G = (V, E)$ where $V = \{1 \dots n\}$ is the vertex set and $E \subset V \times V$ is the edge set. We say that $i \sim j$ if $(i, j) \in E$. G is equipped with edge weight function $w_E : E \to \mathbb{R}_+$. [1] This naturally induces a notion of edge length $\ell_E : E \to \mathbb{R}_+$ given by $\ell_E(i, j) := \frac{1}{w_E(i,j)}$. These functions may then be extended to ones on $V \times V$, namely:

$$\ell : V \times V \to \mathbb{R}_+ \cup \{\infty\}$$
$$w : V \times V \to \mathbb{R}_+ \cup \{0\},$$

given by

$$\ell(i, j) = \begin{cases} \ell_E(i, j) & \text{if } (i, j) \in E \\ \infty & \text{otherwise} \end{cases}$$
$$w(i, j) = \begin{cases} w_E(i, j) & \text{if } (i, j) \in E \\ 0 & \text{otherwise.} \end{cases}$$

This lets us define the *adjacency matrix* as

$$A = (w(i, j))_{1 \le i,j \le n}$$

and the *degree matrix*

$$D = \operatorname{diag}\left(\sum_{j \sim 1} w(1, j) \dots \sum_{j \sim n} w(n, j)\right).$$

We will be concerned with differential operators on functions $u : V \to \mathbb{R}$.
By finiteness of the vertex set, the space of all such functions is finite dimensional and is usually expressed as $\mathcal{L}^2(V)$ (square integrable functions).
We then define the *Discrete Laplacian* to be $\Delta : \mathcal{L}^2(V) \to \mathcal{L}^2(V)$ given by

$$\Delta = D - A. \tag{1}$$

[1] In this case we say (G, w) is a weighted graph.

Note that according to our construction, the constant vector is both a left and right eigenvector of the Laplacian with eigenvalue zero because the terms in any row or column sum to 0. This may be remedied by adding a so called scalar potential, which we will discuss later.

This definition of the discrete Laplacian may appear unnatural, however it is deeply rooted in the definition of the Laplacian seen in differential geometry: $\Delta(f) = -\operatorname{div}(\nabla f)$.

Recall that for a manifold M, the gradient (∇) is a map from square integrable functions on the manifold ($\mathcal{L}^2(M)$) to a vector field tangent to the surface of M (called the tangent bundle, often denoted TM) by assigning to each point p a vector of partial derivatives.

$$\nabla_p : (f, p) \mapsto \left(\frac{\partial f}{\partial x_1} \dots \frac{\partial f}{\partial x_n}\right)\Big|_p,$$

where the x_i's are basis vectors for T_pM, the tangent space at p. The following is a derivation of eq. 1 that illustrates analogs of derivatives of functions on weighted graphs.

On a graph, the tangent space about a vertex is the disjoint union of edges incident to it, so we can define the gradient along one of these edges to be

$$\nabla u(i, j) = \frac{u(j) - u(i)}{\ell(i, j)}.$$

Similarly, the divergence of a vector field is given by the integral of the vector field emanating from a point. We define the divergence on a graph to be the sum of 'flow' emanating from a vertex, that is,

$$\operatorname{div} f(i) = \sum_{j \sim i} f(i, j).$$

It is an easy computation to verify that composing these two definitions gives rise to the formula above.

2 Continuous Courant's Theorem

In this section, we provide the necessary definitions and present the nodal domain theorem of Courant proved in his 1953 paper.
Say $f : \Omega \to \mathbb{R}$ where Ω is some domain.

Definition 6. *The* Nodal Set *($\mathcal{N}(f)$) of a function f is the set of points where f vanishes that is:*

$$\mathcal{N}(f) = \{x \in \Omega : f(x) = 0\}$$

Definition 7. *A* Nodal Domain *of f is a connected component of $\Omega \backslash \mathcal{N}(f)$.*

Theorem 1 (Courant, 1953)**.** *If $S : \mathcal{L}^2(\Omega) \to \mathcal{L}^2(\Omega)$ is a self-adjoint second order elliptic operator with arbitrary boundary conditions [2] with eigenvalues:*

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n \leq \ldots$$

and corresponding eigenfunctions

$$f_1, f_2 \ldots f_n, \ldots$$

then f_n has at most n nodal domains.

Definition 8. *A differential operator S is called a* Schrödinger Operator*, or generalized Laplacian, if it is given by $S = -\Delta + P$ where $P : \mathcal{L}^2(\Omega) \to \mathbb{R}$ is a scalar valued function.*

Corollary 2 (Interlacing Property)**.** *If S is a Schrödinger Operator defined on a one-dimensional domain, then the nodal sets of f_n and f_{n+1} interlace.*

Definition 9. *We say that two sets of points X, Y* interlace *if between any two points of X, there is a unique point in Y.*

This property is of particular interest because it is a very strong condition on the location of nodal domains of successive eigenfunctions, which helps characterize the spectrum and geometry of manifolds.

3 A Graph Theoretic Analog

We can extend our concept of Discrete Laplacian to give us a Discrete Schrödinger Operator by adding a potential function P to the Laplacian, where P is a function from the vertices to the real numbers. Explicitly, $P(i) \in \mathbb{R} \ \ \forall\, i \in V$.

A priori, there is no way to draw a parallel to Courant's Theorem for graphs because we have not built up a notion of eigenfunctions on graphs. However, because we are working on a finite space there is a natural isomorphism from eigenvectors to vertex-valued functions given by

$$u = (u_1 \ldots u_n) \quad \longleftrightarrow \quad f_u : i \mapsto u_i.$$

Because of this, we may treat eigenvectors as eigenfunctions and will use the notation $u(i)$ to denote the function associated with u evaluated at i. This also allows us to index vectors later.

These functions, however, still fail to provide a well-defined notion of nodal set as the eigenvectors may not vanish on any vertex at all. Instead, it is actually more advantageous to extend the notion of nodal domains to graphs.

Definition 10. *A* strong sign graph *of G with respect to $u : V \to \mathbb{R}$ is a maximal subtree $T \subset G$ such that $u|_T$ has constant sign.*

This definition is presented in a paper by Davies *et al* in 2001 where they proved the following theorem.

Theorem 3 (Davies *et al.*, 2001)**.** *Given $\lambda_1 \leq \cdots \leq \lambda_n$, the eigenvalues of a generalized Laplacian, if λ_k has multiplicity r, then any corresponding eigenvector has at most $k + r - 1$ strong sign graphs.*

Corollary 4. *If all the eigenvalues of a generalized Laplacian are simple (multiplicity one), then u_k, the eigenvector associated with λ_k, has at most k strong sign graphs.*

The eigenvectors are discontinuous by the nature of V being discrete, however, by linearly interpolating each eigenfunction u we obtain a continuous extension $\tilde{u} : G \to \mathbb{R}$ given by the formula:

$$\tilde{u}(t) := \nabla(i,j)t + u(i).$$

Given this, we can properly define the nodal set of u to be the zeros of $\tilde{u}$. After parametrizing each edge by $t \in (0,1)$, the zeros can be found explicitly to be:[3]

$$t^\star = -\frac{u(i)}{\nabla u(i,j)}$$
$$\Rightarrow \tilde{u}(t^\star) = \nabla u(i,j)\left(-\frac{u(i)}{\nabla u(i,j)}\right) + u(i) = 0$$

[2] These terms may seem rather daunting, but they are not crucial to understanding this theorem. It is enough to note that self-adjoint ensures real eigenvalues, and elliptic gives positivity of the first eigenvalue.

[3] This notion of a parametrization of the edges as a one dimensional manifold is used in the study of metric graphs and will be crucial to completing our proof of the discretized Courant Nodal Domain Theorem.

Note that the strong sign graphs of the extended function $\tilde{u}$ becomes the strong sign graph union some partial edges, namely subintervals $[0, t^\star]$, of where edges are parametrized from $(0, 1)$ as above.

The only missing piece now is a graph theoretic "one-dimensional" domain. The most important property of single dimensional spaces is that there is a unique shortest path/geodesic between any two points. So we expect to restrict our search to trees.

4 Discrete Green's Theorem

Recall that Green's Theorem from multivariate calculus is as follows

$$\iint_\Omega \left(\frac{\partial P}{\partial x} - \frac{\partial Q}{\partial y}\right) \mathrm{d}x\,\mathrm{d}y = \oint_{\partial\Omega} Q\,\mathrm{d}x + P\,\mathrm{d}y$$

and the corresponding version on manifolds:

$$\int_M (u\Delta v - v\Delta u)\ \mathrm{d}V = \int_{\partial M} (u\,\mathrm{d}\vec{n}(v) - v\,\mathrm{d}\vec{n}(u))\ \mathrm{d}\tilde{V}$$

In order to discuss Green's Theorem, we require a notions of boundary[4] for some nontrivial connected subgraph S with vertex set $V(S)$. The usual boundary is the following:

$$\partial(S) = \{(i,j) \in E : i \in V(S) \text{ and } j \in V \backslash V(S)\}.$$

In a sense we assign an orientation to each edge in the boundary pointing outward from the subgraph. This mimics an outward normal on domains.

Theorem 5 (Green's (Discrete))**.** *For $u, v \in \mathcal{L}^2(V)$ with T, a strong sign graph of u, we have:*

$$\sum_{i\in V(T)} [u(i)\Delta v(i) - \Delta u(i)v(i)] = \sum_{(i,j)\in\partial(T)} [\nabla u(i,j)v(i) - u(i)\nabla v(i,j)]\,. \tag{2}$$

Proof. The proof is a straightforward calculation. Expanding the LHS and regrouping, we get

$$\begin{aligned}
&\sum_{i\in V(T)} [u(i)\Delta v(i) - \Delta u(i)v(i)] \\
&\quad= \sum_{i\in V(T)} [u(i)[-\mathrm{div}\nabla v(i)] - [-\mathrm{div}\nabla u(i)]v(i)] \\
&\quad= \sum_{i\in V(T)} [\mathrm{div}\nabla u(i)v(i) - u(i)\mathrm{div}\nabla v(i)] \\
&\quad= \sum_{i\in V(T)} \left[\sum_{j\sim i} \nabla u(i,j)v(i) - u(i)\sum_{j\sim i}\nabla v(i,j)\right] \\
&\quad= \sum_{(i,j)\in E} [\nabla u(i,j)v(i) - u(i)\nabla v(i,j)]
\end{aligned}$$

Observe that for each $(i,j) \in E\backslash\partial(T)$, the corresponding summand occurs along with it's negative. Hence,

$$= \sum_{(i,j)\in\partial(T)} [\nabla u(i,j)v(i) - u(i)\nabla v(i,j)]\,.$$

□

Notice that by adding and subtracting a potential function from the right side of eq. 2, we obtain the equivalent statement for generalized Laplacians.

5 Interlacing Property on Trees

We now have all the tools we need to complete a proof of Courant's Nodal Domain Theorem for trees.

Theorem 6 (Chapon, 2009)**.** *Given (G, w) a weighted tree with associated Schrödinger Operator S, if $\lambda_1 \leq \cdots \leq \lambda_n$ are the eigenvalues of S and the associated eigenvectors are $u_1 \ldots u_n$, then u_k has exactly k nodal domains. Moreover, the nodal sets of u_n and u_{n+1} interlace.*

Proof. First, the Perron-Frobenius theorem states that λ_1 is simple and it has a corresponding eigenvector everywhere positive. This means that u_1 has one strong sign graph. Now, by Corollary 4, it suffices to show that u_{k+1} has at least one more nodal domain than u_k.

Say $\lambda < \mu$ are eigenvalues of S with associated eigenvectors u, v respectively. Let T be a strong sign graph of u, and assume for the sake of contradiction that v is of constant sign on T, that is, $\mathcal{N}(v) \cap T = \emptyset$. Without loss of generality, assume that[5] $u|_T, v|_T > 0$ and consider the

[4] We can alternatively treat a graph like a one-dimensional metric space and allow the boundary of a strong sign graph to be the points. $\mathscr{B}(T(f)) = \{t \in [0,1) : f(t) = 0\}$. This is the approach used in Chapon's proof of the Discrete Courant's theorem.

[5] The other cases fall out by the same argument possibly reversing a negative sign.

Discrete Green's Theorem on the subgraph T.

$$\sum_{i\in V(T)}(u(i)\Delta v(i)-\Delta u(i)v(i)) = \sum_{(i,j)\in\partial(T)}(\nabla u(i,j)v(i)-u(i)\nabla v(i,j))$$

and

$$(\mu-\lambda)\sum_{i\in V(T)}u(i)v(i) = \sum_{(i,j)\in\partial(T)}\nabla u(i,j)\left(v(i)-\nabla v(i,j)\frac{u(i)}{\nabla u(i,j)}\right). \tag{3}$$

The left hand side of eq. 3 is positive as both $(\mu-\lambda)$ and the summation are. However,

$$\sum_{(i,j)\in\partial(T)}\nabla u(i,j)<0$$

since $u|_T > 0$ and T is a strong sign graph. Notice also that $(v(i)-\nabla v(i,j)u(i)/\nabla u(i,j)) = \tilde{v}(t_u^\star)$ where $t_u^\star\in[0,1)$ is the zero of $\tilde{u}$ along the boundary edge (i,j). Now, by assumption

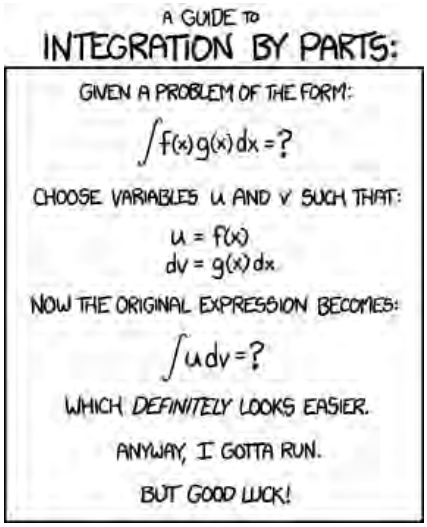
$$\left(v(i)-\nabla v(i,j)\frac{u(i)}{\nabla u(i,j)}\right)=\tilde{v}(t_u^\star)\geq 0$$

since $\mathcal{N}(v)\cap T=\emptyset$ and v is of constant sign on T, so we have reached a contradiction.

Thus, u_{k+1} changes sign on each strong sign graph of u_k, that is $\tilde{u}_{k+1}$ vanishes on each strong sign graph of u_k. Therefore, u_{k+1} has at least one more strong sign graph than u_k and by induction, u_k has exactly k strong sign graphs. Moreover, the above argument demonstrates that the Nodals Sets of consecutive eigenvectors interlace. □

JOKES AND COMICS

A physicist and a mathematician are sitting in a faculty lounge. Suddenly, the coffee machine catches on fire. The physicist grabs a bucket and leap towards the sink, filled the bucket with water and puts out the fire. Second day, the same two sit in the same lounge. Again, the coffee machine catches on fire. This time, the mathematician stands up, got a bucket, hands the bucket to the physicist, thus reducing the problem to a previously solved one.



xkcd 1202: *Integration by parts*
If you can manage to choose u and v such that $u=v=x$, then the answer is just $(\frac{1}{2})x^2$, which is easy to remember. Oh, and add a '+C' or you'll get yelled at...

CREDITS

in alphabetical order

The δelta-εpsilon Editing Team

- Léo Belzile
- Michael Crawford
- Alexander Dieplam
- Guilherme Mansur
- Yi Tian Xu
- Meng Zhao

Reviewing Board

- Kael Dixon
- Bundit Laekhanukit
- Leta Montopoli
- Jason Pollack
- Benjamin Smith
- Paul Tang
- Alexandra Tcheng

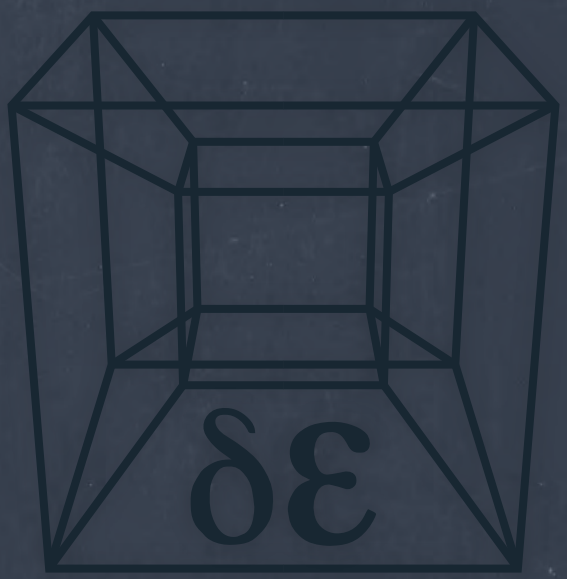
Cover Art and Design

- Joey Litalien

We would like to give credit to the following websites for the comics, the math jokes and the cover art inspiration in this year's issue:
`http://www.xkcd.com`
`geometrydaily.tumblr.com` (Tilman)
`http://www.ams.org/notices/200501/fea-dundes.pdf`

ACKNOWLEDGEMENTS

Special thanks to Dr. Linan Chen and Pr. Mikaël Pichot for accepting to be interviewed about their work and research. We would also like to thank everyone on our reviewing board for their contributions in ensuring the quality of this publication. In addition, thank you to the authors of the articles which appeared in this edition. Thanks to SUMS, the Department of Mathematics and Statistics, the Arts Undergraduate Society for their generous financial support.


δε