Smooth Manifold, Differential Form, and De Rham Cohomology

Fengcheng Lin
University of California Santa Barbara

Smooth Manifold

A topological n-manifold M is a topological space that is locally Euclidean of dimension n and satisfies some other nice properties. This means there are **charts** $\{(U_i, \varphi_i) : i \in I\}$ such that $\{U_i\}$ is an open cover for M and $\varphi_i : U_i \to \hat{U}_i$ is a homeomorphism from U_i to an open subset $\hat{U}_i \subseteq \mathbb{R}^n$. To see how M is locally Euclidean, notice that for any $p \in M$, there exists an open set U_i that contains p s.t. φ_i maps U_i homeomorphically onto a subset of $\mathbb{R}^n$.

For a function $f : M \to \mathbb{R}$ and a chart (U, φ) of M s.t. $p \in U$, we define the **coordinate representation** of the function at p by $f \circ \varphi^{-1} : \hat{U} \to \mathbb{R}$. Since $f \circ \varphi^{-1}$ is a function from a subset of $\mathbb{R}^n$ to $\mathbb{R}$, we can perform ordinary calculus on $f \circ \varphi^{-1}$. However, to the coordinate representation $f \circ \varphi^{-1}$, it turns out that for any topological n-manifold M with charts $\{(U_i, \varphi_i) : i \in I\}$, the function $\varphi_i \circ \varphi_j^{-1}$ must be smooth (infinitely differentiable) for any $i, j \in I$. Topological manifolds that satisfy this smoothness condition are called **smooth manifolds**. For a smooth manifold M, we say that f is **smooth** at $p \in M$ if $f \circ \varphi^{-1}$ is smooth in ordinary calculus. We denote the vector space of all smooth real-valued functions on M by $C^\infty(M)$.

Alternating Tensor

Let V be a finite-dimensional real vector space. A k**-tensor** on V is a real-valued multilinear function of k elements of V:

$$\alpha : \underbrace{V \times ... \times V}_{k \text{ copies}} \to \mathbb{R}. \tag{1}$$

For a k-tensor α, we say that α is **alternating** if its value changes sign whenever two of its inputs are interchanged. That is, for any i, j,

$$\alpha(v_1, ..., v_i, ..., v_j, ..., v_k) = -\alpha(v_1, ..., v_j, ..., v_i, ..., v_k). \tag{2}$$

We denote the vector space of all alternating k-tensor on V by $\Lambda^k(V^*)$. When $k = 1$, $\Lambda^1(V^*)$ is just the vector space of linear functionals on V. When $k = 0$, $\Lambda^0(V^*)$ is the vector space of all constant functions, or just $\mathbb{R}$.
An important property of alternating tensor α is that α gives the value zero whenever two of its inputs are equal. This is because if $v_i = v_j$, then

$$\alpha(v_1, ..., v_i, ..., v_j, ..., v_k) = -\alpha(v_1, ..., v_i, ..., v_j, ..., v_k), \tag{3}$$

which implies $\alpha(v_1, ..., v_i, ..., v_j, ..., v_k) = 0$. Using this property, we can derive that α gives the value zero whenever its inputs are linear dependent.
For any k-tensor α, we can construct an alternating k-tensor $\mathrm{Alt}\,\alpha$ by

$$\mathrm{Alt}\,\alpha(v_1, ..., v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} (\mathrm{sgn}\,\sigma) \alpha(v_{\sigma(1)}, ..., v_{\sigma(k)}). \tag{4}$$

where σ is a permutation of n and $\mathrm{sgn}(\sigma)$ is the sign of the permutation.

An important example of alternating tensor is the determinant in linear algebra. For the space of $n \times n$ matrix, the determinant can be regarded as an alternating n-tensor on the column vector space. That is, we can write $\det(A) = \det(v_1, ..., v_n)$ where v_i is the i-th column vector of the matrix A. This is because the definition of $\det(A)$

$$\det(A) = \sum_{\sigma \in S_n} (\mathrm{sgn}\,\sigma) \prod_{k=1}^{n} a_{k\sigma(k)}, \tag{5}$$

changes sign whenever we interchange its column vectors. Since $\det(A)$ is multilinear on its column vectors, $\det$ is an alternating tensor on the column vector space.
This explains why the determinant can be used as a test for the linear independence of matrix: if the column vectors of a matrix are linear dependent, then we must have $\det(A) = \det(v_1, ..., v_n) = 0$ since $\det$ is an alternating tensor.

Vector Space of Alternating Tensors

For an n-dimensional vector space V and a basis $(\varepsilon^1, ..., \varepsilon^n)$ for the dual space V^*, we define ε^I for each multi-index $I = (i_1, ..., i_k)$ of length k with $i_j = 1, ..., n$ by

$$\varepsilon^I(v_1, ..., v_k) = \det \begin{bmatrix} \varepsilon^{i_1}(v_1) ... & \varepsilon^{i_1}(v_k) \\ ... & ... \\ \varepsilon^{i_k}(v_1) ... & \varepsilon^{i_k}(v_k) \end{bmatrix}. \tag{6}$$

Since $\det$ is an alternating tensor, ε^I is also an alternating tensor, called the **elementary alternating tensor**, and it turns out that $\Lambda^k(V^*)$ is finite dimensional with the basis

$$\mathcal{E} = \{\varepsilon^I : I \text{ is an increasing multi-index of length } k\} \tag{7}$$

Because to construct I is to choose an increasing sequence of length k from n elements, $\dim \Lambda^k(V^*) = \binom{n}{k}$.

Differential Form

For a smooth n-manifold M, we can attach a vector space to every point $p \in M$. Such a vector space is called the **tangent space** at p, denoted by T_pM. Intuitively, for $f \in C^\infty(M)$, the tangent space at p represents the space of all directional derivatives of f at p, i.e., every $v \in T_pM$ is a linear map $v : C^\infty(M) \to \mathbb{R}$. An important fact is that T_pM is homeomorphic to $\mathbb{R}^n$ for all $p \in M$.

Now, we can define what differential form is. A **differential k-form** (or just k-**form**) ω is defined to be an alternating tensor field on M. The value of ω at each point $p \in M$ is an alternating tensor $\omega(p) = \omega_p \in \Lambda^k(T_pM^*)$. For any $p \in M$ and a chart (U, φ) that contains p, because $(x^1, ..., x^n)$ is a basis for $(\mathbb{R}^n)^*$, $\mathcal{E} = \{x^I : I \text{ is an increasing multi-index}\}$ is a common basis for all $\Lambda^k(T_qM^*)$ where $q \in U$. As a result, ω could be locally written as

$$\omega = \sum_I \omega_I x^I, \tag{8}$$

where ω_I is the function of the coefficient of each basis vector. We say ω is **smooth** at p if ω_I is smooth for every I, and the vector space of all smooth k-forms is denoted by $\Omega^k(M)$. Notice that a smooth 0-form is just a smooth function, which implies $\Omega^0(M) = C^\infty(M)$.

For every smooth map $F : M \to N$, there is a **pullback** $F^* : \Omega^k(N) \to \Omega^k(M)$ for every k.

Exterior Derivative

For $f \in C^\infty(M)$, we define the **differential** of f to be

$$df_p(v) = v(f). \tag{9}$$

Therefore, the value of df at $p \in M$ is a linear functional df_p on T_pM, and df is hence a 1-form (actually df is a smooth 1-form)

The **exterior differentiation** is a unique linear operator d that sends a smooth k-form to a smooth $k+1$ form for all k, $d : \Omega^k(M) \to \Omega^{k+1}(M)$. When $k = 0$, the exterior differentiation $d(f)$ coincides with the differential df. The exterior differentiation can be locally written as

$$d(\sum_J \omega_J dx^J) = \sum_J d\omega_J \wedge dx^J. \tag{10}$$

where the **wedge product** $\wedge$ is an operator that combines a smooth m-form and a smooth n-form to produce a smooth $(m+n)$-form, explicitly defined as $(\omega \wedge \eta)_p = \frac{(m+n)!}{m!n!} \mathrm{Alt}\,(\omega_p \otimes \eta_p)$.

We say that a smooth differential form $\omega \in \Omega^k(M)$ is **closed** if $d\omega = 0$, and **exact** if there exists a smooth $(k-1)$-form η on M s.t. $\omega = d\eta$.

De Rham Cohomology

Let M be a smooth manifold. As the exterior derivative $d : \Omega^k(M) \to \Omega^{k+1}(M)$ is linear, its image and kernal are linear subspaces. We define

$$\begin{aligned} \mathcal{Z}^k(M) &= \mathrm{Ker}(d : \Omega^k(M) \to \Omega^{k+1}(M)) = \{\text{all closed } k\text{-forms}\} \\ \mathcal{B}^k(M) &= \mathrm{Im}(d : \Omega^{k-1}(M) \to \Omega^k(M)) = \{\text{all exact } k\text{-forms}\} \end{aligned} \tag{11}$$

Then the k**-th de Rham cohomology group** of M is the quotient vector space

$$H^k_{dR}(M) = \mathcal{Z}^k(M)/\mathcal{B}^k(M) \tag{12}$$

One notable property of de Rham cohomology group is its homotopy invariance. A direct application of the homotopy invariance tells that if M is a contractible smooth manifold, then $H^k_{dR}(M) = 0$ for every k.
For any smooth map $F : M \to N$, the pullback $F^* : \Omega^k(N) \to \Omega^k(M)$ induces a linear map called the **induced cohomology map** from $H^k_{dR}(N)$ to $H^k_{dR}(M)$, also denoted by F^*. This assignment defines a contravariant functor from the category of smooth manifolds to the category of real vector spaces.

Mayer-Vietoris Theorem

For a sequence of vector spaces and linear maps

$$V_1 \xrightarrow{F_1} V_2 \xrightarrow{F_2} V_3 \xrightarrow{F_3} V_4 \xrightarrow{F_4} ..., \tag{13}$$

we say the sequence is **exact** if the image of each map is equal to the kernel of the next, i.e., $\mathrm{Im}\, F_i = \mathrm{Ker}\, F_{i+1}$. An important property of exact sequences is that if the sequence $0 \xrightarrow{i} A \xrightarrow{f} B \xrightarrow{j} 0$ is exact, then f is a bijection from A to B.

Mayer-Vietoris Theorem is similar to Van Kampen's theorem in homotopy theory. It states the relationship between the de Rham cohomology group of open subsets of M and the de Rham cohomology group of M.
Let U, V be open subsets of a smooth manifold M s.t. $U \cup V = M$. Consider the inclusion from $U \cap V$ to U and V, denoted by i_U, i_V respectively, and the inclusion from U and V to M, denoted by j_U, j_V respectively.
Mayer-Vietoris Theorem states that for each integer k, there exists a linear map $\delta : H^k_{dR}(U \cap V) \to H^{k+1}_{dR}(M)$ s.t. the following sequence, called the **Mayer-Vietoris sequence** for the open cover $\{U, V\}$, is exact.

$$\begin{aligned} ... \xrightarrow{\delta} H^k_{dR}(M) \xrightarrow{j_U^* \oplus j_V^*} H^k_{dR}(U) \oplus H^k_{dR}(V) \xrightarrow{i_U^* - i_V^*} H^k_{dR}(U \cap V) \\ \xrightarrow{\delta} H^{k+1}_{dR}(M) \xrightarrow{j_U^* \oplus j_V^*} ... \end{aligned} \tag{14}$$

Below is an example of how to compute $H^k_{dR}(S^n)$ for $n \geq 2$ using the theorem:
Let U, V be S^n with the north and south pole removed, respectively. Then U, V are homotopic to $\mathbb{R}^n$, which is contractible, and $U \cap V$ is homotopic to $\mathbb{R}^n \setminus \{0\}$, which is homotopic to S^{n-1}. Therefore, Mayer-Vietoris Theorem tells us that

$$0 \xrightarrow{i_U^* - i_V^*} H^{k-1}_{dR}(U \cap V) \xrightarrow{\delta} H^k_{dR}(S^n) \xrightarrow{j_U^* \oplus j_V^*} 0 \tag{15}$$

is exact, which implies $H^k_{dR}(S^n) \cong H^{k-1}_{dR}(U \cap V) \cong H^{k-1}_{dR}(S^{n-1})$ for $k > 1$. Since we have $H^1_{dR}(S^1) = \mathbb{R}$, this shows $H^n_{dR}(S^n) = H^1_{dR}(S^1) = \mathbb{R}$.

Acknowledgements and References

I really thank John White for his guidance and the UCSB Directed Reading Program (2024) for the opportunity to work on this project.
Textbook: *Lee, J.M., Introduction to Smooth Manifolds, Graduate Texts in Mathematics, Springer New York, 2013*

Solutions From the Inverse Scattering Transform

Reese Karo
University of California Santa Barbara

Brief Intro to PDEs

Partial Differential Equations (PDEs) are mathematical equations that describe the behavior of systems involving rates of change with respect to temporal and spatial variables. Commonly in the form $u(x,t)$, where u is the field variable describing a value (i.e. temperature in a rod, height of a wave). PDEs relate partial derivatives like $\frac{\partial u}{\partial t} = u_t$ or $\frac{\partial^2 u}{\partial x^2} = u_{xx}$, showing how systems evolve over time and space.
PDEs are categorized into **linear** and **non-linear** types, Linear PDEs include a linear combination of the field variable and its derivatives, whereas non-linear PDEs involve products of the field variable and its derivatives. A few useful equations include:
Defocusing Non-linear Schrödinger Equation (NLS):

$$\frac{\partial \psi}{\partial t} + \frac{1}{2}\frac{\partial^2 \psi}{\partial x^2} - k|\psi|^2\psi = 0 \tag{1}$$

The NLS is a fundamental equation used to describe wave packet dynamics in non-linear media. **Fourier Transform (FT) and Inverse Fourier Transform (IFT)**

$$\text{FT: } \hat{\psi}(k,t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi(x,t)e^{-ikx}\,dx \text{ , IFT: } \psi(x,t) = \int_{-\infty}^{\infty} \hat{\psi}(k,t)e^{ikx}\,dk$$

The Fourier Transform and Inverse are crucial tools for solving the initial value problem of linear PDE's such as the **linear Schrödinger equation**

$$i\psi_t + \frac{1}{2}\psi_{xx} = 0, \psi(x,0) = \psi_0(x) \tag{2}$$

By applying the Fourier Transform to (2) and using the initial condition $\psi(x,0) = \psi_0(x)$, we get $i\hat{\psi}_t - \frac{k^2}{2}\hat{\psi} = 0$ and $\hat{\psi}(k,0) = \hat{\psi}_0(k)$, which is a simple ordinary differential equation. The solution to this ODE is $\hat{\psi}(k,t) = \hat{\psi}_0(k)e^{-i\frac{k^2}{2}t}$. Then, by plugging this into the IFT equation, we find our solution:

$$\hat{\psi}(k,t) = \frac{1}{2\pi}\int_{\mathbb{R}} \psi_0(k)e^{ik(1-\frac{k}{2}t)}dk.$$

Lax Pairs

Lax Pairs [2] provide a framework for integrable systems, enabling the derivation of PDEs such as the Korteweg–De Vries equation, Nonlinear Schrödinger equation, Sine-Gordon equation, and more [1, pp. 9-15]. We define the commutator as $[U,V] = UV - VU$, and the **Lax pair** for our system is given by:

$$\frac{\partial w}{\partial x} = Uw, \qquad \frac{\partial w}{\partial t} = Vw \tag{3}$$

where the spatial linear operator U and the temporal spectral operator V are:

$$U(x,t;\lambda) = \begin{bmatrix} -i\lambda & \psi \\ \psi^* & i\lambda \end{bmatrix}, \tag{4}$$

$$V(x,t;\lambda) = \begin{bmatrix} -i\lambda^2 - i\frac{1}{2}|\psi|^2 & \lambda\psi + i\frac{1}{2}\psi_x \\ \lambda\psi^* + i\frac{1}{2}\psi_x^* & i\lambda^2 - i\frac{1}{2}|\psi|^2 \end{bmatrix}, \tag{5}$$

such that λ is a complex spectral parameter. By cross-differentiating (3), we obtain the *zero-curvature condition*, which leads to the *defocusing Nonlinear Schrödinger equation* [1, p. 9]:

$$\frac{\partial U}{\partial t} - \frac{\partial V}{\partial x} + [U,V] = 0$$

This equation is satisfied if $\psi(x,t)$ solves the NLS equation. In the framework of scattering theory, λ determines the behavior of solutions at infinity. The scattering data, which consist of the reflection and transmission coefficients, depend on λ. These coefficients capture information about how the potential $\psi(x,t)$ affects incoming waves, with different values of λ providing details of the scattering process. Additionally, λ relates to the kernel in the Volterra integral equations used

Inverse Scattering Transform For NLS

The **Inverse Scattering Transform (IST)** is a method used for solving certain nonlinear PDEs by evolving scattered data over time. By using the boundary condition properties, we can invert back to the original potential, hence finding a solution to the nonlinear PDE. To facilitate this transformation, we use **Jost Solutions**, which help determine the scattering data $a(\lambda)$ and $b(\lambda)$. These solutions satisfy the linear Schrödinger equation with boundary conditions approaching limits as $|x| \to \infty$:

$$J_-(x;\lambda) = \begin{bmatrix} e^{-i\lambda x} & 0 \\ 0 & e^{i\lambda x} \end{bmatrix} \quad \text{as} \quad x \to -\infty$$

$$J_+(x;\lambda) = \begin{bmatrix} e^{-i\lambda x} & 0 \\ 0 & e^{i\lambda x} \end{bmatrix} \quad \text{as} \quad x \to \infty$$

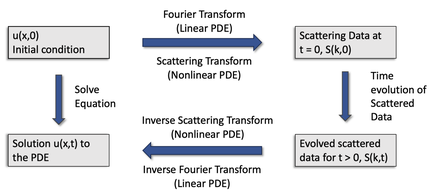
We note that both $J_\pm$ are non-singular matrices. Jost solutions, J_- and J_+ have linearly independent columns; however, all four solutions are not independent of each other.[3, p. 5] Since J_+ and J_- form a basis of the space of solutions, we can write them as a linear combination of each other.

$$J_+ = J_-S(\lambda) \iff J_- = J_+S^{-1}(\lambda)$$

In particular, $J_-^{(1)}(x;\lambda) = a(\lambda)J_+^{(1)}(x;\lambda) + b(\lambda)J_+^{(2)}(x;\lambda)$. If we divide both sides by $a(\lambda)$ we have a solution

$$w(x;\lambda) = J_+^{(1)}(x;\lambda) + R(\lambda)J_+^{(2)}(x;\lambda)$$

where we define $R(\lambda) = \frac{b}{a}(\lambda)$ and $T(\lambda) = \frac{1}{a(\lambda)}$ which is defined as the **reflection and transmission coefficients**. Moreover, the reflection coefficient evolves in time: $R(\lambda;t) = R(\lambda,0)e^{2i\lambda^2 t}$. We now focus on the direct transform, found in [3, p. 5].
Definition: (Direct Transform) For a suitable function $\psi : \mathbb{R} \to \mathbb{C}$ decaying as $|x| \to \infty$, then the direct transform for the defocusing Nonlinear Schrödinger equation is the mapping of $\psi \mapsto R$ associating to ψ its reflection coefficient $R = R(\lambda)$, $\lambda \in \mathbb{R}$
Schematic of the methodology behind Fourier Transform and IST of linear and nonlinear PDE's.



The direct transform gathers scattering data, such as $a(\lambda)$ and $b(\lambda)$ which ultimately yields the reflection coefficient, $R(\lambda)$. We then can solve the **Volterra integral equations** for the generalized Jost solutions:

$$u(x;\lambda) = 1 + \int_x^{-\infty} e^{2i\lambda y}\psi(y)v(y;\lambda)dy \text{ , } v(x;\lambda) = \int_x^{-\infty} e^{2i\lambda y}\psi(y)^* u(y;\lambda)dy$$

Solving for $u(x,\lambda)$ and obtain the following kernel,

$$u(x;\lambda) = 1 + \int_x^{-\infty} K(x,z;\lambda)u(z;\lambda)dz$$

$$K(x,z;\lambda) = \psi^* \int_z^x e^{2i\lambda(y-z)}\psi(y)dy,\ x > z$$

When solved, we can recover the potential, ψ from the direct scattering, analogously to the Fourier Transform. Thus we end up with the following relationship by the IST.[1, p. 20-24]

$$\psi(x,t) = -2K(x,x;\lambda)$$

Alternatively, we can use IST via the Riemann-Hilbert Problem, equivalent to solving the integral equations above.

Riemann-Hilbert Problem

Our specific Riemann-Hilbert problem asks us to find a matrix $M \in \mathbb{R}^{2x2}$ such that M is: Analytic for $\lambda \in \mathbb{C} \setminus \mathbb{R}$, satisfies the jump condition matrix, D, and has a normalizing condition $\lim_{\lambda\to\infty} M(\lambda;x,t) = \mathbb{I}$. Thus, by combining our Jost solutions into a matrix M such that the upper and lower blocks extend into the upper and lower half-planes respectively, we have

$$M_+(\lambda;x) = \begin{bmatrix} \frac{e^{i\lambda x}}{a(\lambda)}J_-^{(1)} & e^{-i\lambda x}J_+^{(2)} \end{bmatrix} \quad \text{for} \quad \Im(\lambda) > 0,$$

$$M_-(\lambda;x) = \begin{bmatrix} e^{i\lambda x}J_+^{(1)} & \frac{e^{-i\lambda x}}{a(\lambda)}J_-^{(2)} \end{bmatrix} \quad \text{for} \quad \Im(\lambda) < 0.$$

Note that M is a single matrix and each Jost solution component is scaled by $e^{\pm i\lambda x}$ to maintain $\det M(\lambda;x) = 1$.

Lemma : (Analyticity of $\mathbf{M(\lambda;x)}$ [3, p. 14]): For every $x \in \mathbb{R}$, the elements of the matrix $\mathbf{M(\lambda;x)}$ are all analytic functions of λ for $\Im(\lambda) \geq 0$.
However, we get a jump condition as λ approaches the real axis from both the upper and lower imaginary planes. According to [3, p. 14-15], we define the following where $\lambda \in \mathbb{R}$, $M_\pm(\lambda;x) := \lim_{\epsilon\downarrow 0} M(\lambda \pm i\epsilon;x)$.
Since Jost solutions can be written as a linear combination of each other, we also have

$$M_+(\lambda;x)\begin{bmatrix} 1 \\ -e^{2i\lambda x}\frac{b}{a}(\lambda) \end{bmatrix} = M_-\begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ , } M_+(\lambda;x)\begin{bmatrix} 1 \\ 0 \end{bmatrix} = M_-\begin{bmatrix} -e^{-2i\lambda x}\frac{b^*}{a^*}(\lambda) \\ 1 \end{bmatrix}$$

which can be rewritten as

$$M_+(\lambda;x)\begin{bmatrix} 1 & 0 \\ -e^{2i\lambda x}\frac{b}{a}(\lambda) & 1 \end{bmatrix} = M_-\begin{bmatrix} 1 & -e^{-2i\lambda x}\frac{b^*}{a^*}(\lambda) \\ 1 & 0 \end{bmatrix}$$

By multiplying by the inverse of the right matrix on the left-hand side, we get the relationship for $M_+(\lambda;x)$

$$M_+(\lambda;x) = M_-(\lambda;x)D(\lambda;x)$$

where the jump matrix $D(\lambda;x)$ is

$$D(\lambda;x) = \begin{bmatrix} 1 - |R(\lambda)|^2 & -e^{-2i\lambda x}R(\lambda)^* \\ e^{2i\lambda x}R(\lambda) & 1 \end{bmatrix}$$

We find the solution, $\psi(x,t)$ by differentiating the second column of M with respect to x, $\frac{\partial}{\partial x}M^{(2)}(\lambda,x)$. From (3), we have

$$M_x^{(2)} = \begin{bmatrix} -2i\lambda & \psi \\ \psi^* & 0 \end{bmatrix} M^{(2)} \implies \psi(x,t) = 2i\lim_{\lambda\to\infty} M_{12}(\lambda;x,t)$$

Therefore, whether we use the Riemann-Hilbert problem, or the integral equations using techniques such as Neumann series to solve for u and v to collect scattering data, we can successfully return the potential $\psi(x,t)$ from the evolved scattered data.

Acknowledgements

I would like to thank my mentor, Christian Hong, for his guidance and teachings in the subject of PDEs, and UCSB DRP for this unforgettable experience.

References

[1] M. J. Ablowitz and H. Segur. *Solitons and the Inverse Scattering Transform*. Vol. 4. Classics in Applied Mathematics. Philadelphia: SIAM, 1981.

[2] Peter D. Lax. "Integrals of Nonlinear Equations of Evolution and Solitary Waves". In: *Communications on Pure and Applied Mathematics* 21.5 (1968), pp. 467–490.

[3] Peter D. Miller. "What is ... the Inverse Scattering Transform?" In: *Notices of the American Mathematical Society* 59.10 (Nov. 2012), pp. 1340–1341.

UC **SANTA BARBARA**

The BQP Solar System

Anikait Mundhra Mentor: Sawyer Dobson

Introduction

A *complexity class* consists of a set of computational problems (languages) adhering to a rule. The difficulty of computational problems can be classified by the complexity classes they lie in; they help us understand which problems are tractable and which require more resources. Classical complexity classes consist of problems that use *Turing Machines* (TMs). We list below some relevant classical complexity classes.

Let L be a language. We write poly(n) to refer to any polynomial in n, and $\mathcal{D}$ as the uniform distribution over $\{0,1\}^{\text{poly}(n)}$. We assume any TM to have binary output, 0 (*rejecting*) or 1 (*accepting*).

- $L \in$ **P** iff there exists a poly-time TM M such that $M(x) = 1$ iff $x \in L$.
- $L \in$ **NP** iff there exists a poly-time TM M such that for all x, there exists a certificate $y \in \{0,1\}^{\text{poly}(n)}$, $M(x,y) = 1$ iff $x \in L$.
- $L \in$ **BPP** iff there exists a poly-time TM M such that for $\Pr_{r\sim\mathcal{D}}[M(x,r) = 1] \geq \frac{2}{3}$ if $x \in L$ and $\Pr_{r\sim\mathcal{D}}[M(x,r) = 1] \leq \frac{1}{3}$ if $x \notin L$.
- $L \in$ **#P** iff L is a function mapping an input x of a poly-time non-deterministic TM M to the number of accepting paths of $M(x)$. In other words, L is a function mapping x to $|\{r \in \mathcal{D} : M(x,r) = 1\}|$.
- $L \in$ **PP** iff there exists a poly-time TM M such that $\Pr_{r\sim\mathcal{D}}[M(x,r) = 1] > \frac{1}{2}$ if $x \in L$ and $\Pr_{r\sim\mathcal{D}}[M(x,r) = 1] < \frac{1}{2}$ if $x \notin L$.
- $L \in$ **P-SPACE** iff there exists a poly-space TM M such that $M(x) = 1$ iff $x \in L$.
- $L \in$ **EXP** iff there exists a exponential-time TM M such that $M(x) = 1$ iff $x \in L$.

It is believed that quantum computation is more powerful than classical computation, giving rise to quantum complexity classes. These classes consist of problems that use *Quantum Circuits* (QCs). Thus, we are interested in the limits of quantum computation: which problems can and cannot be solved efficiently by quantum computers? The complexity class **BQP** aims to capture problems that quantum computers can solve efficiently. We first formally define **BQP**, and then reason about where it stands in relation to the classical complexity classes.

Bounded-Error Quantum Polynomial-Time (BQP)

Similar to the notion of the classic complexity class **BPP**, we define the quantum complexity class **BQP**. We say a family of QCs $\{C_k\}$ is poly-time uniform there exists a poly-time TM which given k, outputs a description of C_k.

- $L \in$ **BQP** iff there exists a family of poly-time uniform QCs $\{C_x\}$ with $O(\text{poly}(n))$ gates ($n = |x|$) such that the probability that the **first** qubit of $C_x|0\rangle^{\text{poly}(n)}$ is 1 is at least $\frac{2}{3}$ if $x \in L$ and is at most $\frac{1}{3}$ if $x \notin L$.

Oracles

Many times, for two complexity classes $\mathbf{C_1}$ and $\mathbf{C_2}$, it can be hard to prove or disprove any relation between them directly. However, we can sometimes show these classes are separate with respect to an oracle.

Formally, an *oracle* $\mathcal{O}$ is just a language, i.e. a set of binary strings. A *black-box function* is a function whose output can be computed in a single time-step without any other knowledge of the function. An *oracle TM* or *oracle QC* with an access to an oracle $\mathcal{O}$ is a TM or QC given access to a black-box function f which computes if any binary string lies in $\mathcal{O}$. For any complexity class **C** that is defined in terms of Turing machines, we obtain the definition of the complexity class $\mathbf{C}^{\mathcal{O}}$ by replacing, in the definition of **C**, every occurrence of "TM" (or "QC") with "oracle TM (or QC) with access to $\mathcal{O}$". For example, we can think of $\mathbf{NP}^{\mathcal{O}}$ as the following.

- $L \in \mathbf{NP}^{\mathcal{O}}$ iff there exists a poly-time oracle machine $M^{\mathcal{O}}$ with access to $\mathcal{O}$ such that for all x, there exists a certificate $y \in \{0,1\}^{\text{poly}(n)}$ such that $M^{\mathcal{O}}(x,y) = 1$ iff $x \in L$.

Similarly, we define $\mathbf{P}^{\#\mathbf{P}}$ to be the class of all languages decidable by a poly-time TM furnished with the ability to solve any problem in **#P** in unit time.

So, sometimes, we are unable to show $\mathbf{C_1} \not\subseteq \mathbf{C_2}$ yet we can show there exists an oracle $\mathcal{O}$ for which $\mathbf{C_1}^{\mathcal{O}} \not\subseteq \mathbf{C_2}^{\mathcal{O}}$. Note this second statement does not imply the first, but only gives us "evidence" towards believing the first may be true.

As any quantum circuit can be classically simulated in time exponential to its input, we have **BQP** $\subseteq$ **EXP**

$\mathbf{NP}^{\mathcal{O}} \not\subseteq \mathbf{BQP}^{\mathcal{O}}$.
Consider the black-box search problem:

> **In**: A function $f : \{0,1\}^n \to \{0,1\}$ as a black-box $\mathcal{O}_f$, where $\mathcal{O}_f(x) = f(x)$ and $\mathcal{O}_f|x\rangle = (-1)^{f(x)}|x\rangle$.
> **Out**: 1 iff $\exists x \in \{0,1\}^n$ s.t. $f(x) = 1$.

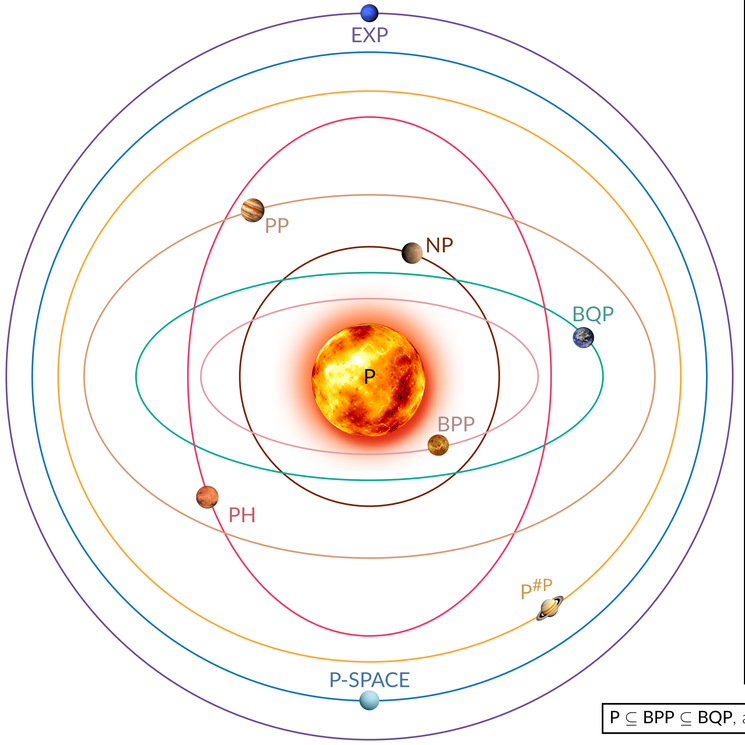
Notice if for every f in the language, we set y to be such that $f(y) = 1$, then the TM M that takes in (f, y) and returns the query from $f(y)$ solves the problem in the **NP** definition with black-box access.
For **BQP**, we start by claiming claim any QC solving the black-box search problem for any general function f needs to query $\mathcal{O}_f$ at least $\Omega(2^{n/2})$ times. Let $\vec{v} \in \{0,1\}^{2^n}$ be such that $\vec{v}_x = f(x)$ for $x \in \{0,1\}^n$. Notice we want to compute $\vee_x \vec{v}_x$ (logical OR).
Theorem: If we use T queries to compute $\vee_x \vec{v}_x$, then there exists a polynomial p with degree $2T$ such that

$$|p(\vec{v}) - \vee_x \vec{v}_x| \leq \tfrac{1}{3}.$$

Proof: We can write our QC as $U_T\mathcal{O}_f U_{T-1}\dots U_1\mathcal{O}_f U_0$, where each U_i is a unitary. We start by induction, to show each amplitude of the state after k queries is a polynomial in $\vec{v}$ of degree k. The base case is clear. Suppose after k queries, the state of our QC is $\sum_{y\in\{0,1\}^n} \alpha_y^k|y\rangle$ where α_y^k are polynomials of degree k in $\vec{v}_y$. Before the $k+1$ query, our QC will apply some unitary operation U_k, which will preserve the degrees of α_y^k and make them some new α_y^k. We will then apply our black-box function $\mathcal{O}_f$ to get

$$\mathcal{O}_f \sum_y \alpha_y^k|y\rangle = \sum_y (-1)^{f(y)}\alpha_y^k|y\rangle = \sum_y (1-2\vec{v}_y)\alpha_y^k|y\rangle$$

So, the induction holds. Now, notice if we let p be $\Pr[1\text{st qubit} = 1]$, then $p = \sum_{1y'} |\alpha_{1y'}^k|^2$, which is a polynomial in $\vec{v}$ of degree $2T$. Thus, the theorem holds as p is the output of the QC computing OR.
Fact: Any polynomial approximating OR needs degree $\Omega(2^{n/2})$. This can be proven with symmetrization and the Markov brothers' inequality.
This tells us, for some QC C to solve this problem in general for any function f, it needs $\Omega(2^{n/2})$ queries. This gives us a black-box lower bound for **BQP**, and ultimately tells us $\mathbf{NP}^{\mathcal{O}} \not\subseteq \mathbf{BQP}^{\mathcal{O}}$.



BQP $\subseteq$ **P-SPACE**.
Let $L \in$ **BQP** and let x be our input. We make the following poly-space TM:
For each $y \in \{0,1\}^{n-1}$, we compute α_P for each Feynman Path $P \in \mathcal{P}_{|1y\rangle}$ of C_x and sum them, discarding old values. We square each sum, and add it to a rolling sum over the y's, discarding old values. Notice we have computed $\sum_{y\in\{0,1\}^{n-1}} \left|\sum_{P\in\mathcal{P}_{|1y\rangle}} \alpha_P\right|^2$ in poly-space. We output $x \in L$ iff our sum over 2^H is at least $2/3$.

BQP $\subseteq$ **PP**.
Let $L \in$ **BQP** and let x be our input. We make the following poly-time TM:
Compute two Feynman Paths P, α_P and $P', \alpha_{P'}$ of C_x at random (in poly-time). If P and P' do not end in the same state, flip a coin for the output. Else, $P, P' \in \mathcal{P}_{|by\rangle}$, $b \in \{0,1\}$. Output $x \in L$ iff $(-1)^{(1-b)}\alpha_P\alpha_{P'} > 0$. Notice we output 1 iff the two paths contribute towards a higher $\Pr[1\text{st qubit} = 1]$, so our probability of guessing right is greater than $1/2$. Thus, $L \in$ **PP**.

BQP $\subseteq \mathbf{P}^{\#\mathbf{P}}$.
Let $L \in$ **BQP** and let x be our input. Define $\#_{+1}, \#_{-1}$ to be the number of Feynman Paths $P, P' \in \mathcal{P}_{|1y\rangle}$ in C_x such that $\alpha_P\alpha_{P'} = 1, -1$ respectively. We have $\Pr[1\text{st qubit} = 1] = \frac{1}{2^H}(\#_{+1} - \#_{-1})$. Notice as calculating $\alpha_P\alpha_{P'}$ for two paths P, P' takes polynomial time, so finding $\#_{+1}, \#_{-1}$ is in **#P**. Thus, let our machine take x as input, find $\#_{+1}, \#_{-1}$ with oracle calls, and output 1 iff $\frac{1}{2^H}(\#_{+1} - \#_{-1}) \geq \frac{2}{3}$. It follows $L \in \mathbf{P}^{\#\mathbf{P}}$.

$\mathbf{BQP}^{\mathcal{O}} \not\subseteq \mathbf{NP}^{\mathcal{O}}$.
Consider Simon's problem:

> **In**: $f : \{0,1\}^n \to \{0,1\}^n$ as a black-box $\mathcal{O}_f$, where $\mathcal{O}_f(x) = f(x)$ and $\mathcal{O}_f|x\rangle|y\rangle = |x\rangle|y \oplus f(x)\rangle$.
> **Out**: 1 if $\exists s \in \{0,1\}^n \setminus \{0^n\}$ s.t. $f(x) = f(y)$ and $x \neq y \iff x = y \oplus s$, or 0 if f is one to one.
> We are promised f falls into one of those two cases.

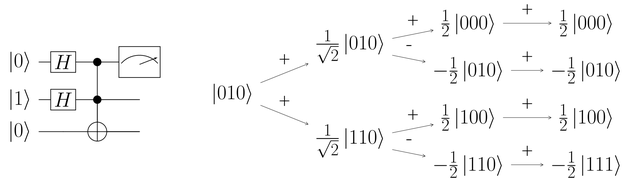
It is known Simon's problem does not give any oracle separation between **BQP** and **NP**. Let's define Co-Simon's problem as Simon's problem with the outputs flipped (output 1 iff f is one to one).
Notice Co-Simon's problem has a **BQP** circuit with black-box access (by flipping the outputs of the **BQP** circuits for Simon's Problem). However, Co-Simon's problem cannot be solved in the **NP** definition with black-box access across all functions f:
Assume for contradiction there is some poly-time TM M such that for all promised functions f, there exists a certificate $y \in \{0,1\}^{\text{poly}(n)}$ such that $M(f,y) = 1 \iff f$ is one-to-one. We know M makes some $T = O(\text{poly}(n))$ queries on inputs $(q_1, \dots, q_T)$ to $\mathcal{O}_f$ so it knows s cannot be $\binom{T}{2}$ different numbers. Yet, s can be exponentially $(2^n - 1)$ many numbers, so we can construct an f' with $f'(x) = f'(x \oplus s)$ where $f'(q_i) = f(q_i)$. Since M is deterministic, $M(f', y) = 1$, which is a contradiction as f' is not one to one.
Notice by this contradiction, we also realize M must make $T = \Sigma(2^{n/2})$ queries for $\binom{T}{2} \geq 2^n - 1$. This gives us a black-box lower bound for **NP**, and ultimately tells us $\mathbf{BQP}^{\mathcal{O}} \not\subseteq \mathbf{NP}^{\mathcal{O}}$.

P $\subseteq$ **BPP** $\subseteq$ **BQP**, as any poly-time TM can be written as a poly-time QC.

Feynman Paths

A common trick when working with arbitrary poly-time QCs is to decompose them into an equivalent poly-time QC only consisting of Hadamard and Toffoli gates.

A Feynman Path of a circuit is the path starting at some initial state and ending at a final state, with some amplitude. When the circuit composed solely of Hadamard and Toffoli gates, each state does not split, or either splits into two paths of equal magnitude.



For a quantum circuit C in **BQP**, its output depends on the probability of the first qubit being in state $|1\rangle$. Let $\mathcal{P}_{|x\rangle}$ be the set of Feynman Paths ending in state $|x\rangle$, let $\alpha_P = \pm 1$ be the sign of a path P and let H be the number of Hadamards in C. We have

$$\Pr[1\text{st qubit} = 1] = \sum_{y\in\{0,1\}^{n-1}} \left|\frac{1}{\sqrt{2^H}} \sum_{P\in\mathcal{P}_{|1y\rangle}} \alpha_P\right|^2 = \frac{1}{2^H} \sum_{y\in\{0,1\}^{n-1}} \sum_{P,P'\in\mathcal{P}_{|1y\rangle}} \alpha_P\alpha_{P'}.$$

Feynman Paths are incredibly useful in understanding and simulating QCs.

From Black-Box Lower Bounds to Oracle Separations

Let $L \subseteq \mathcal{F}$ be some language, where $\mathcal{F}$ is a set of black-box functions. Let's say there is some poly-time TM which solves this problem in the **NP** definition across all $f \in \mathcal{F}$, but for all QCs solving this problem across all $f \in \mathcal{F}$, they need some exponential number of queries to the black-box. We can show an oracle separation between **NP** and **BQP**:

We know for all poly-time QC C solving this problem, they can only make polynomially many queries. By the lower bound, for each poly-time C, there are infinitely many functions f of large enough size, which C fails on. As QCs are enumerable, for each poly-time QC C, pick a function f_C (that foils C) of unique size. Let $L' = \{1^{|f|} \mid f \in L\}$ and let $\mathcal{O}$ be the set of f that were picked. Notice that $L' \in \mathbf{NP}^{\mathcal{O}}$ by our poly-time TM that can solve the black-box problem across all $f \in \mathcal{F}$. Yet, $L' \notin \mathbf{BQP}^{\mathcal{O}}$ as for each poly-time QC C, there is some f in the language which C fails on. Thus, we have shown $\mathbf{NP}^{\mathcal{O}} \not\subseteq \mathbf{BQP}^{\mathcal{O}}$.

Note we can switch **NP** and **BQP** above (switching TMs and QCs) to argue $\mathbf{BQP}^{\mathcal{O}} \not\subseteq \mathbf{NP}^{\mathcal{O}}$.

Acknowledgements and References

I am profoundly grateful to my mentor, Sawyer Dobson, whose discrete knowledge and continuous support made this possible. I would like to thank him for his patience and effort throughout the reading. I would be stuck in orbit **BQP** without his guidance.

I would also like to thank the DRP Committee and organizers for this opportunity.

Majority of this information comes from scribe notes of CS 359D at Stanford.

The First 197 Alternating Knots

Alex Gaither and Mihir Mantri
University of California Santa Barbara

What is a knot?

Imagine a knot. What might come to your mind is the bow of a shoelace. This is a knot in the traditional sense, but it is not a mathematical knot. Now imagine that the string has arbitrarily small thickness and the ends of the shoelace are fused together so that the string forms one tangled loop. Now, this is a mathematical knot.

Definitions

Knot: A circle embedded in $\mathbb{R}^3$
Knot diagram: A projection of a knot onto $\mathbb{R}^2$ that includes crossing information of the knot. A **crossing** is a point whose preimage under the projection has two points; whichever point is closer to the projection is the crossing information.

Some knots look different but they are just deformations of each other. We say that two knots are equivalent if they can be deformed to each other by an *ambient isotopy*.

Ambient isotopy: A continuous deformation of a knot that does not intersect itself
Knot invariant: A property of a knot that does not change under ambient isotopy

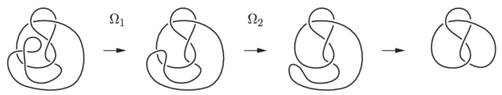


Figure 1: A sequence of ambient isotopies on a knot diagram

With respect to some fixed projection, ambient isotopy can be categorized into deformations that do not affect the crossings and three deformations that do, the Reidemeister moves:

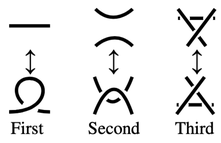


Figure 2: The Reidemeister moves

A knot diagram is **reduced** if it is a diagram with the least number of crossings possible.
Connect sum: The connect sum $J + K$ of two knots J and K is obtained by cutting an arc from each knot then connecting the newly made endpoints.

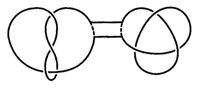


Figure 3: Connect sum of two knots.

Prime knot: A knot which cannot be obtained as a connect sum of nontrivial knots

---

Mazur Swindle: $J + K = 0$ if and only if one of $J = 0$ or $K = 0$.
Prime decomposition (Schubert): Every knot decomposes into a unique connect sum of primes.
Corollary: The set of knots with the operation of connect sum (the knot monoid) is a free $\mathbb{N}$-module generated by the prime knots isomorphic to $\bigoplus_{i\in\mathbb{N}} \mathbb{N}$.

Thus finding a basis for the knot monoid amounts to tabulating the prime knots. So, we aim to tabulate prime knots.

See our repository here!

Dowker code

Dowker code: A way to describe a knot diagram. Take a knot diagram with n crossings. Pick a starting crossing and traverse around the knot, labeling each crossing from 1 to $2n$ until you have returned to where you started. This will always pair odd numbers with even numbers on crossings. Thus we denote a Dowker code by a list of even numbers in order of the size of their corresponding odd number.

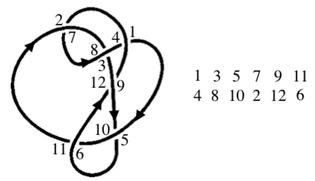


Figure 4: Constructing a Dowker Code the 6_3 knot

Consider the inverse problem: Obtaining knots from Dowker codes. It is possible that a Dowker code is **unrealizable**. That is, it cannot actually be made into a knot because it implies an extra crossing.

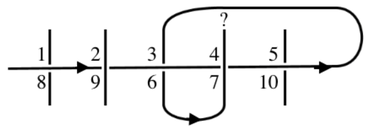


Figure 5: [8, 6, 10, 4, 2] is unrealizable

Theorem (Dowker-Thistlewaite): A realizable Dowker code specifies a unique knot, up to reflection.

Notice that this correspondence is far from injective. For instance, there are $4n$ ways to obtain a code from a knot projection with n crossings.

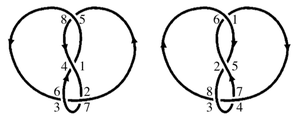


Figure 6: Two different codes for the square knot

Alternating knots

Alternating knot: A knot that has an alternating knot diagram. The Dowker notation usually includes signage on crossings, but for alternating knots, every crossing is positive. According to **Tait's Flyping Theorem**, which was proved in 1991, all projections of a prime, reduced, alternating knot are related through a sequence of moves called flypes.

Because of this theorem, we decided to focus on tabulating prime alternating knots.

A **tangle** is an area within a knot that when enclosed with a dotted line, the dotted line is crossed by the string exactly four times.
A **flype** involves the movement of two connected parts: a crossing and a tangle.

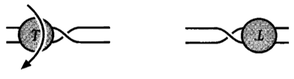


Figure 7: A flype on a tangle T

Outline of Tabulation

Because of Tait's Flyping Theorem, to tabulate alternating knots with n crossings we generate all Dowker codes of up to length $2n$ and eliminate ones that don't satisfy:

1. The code is realizable
2. The code is lexicographically minimal among equivalent codes
3. The code specifies a reduced prime knot
4. The knot diagram specified by the code cannot be obtained by flyping a knot diagram whose code is preferred according to the previous criteria

Tabulating alternating knots

Step 1:

To ensure realizability, we represent a knot diagram as a graph by replacing each crossing with a square of four nodes, and remove those codes for which this graph is nonplanar.

Step 2:

We conventionally choose the lexicographically minimal Dowker code as a representative for each isotopy class, and remove all other equivalent codes. For example, from Figure 6, the abbreviated code of the right diagram [4, 6, 8, 2] is preferred to that of the left [6, 8, 2, 4].

Step 3:

Notice that the Dowker codes of composite knots can be separated into two separate subsequences representing their two factors. Thus, we remove any code exhibiting such subsequences.
This also removes all possible Type I Reidemeister moves as a loop will be detected as a single crossing subsequence.

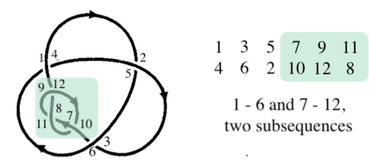


Figure 8: Subsequences in a Composite Knot

Step 4:

We perform all possible flypes on the diagrams corresponding to each Dokwer code to identify equivalent codes. The tangle-crossing pair of a flype is identifiable by two sequences of numbers and is followed by a crossing. (See crossings 2, 3 and crossing 1 in *Figure 4*.)
To execute a flype on a knot, we again convert the knot to a graph. Next, we cut out the subgraph corresponding to a tangle-crossing pair, rotate it and reattach it to the graph. Finally, we reconstruct the Dowker code from the resulting graph.

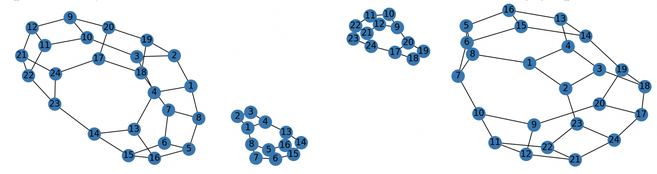


Figure 9: The 6_3 knot (left), flyped to an equivalent knot through this process

Acknowledgements

We would like to thank our mentor Choomno Moos for his encouragement and guidance. We appreciate this opportunity that was provided by the 2024 DRP program.

References

[1] C. Adams. *The Knot Book*. Providence, RI: W.H. Freeman and Company, 2000.

[2] Andrey Boris Khesin. "The 250 Knots with up to 10 Crossings". In: *arXiv* (2017). DOI: 10.48550/ARXIV.1705.10319.

Training a Convolutional Neural Network to Recognize Instruments

Anna Maximova, Mentored by Alan Raydan

2024 Mathematics Directed Reading Program. Department of Mathematics, University of California, Santa Barbara

The Vision

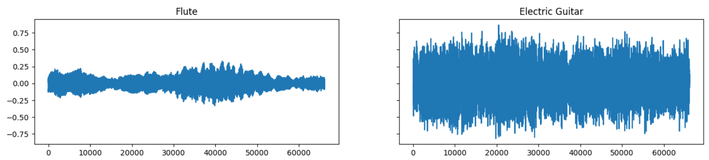
When Alan and I had our first meeting, we came up with a vision of a generative music AI that could take a well known composer as input and generate a few seconds of classical music of that composers style. While exciting, that is a huge project and to get started we needed to come up with something a little more doable. So as the first stepping stone in this vision was to be a neural network that could be trained to identify the instrument in an audio clip.

The Data

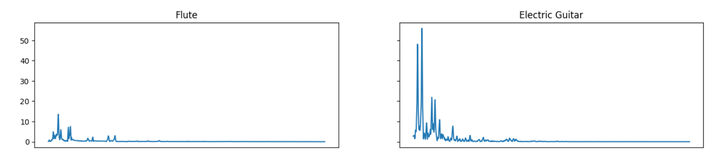
The data I chose for this project came from Universitat Pompeu Fabra and was compiled by Ferdinand Fuhrmann for his PhD thesis. It contains, "6705 audio files in 16 bit stereo wav format sampled at 44.1kHz." [1] Each recording is 3 seconds long and is labeled by the predominant instrument in each sound. The instruments are: cello, clarinet, flute, acoustic guitar, electric guitar, organ, piano, saxophone, trumpet, violin, and human singing voice.

The Melspectrograms

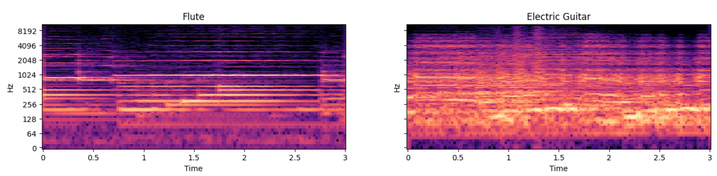
For our model to be able to work with the input we needed to convert each of the wav files to some collection of numbers. The most common practice is transforming audio to images. Below is a sequence of images for the same two audio files, where the first audio files contains flute as the predominant instrument and the second contains electric guitar as the predominant instrument. First we take a look at the raw signal.



We use the Fourier transform to get a better understanding of which frequencies are present for each of the instruments.



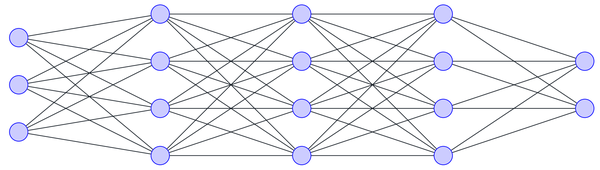
And finally the melspectrograms. A spectrogram is a plot of time vs frequency. A brighter color means a greater amplitude/presence of that frequency at that time. **Fun fact:** humans perceive sound logarithmically, which means that "we can easily tell the difference between 500 and 1000 Hz, but we will hardly be able to tell a difference between 10,000 and 10,500 Hz."[2]



The Convolutional Neural Network

What is a neural network?

To understand out instrument recognition model, we must first understand what a neural network is. In the simplest terms, it is a machine learning model that is modeled after neurons in the brain. (Is it actually a good representation of how the brain works? Not particularly, so I wouldn't be too worried about this model taking over the world anytime soon.)
The basic skeleton of a neural network is a series of layers of nodes. In the example below, we have one input layer of size 3, three hidden layers each of size 4, and one output layer of size 2.



Each node receives information from the nodes of the previous layer, manipulates that data in some way and then passes that information on to the next layer of nodes. Using a loss function and back propagation the neural network is able to adjust the way each node manipulates the data so that the neural network can improve its prediction.

What is a convolutional neural network?

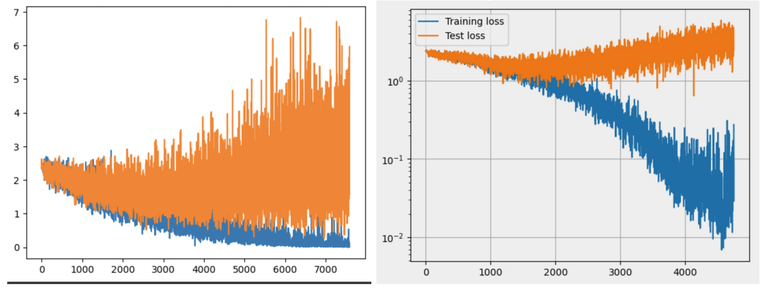
A neural network with convolutional and pooling layers. The purpose of a convolutional layer is to extract features. We take segments of the matrix at a time and perform some kind of operation on just those entries. To illustrate, a basic example would be looking just at the top left 2 by 2 matrix in the image, taking a linear combination of those values and storing that in a new matrix. Then shifting by one column to the right and repeating this procedure until you repeat that on every submatrix of size 2 by 2 of that image.
The purpose of a pooling layer is to reduce the size of the image and as a result summarize the data a little bit. An example would be similar to the convolutional layer example above but instead of taking a linear combination, simply taking the average or the maximum of the values in those submatrices.
Note that there are weights in the convolutional layer that are adjusted throughout the training process while there are no weights in the pooling layer.

Structure of our neural network

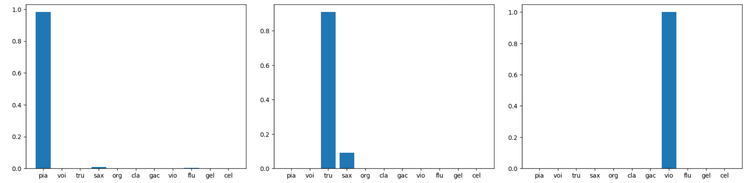
Our CNN is made up of four convolutional layers all with kernel size 3, the first taking in an image with one channel and outputting an image with 16 channels, the second taking in an image with 16 channels and outputting an image with 32 channels. The third convolutional layer output an image with 64 channels and the last layer outputs an image with 128 channels. In between each pair of convolutional layers, we have a ReLU layer (a type of linear activation function) and a pooling layer with kernel size 2. The final layer in the CNN is to unravel the image into a one dimensional array to feed through a linear layer. First iteration of model was based on example by Syed Abdul Gaffar Shakhadri [3]

Plot of the Loss

The left graph below shows the result of training the model using batch size 16. This caused the huge variance in test loss and the adjustment for this problem was setting the batch size to 128. The right graph below shows the result of overfitting. We had hoped that the issue of overfitting would disappear once we started using the entire dataset. Unfortunately it did not so the next steps we took were to introduce dropout layers after every pooling layer which randomly chooses channels to zero out.



To better illustrate the predictions of our model, below are three probability plots. We randomly chose three data points from the testing dataset. The first data audio file had piano, the second had a trumpet, and the third had a violin, had the model predict which instrument it believed was predominant and scaled the prediction array to get a probability distribution. As we can see, the model did pretty well.



Next Steps

Moving forward, I would be interested in using a dataset of classical music and starting with this pretrained model that is already familiar with music files to see how well a CNN could predict the composer based off a few second clip of a composition. And from there we could start thinking about building a generative music AI that could take the name of a famous composer and create a few second audio clip of a new composition.

Acknowledgements

I would like to thank my mentor Alan Raydan for his incredible support and mentorship (and his computational power and debugging skills) throughout this program. I would also like to thank the organizers of the of the 2024 Directed Reading Program for this opportunity.

References

[1] Ferdinand Fuhrmann. ""A Comparison of Sound Segregation Techniques for Predominant Instrument Recognition in Musical Audio Signals". PhD thesis. Universitat Pompeu Fabra, 2012.

[2] Leland Roberts. "Understanding the Mel Spectrogram". In: *Medium* (2020).

[3] Syed Abdul Gaffar Shakhadri. "Guide to Audio Classification Using Deep Learning". In: *Analytics Vidhya* (2023).

Word Theory Applied to Musical Scales

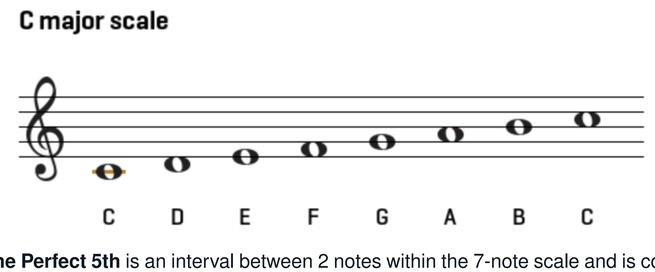
Jackson Rockmael and Tim Guan
University of California Santa Barbara DRP

A Brief History of the Ionian Scale

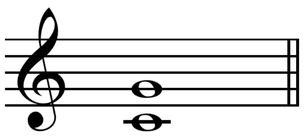
The Ionian scale, known today as the major scale, originated in ancient Greek music and became a cornerstone of Western music due to its bright and harmonious sound. Today, the major scale remains fundamental in classical, pop, rock, and jazz music, underscoring its timeless and universal appeal. Out of all the 792 seven-note scales, we set out to find what makes Ionian unique.

Fundamentals of Music Theory

What is a Scale? A musical scale is a sequence of notes arranged in ascending descending order of pitch within an octave. Scales form the foundation of musical compositions and are used to define the tonality of a piece of music. Each scale is made up of notes that follow a specific pattern of intervals, which are the tonal gaps between the notes.



The Perfect 5th is an interval between 2 notes within the 7-note scale and is considered highly consonant, meaning it sounds stable and pleasant to the ear. This consonance is due to the simple frequency ratio of 3:2 between the two notes, which creates a sense of balance and resolution, making it the most important interval in music. For example, G is the perfect 5th of C.



When looking for the scale that sounds the best, we turn our attention to a scale that has the maximum amount of perfect 5th. This narrows down our search to 7 scales commonly known as, Ionian, Dorian, Phrygian, Lydian, Mixolydian, Aeolian, and Locrian; referred to as the *modes*.

Monoids

The **free monoid** generated by $\{x, y\}$, denoted as $\{x, y\}^*$, is the set of possible finite sequences of x and y and these finite sequences are called strings. $\{x, y\}^*$ is closed under the concatenation of strings and contains the identity, the empty string.
We can represent each musical mode as a sequence of intervals, with each interval either ascending by one or two notes. For example, the Ionian mode can be represented as $'aabaaab'$ where a denotes ascent of two notes and b denotes an ascent of one note. When expressed in this way, each mode is simply a rotation of the others in the sequence.

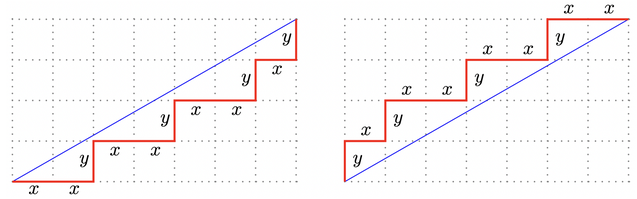
Christoffel Words

The notation $a \perp b$ refers to a and b being relatively prime. Suppose $a, b \in N$ and $a \perp b$. The lower Christoffel path of slope $\frac{b}{a}$ is the path from $(0, 0)$ to (a, b) in the integer lattice $\mathbf{Z} \times \mathbf{Z}$ that satisfies the following two conditions.

(i) The path lies below the line segment that begins at the origin and ends at (a, b).

(ii) The region in the plane enclosed by the path and the line segment contains no other points of $\mathbf{Z} \times \mathbf{Z}$ besides those of the path.

The following diagram represents the upper and lower Christoffel words of slope $\frac{4}{7}$ which are $'yxyxyxyxyxx'$ and $'xxyxyxyxyxy'$ making them strings in $\{x, y\}$.



Definition: Two elements w and w' of $\{a, b\}^*$ are *conjugate* if and only if there exist words u and v such that $w = uv$ and $w' = vu$.

In the illustration below, our words (or scales) are just rotations of each other. This is the case for all conjugates in the free monoid $\{a, b\}^*$. Moreover, in the free group, $\{a, b\}$, the set of all reduced words on the alphabet $\{a,\ b,\ a^{-1},\ b^{-1}\}$ and the inverse of any word is constructed by taking the reverse spelling and inverting each element. For example, $(aab)^{-1} = b^{-1}a^{-1}a^{-1}$.

Conjugation on Lydian	Result	Mode Name
$b \circ aaabaab \circ b^{-1}$	$baaabaa$	Phrygian
$ab \circ aaabaab \circ b^{-1}a^{-1}$	$abaaaba$	Dorian
$aab \circ aaabaab \circ b^{-1}a^{-1}a^{-1}$	$aabaaab$	Ionian
$baab \circ aabaaab \circ b^{-1}a^{-1}a^{-1}b^{-1}$	$baabaaa$	Locrian
$abaab \circ aabaaab \circ b^{-1}a^{-1}a^{-1}b^{-1}a^{-1}$	$abaabaa$	Aeolian
$aabaab \circ aabaaab \circ b^{-1}a^{-1}a^{-1}a^{-1}b^{-1}a^{-1}$	$aabaaba$	Mixolydian
$aaabaab \circ aabaaab \circ b^{-1}a^{-1}a^{-1}a^{-1}b^{-1}a^{-1}a^{-1}$	$aaabaab$	Lydian

Sturmian Morphisms

Definition: A **Sturmian Morphism** is a monoid homomorphism $\{a, b\}^* \to \{a, b\}^*$ that sends every Christoffel word to a conjugate of a Christoffel word. The set of Sturmian morphisms forms a monoid under function composition. We denote the monoid of Sturmian morphisms by St.

If $A \in St$ and $z_1 z_2 ... z_n \in \{a, b\}^*$, then $A(z_1 z_2 ... z_n) = A(z_1)A(z_2)...A(z_n)$ such that any Sturmian Morphism of $\{a, b\}$ is determined by the images of a and b so we identify A with the ordered pair $(A(a), A(b))$.

Generation of Scales: Using the notation above the monoid, St, of Sturmian morphisms is generated by the following Sturmian Morphisms.

$$G = (a, ab), \tilde{G} = (a, ba), D = (ba, b), \tilde{D} = (ab, b), E = (b, a)$$

Special Sturmian Morphisms

An important sub-monoid of this, St_0, is the monoid generated by **G, $\tilde{\text{G}}$, D and $\tilde{\text{G}}$**(note the absence of ***E***). St_0 is called the moniod of *special Sturmian Morphisms*, and these play a distinguished role in the *Divider Incidence Theorem*.

Divider Incidence Theorem: In the conjugacy class of a Christoffel word of length n, there are $n - 1$ words that can be obtained as images $f(ab) = f(a)(b) = f(a)f(b)$ of the initial word ab where $f \in St_0$. We separate this word $f(ab)$ into factors giving us a divided word $f(a)|f(b)$.

Mode	Sturmian Representation on (ab)
Ionian	$GGD(ab) = G\tilde{G}D(a)(b) = (aaba)\|(aab)$
Dorian	$G\tilde{G}\tilde{G}D(ab) = G\tilde{G}\tilde{D}(a)(b) = (abaa)\|(aba)$
Phrygian	$\tilde{G}\tilde{G}\tilde{G}D(ab) = \tilde{G}\tilde{G}\tilde{D}(a)(b) = (baaa)\|(baa)$
Lydian	$GG\tilde{D}(ab) = GG\tilde{D}(a)(b) = (aaab)\|(aab)$
Mixolydian	$G\tilde{G}\tilde{D}(ab) = G\tilde{G}\tilde{D}(a)(b) = (aaba)\|(aba)$
Aeolian	$\tilde{G}\tilde{G}\tilde{D}(ab) = \tilde{G}\tilde{G}\tilde{D}(a)(b) = (abaa)\|(baa)$

Conclusion

The following table gives the six possible diatonic words (or scales) that can be obtained through our *special Sturmian Morphisms.* We should notice that there is one conjugate missing from this list, which is the *Locrian* mode, otherwise known as **'E'**, represented by $'(baab)|(aaa)'$. This is the only conjugate that cannot be generated by $f(ab)$ with $f \in St_0$, which we call the *'bad conjugate.'*

Out of the generators of St_0 we can determine that D and G best preserve the scale integrity. So when choosing a scale out of the 6 modes that are generated by St_0, we would choose the scale generated by D and G, which is Ionian. So, through Christoffel words and Sturmian Morphisms, we can fully characterize Ionian as the universally best 7-note scale.

Acknowledgements

We thank Alexander Sabater for his guidance as well as the UCSB Directed Reading Program and Math Department for the opportunity to work on this project.

References

(1) Berstel, Jean, et al. "Codes and Automata." Université Gustave Eiffel, 2008, www-igm.univ-mlv.fr/ berstel/Articles/2008wordsbookMtlUltimate.pdf.

(2) Fiore, Thomas M. "Music and Mathematics: From Pythagoras to Fractals." University of Michigan-Dearborn, 2006, https://www-personal.umd.umich.edu/ tmfiore/1/musictotal.pdf.

(3) Trotter, Matthias. "Group Theory in Cryptography." University of Chicago, 2009, www.math.uchicago.edu/may/VIGRE/VIGRE2009/REUPapers/Trotter

2D Navier-Stokes Equations: Existence and Uniqueness of Solutions

Jack Wan
University of California Santa Barbara

Derivation of the Model

Let $\Omega \subset \mathbb{R}^d$ be a domain occupied by a viscous fluid. Denote by $X(\cdot,t) : \Omega \to \Omega$, $a \mapsto X(a,t)$, the flow map; $u(x,t) = (u_1(x,t),\dots,u_d(x,t))$ the velocity field; $\rho(\cdot,t)$ the density scalar function; V a volume element in the fluid.

Fluid assumptions: incompressible ($\nabla \cdot u = 0$); homogeneous (ρ is constant in space).

Physical Laws:

- Conservation of mass: ρ is constant in time, so ρ is just a constant ρ_0.
- Newton's second law of motion:

$$\int_{X(V,t)} \rho_0(\partial_t u + u \cdot \nabla u)dx = \int_{X(V,t)} -\nabla p + \nu \Delta u dx$$

where p is the pressure of the fluid, ν is the dynamic viscosity and n is the outward unit normal to $\partial X(V,t)$

Thus, we have

$$\partial_t u + u \cdot \nabla u + \frac{1}{\rho_0}\nabla p - \frac{\nu}{\rho_0}\Delta u = 0.$$

We address the existence and uniqueness of solutions on $\mathbb{T}^2 = [0, 2\pi]^2$ with periodic boundary conditions. We take $\rho_0 = \nu = 1$ as these constants don't contribute to the analysis of the problem.

Weak Solutions

We say that u is a weak solution to the 2D periodic Navier-Stokes equations if

$$\int_{[0,\infty)\times\mathbb{T}^2} (u \cdot \partial_t \phi + u \cdot \Delta\phi + u \cdot (u \cdot \nabla\phi))dxdt = \int_{\mathbb{T}^2} u(0,x) \cdot \phi(0,x)dx$$

for all $\phi \in C_0^\infty([0,\infty) \times \mathbb{T}^2)$ such that $\nabla \cdot \phi = 0$, and

$$\int_{\mathbb{T}^2} u \cdot \nabla\phi dx = 0$$

for all $\phi \in C_0^\infty(\mathbb{T}^2)$.

Main Theorem

Let $u_0 \in L^2$ be divergence-free and mean-free. The 2D periodic Navier-Stokes equation has a unique global weak solution with regularity $u \in L^\infty(0,\infty; L^2(\mathbb{T}^2)) \cap L^2(0,\infty; H^1(\mathbb{T}^2))$.

Proof of the Existence of Solutions

The proof is divided into several steps:

Step 1. Consider the mollified Navier-Stokes equations

$$\partial_t u^\varepsilon + \mathcal{J}_\varepsilon u^\varepsilon \cdot \nabla u^\varepsilon - \Delta u^\varepsilon + \nabla p^\epsilon = 0$$

$$u^\varepsilon(x,0) = \mathcal{J}_\varepsilon u_0, \nabla \cdot u^\varepsilon = 0,$$

where $\mathcal{J}_\varepsilon$ is a standard mollification operator. For each $\varepsilon \in (0,1)$, this mollified system has a global unique smooth solution. This follows from the contraction mapping principle.

Step 2. In view of the divergence-free condition obeyed by $\mathcal{J}_\varepsilon u^\varepsilon$, the following cancellation law

$$\int_{\mathbb{T}^2} \mathcal{J}_\varepsilon u^\varepsilon \cdot \nabla u^\varepsilon \cdot u^\varepsilon dx = 0$$

holds, and consequently, u^ε obeys the energy equality

$$\frac{1}{2}\frac{d}{dt}\|u^\varepsilon\|_{L^2}^2 + \|\nabla u^\varepsilon\|_{L^2}^2 = 0.$$

Integrating in time from 0 to t, we obtain

$$\|u^\varepsilon(t)\|_{L^2}^2 + 2\int_0^t \|\nabla u^\varepsilon(s)\|_{L^2}^2 ds \le \|u_0\|_{L^2}^2$$

and deduce that $\{u^\varepsilon\}_{\varepsilon\in(0,1)}$ is uniformly bounded in $L^2(0,\infty; H^1(\mathbb{T}^2))$ and $L^\infty(0,\infty; L^2(\mathbb{T}^2))$.

Proof of the Existence of Solutions

Step 3. We denote by $\mathbb{P}$ the Leray-Hodge projector, which is the orthogonal projection from L^2 onto the closed subset of divergence-free vector fields in L^2. In view of the boundedness of $\mathbb{P}$ on L^2 and the Ladyzhenskaya interpolation inequality, we have

$$\left|\int_{\mathbb{T}^2} (\mathbb{P}(\mathcal{J}_\varepsilon u^\varepsilon \cdot \nabla u^\varepsilon) - \Delta u^\varepsilon)\,\phi dx\right| \le \left(\|u^\varepsilon\|_{L^4}^2 + \|\nabla u^\varepsilon\|_{L^2}\right)\|\nabla\phi\|_{L^2} \le C\left(\|u^\varepsilon\|_{L^2}\|\nabla u^\varepsilon\|_{L^2} + \|\nabla u^\varepsilon\|_{L^2}\right)$$

for all $\phi \in H^1(\mathbb{T}^2)$ with $\|\phi\|_{H^1} \le 1$. Consequently,

$$\int_0^T \|\partial_t u^\varepsilon(t)\|_{H^{-1}}^2 dt \le C\int_0^T \left(\|u^\varepsilon(t)\|_{L^2}^2\|\nabla u^\varepsilon(t)\|_{L^2}^2 + \|\nabla u^\varepsilon(t)\|_{L^2}^2\right)dt \le C_0$$

where C_0 is a constant depending on the size of the initial velocity in $L^2(\mathbb{T}^2)$. Therefore, the family $\{\partial_t u^\varepsilon\}_{\varepsilon>0}$ is uniformly bounded in $L^2(0,T; H^{-1}(\mathbb{T}^2))$ for any $T > 0$.

Step 4. Since

$$H^1(\mathbb{T}^2) \underset{compact}{\subseteq} L^2(\mathbb{T}^2) \underset{continous}{\subseteq} H^{-1}(\mathbb{T}^2),$$

the Aubin-Lions lemma implies the existence of a subsequence that converges strongly in $L^2(0,T; L^2(\mathbb{T}^2))$ to a weak solution u of the periodic Navier-Stokes equations. The regularity of weak solutions follows from the Banach Alaoglu theorem.

Proof of the Uniqueness of Solutions

Let u and v be two weak solutions of the Navier-Stokes equation with initial data $u_0 = v_0$.
The difference $\omega = u - v$ obeys

$$\partial_t \omega - \Delta\omega + u \cdot \nabla\omega + \omega \cdot \nabla v + \nabla q = 0.$$

We multiply by ω, integrate over $\mathbb{T}^2$, estimate using Ladyzhenskaya's interpolation inequality, and obtain

$$\frac{1}{2}\frac{d}{dt}\|\omega\|_{L^2}^2 + \|\nabla\omega\|_{L^2}^2 \le C\|\omega\|_{L^4}^2\|\nabla v\|_{L^2} \le C\|\omega\|_{L^2}\|\nabla\omega\|_{L^2}\|\nabla v\|_{L^2} \le \frac{1}{2}\|\nabla\omega\|_{L^2}^2 + c\|\nabla v\|_{L^2}^2\|\omega\|_{L^2}^2.$$

Consequently, we infer that

$$\frac{d}{dt}\|\omega\|_{L^2}^2 \le c\|\nabla v\|_{L^2}^2\|\omega\|_{L^2}^2.$$

As $v \in L^2(0,\infty; H^1(\mathbb{T}^2))$, it follows from Gronwall's inequality that $\|\omega(t)\|_{L^2} = 0$. Therefore $\omega(t) = 0$ for a.e. $x \in \mathbb{T}^2$. This proves the uniqueness of solutions.

3D Problem

In the three-dimensional case, the global well-posedness is an open problem, called "the Navier-Stokes existence and smoothness" problem. It is a Millennium Problem selected by the Clay Mathematics Institute of Cambridge for an award that is worth 1 million dollars!

References

Jacob Bedrossian, Vlad Vicol. The Mathematical Analysis of the Incompressible Euler and Navier-Stokes Equations: An Introduction. AMS Graduate Studies in Mathematics 225 (2022).

Acknowledgments

I would like to thank my mentor Elie Abdo for his support and guidance and the Directed Reading Program at UCSB that made this project possible.

A Brief Introduction to Algebraic Topology

Darige Xu

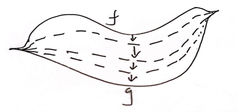
University of California Santa Barbara

Introduction

Algebraic topology is the study of topological spaces by using the tools from algebra. People are interested in finding algebraic invariants of topological spaces in order to classify them. Algebraic topology usually splits into two parts: Homotopy and Homology.

Concepts in Homotopy

Homotopy could be thought as a "path" between continuous functions. Let's say there are two paths f, g in a topological space X. We want to find a way to continuously deform one path to the other. If such a way exists, then we are going to say those two functions are homotopic. Moreover, if a space could continuously deform into another space, we will say two spaces are homotopy equivalent or have the same homotopy type [1].



Here is an example: If f, g are two loops that start and end at the same points in $\mathbb{R}^n$, we find that f could continuously deform through the dotted lines and finally become g.

Definition: Let $f, g : X \to Y$ be two continuous functions between two topological spaces. Then a homotopy from f to g is a continuous function

$$\begin{aligned} F : X \times [0,1] &\to Y \\ (x, 0) &\mapsto f(x) \\ (x, 1) &\mapsto g(x) \end{aligned}$$

for all $x \in X$.

Based on this definition, we can see that homotopy could build an equivalent class for a path at a fixed point.

We can also define an operation between loops. If f, g are two loops in topological space X such that g starts at the ending point of f, ie $f(1) = g(0)$, then we are going to say f compose g, or $f \cdot g$, to be a new loop that starts at $f(0)$, go through $f(1),\ g(0)$, and ends at $g(1)$.

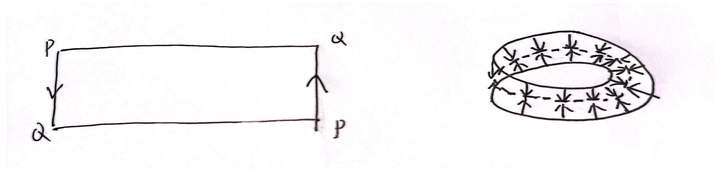
By those two definitions, we can say that all the homotopy classes of the loops in X at a fixed based point x_0 is going to build a group. We call it a Fundamental group.

Fundamental Groups:

$$\pi_1(X, x_0) = \{\text{Loops in } X \text{ based at } x_0\}/\text{Homotopy betweeen loops}$$

let's see some examples:

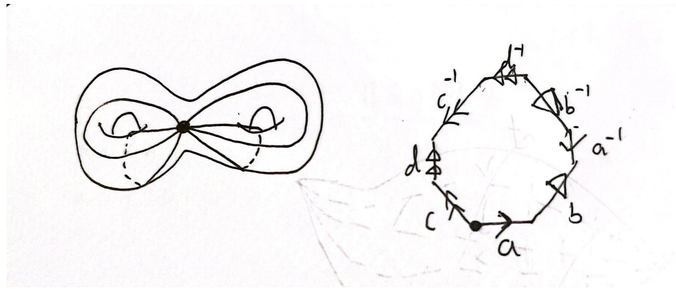
- Points and Straight lines: they have trivial fundamental groups since those spaces are simple-connected.
- Circle: $\pi_1(S^1, x_0)$ is $\mathbb{Z}$. the elements of $\pi_1(S^1)$ are just loops that go around n times, $n \in \mathbb{Z}$
- Sphere: its fundamental group is also trivial since it is simple-connected.
- Möbius strip: its fundamental group is also $\mathbb{Z}$, as Möbius strip is homotopy-equivalent to a circle.



the left side is the cell-complex for the Möbius strip, and the right side explains why the Möbius strip is homotopy-equivalent to the circle

Another less trivial example is the Genus-two surface whose fundamental group has the following presentation.

$$\pi_1(\Sigma_2, x_0) \cong \langle a, b, c, d \mid [a, b][c, d] = 1 \rangle$$



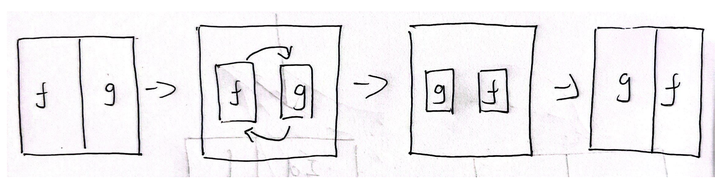
The left side is a picture of the Genus-two surface, and the right side is its cell-complex picture. (If you glue each side of the octagon with the corresponding letter then you will get the left picture)

Higher Homotopy Groups

Fundamental groups are usually not abelian, but their higher dimensional analogs, which we call higher homotopy groups, are abelian.

The idea here is to generalize the one-dimension loop into higher-dimension. For a n^{th} sphere S^n and topological space X, we say that $f : S^n \to S$ to be, informally, an "n^{th} dimension loop". And then for all the "n^{th} dimension loop" at a fixed point, we can get $\pi_n(X, x_0)$ by quotient it under homotopy between loops.

After giving the definition of homotopy groups, we can go through why for $n > 1$, homotopy groups are abelian. Here is a visual proof:

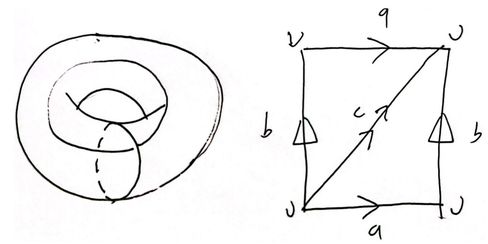


f and g here represent the two S^n that maps to X. The boundary of those S^n is mapped to the base point, x_0, in X. So we could "shrink down" f and g to have enough space to move it around without changing the connections of the base point. However, the same trick could not work for one dimension, as we cannot move two intervals around.

Concepts in Homology

Another algebraic invariant is homology. Most spaces can be decomposed into simpler objects called simplices and the way they fit together tells us a lot about the topology of the space. For example, the torus can be decomposed into two triangles as seen below.

Example: Torus:



We will not give the formal definition, but by looking at the way the 0-, 1- and 2-dimensional simplices fit together, we can write down the following groups:

$$H_0(T) = \mathbb{Z},\ H_1(T) = \mathbb{Z} \oplus \mathbb{Z},\ H_2(X) = \mathbb{Z}$$

One important distinction here is that homology groups are always abelian unlike fundamental groups.

Hurewicz Theorem

As we define the homotopy groups and homology groups, we will naturally wonder if there are connections between them. Hurewicz Theorem tells us that we can sometimes find an isomorphism between them.

Let X be a pointed path-connected topological space, and $\pi_n(X)$, $H_n(X)$ be the n^{th} homotopy group and homology group.

If $n = 1$, then H_1 is isomorphic to the abelianization of π_1. Here abelianization of π_1 means that we are going to quotient all the non-abelian parts. Formally, $\pi_{ab}(X) := \pi_1(x)/[\pi_1(x), \pi_1(x)]$, where $[\pi_1(x), \pi_1(x)]$ is a subgroup of $\pi_1(X)$ and generated by all the elements of the form $aba^{-1}b^{-1}$ in $\pi_1(X)$.

Similar things happen for higher homotopy and homology groups (with some conditions) and that is the statement of Hurewicz theorem.

Acknowledgements

Thanks for the guidance from Benedict as well as the UCSB Directed Reading Program for the opportunity to work on this project.

References

[1] Allen Hatcher. *Algebraic topology*. Cambridge University Press, 2005.

A Brief Introduction to Markov Chains and PageRank

Kathryn Lyu Mentor: Alex Bisnath

University of California Santa Barbara - Directed Reading Program 2024

Intro to Markov Chains

In the discrete state space $\mathbb{S}$, we say the process $(Z_n)_{n\in\mathbb{N}}$ is Markov that $\forall n \geq 1$ the probability distribution of Z_{n+1} is determined by the state Z_n of the process at time n. In simpler terms, we say that the future, given the present, is independent of the past. Since the probability does not depend on n, A Markov process is time homogeneous that is to say:

$$\mathbb{P}(X_{n+1} = j|X_n = i) = \mathbb{P}(X_1 = j|X_0 = i)$$

Also, $\mathbb{P}(X_{n+1} = j|X_0 = x_0, X_1 = x_1 \dots X_n = i) = \mathbb{P}(X_{n+1} = j|X_n = i)$. Therefore, for each step, we just need to consider the previous state.
It is easier to find the probability that the system moves from state i to state j by finding the corresponding entry. The row corresponds to the current state and the column corresponds to the next state. Notice that the sum of each row should be 1.

$$[P_{i,j}]_{0\leq i,j\leq N} = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \dots & P_{0,N} \\ P_{1,0} & P_{1,1} & P_{1,2} & \dots & P_{1,N} \\ P_{2,0} & P_{2,1} & P_{2,2} & \dots & P_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{N,0} & P_{N,1} & P_{N,2} & \dots & P_{N,N} \end{bmatrix}$$

In order to investigate the relation between the states, we define the **Hitting time** $T_A = min\{n > 0, S_n = A\}$ which corresponds to the first time the process will reach or "hit" a state in the subset A, and the probability of hitting the set through state $l \in A$ from $k \in \mathbb{S}$ is $g_l(k) = \sum_{m\in\mathbb{S}} P_{k,m}g_l(m)$. Therefore, if the state j is absorbing, we have that $\mathbb{P}(T_j < \infty|Z_0 = i) = \lim_{n\to\infty} \mathbb{P}(Z_n = j|Z_0 = i)$. We can also define the **Return Probability**, which is the probability of returning to state j in finite time starting from state i is

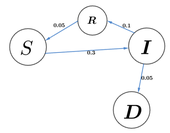
$$p_{ij} = \mathbb{P}(T_j^r < \infty|X_0 = i) = \mathbb{P}(X_n = j(n \geq 1)|X_0 = i)$$

. With this definition, we can introduce the notion of recurrent states. :
State j is said to be **recurrent**: if $P_j(T_j < +\infty) = 1$, which implies that the chain will return to j eventually . Also, $\sum_{n=0}^{+\infty} P_{j,j}^n = +\infty$
State j is said to be **transient** if $P_j(T_j < +\infty) < 1$, which implies that the chain will not return back and $\sum_{n=0}^{+\infty} P_{j,j}^n < +\infty$). Now, we can explore more about the relations between states:
The state j is **accessible** from state i if $P_{i,j}^n > 0$.
When both $P_{i,j}^n > 0$ and $P_{j,i}^n > 0$, we say the states i, j **communicate**.
The sets of states which communicate with each other partition the set of states, if this partition only contains one subdivision, the chain is called **irreducible**, and all states communicate with each other.
Recurrence and transience are **class properties**: all the states in a communication class are either recurrent or transient.

We define the **period** of a state i as $d(i) = gcd\{n > 0, P_{ii}^n > 0\}$. If $d(i) = 1$, the state i is said to be **aperiod**.
Since for an absorbing state, $P_{ii} = 1$, an absorbing state is aperiodic and recurrent.
A recurrent state is said to be *ergodic*, if it is both *positive recurrent (irrducible)* and *aperiodic*. If $[P^n]_{i,i} = 0$ for all $n \geq 1$, then the period of state i is 0, and it is transient.
Note that that if the sequence of n contains two distinct numbers that are relatively prime to each other, the state is aperiodic.
Periodicity is also a **class property**: all states in a given communicating class have the same periodicity (periodic versus aperiodic), if they are periodic, they will all have the same period.

Also, A Markov Chain is aperiodic when all of its states are aperiodic.

SIRD Model



S, R, I, D stand for Susceptible, Infected, Recovered, and Deceased states respectively.
Hitting time: $T_S = 3$ as the minimum step is through $S \to I \to R \to S$.
Absorbing: Since D is isolated, that all $P_{DD} = 1$, the state is absorbing.
Recurrent transient states: S, R, I are all recurrent
Since S, R, I are accessible from each other, we say that all of them communicate with each other and they form a communication class.

Introduction of PageRank

The PageRank algorithm is based on using Markov Chains to analyze the web pages and measure the importance of website pages, it is used by search engines like Google. For each page, its PageRank is the sum of all the PageRank it receives from pages linking to it. The more incoming links a page has, the more important the page is, and back links from more important pages carry more weight than back links from less important pages. In Markov Chain, each page is a state in the chain and links between pages represent transitions from one state to another, with transition probabilities that are typically uniform across all outbound links from a given page.

Stationary Distribution

The ***probability distribution*** π on $\mathbb{S}$ with transition matrix P is said to be *stationary* if and only if the vector π is invariant by the matrix P, which implies it remains unchanged in the Markov chain as time progresses, that

$$\pi = \pi P$$

Then, for $\mathbb{S} = \{0, 1 \dots N\}$ is finite and $\pi_{ij} := \lim_{n\to\infty} \mathbb{P}(X_n = j|X_0 = i)$ exists for all $i, j \in \mathbb{S}$, we have

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \pi_{0,0} & \pi_{0,1} & \dots & \pi_{0,N} \\ \pi_{1,0} & \pi_{1,1} & \dots & \pi_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{N,0} & \pi_{N,1} & \dots & P_{N,N} \end{bmatrix}$$

and therefore

$$\pi_i = \sum_{j=1}^{N} \pi_{ij} P$$

For the Markov Chain that is **ergodic**, the chain $(X_n)_{n\in\mathbb{N}}$ holds the limiting distribution

$$\pi_i := \lim_{n\to\infty} \mathbb{P}(X_n = i|X_0 = j) = \lim_{n\to\infty} [P^n]_{j.i} = \frac{1}{\mu_i(i)} = \frac{1}{\mathbb{E}_i(T_i)}$$

Perron-Frobenius Theorem
If A is a positive stochastic matrix, then the eigenvalues satisfy $\lambda_1 = 1$ and $|\lambda_j| < 1$, for $j > 1$. This means that A has a unique positive, steady-state vector q and that every Markov chain defined by A will converge to q.
Markov chains and the Perron-Frobenius theorem are the central ingredients in Google's PageRank algorithm.

PageRank

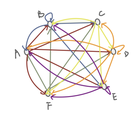
The existence of stationary distributions is key to assigning a rank to a web page First, we can create the transition matrix is formed where each entry P_{ij} represents the probability of moving from page i to page j. Then, with the property of Markov Chain, $x_{k+1} = Px_k$, where x is the PageRank vector that represents the probability of each state in step k. However, there is no guarantee that every page rank process will converge to a stationary distribution. With the previous definition, we know that this can be fixed by making the Markov chain ergodic, and one way to make a Markov chain ergodic is to insert an edge between every two nodes.
Also, the **Perron-Frobenius theorem** implies that a Markov chain $x_{k+1} = Px_k$ converges to a unique steady-state vector when the matrix P is positive. This theorem provides the idea to obtain a new positive matrix P' by imagining that users randomly jump to any other page with a small probability, making sure that all states are connected. Assume that there exists a matrix R_n such that

$$R_n = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \vdots & \ddots & \\ \frac{1}{n} & \frac{1}{n} & \dots & \frac{1}{n} \end{bmatrix}$$

Then we create a new **Google Matrix** P' by mixing P and R. Choosing an appropriate parameter α, we set $P' = \alpha P + (1 - \alpha)R_n$, then P' is a positive stochastic matrix. While the R-component in M ensures that the Markov chain converges, it also changes the stationary distribution. To ensure the impact is not too large, α should be chosen close to 1. A typical value for α is 0.85.

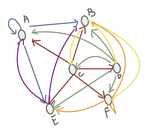
Example of PageRank

There are two models for PageRank.
Example 1: Most website are similar
In this case, all nodes have 4-5 outgoing links.



$$P_{ij} = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & \frac{1}{4} \\ \frac{1}{6} & 0 & \frac{1}{12} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & \frac{1}{8} & 0 & \frac{1}{8} & \frac{1}{2} & \frac{3}{16} \\ \frac{1}{5} & 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{2}{5} \\ \frac{1}{12} & \frac{1}{4} & \frac{1}{4} & \frac{5}{12} & 0 & 0 \end{bmatrix}$$

Then, we can find the stationary vector $[0.15, 0.165, 0.14, 0.13, 0.26, 0.15]$
Example 2: One or two websites are 'hubs'
There are two 'hubs' that has outgoing links with remaining nodes.



$$P_{ij} = \begin{bmatrix} \frac{1}{6} & \frac{2}{3} & 0 & 0 & \frac{1}{6} & 0 \\ 0 & 0 & \frac{1}{12} & \frac{1}{4} & 0 & \frac{2}{3} \\ \frac{1}{5} & \frac{1}{10} & 0 & \frac{1}{5} & \frac{3}{10} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{10} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

Then, we can find the stationary vector $[0.17, 0.27, 0.04, 0.08, 0.15, 0.29]$

References

[1] David Austin. *Markov chains and Google's PageRank algorithm*. URL: https://math.libretexts.org/Bookshelves/Linear_Algebra/Understanding_Linear_Algebra_(Austin)/04%3A_Eigenvalues_and_eigenvectors/4.05%3A_Markov_chains_and_Google's_PageRank_algorithm.

[2] Nicolas Privault. *Understanding Markov Chains*. Springer Singapore, 2018.

A Combinatorial Approach to The Free Central Limit Theorem

Jeremy Li, mentored by Quinn Kolt

Department of Mathematics, University of California Santa Barbara (DRP 2024)

Abstract

Free probability theory is a relatively new area of study, bringing together many different fields of mathematics, such as operator algebras, random matrices, combinatorics, and representation theory of symmetric groups. In this poster, we will explore the basic groundwork for free probability and provide a combinatorial proof of the free central limit theorem.

Background

$*$-Probability Space $(\mathcal{A}, \varphi)$

A $*$-probability space consists of a unital $*$-algebra $\mathcal{A}$ and an expectation $\varphi : \mathcal{A} \to \mathbb{C}$.

A **unital $*$-algebra** $\mathcal{A}$ is a vector space over $\mathbb{C}$ equipped multiplication that is associative but not necessarily commutative, a multiplicative identity $1_{\mathcal{A}}$, and an adjoint operation $*$ (e.g., adjoint of linear maps, conjugate transpose of matrices).

- Elements $a \in \mathcal{A}$ are called **(non-commutative) random variables**.
- $\mathcal{A}_0$ is called a **subalgebra** of $\mathcal{A}$ if $\mathcal{A}_0 \subseteq \mathcal{A}$ and it is itself a unital $*$-algebra.
- The expectation φ is a **linear functional** with $\varphi(1_{\mathcal{A}}) = 1$ and $\varphi(a^*a) \geq 0, \forall a \in \mathcal{A}$.

Classical Independence

Let $(\mathcal{A}, \varphi)$ be a $*$-probability space. Elements $a, b \in \mathcal{A}$ are **(classically) independent** if $ab = ba$ and $\varphi(a^n b^m) = \varphi(a^n)\varphi(b^m)$ for all $n, m \in \mathbb{N}$.

Free Independence

Let $(\mathcal{A}, \varphi)$ be a $*$-probability space and I be a fixed index set. For each $i \in I$, let $\mathcal{A}_i \subseteq \mathcal{A}$ be a subalgebra. The subalgebras $(\mathcal{A}_i)_{i \in I}$ are **freely independent (FI)** if

$$\varphi(a_1 a_2 \dots a_k) = 0$$

whenever we have the following:

- $a_j \in \mathcal{A}_{i(j)}$, with $i(j) \in I$, for all $j = 1, 2, \dots, k$ and $k \in \mathbb{N}$.
- $\varphi(a_j) = 0$ for all $j = 1, 2, \dots, k$.
- Neighboring elements are from different subalgebras, i.e., $i(1) \neq i(2)$, ..., $i(k-1) \neq i(k)$.

Elements $a, b \in \mathcal{A}$ are **freely independent** if the subalgebras they generate are FI.

Convergence in Distribution

Let $(\mathcal{A}_N, \varphi_N)_{N \in \mathbb{N}}$ and $(\mathcal{A}, \varphi)$ be $*$-probability spaces. Consider $a_N \in \mathcal{A}_N$ for each $N \in \mathbb{N}$ and $a \in \mathcal{A}$. We say that a_N **converges in distribution** to a as $N \to \infty$ and denote this by $a_N \xrightarrow{\text{distr}} a$, if we have convergence of all moments: $\lim_{N \to \infty} \varphi_N(a_N^n) = \varphi(a^n)$ for all $n \in \mathbb{N}$

The Classical Central Limit Theorem

Theorem 1. (Classical Central Limit Theorem) Let $(\mathcal{A}, \varphi)$ be a $*$-probability space and $a_1, a_2, \dots \in \mathcal{A}$ be a sequence of independent and identically distributed self-adjoint random variables. Assume all random variables are centered: $\varphi(a_r) = 0, \forall r \in \mathbb{N}$ and denote by $\sigma^2 := \varphi(a_r^2)$ the common variance of the random variables. Then, we have

$$\frac{a_1 + \ldots + a_N}{\sqrt{N}} \xrightarrow{\text{distr}} x,$$

where x is a normal random variable with mean 0 and variance σ^2.

Remark. This statement means explicitly

$$\lim_{N \to \infty} \varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} t^n e^{-\frac{t^2}{2\sigma^2}} dt, \qquad \forall n \in \mathbb{N}.$$

The Free Central Limit Theorem

Theorem 2. (Free Central Limit Theorem) Let $(\mathcal{A}, \varphi)$ be a $*$-probability space and $a_1, a_2, \dots \in \mathcal{A}$ be a sequence of **freely independent and identically distributed (FIID)** self-adjoint random variables. Assume all random variables are centered: $\varphi(a_r) = 0, \forall r \in \mathbb{N}$ and denote by $\sigma^2 := \varphi(a_r^2)$ the common variance of the random variables. Then, we have

$$\frac{a_1 + \ldots + a_N}{\sqrt{N}} \xrightarrow{\text{distr}} s,$$

where s is a semicircular random variable with variance σ^2.

Remark. This statement means explicitly, $\forall n \in \mathbb{N}$ and $\sigma := \sqrt{\sigma^2}$,

$$\lim_{N \to \infty} \varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = \int_{-2\sigma}^{2\sigma} \frac{t^n}{2\pi\sigma^2} \sqrt{4\sigma^2 - t^2} dt = \begin{cases} \frac{\sigma^{2k}}{k+1}\binom{2k}{k}, & \text{if } n = 2k \text{ is even} \\ 0, & \text{if } n \text{ is odd} \end{cases}$$

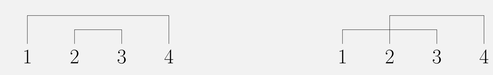
We provide a combinatorial proof of **Theorem 2 (Free CLT)** using **Lemmas 1-3**.

Lemma 1. Under the condition of **Theorem 2**,

$$\lim_{N \to \infty} \varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = D_n \sigma^n,$$

where $D_n := \left|\{\pi : \pi \text{ non-crossing pair partition of } \{1, \dots, n\}\}\right|$.

Figure 1. Non-crossing / crossing pair partitions of $\{1, 2, 3, 4\}$



a **non-crossing** pair partition — a **crossing** pair partition

Proof of Lemma 1. By definition of FIID, we have for large N,

$$\varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = \sum_{1 \leq r(1), \ldots, r(n) \leq N} N^{-n/2} \varphi(a_{r(1)} \cdots a_{r(n)}) \approx \sum_{\substack{\pi \text{ partition} \\ \text{of } \{1, \ldots, n\}}} N^{|\pi| - n/2} \kappa_\pi$$

where π has index tuple $(r(1), \dots, r(n))$ and $\kappa_\pi := \varphi(a_{r(1)} \cdots a_{r(n)})$.

If π contains a subset with only one element, then the resulting κ_π must be zero, since the $\varphi(a_{r(m)}) = 0, \forall a_{r(m)}$. Moreover, if π contains a subset with at least three elements, then $|\pi| < n/2$, so $N^{|\pi| - n/2}$ vanishes as $N \to \infty$. Therefore, we have

$$\lim_{N \to \infty} \varphi\left(\left(\frac{a_1 + \ldots + a_N}{\sqrt{N}}\right)^n\right) = \sum_{\substack{\pi \text{ pair partition} \\ \text{of } \{1, \ldots, n\}}} \kappa_\pi$$

Case 1. Consider the case when all consecutive indices are different: $r(1) \neq \ldots \neq r(n)$. Since $\varphi(a_{r(m)}) = 0$ for all $m = 1, \dots, n$, we have by the definition of free independence

$$\kappa_\pi = \varphi(a_{r(1)} \cdots a_{r(n)}) = 0$$

Case 2. Consider the case when two consecutive indices coincide, i.e., $r(m) = r(m+1)$, for some $m = 1, \dots, n-1$. Since $a_{r(m)} a_{r(m+1)}$ is freely independent from the subalgebra generated by $\{a_{r(1)}, \dots, a_{r(m-1)}, a_{r(m+2)}, \dots, a_{r(n)}\}$, we have the following factorization

$$\kappa_\pi = \varphi(a_{r(1)} \cdots a_{r(m-1)} a_{r(m+2)} \cdots a_{r(n)}) \cdot \varphi(a_{r(m)} a_{r(m+1)}) \cdot \sigma^2$$

We repeat this iteration until we either get zero in one of the steps or arrive at the moment $\varphi(1_{\mathcal{A}}) = 1$, in which the corresponding partition will give a contribution of σ^n. This occurs precisely when π is a non-crossing pair partition of $\{1, \dots, n\}$. □

Non-crossing Pair Partitions and Dyck Paths

The n-th **Catalan number** C_n for $n \geq 0$ is given by

$$C_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{n!(n+1)!}$$

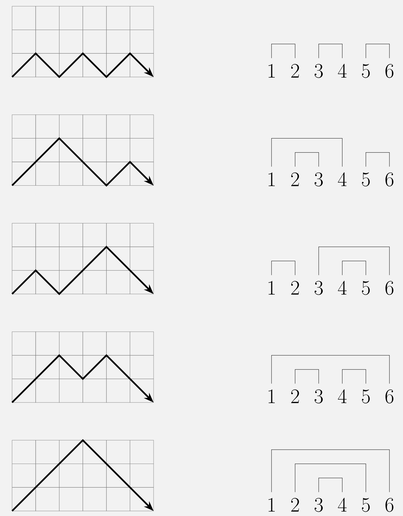
An equivalent representation of the Catalan numbers is the following recurrence relation

$$C_0 = C_1 = 1, \quad C_n = \sum_{k=1}^{n} C_{k-1} C_{n-k}, \ n \geq 2$$

We define **Dyck paths** in the lattice $\mathbb{Z}^2$ to be walks from $(0, 0)$ to $(n, 0)$, for $n \in \mathbb{N}$ even, with steps either of the form $(+1, +1)$ or $(+1, -1)$, keeping y-coordinate nonnegative.

Lemma 2. There is a **one-to-one correspondence (bijection)** between Dyck paths on $2n$ steps and non-crossing pair partition of $\{1, 2, ..., 2n\}$.

Figure 2. Dyck paths on 6 steps / non-crossing pair partition of $\{1, 2, 3, 4, 5, 6\}$



Bijection between Dyck paths and non-crossing pair partitions

Lemma 3. $C_n = D_{2n}$, the number of Dyck paths on $2n$ steps.

Proof of Lemma 3. Consider a Dyck path on $2n$ steps and suppose it first hits $y = 0$ at the $2k$-th step with $1 \leq k \leq n$. The path has to move $(+1, +1)$ (go up) on the first step, and it must move $(+1, -1)$ (go down) at the $2k$-th step. The middle $2(k-1)$ steps are arbitrary as long as $y \geq 0$, so the first $2k$ steps have $D_{2(k-1)}$ combinations. The remaining $2(n-k)$ steps also have no restriction and have $D_{2(n-k)}$ combinations. Thus, this case (first hits at $2k$) contributes $D_{2(k-1)} D_{2(n-k)}$ to D_{2n}. By induction, we have

$$D_{2n} = \sum_{k=1}^{n} D_{2(k-1)} D_{2(n-k)} = \sum_{k=1}^{n} C_{k-1} C_{n-k} = C_n \quad \text{(with the convention } D_0 = 1\text{)}.$$

This shows Catalan numbers count the number of Dyck paths. Together with **Lemma 2**, we proved that Catalan numbers count the number of non-crossing pair partitions. □

Reference

A. Nica, R. Speicher, Lectures on the Combinatorics of Free Probability Theory

A Hypercondensed Introduction to Hyperbolic Geometry

Tracy Sun, Cindy Zhao
University of California Santa Barbara

Introduction

Over 2000 years ago, Euclid derived geometry from four intuitive axioms and an arbitrary "parallel postulate". Getting rid of it became the focus of mathematicians for millennia. Taking a line L and a point p not on L, mathematician János Bolyai supposed that L had several parallel lines going through p for the sake of contradiction. He instead discovered a consistent geometry distinct from Euclid's. Non-Euclidean geometry proved that the parallel postulate cannot be derived from the other four, diverging into two well-studied branches: **hyperbolic** and **spherical**. We will focus on hyperbolic geometry for this poster.

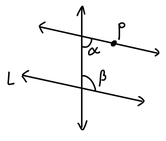


Figure 1: Only one parallel line through p such that $\alpha+\beta = 180\,\text{deg}$.

Figure 2: In spherical geometry, all lines (great circles) intersect.

Figure 3: Multiple parallel lines (Credit: Damia Taimina)

Hyperbolic geometry occurs on surfaces with negative curvature. In nature, these are corals, organs, cells, kale, jellyfish, and perhaps even the universe! Furthermore, hyperbolic structures apply to the theory of relativity and machine learning models for hyperbolic datasets, e.g. historical-linguistics data. To better understand hyperbolic space, lettuce dive into two different models for it.

The Upper Half-Plane Model

How about adapting from a space we already know? Let's take the top half of the complex plane defined as $\mathbb{H} = z \in \mathbb{C} : Im(z) > 0$. For two distinct points p and q in $\mathbb{H}$, there is a hyperbolic line l between them as follows:

$$l = \begin{cases} z \in \mathbb{C} : Re(z) = Re(p) & \text{if } Re(p) = Re(q) \\ z \in \mathbb{C} : |z - c| = r & \text{if } Re(p) \neq Re(q) \end{cases} \quad (1)$$

where $c = \frac{|p^2|-|q|^2}{2(Re(p)-Re(q))}$ and $r = |c - p|$.
For each pair p and q of distinct points in $\mathbb{H}$, there exists a unique hyperbolic line l in $\mathbb{H}$ passing through p and q.

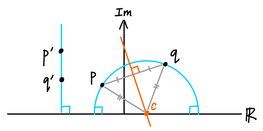


Figure 4: Constructing Euclidean circles with perpendicular bisectors.

One property of interest is **parallel lines**, as they behave differently from Euclidean space. Let l be a hyperbolic line in $\mathbb{H}$ and p a point in $\mathbb{H}$ not on l. Then, there exist infinitely many distinct hyperbolic lines through p that are parallel to l.

Since there are infinitely many points on R between K and L, we can construct an Euclidean circle passing through x and p with center on R for each x, then L has infinitely many parallel lines through point p.

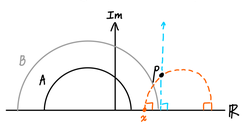


Figure 5: Infinite orange parallel lines to A.

The hyperbolic plane can be modeled as the upper half of the extended complex numbers $\bar{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. A circle in $\bar{\mathbb{C}}$ is either a Euclidean circle in $\mathbb{C}$ or the union of a Euclidean line in $\mathbb{C}$ with $\{\infty\}$. A disc in $\mathbb{C}$ is one of the complements of a circle in $\mathbb{C}$: the area either outside or inside a circle, not including the circle itself.

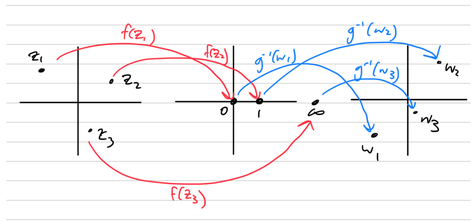
Crucially, $\mathbb{H}$ can be modeled as a disc in $\bar{\mathbb{C}}$! This is because the real axis (the boundary at infinity for our half plane model) is considered a circle in $\bar{\mathbb{C}}$. The boundary at infinity for any set in $\mathbb{H}$ is where it intersects the real axis.

Möbius Transformations

Definition
A Möbius transformation is a function of the form $m(z) = \frac{az+b}{cz+d}$, where $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$.
They have some very useful properties, including:

- Möbius transformations are continuous and bijective
- Inverse of a Möbius transformation is also a Möbius transformation
- Composition of Möbius transformations is also a Möbius transformation
- Möbius transformations preserve lengths, angles, lines, circles, and discs in $\bar{\mathbb{C}}$

Fixed Points
A **fixed point** of a Möbius transformation is defined as a point where $m(z) = z$.
A Möbius transformation that has more than two fixed points is equivalent to the identity transformation, where $m(z) = z$ for all z. Because of this, a Möbius transformation can be specified by how it acts on any three distinct ordered points in $\bar{\mathbb{C}}$.



Classifications
Two Möbius transformations m_1, m_2 are considered **conjugate** if there exists a Möbius transformation p such that $m_2 = p \circ m_1 \circ p^{-1}$. Similar to matrix diagonalization in linear algebra, we can simplify analysis of Möbius transformations by conjugating them into three **standard forms**: parabolic, elliptic, and hyperbolic.

For elliptical and hyperbolic Möbius transformations, the coefficient of the standard form is considered the **multiplier**.

Matrix Classification
Since Möbius transformations have a distinct structure, their coefficients can be written as distinct matrices: $m(z) = \frac{ax+b}{cx+d} \sim A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Then we can endow Möbius transformations with matrix characteristics. From the example above, we can **normalize** m by multiplying it by $\alpha = \frac{1}{\det(m)}$. However, normalized m has ambiguity at $\alpha = 1$, so we can construct a well-defined set of the **trace** $T(m) = (a + d)^2$. Crucially, the trace of a Möbius transformation is invariant under conjugation, since

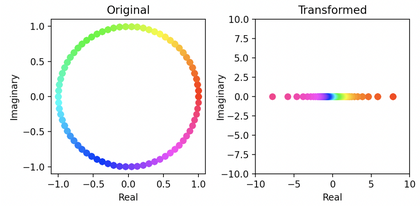
$$\begin{aligned} T(p \circ m \circ p^{-1}) &= T((p \circ m) \circ p^{-1}) \\ &= T(p^{-1} \circ (p \circ m)) \\ &= T(p^{-1} \circ p \circ m) \\ &= T(m). \end{aligned} \quad (2)$$

Thanks to the associativity and commutativity of function compositions, instead of conjugating Möbius transformations to classify them, we can instead simply normalize them and compute the trace.

Type	Standard Form	Trace	Number of Fixed Points
Parabolic	$n(z) = z + 1$	4	One
Elliptical	$n(z) = e^{2i\theta}z$	$[0, 4)$	Two
Hyperbolic	$n(z) = p^2e^{2i\theta}z$	$(-\infty, 0) \cup (4, \infty)$.	Two

The Poincaré Disc Model

From Disc to Half Plane

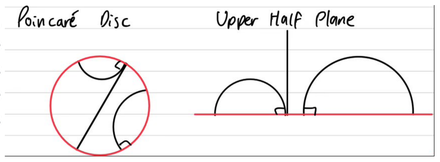


As both the upper half plane and the unit disc of the Poincaré model are considered discs in $\bar{\mathbb{C}}$ and Möbius transformations preserve discs in $\bar{\mathbb{C}}$, we can use Möbius transformations to convert between the two different model types. This will also preserve lines, circles, and angles between the two models.

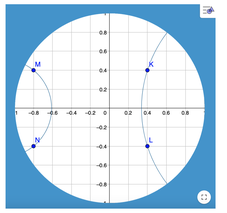
Figure 6: How one example of a Möbius transformation that takes the unit disc in $\bar{\mathbb{C}}$ to the real axis ($m(z) = \frac{(z+1)(i-1)}{(z-1)(i+1)}$) transforms the boundary at infinity.

Lines in Poincaré Disc Model

In the Poincaré disc model, lines in hyperbolic space are represented as Euclidean lines that are the diameters of the unit circle, or Euclidean circles perpendicular to it.



Hyperbolic Distance in Poincaré Disc Model



In the Poincaré disc model of hyperbolic space, distance works differently than in Euclidean space. Two line segments that appear of similar length on different parts of the disc can actually represent vastly different distances.

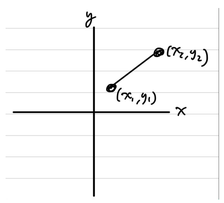


Figure 7: For a continuous and differentiable path $f(t)$, the length of $f(t)$ can be expressed as an integral, $\int_f \sqrt{\frac{dx^2}{dt} + \frac{dy^2}{dt}}dt$.

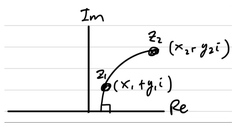


Figure 8: For the upper half plane model, the Euclidean length is scaled by a factor of $\frac{1}{Im(z)}$, reflecting how the real axis ($Im(z) = 0$) represents infinity.

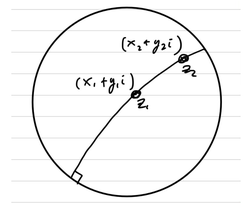


Figure 9: For the Poincaré disc model, the Euclidean length is scaled by a factor of $\frac{2}{1-|z|^2}$.

Acknowledgements

We thank Katherine Merkl for her guidance, as well as the UCSB Directed Reading Program for the opportunity to work on this project.

References

[1] James W. Anderson. *Hyperbolic Geometry*. London: Springer, 2005.
[2] *History of Hyperbolic Geometry*. Online: Models and projections of hyperbolic geometry.

A Survey of Results in Homology

Geordie Taber – Mentored by Troy Kling

University of California, Santa Barbara

Introduction

In order to study different topological spaces, it is useful to assign invariants to each space that are preserved under some notion of equivalence. Although homeomorphism is a natural candidate, the weaker property of homotopy equivalence is more flexible. Homology is a branch of algebraic topology that associates a sequence of homology groups to a space. If two spaces are homotopy equivalent, then the corresponding entries in each sequence of homology groups are isomorphic. Homology is an important tool for studying topological spaces because unlike the homotopy groups, homology groups are easily computable.

The Idea of Homology

The homology groups of a space provide a description of how high-dimensional features attach to lower-dimensional features in a space. Homology provides a systematic way of describing holes, connected components, and other features of a space. Several different homology theories have been developed for particular kinds of spaces, such as cellular and simplicial homology. **Singular homology** is a more general approach defined for arbitrary topological spaces. The construction is very detailed, but the idea can be axiomatized in order to avoid many of the geometric arguments used in proving its fundamental results.

Preliminaries

- The n-**simplex** is the smallest convex set $\Delta^n \subseteq \mathbf{R}^{n+1}$ that contains the standard basis vectors $\{v_0, v_1, \ldots, v_n, v_{n+1}\}$.

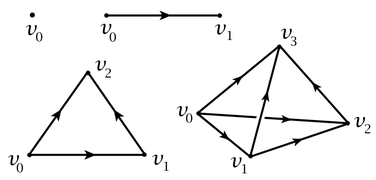


Figure 1. The 0-, 1-, 2-, and 3-simplices [1]

- A **singular** n**-simplex** in a space X is a continuous map $\sigma : \Delta^n \to X$.
- The nth **chain group** of a space X, denoted by $C_n(X)$, is the free abelian group whose basis is the set of singular n-simplices of X, and each element of this group is an n-**chain**, a formal sum of singular n-simplices.
- The **boundary maps** $\partial_n : C_n(X) \to C_{n-1}(X)$ are defined

$$\partial_n(\sigma) = \sum_i (-1)^i \sigma \mid [v_0, \ldots, \hat{v_i}, \ldots, v_n],$$

where $\sigma \mid [v_0, \ldots, \hat{v_i}, \ldots, v_n]$ denotes the restriction of the singular n-simplex σ to a face consisting of all the v_j except for v_i. The map extends linearly to n-chains. Thus the restriction can be seen as a singular $n-1$-simplex.

- A sequence of homomorphisms of abelian groups

$$\cdots \longrightarrow C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \longrightarrow \cdots \longrightarrow C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0$$

is called a **chain complex** if $\partial_n \partial_{n+1} = 0$ for all n, which is succinctly expressed $\partial^2 = 0$

- It can be shown that for the boundary maps $\operatorname{im} \partial_{n+1} \subseteq \ker \partial_n$. The nth **singular homology group** $H_n(X)$ is the quotient $\ker \partial_n \,/\, \operatorname{im} \partial_{n+1}$.

Exact Sequences and Category-theoretic Results

- A sequence of homomorphisms

$$\cdots \longrightarrow A_{n+1} \xrightarrow{\alpha_{n+1}} A_n \xrightarrow{\alpha_n} A_{n-1} \longrightarrow \cdots$$

is an **exact sequence** if for each adjacent pair $\operatorname{im} \alpha_{n+1} = \ker \alpha_n$.

- For spaces X and Y, any map $f : X \to Y$ induces a map on homology $f_* : H_n(X) \to H_n(Y)$. By composing f with $\sigma \in X^{\Delta^n}$, an element $f\sigma \in Y^{\Delta^n}$ is obtained. Thus f extends linearly to a map $f_\# : C_n(X) \to C_n(Y)$.
- Furthermore, $f_\#$ is called a **chain map**, because $f_\#$ commutes with applying the boundary maps, written $f_\# \partial = \partial f_\#$. We have the following **commutative diagram**.

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & C_{n+1}(X) & \xrightarrow{\partial} & C_n(X) & \xrightarrow{\partial} & C_{n-1}(X) & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle f_\#} & & \downarrow{\scriptstyle f_\#} & & \downarrow{\scriptstyle f_\#} & & \\
\cdots & \longrightarrow & C_{n+1}(Y) & \xrightarrow{\partial} & C_n(Y) & \xrightarrow{\partial} & C_{n-1}(Y) & \longrightarrow & \cdots
\end{array}$$

- A chain map between chain complexes induces a homomorphism on homology. Thus there exists $f_* : H_n(X) \to H_n(Y)$.

The Five-Lemma

The Five-Lemma states that in the following commutative diagram over an abelian category, if the rows are exact and α, β, δ, and ϵ are isomorphisms, then γ is also an isomorphism.

$$\begin{array}{ccccccccc}
A & \xrightarrow{i} & B & \xrightarrow{j} & C & \xrightarrow{k} & D & \xrightarrow{\ell} & E \\
\downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \downarrow{\scriptstyle \delta} & & \downarrow{\scriptstyle \varepsilon} \\
A' & \xrightarrow{i'} & B' & \xrightarrow{j'} & C' & \xrightarrow{k'} & D' & \xrightarrow{\ell'} & E'
\end{array}$$

- The **simplicial homology groups** are defined for a specific type of topological spaces known as Δ-complexes, and the Five-Lemma is used to prove that for such spaces, the singular and simplicial homology groups are isomorphic.

The Splitting Lemma

For a short exact sequence

$$0 \longrightarrow A \xrightarrow{i} B \xrightarrow{j} C \longrightarrow 0,$$

the following are equivalent:

- There exists a homomorphism $p : B \to A$ such that $pi = \mathbb{1}_A$.
- There exists a homomorphism $s : C \to B$ such that $js = \mathbb{1}_C$.
- There is an isomorphism $B \approx A \oplus C$ that creates the commutative diagram below, with the lower maps being $a \mapsto (a, 0)$ and $(a, c) \mapsto c$.

$$\begin{array}{ccccccccc}
0 & \longrightarrow & A & \xrightarrow{i} & B & \xrightarrow{j} & C & \longrightarrow & 0 \\
 & & & \searrow & \downarrow{\scriptstyle \approx} & \nearrow & & & \\
 & & & & A \oplus C & & & &
\end{array}$$

Degree Theory

- For a map $f : S^n \to S^n$, the induced map $f_* : H_n(S_n) \to H_n(S_n)$ is a homomorphism of an infinite cyclic group onto itself; hence, $f_*(\alpha) = d\alpha$ for some integer d.
- This integer depends only on f and is called the **degree** of f, denoted $\deg f$.
- The idea of degree is an important application in algebraic topology because it is the original method for proving **Brouwer's fixed-point theorem**.

Brouwer's Fixed-Point Theorem

Every continuous map $h : D^n \to D^n$ has a **fixed point**, a point $x_* \in D^n$ such that $h(x_*) = x_*$.

Let $f, g : S^n \to S^n$ be continuous maps. The degree has several interesting properties:

- If f is not surjective, then $\deg f = 0$.
- If f and g are homotopic, then $\deg f = \deg g$. This is not difficult to show because f and g must satisfy $f_* = g_*$. Surprisingly, the converse is also true.
- Degree is multiplicative: $\deg fg = \deg f \cdot \deg g$. This is because $(fg)_* = f_* g_*$.

The Hairy Ball Theorem

The space S^n has a continuous field of non-zero tangent vectors if and only if n is odd.

- If $n = 2k-1$, then one can construct the vector field by $(x_1, x_2, \ldots, x_{2k-1}, x_{2k}) \mapsto (-x_2, x_1, \ldots, -x_{2k}, x_{2k-1})$.
- Conversely, if the vector field exists, one can show that a homotopy exists between the identity $\mathbb{1}$ and the antipodal map $-\mathbb{1}$. But if n is even, the antipodal map has degree $(-1)^{n+1} = -1 \neq 1$.

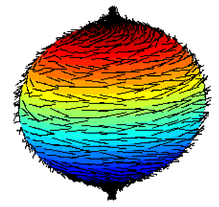


Figure 2. A failed attempt to comb S^2. Source: Wikipedia.

Acknowledgements

I would like to thank my mentor Troy for his enthusiasm and support over the past few months. Reading Hatcher's text was a challenge, but our weekly meetings helped me to understand the material much more intuitively.

References

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

A Symmetry Problem

Ethan Martirosyan and Yingpeng He Mentor: Jihye Lee

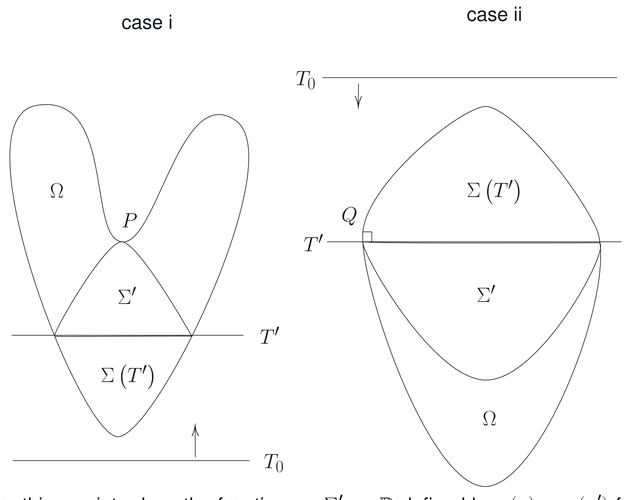
University of California Santa Barbara

Serrin's Overdetermined Problem

Let us consider the following problem. Let $\Omega \subseteq \mathbb{R}^n$ be a domain that is bounded, open, and connected. Furthermore, suppose that the boundary $\partial\Omega$ is smooth. Let $u : \Omega \to \mathbb{R}$ be a C^2 function that satisfies the following conditions: $\Delta u = -1$ in Ω and $u = 0$ and $\frac{\partial u}{\partial \nu} = c$ on $\partial\Omega$ for some constant c and ν is the outward normal vector to $\partial\Omega$. Then, Ω must be a ball. Furthermore, we know that $u(x) = (b^2 - r^2)/2n$, where b is the ball's radius and r is the distance to its center.

First Proof

The first proof we present is from Professor James Serrin himself [3]. This proof utilizes the **moving plane method**. Let T_0 be a $n-1$ dimensional hyperplane in $\mathbb{R}^n$ that does not intersect the domain Ω. We begin to move this plane normal to itself until it intersects Ω. When this occurs, the new plane T splits Ω into two parts. The part of Ω that lies on the same side of T as our initial plane T_0 is denoted by $\Sigma(T)$. We reflect $\Sigma(T)$ in T to obtain $\Sigma' := \Sigma'(T)$. As T moves through Ω, Σ' will remain in Ω until it becomes internally tangent to Ω at a point P (case i) or T becomes orthogonal to Ω at some point Q (case ii). When either of these occurs, we stop moving the plane T, and we denote the resulting plane by T'. We claim that Ω is symmetric about T'. Showing this would prove the theorem. To see how, we recall that the plane T_0 was chosen arbitrarily. If Ω is symmetric about T', then Ω is symmetric in all possible directions. Since Ω is simply connected and has this strong symmetry property, it must be a ball.



To prove this, we introduce the function $v : \Sigma' \to \mathbb{R}$ defined by $v(x) = u(x')$ for $x \in \Sigma'$, where x' is the reflection of x across T'. By the maximum principle, we deduce that $u - v > 0$ or $u - v = 0$ in Σ'. For the sake of contradiction, suppose that $u - v > 0$. If Σ' is internally tangent to Ω at some point P, then we may appeal to the boundary point maximum principle to deduce that $\frac{\partial}{\partial \nu}(u - v) > 0$ at P [1]. However, we know that $\partial u/\partial \nu = \partial v/\partial \nu = c$. Thus we have reached a contradiction. If T' is orthogonal to the boundary of Ω at some point Q, then we show that u and v have the same first and second derivatives at Q. Using a modified version of the boundary point maximum principle, we can also show that $\frac{\partial}{\partial s}(u - v) > 0$ or $\frac{\partial^2}{\partial^2 s}(u - v) > 0$ for any direction s that enters Σ' non-tangentially at Q. However, this directly contradicts the fact that u and v have the same first and second derivatives at Q. We may thus conclude that $u = v$ and that Ω is symmetric about T'.

Second Proof

The second proof we present is from Weinberger [2]. To start, we first compute

$$\Delta\left(r\frac{\partial u}{\partial r}\right) = r\frac{\partial}{\partial r}(\Delta u) + 2\Delta u = -2,$$

where r is the distance to the origin. Using this and the fact that $\Delta u = -1$, we obtain

$$\int_\Omega \left[2u - r\frac{\partial u}{\partial r}\right]dx = \int_\Omega \left[-u\Delta\left(r\frac{\partial u}{\partial r}\right) + r\frac{\partial u}{\partial r}\Delta u\right]dx$$

Using Green's identity yields

$$\int_\Omega \left[-u\Delta\left(r\frac{\partial u}{\partial r}\right) + r\frac{\partial u}{\partial r}\Delta u\right]dx = \int_{\partial\Omega} \left[-u\frac{\partial}{\partial \nu}\left(r\frac{\partial u}{\partial r}\right) + r\frac{\partial u}{\partial r}\frac{\partial u}{\partial \nu}\right]dS$$

By assumption, we have $u = 0$ on $\partial\Omega$. Thus, we find that

$$\int_{\partial\Omega} \left[-u\frac{\partial}{\partial \nu}\left(r\frac{\partial u}{\partial r}\right) + r\frac{\partial u}{\partial r}\frac{\partial u}{\partial \nu}\right]dS = \int_{\partial\Omega} r\frac{\partial r}{\partial \nu}\left(\frac{\partial u}{\partial \nu}\right)^2 dS$$

By assumption, we know that $\partial u/\partial \nu = c$ on $\partial\Omega$. Thus, we find that

$$\int_{\partial\Omega} r\frac{\partial r}{\partial \nu}\left(\frac{\partial u}{\partial \nu}\right)^2 dS = c^2\int_{\partial\Omega} r\frac{\partial r}{\partial \nu}dS = c^2 n\int_\Omega dx = nc^2V$$

where V is the volume of Ω. Green's theorem also implies

$$\int_\Omega r\frac{\partial u}{\partial r}dx = -n\int_\Omega u dx$$

so that substitution yields

$$(n+2)\int_\Omega u dx = nc^2V$$

However, we also note that

$$1 = (\Delta u)^2 \le n\sum_{i=1}^n u_{ii}^2 \le n\sum_{i,j} u_{ij}^2$$

by the Cauchy-Schwarz inequality. From this, we deduce that

$$\Delta\left(|\nabla u|^2 + \frac{2}{n}u\right) = 2\sum_{i,j} u_{ij}^2 - \frac{2}{n} \ge 0$$

Using this and the fact that $|\nabla u|^2 + (2/n)u = c^2$ on $\partial\Omega$, we may appeal to the maximum principle to deduce that $|\nabla u| + (2/n)u < c^2$ in Ω or $|\nabla u| + (2/n)u = c^2$ in Ω. If the inequality held, then we could integrate over Ω to deduce that

$$(n+2)\int_\Omega u dx < nc^2V$$

This contradiction informs us that $|\nabla u|^2 + (2/n)u = c^2$ in Ω so that

$$\Delta\left(|\nabla u|^2 + \frac{2}{n}u\right) = 2\sum_{i,j} u_{ij}^2 - \frac{2}{n} = 0$$

and

$$1 = n\sum_{i=1}^n u_{ii}^2 = \sum_{i,j} u_{ij}^2$$

which implies that $u_{ij} = -\delta_{ij}/n$. Solving the corresponding partial differential equations yields

$$u = \frac{1}{2n}(B - r^2)$$

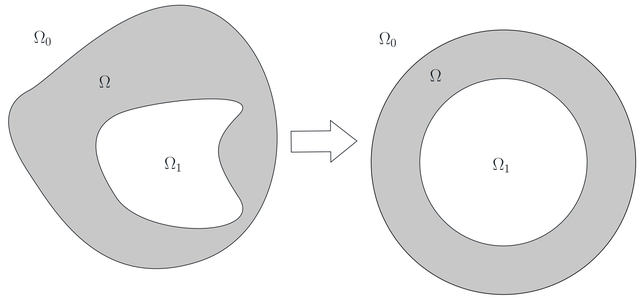
where B is a constant. Since $u = 0$ on $\partial\Omega$, B is positive and Ω is a ball of radius $B^{1/2}$.

Applications

This theorem is significant because it allows us to determine the shape of Ω from properties of u. It also has many applications in physics. For example, we may consider an incompressible viscous fluid moving through a straight pipe of cross sectional form Ω. If we fix a rectangular coordinate system with the z-axis directed along the pipe, then the velocity u depends only on x and y, and it satisfies the differential equation $\Delta u = -A$ for some constant A. Furthermore, because the fluid is viscous, we know that $u = 0$ on $\partial\Omega$; that is, there is no movement on the boundary of the pipe. Finally, we note that $\mu\partial u/\partial\nu$ is the tangential stress on the pipe wall, where μ is the viscosity constant. If the tangential stress is constant, then we may apply the above theorem to conclude that Ω is a circular cross section.

Further Results

A similar result was proved by Wolfgang Reichel [4]. Let Ω_0 and Ω_1 be smooth domains in $\mathbb{R}^n$ and let $\Omega = \Omega_0 \setminus \overline{\Omega}_1$ be connected. Suppose that $f \in C^1$ is a function satisfying $\Delta u + f(u, |\nabla u|) = 0$ in $\overline{\Omega}$, $0 < u < a$ in Ω, $u = 0$ on $\partial\Omega_0$, $u = a$ on $\partial\Omega_1$, and $\partial u/\partial\nu = c_i$ on $\partial\Omega_i$. Then, we conclude that Ω is an annulus and u is radially symmetric and decreasing in r.



The above picture demonstrates the theorem. On the left-hand side, we see our hypotheses. On the right-hand side, we see the conclusion.

Acknowledgements

We would like to thank Jihye Lee for mentoring us. Furthermore, we express gratitude to the 2024 UCSB Directed Reading Program for giving us this opportunity.

References

[1] Hans Weingberger. Maximum Principles in Differential Equations. 1984.
[2] Hans Weingberger. Remark on the Preceding Paper of Serrin. *Arch. Rational Mech. Anal.* 1971.
[3] James Serrin. A Symmetry Problem in Potential Theory. *Arch. Rational Mech. Anal.* 1971.
[4] Wolfgang Reichel. Radial Symmetry by Moving Planes for Semilinear Elliptic BVPs on annuli and other non-convex domains. *Elliptic and parabolic problems* 1995.

An Application of the Direct Method of the Calculus of Variations

William Mahnke, Elizabeth Thomson and Sihan Jiang

Department of Mathematics, University of California, Santa Barbara

Introduction

In this project, we aim to solve the PDE describing the **ground state** for Hydrogen (1) by using **the direct method of the calculus of variations**.

$$\left(-\frac{1}{2}\Delta - \frac{1}{|x|}\right)\psi(x) = \epsilon\psi(x) \text{ (with } x \in \mathbb{R}^3) \quad (1)$$

$$||\psi||_2 = 1 \quad (2)$$

This is an eigenvalue problem for a time-independent Schrodinger equation. Since we are solving this equation for the ground state, seek to find the minimal ϵ in the spectrum of the operator in (1).

This PDE comes from finding the **Euler-Lagrange equation** for the energy functional,

$$F(\psi) := ||\nabla\psi||_2^2 - \int_{\mathbb{R}^3}\frac{|\psi(x)|^2}{|x|}dx \quad (3)$$

The Euler-Lagrange equation (3) connects a potential solution to the PDE (1) to a minimization problem over an appropriate domain. In this case, we minimize over a relaxed domain $||\psi||_2 \leq 1$.

There are several difficulties regarding solving the PDE.

1. The second-order differentiability of $\psi(x)$ is too strong to be used to find the solution directly.
2. The behavior of coulomb potential $\frac{1}{|x|}$ near x = 0 needs to be addressed.
3. The problem is posed over $\mathbb{R}^3$ rather than a bounded domain, so typical functional analysis tools on bounded domains do not apply to this problem.

PDEs that resemble equation (1) can be solved analytically for certain potentials, but it is not possible to solve them analytically for many-body systems. Nonetheless, we can use the calculus of variations to analyze atoms in a multi-electron system, using a similar approach to that of Hydrogen in this project.

Overview of the Direct Method

The direct method changes the focus of solving (1) to finding a critical points of it's corresponding energy functional F, as defined in (3). The direct method requires we prove,

1. The energy functional F is coercive.
2. F is weakly lower semi-continuous. The kinetic term is weakly lower semi-continuous, but in fact the potential term is weakly continuous.

Then, given a minimizing sequence $\{\psi_n\}$ associated with (3), the Banach-Alaoglu theorem along with our coercivity condition implies that, since $\{\psi_n\}$ is bounded in $H^1(\mathbb{R}^3)$, it admits a weak limit ψ, up to a subsequence. We then use lower semi-continuity of (3) to prove that ψ is in fact the minimizer. The last step of showing the existence of the minimizer which will solve (1) with condition (2) will be proving the minimizer necessarily has unit norm.

Coercivity

To show F is coercive, we want to show there exists $a > 0, b \geq 0$ such that:

$$\text{For all } f \text{ in our domain of } F : F(f) \geq a||\nabla f||_2^2 - b \quad (4)$$

Showing coercivity requires another result, Hardy's inequality:

$$\int_{\mathbb{R}^3}\frac{|\psi(x)|^2}{|x|^2}dx \leq 4||\nabla\psi||_2^2 \quad (5)$$

Using Holder's inequality, we can express the integral in (3) as:

$$\int_{\mathbb{R}^3}\frac{|\psi(x)|^2}{|x|}dx \leq ||\psi||_2\left|\left|\frac{\psi}{|x|}\right|\right|_2 \quad (6)$$

Young's inequality and (6) implies for all positive ε:

$$\int_{\mathbb{R}^3}\frac{|\psi|^2}{|x|}dx \leq \frac{\varepsilon}{2}\left|\left|\frac{\psi}{|x|}\right|\right|_2^2 + \frac{1}{2\varepsilon}||\psi||_2^2 \quad (7)$$

Since our domain for ψ is constricted to $||\psi||_2 \leq 1$, combining (5) and (7) implies:

$$\int_{\mathbb{R}^3}\frac{|\psi|^2}{|x|}dx \leq 2\varepsilon||\nabla\psi||_2^2 + \frac{1}{2\varepsilon} \quad (8)$$

Thus, setting $\varepsilon = \frac{1}{4}$ we obtain:

$$F(f) \geq \frac{1}{2}||\nabla\psi||_2^2 - 2 \quad (9)$$

Weak Lower Semi-Continuity

Showing weak lower semi-continuity of F requires showing weak lower semi-continuity of both the kinetic and potential terms.

Given our minimizing sequence $\{\psi_n\}$ combined with a result from [2], we get $||\psi_n||$ is bounded in H^1, where:

$$||\psi_n||_{H^1(\mathbb{R}^3)}^2 = ||\nabla\psi_n||_2^2 + ||\psi_n||_2^2 \quad (10)$$

Given $||\nabla\psi||_2$ is also convex, another result from [2] implies the kinetic term is weakly lower semi-continuous.

To show the weak continuity of the potential term, we'll split the integral into two pieces to evaluate how the integrand behaves close to and away from the origin.

$$\int_{\mathbb{R}^3}\frac{|\psi_n(x)|^2}{|x|}dx = \int_{|x|\leq\frac{1}{\varepsilon}}\frac{|\psi_n(x)|^2}{|x|}dx + \int_{|x|>\frac{1}{\varepsilon}}\frac{|\psi_n(x)|^2}{|x|}dx \quad (11)$$

Since ψ is restricted to $||\psi||_2 \leq 1$, away from the origin:

$$\int_{|x|>\frac{1}{\varepsilon}}\frac{|\psi_n(x)|^2}{|x|}dx < \varepsilon\int_{|x|>\frac{1}{\varepsilon}}|\psi_n(x)|^2dx \leq \varepsilon \quad (12)$$

For the region close to the origin, we first observe that

$$\int_{|x|\leq\frac{1}{\varepsilon}}\frac{1}{|x|^p}dx = \int_0^{2\pi}\int_0^{\pi}\int_0^{1/\varepsilon}\frac{1}{p^p}\,p^2\sin\phi d\rho d\phi d\theta \quad (13)$$

The integral converges for p < 3, which implies $\frac{1}{|x|} \in L^p(|x| \leq \frac{1}{\varepsilon})$ for $1 < p < 3$. Using Sobolev's inequality, we observe:

$$||\psi_n||_{L^6(|x|\leq\frac{1}{\varepsilon})} \leq C||D\psi_n||_{L^2(|x|\leq\frac{1}{\varepsilon})} \leq C||\psi_n||_{H^1(|x|\leq\frac{1}{\varepsilon})} \quad (14)$$

(14) implies ψ_n's are uniformly bounded in L^6, implying $|\psi_n|^2$'s are uniformly bounded in L^3.

Since our minimizing sequence is weakly convergent, i.e. $\psi_n \rightharpoonup \psi$ (in L^6, not just H^1), $\langle f, \psi_n\rangle \to \langle f, \psi\rangle$ for all f in ψ's dual space, in this case $L^{3/2}$. (13) implies $\frac{1}{|x|} \in L^{3/2}(|x| \leq \frac{1}{\varepsilon})$, so the previous result can applied with $\frac{1}{|x|}$, implying integral is strongly convergent.

Since F has been shown to be weakly lower semi-continuous and coercive, a minimizer for the functional exists. The last step is showing the minimizer necessarily has unit norm. Using a smooth, compact test function:

$$\psi_\lambda(x) := \frac{1}{\lambda^{3/2}}\psi(x) \quad (15)$$

It can be shown a large enough λ will eventually make F negative. Combined with the fact $F(\frac{\phi}{||\phi||_2}) = \frac{1}{||\phi||^2}F(\phi)$, any minimizer with a norm less than one would cause a contradiction.

Conclusion

Using the Direct Method, we have shown through the last few steps that a solution to the problem layed out in the introduction exists.

An important take-away from this result is given a PDE with difficult properties such as differentiability it is possible to show the existence of a solution by weakening the properties of the solution and working within a different space of functions. Using the Direct Method of the Calculus of Variations, we were able to reframe the problem and use properties of the weak topology to fulfill the original goal.

References

[1] Lawrence C. Evans.
Partial differential equations.
Graduate studies in mathematics. American Mathematical Society, Providence, R.I, 2nd ed edition, 2010.

[2] Haim Brezis.
Functional Analysis, Sobolev Spaces and Partial Differential Equations.
Springer New York, New York, NY, 2011.

[3] Elliot Lieb and Michael Loss.
Analysis.
American Mathematical Society, 2nd ed edition, 2001.

Acknowledgements

We would like to thank our mentor Zach for his support throughout the project and the Directed Reading Program for allowing us to participate in the program and providing additional support to assist our learning.

An Introduction to Knot Theory

Joy Chang, Mizuki Shitomi

University of California - Santa Barbara

What is Knot Theory?

Knot theory is a sub-field of Topology that deals with the study of mathematical knots. It allows us to classify knots based on their properties and helps us explain how knots are transformed within space. Although there are many definitions of a knot, in knot theory, a **knot** is the cross-section of a single point. To help us study knots, knots are drawn into **projections** where we can clearly see the crossing. There are several ways to project a knot whether that be through sticks, tricolorability, or planar graph (mentioned later on). These knot projections help us classify the types of knots such as the unknot/trivial knot which is the simplest knot (Figure 1.1) or trefoil knot which is the simplest non-trivial knot (figure 1.2).

Figure 1.1

Figure 1.2

Reidemeister Moves

More often than knot (pun intended), a knot is not as simple as the figure above. When knots are more complicated, it is harder to determine the types of know they are including whether it is a knot or an unknot. Hence, the **Reidemeister moves** is a method used to help us classify whether a knot is an unknot. It allows us to alter the knot without changing its properties. There are three Reidemeister moves.

- The **first** Reidemeister move allows us to put in or take out a twist in the knot (Figure 2.1).

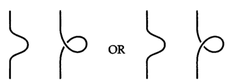


Figure 2.1

- The **second** Reidemeister move allows us to either add two crossings or remove two crossings (Figure 2.2).

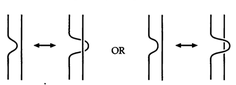


Figure 2.2

- The **third** Reidemeister move allows us to slide a strand of the knot from one side of a crossing to the other side of the crossing (Figure 2.3).

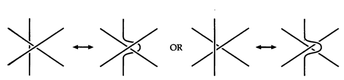
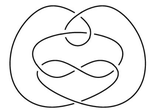


Figure 2.3

Try It Yourself

Prove that the knot below is an unknot using the Reidemeister moves.



Planar Graphs

A **planar graph** explains itself in its name - a graph that lies in the plane. It can be created from a projection of a knot or link in the following steps. A link is a set of knots all tangled up together.

- Shade every other region of the link projection (Figure 3.1).
- Put a vertex at the center of each shaded region and connect with an edge any two vertices that are in regions that share a crossing (Figure 3.2).
- Define crossings to be positive or negative (Figure 3.4).
- The result is a **signed planar graph** (Figure 3.3).

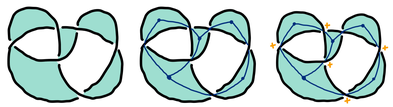


Figure 3.1 Figure 3.2 Figure 3.3 Figure 3.4

We can also go in the other direction - turning a signed planar graph into a knot projection. Just follow these steps:

- Put an X across each edge in the singed planar graph (Figure 4.1).
- Connect the edges formed by X inside each region (Figure 4.2).
- Shade the areas that contain a vertex (Figure 4.3).
- At each of the X's, put in a crossing corresponding to whether the edge is a + or a - edge (Figure 4.4).

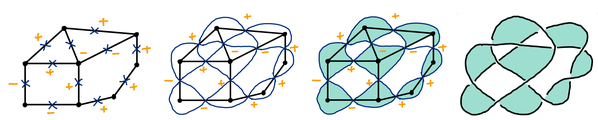


Figure 4.1 Figure 4.2 Figure 4.3 Figure 4.4

Try it yourself: Turn the signed planar graph in Figure 4.5 into the corresponding link projection.

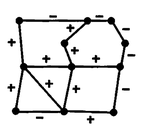


Figure 4.5

Knots vs. Graphs

Why do we want to convert knot projections into planar graphs and vice versa? Sometimes the problems in knot theory can be easier to solve if we turn the knot projections into signed planar graphs. For instance, there is an open problem that aims to find a practical algorithm for determining if a projection is a projection of the unknot. This is equivalent to asking if there is a sequence of Reidemeister moves that can convert the given projection to the projection of the unknot. By turning knot projections into signed planar graphs, this problem becomes determining the induced Reidemeister moves in the signed planar graph.

The planar graphs also have real-world applications in fields such as Chemistry, machine learning, statistical mechanics, and hydraulic engineering. Here we discuss a mathematical model of ferromagnetism in statistical mechanics known as the **Ising model**. It models a system where particles only interact with nearby ones. Two particles that are not neighbors have no effect on one another.

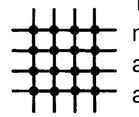


Figure 5.1

Take the magnetization of a metal as an example: each molecule of the metal is considered to be a vertex of a graph. The interactions between adjacent molecules are represented by the edges of the planar graph. Only two molecules connected by an edge can interact.

Figure 5.2

Lattice is a particular type of the Ising model, where the vertices and edges form a regular repeating in two-dimensional space (Figure 5.1). To relate this concept to the real world, metals consist of molecules that are at the vertices of a lattice in three-dimensional space (Figure 5.2).

Conclusion

Although we only covered Reidmeister moves and planar graphs, the world of knot theory is endless. There are several other methods to understand how knots interact with space that we unfortunately cannot cover today. Moreover, knot theory can apply to graph theory, quantum theory, DNA modeling, and in everyday interactions. So the next time you are tying up your shoelaces or your charging cords, try thinking through the lenses of a topologist.

References

"The Knot Book" by Colin C. Adams

We thank our mentor Mychelle Parker for making this project possible as well as the UCSB Directed Reading Program.

UC **SANTA BARBARA**

An Introduction to Seiberg-Witten Invariants Background Knowledge

Ziheng Cao [2] Mentor: Debin Liu [1]

[1]UC Santa Barbara, Department of Mathematics

[2]Fudan University, Department of Physics

Introduction

Seiberg-Witten theory in physics was developed in the mid-1990s as an exact solution to $4d$ $\mathcal{N}=2$ supersymmetric gauge theory at IR fixed point. The IR Wilson effective action is obtained by integrating over heavy modes. In gauge theory there are infinite gauge inequivalent vacua which form the moduli space of the theory. Singularities exist in the moduli space. To compute the infrared effective action, we can study the behavior of the functions of interest (prepotential and vacuum expectation value of the scalar field) on the moduli space, particularly focusing on the monodromy near the singular points, i.e., the variations around the singularity as one goes around it. Seiberg and Witten interpret the singularities in moduli space as magnetic monopoles, a type of soliton. A set of functions satisfying specific monodromy conditions can be found using Riemann-Hilbert correspondence, which involves the application of elliptic curves. Thus, through the aforementioned method, we obtain exact solutions for the IR dynamics.

In mathematics, searching for topological invariants of manifolds can be used to classify and characterize them. Utilizing equivalence classes of solutions to PDEs to derive topological invariants is an important approach in this regard.For example, the Atiyah-Singer index theorem analyzes the solution space of linear partial differential equations to obtain topological invariants, which subsequently have geometric applications. Similarly, employing nonlinear PDEs would yield additional invariants. Donaldson invariants describe the topological properties of four-dimensional compact oriented manifolds, but computing them can be quite involved in some cases. Seiberg-Witten invariants, to some extent, offer a simpler means of obtaining topological invariants.

Different Structures of Manifolds

- **Riemannian Structure**
  The structure group of the tangent bundle $T\mathcal{M}$ can be reduced from $GL(n,\mathbb{R})$ to $O(n)$ due to Riemann metric.
- **Orientable Structure**
  The structure group can be reduced form $O(n)$ to $SO(n)$.
- **Symplectic Structure**
  The structure group (of an even dimensional manifold) of the tangent bundle $T\mathcal{M}$ can be reduced form $GL(2n,\mathbb{R})$ to $Sp(2n,\mathbb{R})$.
- **Almost Complex Structure**
  The structure group (of an even dimensional Riemannian manifold) of the tangent bundle $T\mathcal{M}$ can be reduced form $SO(2n,\mathbb{R})$ to $U(n)$.
- **Spin Structure** The structure group of the tangent bundle $T\mathcal{M}$ can be reduced form $GL(n,\mathbb{R})$ to $Spin(n)$. Alternatively, the spin structure of an oriented Riemannian manifold allows us to lift the structure group of the manifold's tangent bundle from $SO(n)$ to $Spin(n)$.

To ascertain whether different manifolds possess certain structures, or, in other words, to detect the topological obstructions when reducing the structure group from $GL(n)$ to a subgroup, we can employ characteristic classes.

Acknowledgement

I am very grateful to my mentor Debin Liu for his help. As an EAP exchange student, I thoroughly enjoyed my time at UC Santa Barbara, and DRP was an unforgettable experience. Thank you all for your help!

$Spin$ Structure and $Spin^c$ Structure

It is well known that we can establish a bijective correspondence between $\mathbb{R}^2$ and $\mathbb{C}$ by selecting the basis vectors (e_1,e_2) in $\mathbb{R}^2$ and setting $e_2=ie_1$. Similarly, we can utilize quaternions to establish a one-to-one correspondence between $\mathbb{R}^4$ and 2×2 anti-Hermitian matrices.

$$Q(t,x,y,z)=\begin{pmatrix} t+iz & -x+iy \\ x+iy & t-iz \end{pmatrix}$$

Hence, we can define $Spin(4)=SU_+(2)\times SU_-(2)$ as the direct product of two copies of $SU(2)$, and rewrite the action of $SO(4)$ on $\mathbb{R}^4$ as the adjoint action of two copies of $SU(2)$ on the quaternions.

$$\rho_g(Q)=A_-Q(A_+)^{-1},\quad \text{where}\quad g=(A_-,A_+)\in Spin(4).$$

Furthermore, we can demonstrate that each element of $SO(4)$ correspond to two elements in $Spin(4)$, which is known as the double covering of $SO(4)$ by $Spin(4)$. Specifically, the identity element e of $SO(4)$ corresponds to the elements $\{(I,I),(I,I)\}$ in $Spin(4)$.

If we multiply both $SU(2)$ factors of $Spin(4)$ by the same factor $\lambda\in U(1)$, we obtain $Spin^c(4)$.

$$Spin^c(4)=\{(\lambda A_-,\lambda A_+),A_\pm\in SU(2),\lambda\in U(1)\}.$$

Thus, we can obtain short exact sequences for $Spin(4)$ and $Spin^c(4)$.

$$0\to\mathbb{Z}_2\to Spin(4)\to SO(4)\to 0$$
$$0\to\mathbb{Z}_2\to Spin^c(4)\to SO(4)\times U(1)\to 0$$

Instead of using adjoint representation, we can consider fundamental representation of $Spin(4)$ which acts on two copies of $\mathbb{C}^2$. We denote W_+ and W_- as the representation space which acted by $SU_+(2)$ and $SU_-(2)$. Similarly, the representation of $Spin^c(4)$ acts on $W_+\otimes L$ and $W_-\otimes L$.

Suppose that $\mathcal{M}$ is an $4d$ oriented Riemannian manifold. A $Spin$ structure is given by local trivialization with a collection of transition functions satisfying proper conditions such as cocycle condition.

$$\tilde{g}_{\alpha\beta}:U_\alpha\cap U_\beta\to Spin(4).$$

But since the spin group is the double cover of the $SO(4)$ group guaranteed by Riemannian structure, the cocyle condition is only satisfied up to $\mathbb{Z}_2$.

$$\tilde{g}_{\alpha\beta}\tilde{g}_{\beta\gamma}\tilde{g}_{\gamma\alpha}=\pm 1\quad\text{on}\quad U_\alpha\cap U_\beta\cap U_\gamma.$$

Thus, there should be some conditions for a Riemannian manifold to have a $Spin$ structure. And as we mentioned before, characteristic class might be a good choice. Topologists have found a nice necessary and sufficient condition for it: the second Stiefel-Whitney class of tangent bundle $w_2(T\mathcal{M})=0$.

Similarly, the manifold is said to have a $Spin^c$ structure if we have transition functions with proper condition.

$$\tilde{g}_{\alpha\beta}:U_\alpha\cap U_\beta\to Spin^c(4).$$

It comes to our attention that for $Spin^c(4)$ which has one more degree of freedom than $Spin(4)$. Thus we can expect that it is more likely to satisfy the cocycle condition by adjusting the overall phase comparing with $Spin$ structure.

Indeed, every compact oriented four-manifold possesses a $Spin^c$ structure.

Spin Connection and Dirac Operators

A element of Euclidean space V can be regarded as a element in $Hom(W_-,W_+)$. Then, we can define a map $\theta:V\otimes\mathcal{C}\to\mathrm{End}(W=W_+\oplus W_-)$

$$\theta(Q)=\begin{pmatrix} 0 & -Q^\dagger \\ Q & 0 \end{pmatrix}.$$

We can construct a set of basis in $\mathrm{End}(W)$ using Clifford algebra. Let's define complex 4×4 matrices satisfying relation $\{e_i,e_j\}=-2\delta_{ij}$. Here, $\{\}$ denotes anti-commutator. And these matrices are the image of basis of V. These matrices with matrix multiplication generate the basis of linear space $\mathrm{End}(W)$.

$$I,\quad e_i,\quad e_ie_j,\quad e_ie_je_k,\quad \text{for } i<j<k,\quad e_1e_2e_3e_4$$

Thus we can identify the exterior power of V and the complex subspace of $\mathrm{End}(W)$ by identifying the image of wedge product as Clifford multiplication. Indeed, we have a direct sum decomposition

$$\mathrm{End}(W)=\bigoplus_{k=0}^{4}\Lambda^kV.$$

With the direct sum decomposition, we can easily construct a connection on $\mathrm{End}(W)$ using Levi-Civita connection on $V=T^*\mathcal{M}$. Furthermore, one can claim that given a $Spin(4)$ connection on W, there is a unique corresponding connection on $\mathrm{End}(W)$ satisfies the Leibniz rule. Thus, we can give a unique $Spin(4)$ connection on W.

Similarly, for a $Spin^c$ manifold with a connection d_{2A} on the line bundle, there is also a unique connection on $W\otimes L$.

Also we can define a quadratic map $\sigma:W_+\to\Lambda^2_+V$ by

$$\sigma(\psi)=-\frac{i}{2}\sum_{i<j}(e_i\cdot e_j\cdot\psi,\psi)e_i\cdot e_j.$$

Here, Λ^2_+V is the self-dual form.

With a given connection d_A on $Spin^c$ bundle $W\otimes L$, we can define the Dirac operator $D_A:\Gamma(W\otimes L)\to\Gamma(W\otimes L)$

$$D_A(\psi)=\sum_{i=1}^{4}e_i\cdot\nabla^A_{e_i}\psi.$$

For the special case where $\mathcal{M}$ is a four-dimensional Euclidean space and L as the trivial line bundle, the Dirac operator becomes the same form as it in Dirac equation in physics $D_A\to\not{\partial}$.

The Dirac operator divides into two pieces connected by adjoint.

$$D_A^+:\Gamma(W_+\otimes L)\to\Gamma(W_-\otimes L),\quad D_A^-:\Gamma(W_-\otimes L)\to\Gamma(W_+\otimes L).$$

And the Seiberg-Witten equations are

$$D_A^+\psi=0,\quad F_A^+=\sigma(\psi)+\phi.$$

Here, $F_A^+=(1/2)F_{2A}^+$ and F_{2A}^+ is the self-dual part of the curvature on the line bundle L^2 (just the square of L). The solutions to the equations form the moduli space (A,ψ).

References

[1] John Douglas Moore.
Lectures on Seiberg-Witten Invariants.
Springer Berlin, Heidelberg, 2001.

An Introduction to Stochastic Calculus

John Lain and Dalina Sinn
2024 Mathematics Directed Reading Program - UC Santa Barbara

Probability Spaces

On a measurable space (Ω, F), the probability measure is $P : F \to [0, 1]$. The following conditions apply:

a. $P(\emptyset) = 0$, $P(\Omega) = 1$

b. if $A_1, A_2, ... \in F$ and $(A_i)_{i=1}^{\infty}$ is disjoint, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

A probability space contains (Ω, F, P) which the variables indicating:

- P is the exact probability measure
- Ω is the a space with all the possible outcomes
- F is the collection of possible events where each event is a subset of Ω

Stochastic Processes and Brownian Motion

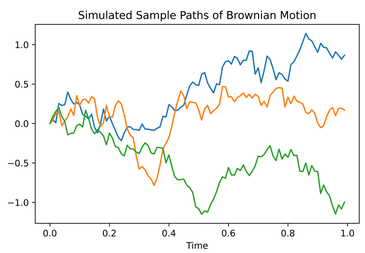
A stochastic process is a parameterized collection of random variables $\{X_t\}$, defined on a probability space (Ω, F, P) with values in $\mathbb{R}^n$. We often have $t \in [0, \infty)$ for the case of continuous stochastic processes. This can be thought of a function of time, where the outcome at each time is a random variable.

Brownian Motion

Brownian motion was observed by botanist Robert Brown while studying pollen grains, which moved in liquid in a jittery motion. This movement can be described mathematically by a 2 dimensional Brownian motion.

A sequence of random variables, B_t, for $t \geq 0$, is defined as a standard Brownian motion if:

1. $B_0 = 0$
2. B_t has continuous sample paths
3. For every t and s, with $s < t$, we have that $B_t - B_s$ is has a normal distribution with variance $t - s$ and mean 0.
4. The distribution of $B_t - B_s$ is independent of the behavior of B_r, for $r < s$. The result of these properties is that Brownian Motion has independent, stationary increments with mean zero.



Quadratic Variation

Let $0 = t_0 < t_1 < t_2 < ... < t_n = T$ be a partition of a time interval $[0, T]$. For some stochastic process X_t, let

$$Q_n(T, X) = \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2$$

The quadratic variation of X_t on the interval is the limit of $Q_n(T, X)$ as n gets large (or as Δt gets small). For a Brownian motion $\{B_t\}_{t \geq 0}$ the quadratic variation is equal to T with probability 1, as the expected value of quadratic variation is T and the limit of the variance of $Q_n(T, B)$ approaches 0. One can show that the total variation of the path is infinite, with probability 1, and the paths $t \to B_t$ of Brownian motion are nowhere differentiable. The total variation of a process, X_t, on $[0, T]$, is defined as

$$\lim_{\Delta t \to 0} \sum_{i=1}^{n-1} |X_{t_{i+1}} - X_{t_i}|$$

Stochastic Integration and the Itô Integral

We will now define the Itô Integral of a stochastic process (under certain conditions that we omit for simplicity). It is possible to generalize the following definitions to multiple dimensions, but we will only focus on the one dimensional case. It is also important to note that this integral is a random variable.

Definition of the Itô Integral

A elementary function h has the form $h_t = \sum_i e_i I_{[t_i, t_{i+1})}(t)$, where I is the indicator function. Note that h_t is a piece wise continuous random process. We define the Itô integral of elementary functions as

$$\int_S^T h_t dB_t = \sum_{i=0}^{n-1} e_i (B_{t_{i+1}} - B_{t_i})$$

Then, for a more general process, X_t, we define

$$\int_S^T X_t dB_t = \lim_{n \to \infty} \int_S^T X_t^{(n)} dB_t$$

where $X_t^{(n)}$ is a elementary function such that

$$\mathbb{E}[\int_S^T (X_t - X_t^{(n)})^2 dt] \to 0 \text{ as } n \to 0$$

When computing the Itô integral, the e_i term for our elementary process $X_t^{(n)}$ becomes X_{t_i}, which is a left endpoint definition. Another seemingly reasonable choice would be to use to use $X_{t_{i+1}}$, which is the right endpoint. Under the Riemann-Stieltjes integral for a real valued function, this choice does not change the result of the integral. However, due to the large variations of the paths of B_t this choice results in different solutions to the integrals. The choice of a midpoint $\frac{t_i + t_{i+1}}{2}$ leads to the Stratonovich integral, which has different properties than the Itô integral. For example, the Itô integral is a martingale whereas the Stratonovich integral is not.

Computing the Itô integral of Brownian motion: As an example, we compute the value of $\int_0^T B_t dB_t$. Let $B_t^{(n)} = \sum_{i=0}^{n} B_{t_i} I_{[t_i, t_{i+1})}(t)$, then

$$\begin{aligned}
\int_0^T B_t dB_t &= \lim_{n \to \infty} \int_0^T B_t^{(n)} dB_t && (1)\\
&= \lim_{n \to \infty} \sum_{i=0}^{n-1} B_{t_i}(B_{t_{i+1}} - B_{t_i}) && (2)\\
&= \lim_{n \to \infty} (\frac{1}{2} B_T^2 - \frac{1}{2} \sum_{i=0}^{n-1} (B_{t_{i+1}} - B_{t_i})^2) && (3)\\
&= \frac{1}{2} B_T^2 - \frac{1}{2} T && (4)
\end{aligned}$$

because the quadratic variation of Brownian motion is T almost surely. In line (3) we also use that:

$$B_T^2 = \sum_{i=0}^{n-1} (B_{t_{i+1}}^2 - B_{t_i}^2) = \sum_{i=0}^{n-1} ((B_{t_{i+1}} - B_{t_i})^2 - 2B_{t_i}(B_{t_{i+1}} - B_{t_i}))$$

Properties of the Itô Integral:

For f and g that are stochastic processes, and $0 \leq S < U < T$, we have

i. $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$

ii. $\int_S^T (cf + g) dB_t = c \int_S^T f dB_t + \int_S^T g dB_t$, where c is a constant

iii. $\mathbb{E}[\int_S^T f dB_t] = 0$

iv. $\int_S^T f dB_t$ is F_T-measurable

(Intuitively, a function is F_t-measurable if its value can be determined from the path of a Brownian Motion up to t. For example, B_{2t} is not F_t-measurable.)

Itô Processes and Stochastic Differential Equations

An Itô process X_t is a stochastic process that can be written as

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t b_s dB_s,$$

with special conditions on a_t and b_t which are random functions of time. We can write the above equality in a shorthand differential form:

$$dX_t = a_t dt + b_t dB_t$$

Itô's Lemma

Given a 1-dimensional Itô Process X_t and $f(t, x) : [0, \infty) \times \mathbb{R} \to \mathbb{R}$ a twice continuously differentiable function, if $Z_t := f(t, X_t)$, then we have that:

$$\begin{aligned}
dZ_t &= \frac{\partial f}{\partial t}(t, X_t) dt + \frac{\partial f}{\partial x}(t, X_t) dX_t + \frac{1}{2} \frac{\partial^2 f}{\partial x^2}(t, X_t)(dX_t)^2\\
&= \left(\frac{\partial f}{\partial t}(t, X_t) + a_t \frac{\partial f}{\partial x}(t, X_t) + \frac{1}{2} b_t^2 \frac{\partial^2 f}{\partial x^2}(t, X_t) \right) dt + b_t \frac{\partial f}{\partial x}(t, X_t) dB_t
\end{aligned}$$

Stochastic Differential Equations

A stochastic differential equation (SDE) describes a stochastic process which is equal to an Itô integral of a function of that process. A SDE has the form:

$$X_t = X_0 + \int_0^t a(X_s, s) ds + \int_0^t b(X_s, s) dB_s$$

This can be written in differential notation as:

$$dX_t = a(X_t, t) dt + b(X_t, t) dB_t; \; X_0 = x$$

Geometric Brownian Motion: A SDE which models asset prices in finance is

$$S_T = S_0 + \int_0^T r S_t dt + \int_0^T \sigma S_t dB_t$$

which is written in differential form as

$$dS_t = r S_t dt + \sigma S_t dB_t$$

This model, called geometric Brownian motion, describes a stochastic process which grows at a rate of r plus some random "noise". In finance, this r term represents an interest rate and σ represents the volatility of the asset. If we let $Y_t = Y_0 e^{(r - \frac{1}{2}\sigma^2)t + \sigma B_t}$ and apply Itô's Lemma, we will see that Y_t satisfies the SDE described above.

$$\begin{aligned}
dY_t &= (r - \frac{1}{2}\sigma^2) Y_0 e^{(r - \frac{1}{2}\sigma^2)t + \sigma B_t} dt + \sigma Y_0 e^{(r - \frac{1}{2}\sigma^2)t + \sigma B_t} dB_t && (5)\\
&\quad + \frac{1}{2}\sigma^2 Y_0 e^{(r - \frac{1}{2}\sigma^2)t + \sigma B_t} (dB_t)^2) && (6)\\
&= (r Y_0 e^{(r - \frac{1}{2}\sigma^2)t + \sigma B_t} dt + \sigma Y_0 e^{(r - \frac{1}{2}\sigma^2)t + \sigma B_t} dB_t) && (7)\\
&= r Y_t dt + \sigma Y_t dB_t && (8)
\end{aligned}$$

where we use that $(dB_t)^2 = dt$. So, $Y_t = Y_0 e^{(r - \frac{1}{2}\sigma^2)t + \sigma B_t}$ solves the SDE.

Acknowledgements

Special thanks to Daniel Ralston for his mentorship and to the UCSB Directed Reading Program for this opportunity.

References

M Haugh. *A Brief Introduction to Stochastic Calculus*. 2016.

B Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 2003.

Arithmetic Functions in Analytic Number Theory

Garrett Kay, Ryan Ashraf, Baiming Wang

University of California Santa Barbara

Interesting Arithmetic Functions

"A real- or complex-valued function defined on the positive integers is called an arithmetical function or a number-theoretic function." [1] Here, we will explore several arithmetical functions which help us better understand the properties of the natural numbers.

Mobius Function

$$\mu(1) = 1$$

For n > 1, write $n = p_1^{\alpha_1} p_2^{\alpha_2} ... p_k^{\alpha_k}$

$\mu(n) = (-1)^k$, if $a_1 = a_1 = ... = a_k = 1$
$\mu(n) = 0$, otherwise

Something interesting occurs when we sum μ over the divisors of a number n. For $n \geq 1$,

$$\sum_{d|n} \mu(d) = \left\lfloor \frac{1}{n} \right\rfloor = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{otherwise} \end{cases}$$

Euler Totient Function

This function counts the positive integers up to n, that are relatively prime to n.

$$\phi(n) = \sum_{k=1,(k,n)=1}^{n} 1$$

Taking the sum of ϕ over the divisors of n gives, for $n \geq 1$,

$$\sum_{d|n} \phi(d) = n$$

Mangoldt Function

$$\Lambda(1) = 0$$

$\Lambda(n) = log(p)$, if $n = p^k$ for some prime p and $m \geq 1$
$\Lambda(n) = 0$, otherwise

Taking the sum of Λ over divisors of n gives, for $n \geq 1$

$$\sum_{d|n} \Lambda(d) = log(n)$$

Liouville Function

$$\lambda(1) = 1$$

For n > 1, write $n = p_1^{\alpha_1} p_2^{\alpha_2} ... p_k^{\alpha_k}$

$$\lambda(n) = (-1)^{a_1 + a_2 + ... + a_k}$$

Taking the sum of λ over divisors of n gives, for $n \geq 1$

$$\sum_{d|n} \lambda(d) = \begin{cases} 1 & \text{if n is a square} \\ 0 & \text{otherwise} \end{cases}$$

Divisor Function

$$\sigma_\alpha(n) = \sum_{d|n} d^\alpha$$

We can define the divisor function for any complex number α. The two most common ones are for $\alpha = 0$ and $\alpha = 1$. They are represented as $d(n) = \sigma_0(n)$, which counts the number of divisors of n and $\sigma(n) = \sigma_1(n)$, which is the sum of the divisors of n.

Introduction to Analytic Number Theory

Analysis and Number Theory at first glance seem antithetical. Analysis is concerned with the continuous, while Number Theory is concerned with the discrete. However, mathematicians never take things at first glance. An active field of research for centuries, Analytic Number Theory has employed the work of mathematical giants such as Euclid, Gauss, Dirichlet, Riemann, and Chebyshev. Many problems remain unsolved today, the most famous being the Riemann hypothesis, rendering this sect of mathematics a curious field for newcomers. Beyond providing deep insights into the structure of primes and integers, Analytic Number Theory's rich tapestry extends to algebraic geometry and theoretical physics, making it evident that these topics are at the crux of many contemporary problems. In this project, we discuss some of the critical results of Analytic Number Theory that hold up much of the modern research in the field.

Average Order of Arithmetic Functions

We introduce an important notation, the "big-oh." For real functions f and g, we denote

$$f(x) = O(g(x))$$

If there exists some $M > 0$ and a such that $|f(x)| \leq Mg(x), \forall x \geq a$. This gives us a good way to compare the growth rates of functions.
If f has a continuous derivative f' on the interval $[y, x]$ with $0 < y < x$, then

$$\sum_{y<n\leq x} f(n) = \int_y^x f(t)dt + \int_y^x (t - [t])f'(t)dt - f(x)(x - [x]) + f(y)(y - [y])$$

This is Euler's summation formula, which allows us to approximate a finite or infinite sum with integrals. Using this formula, we obtain the following results: for $x \geq 1$,

$$\sum_{n\leq x} \frac{1}{n} = \log x + C + O(\frac{1}{x})$$
$$\sum_{n\leq x} \frac{1}{n^s} = \frac{x^{1-s}}{1-s} + \zeta(s) + O(x^{-s}) \text{ if } s > 0, s \neq 1$$
$$\sum_{n\leq x} \frac{1}{n^s} = O(x^{1-s}) \text{ if } s > 1$$
$$\sum_{n\leq x} \frac{1}{n^s} = \frac{x^{1-s}}{1-s} + O(x^{-s}) \text{ if } s \leq 0$$

where $C = \lim_{n\to\infty}(1 + \frac{1}{2} + ... + \frac{1}{n} - \log n)$ and ζ is the Riemann-zeta function.
We can now derive the asymptotic formula for the partial sums of $\sigma_\alpha(n)$. For all $x > 1$,

$$\sum_{n\leq x} \sigma_0(n) = \sum_{n\leq x} d(n) = x \log x + (2C - 1)x + O(\sqrt{x})$$
$$\sum_{n\leq x} \sigma_1(n) = \sum_{n\leq x} \sigma(n) = \frac{1}{2}\zeta(2)x^2 + O(x \log x)$$
$$\sum_{n\leq x} \sigma_\alpha(n) = \frac{\zeta(\alpha + 1)}{\alpha + 1} x^{\alpha+1} + O(x^\beta) \text{ , where } \alpha > 0, \alpha \neq 1, \beta = \max\{1, \alpha\}$$
$$\sum_{n\leq x} \sigma_{-1}(n) = \zeta(2)x + O(\log x)$$
$$\sum_{n\leq x} \sigma_\alpha(n) = \zeta(1 - \alpha)x + O(x^\delta) \text{ , where } \alpha < 0, \alpha \neq -1, \delta = \max\{0, 1 + \alpha\}$$

We may also obtain an asymptotic formula for the Euler function, $\varphi(n)$. For $x > 1$, we have

$$\sum_{n\leq x} \varphi(n) = \frac{3}{\pi^2} x^2 + O(x \log x)$$

With the above result, we may show that
Theorem The set of lattice points visible from the origin has density $\frac{6}{\pi^2}$.

The Distribution of Prime Numbers

Counting the prime numbers and finding the distribution they hold across the integers has been studied extensively for many centuries, Prime Number Theorem currently being one of the crown jewels of all of these mathematical endeavours. To start, we must begin with the analysis of certain arithmetical functions such as Chebyshev's $\theta(x)$ and $\psi(x)$ functions. Chebyshev's funtions are defined as:

$$\vartheta(x) := \sum_{p\leq x} \log p \qquad \psi(x) := \sum_{n\leq x} \Lambda(n)$$

The majority of our results come from observing the asymptotic behaviours of said functions, those being:

$$\lim_{x\to\infty} \left(\frac{\psi(x)}{x}\right) = 1 \qquad \lim_{x\to\infty} \left(\frac{\theta(x)}{x}\right) = 1$$

We also take use of Abel's Identity, which is a powerful method of relating these arithmetical functions:

$$\sum_{y<n\leq x} a(n)f(n) = A(x)f(x) - A(y)f(y) - \int_x^y A(t)f'(t)\,dt$$

Now employing our **vast** knowledge of the subject matter and a little bit of rudimentary algebra, we get the following equivalences:

$$\lim_{x\to\infty} \left(\frac{\pi(x)log(x)}{x}\right) = 1 \qquad \lim_{x\to\infty} \left(\frac{\pi(x)log(\pi(x))}{x}\right) = 1 \qquad \lim_{x\to\infty} \left(\frac{p_n}{n\,log(n)}\right) = 1$$

where $\pi(x)$ represents the prime number counting function and p_n is the n^{th} prime number. All of the limit relations stated are equivalent to Prime Number Theorem, meaning showing one of these relations proves all. From all of this analysis, we get a few rewards in the form of the Shapiro's Theorem, partial sums of the Möbius function, Selberg's asymptotic formula, etc. With these tools we can know draw a sketch for an elementary proof for Prime Number Theorem. We start by defining the function

$$\sigma(x) := e^{-x}\psi(e^x) - 1$$

And now with Selberg's formula we can characterize this as

$$|\sigma(x)|x^2 \leq 2\int_0^x \int_0^y |\sigma(u)|\,du\,dy + O(x)$$

Because we have shown that prime number theorem is equivalent to showing $\sigma(x) \to 0$ as $x \to \infty$ if we show that

$$C := \limsup_{x\to\infty} |\sigma(x)| = 0$$

we are done. Now, let $C > 0$ by assumption and by definition we get that

$$|\sigma(x)| \leq C + g(x) \qquad \text{where } g(x) \to 0 \text{ as } x \to \infty$$

Our earlier characterization of $\sigma(x)$ gives a similar inequality

$$|\sigma(x)| \leq C' + h(x) \qquad \text{where } 0 < C' < C \text{ and } h(x) \to 0 \text{ as } x \to \infty$$

Letting $x \to \infty$ we find that $C \leq C'$, meaning we have run into a contradiction.

Acknowledgements

We thank our mentor Rusiru Gambheera for his guidance as well as the UCSB Directed Reading Program for the opportunity to work on this project.

References

[1] Tom M. Apostol. *Introduction to Analytic Number Theory*. Pasadena, CA: Springer-Verlag, 1976.

Cracking the Code: How to Break Encryption

Riley Paddock, Mentor - Mitchell Jubeir

2024 DRP, UCSB

Cryptography Before Computers

As long as people have needed to share secrets, there has been some **cryptosystem**, a way of encrypting a message so no one except specific parties with a **key** can decrypt it. Then, as long as there have been cryptosystems, there have been **attacks** to decrypt these messages without a key. Here are some of the original examples of cryptosystems. Note, we encode letters of the alphabet as numbers modulo 26 so we can add them and describe permutations of them.

1. Caesar Cipher - Shift every letter by some fixed number k. We denote this encryption by ϕ_k. Here's an example:
$$\phi_1(SECRET) = TFDSFU$$
2. Substitution Cipher - Create some random permutation of your alphabet and apply that permutation to each letter.
$$\sigma \in S_{26} \quad SECRET \to \sigma(S)\sigma(E)\sigma(C)\sigma(R)\sigma(E)\sigma(T)$$
3. Vigenère Cipher - Use a repeated key as a more complex Caesar cipher:
$$\begin{array}{r} SECRET \\ +KEYKEY \\ \hline CIABIR \end{array}$$

The key to breaking these ciphers was **frequency analysis** and the **index of coincidence**.

Frequency Analysis and Index of Coincidence

Frequency Analysis

Comparing the frequencies of letters and letter combinations in normal language to those of a ciphertext to break a substitution cipher.

For example, e is the most common English letter with a frequency of 11.16%. If we had a ciphertext from a substitution cipher where g has a frequency of around 11.6% we can deduce that $\sigma(e) = g$.

Even though a substitution cipher has $26! \approx 4 \cdot 10^{26}$ keys, but frequencey makes solving the key quite easy.

Index of Coincidence

Notated I_C, it is the probability that two randomly chosen letters from a sample are the same.

If we had a true random sample then $I_R = 1/26 \approx 0.0385$, but English isn't random so we get $I_E = 0.0656$. Since this only depends on probabilities, it is invariant under a substitution cipher. With this fact we can find the key size of a vigenere cipher:

$$key\ len = \min_{k \in \mathbb{N}} \left| I_E - \frac{\sum_k I_k}{k} \right|$$

This works because if we have the right key size, each group should be a Caesar cipher, so the I_C should be close to I_E for every group, so our average should be close to zero.

The Enigma

The enigma was a complex machine that performed multiple substitution ciphers on each character, and then changed the substitutions with each letter! In it's military usage by the Nazi's the number of possible keys was...

$$159,000,000,000,000,000,000$$

This number on it's own is infeasible even for modern computers. But through some clever work by British cryptographers this task could be broken into two parts. One which was broken by the first computers and Alan Turing and the other could be broken by **frequency analysis**

Cryptograhy With Computers and Public Keys

As the technology to communicate, encrypt, and decrypt progressed, there became a need to have a secure exchange with someone you have not met, with no pre-established key. Ex. sending credit card info to an online store.

This led to the development of **Public Key Cryptography**. These are what we use today, and their security relies on "hard problems" which we assume people can't solve. So if we can solve the following "hard problems" we can decrypt the cipher.

Discrete Log/ Diffie-Hellman Problem: "Given a cyclic group G, a generator g, and an element x, find n such that $x = g^n$."

Factoring Problem: "Given $n = p \cdot q$, find primes p and q"

Index Calculus

If we wish to find x such that $g^x \equiv h \ (mod\ p)$ We start by picking a factor base, which is typically chosen to be the first r primes: $F = \{2, 3, 5, 7, ..., p_r\}$. Then, we search for values of k such that:

$$g^k (mod\ p) = 2^{e_2} 3^{e_3} 5^{e_5} \dots p_r^{e_{p_r}}$$

Such values of k are rare but is discoverable. We keep track of the $g^k (mod\ p)$ which do factor in our base. Once we have enough of these relations, we can solve a system of equations to determine, l_f for every $f \in F$ such that

$$g^{l_f} \equiv f \ (mod\ p)$$

Once that is done, we start again, trying possible m such that:

$$g^m h (mod\ p) = 2^{e_2} 3^{e_3} 5^{e_5} \dots p_r^{e_{p_r}}$$

If we find a factorization of such an element, then it is easy to compute the discrete logarithm of h because we know all the e_i and the l_i, and once we have an m then we can finish with the following:

$$\begin{aligned} g^m h (mod\ p) = g^{m+x} (mod\ p) &\equiv 2^{e_2} 3^{e_3} 5^{e_5} \dots p_r^{e_{p_r}} \\ g^{m+x} (mod\ p) &\equiv g^{l_2 e_2} g^{l_2 e_3} g^{l_5 e_5} \dots g^{l_{p_r} e_{p_r}} \\ x &= -m + l_2 e_2 + l_3 e_3 + \dots + l_{p_r} e_{p_r} \end{aligned}$$

The Quad-Sieve

The quad-sieve tries to factor numbers by trying to describe them as a difference of squares. For example, we can factor 899 as follows:

$$899 = 900 - 1 = 30^2 - 1^2 = (30 - 1)(30 + 1) = 29 \cdot 31$$

We start by choosing some bound B and we want to consider all the primes smaller than B. We denote the number of primes smaller than B as $\pi(B)$. Then we find $\pi(B) + 1$ numbers a_i such that $b_i \equiv (a_i)^2 \ (mod\ n)$ and b_i only has prime factors smaller than B.

$$b_i = 2^{i_2} 3^{i_2} \dots p_{\pi(B)}^{i_{\pi(B)}}$$

If we can find two or more b_i and b_j which multiply to a square then we have that $b_i b_j = c^2$ and $b_i b_j \equiv (a_i a_j)^2 \ (mod\ n)$ therefore we have that $c^2 \equiv (a_i a_j)^2 \ (mod\ n)$ which means:

$$(c - a_i a_j)(c + a_i a_j) = k \cdot n$$

To find which of these b_i we can multiply to a square we use linear algebra. We can describe each of our b_i in terms of the $\pi(B)$ prime factors and we have $\pi(B) + 1$ of them so our matrix must have a linear dependence, which will be a product that gives a square.

References

Simon Rubenstein-Salzedo
Cryptography
Springer, 2010

Jason Howell
The Index Calculus Algorithm
Clemson University, 1998

Simon Singh
The Code Book
Anchor Books, 2000

Quantum Cryptography

Quantum computing shakes up classical cryptography with two major results; "perfect" key exchange and Shor's algorithm. Before we can describe those, here's some background on **quantum computing**.

Where classical computers use a bit being a 0 or 1, quantum computers use qubits which can be anywhere in between. Let q be a qubit, then we can describe it in some basis:

$$q = a|0\rangle + b|1\rangle \quad a, b \in \mathbb{C} \quad a^2 + b^2 = 1$$

But we can also describe it in terms of some other basis:

$$q = c|+\rangle + d|-\rangle \quad c, d \in \mathbb{C} \quad c^2 + d^2 = 1$$

Importantly if we "observe" our qubit we can only do so in terms of one basis. So either$(|+\rangle, |-\rangle)$ or $(|0\rangle, |1\rangle)$ and if we do this and find out that $q = |+\rangle$ then the other basis resets to $q = \frac{1}{\sqrt{2}}|1\rangle + \frac{1}{\sqrt{2}}|0\rangle$. This property of "resetting" is the basis of quantum key exchange.

Quantum Key Exchange

Alice starts by taking two random strings of 0's and 1's. Then she uses the first set to pick between the two possible basis $(|+\rangle, |-\rangle)$ vs $|1\rangle, |0\rangle)$, and the second to pick which state $(|+\rangle, |0\rangle)$ or $(|-\rangle, |1\rangle)$ this describes a string of qubits, for example:

$$\begin{cases} \text{Basis Choice:} & 110010 \\ \text{State Choice:} & 101000 \end{cases} \longrightarrow |1\rangle|0\rangle|-\rangle|+\rangle|0\rangle|+\rangle$$

After Bob observes the qubits, Alice shares her basis bits. This way Bob and Alice know which of the qubits they have are the same, therefore Bob can deduce what Alice's state bits are for the places where the basis bits match. These matching bits are the key!

If Eve had intercepted Alice's qubits, she can't make a copy and wait for the basis bits since you can't clone qubits. So Eve makes random guesses like bob. After Bob and Alice have their matching state bits, they can share a couple of them, if they ever share two bits that are different, they know Eve must have observed that qubit in the wrong basis and reset it's value.

Shor's Algorithm

Shor's algorithm uses a subroutine that can find the order of an element. So given an abelian group G and a element $g \in G$ find $x \in \mathbb{N}$ such that $g^x = 1$

The algorithm starts by describing a periodic function as a sum of many simpler functions. (A quantum version of the Fourier transform). Then it maximizes a particular inner product on this Fourier transform. Similar to the index of coincidence, this maximum value will be where all the components of your Fourier transform have similar values, which will be the period of your function!

Suppose we wanted to factor $n \in \mathbb{Z}$. Choose a random:

$$a \in \mathbb{Z}/n\mathbb{Z}$$

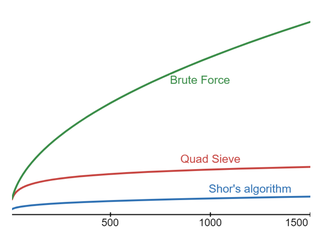
Then we use the above to get r such that $a^r \equiv 1 \ (mod\ n)$. If r is odd, start over. If r is even then we have that

$$\left(a^{r/2}\right)^2 \equiv 1 \ (mod\ n)$$

$$(a^{r/2} - 1)(a^{r/2} + 1) = kn$$

Unless a factor is 1, which has low odds, we have factored n.

Runtime of factoring algorithms
# of bits vs runtime



Curvature Done Optimally

Dongming (Merrick) Hua [1]
Mentor: Fabio Ricci [1]

[1]University of California - Santa Barbara
Department of Mathematics - Directed Reading Program 2024

Introduction

Curvature is a central concept in Riemannian geometry, and bounds on the various curvatures of a manifold M translate into useful constraints on the geometry and topology of M. In particular, lower bounds on the Ricci curvature Ric_M of M play a key role in many important theorems.

We discuss an alternate notion of "curvature bounded below by K" for compact Riemannian manifolds, which only involves the distance on the manifold and the volume measure of the manifold. We will show that in the Riemannian setting, a manifold has Ricci curvature $\geq K$ if and only if it satisfies this alternate condition. This new definition does not explicitly use the Riemannian structure, and thus can be generalized to a broader, nonsmooth class of metric measure spaces.

Comparison Geometry

To get an idea of manifolds with $\mathrm{Ric}_M \geq K$, we can look at the model spaces M_K^n of constant sectional curvature K, where

$$M_K^n = \begin{cases} \text{the sphere } S^n(K), & K > 0 \\ \text{Euclidean space } \mathbb{R}^n, & K = 0 \\ \text{hyperbolic space } \mathbb{H}^n(K), & K < 0. \end{cases}$$

Intuitively, in positive curvature geodesics diverge then converge, in zero curvature they diverge at a constant rate, and in negative curvature they diverge increasingly rapidly.

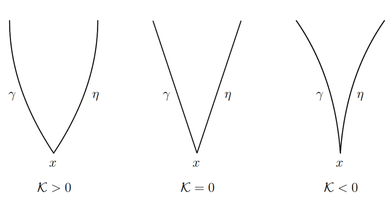


Figure 1. Geodesics in positive, zero, and negative curvature. Image from [3].

Optimal Transport

Optimal transport studies the most efficient way to transport some amount of mass from one configuration to another, such as moving a pile of sand to build a sandcastle:

We can view the configurations of masses as probability measures μ and ν on M and measure efficiency via minimizing the cost

$$\int_M d(x, T(x))^2 \mathrm{d}\mu(x)$$

over maps $T\colon M \to M$ satisfying $T_\sharp \mu = \nu$. McCann showed that if M is compact and $\mu = \rho_0 \,\mathrm{vol}, \nu = \rho_1 \,\mathrm{vol}$, where vol is the normalized volume measure on M, then there exists a unique **optimal transport map** T minimizing the above cost. Moreover, T is of the form

$$T(x) = \exp_x(\nabla\varphi(x))$$

for some semiconvex $\varphi\colon M \to \mathbb{R}$, and the Jacobian determinant of T at x is equal to $\rho_0(x)/\rho_1(T(x))$ μ-almost everywhere.

The Wasserstein 2-Distance

Let $\mathcal{P}_2^{ac}(M)$ be the set of all probability measures on a compact manifold M which are **absolutely continuous** with respect to vol, meaning measures for which we can write $\mu = \rho \,\mathrm{vol}$ for some density ρ. We can give this space a metric by defining

$$W_2(\mu, \nu) = \left(\int_M d(x, T(x))^2 \mathrm{d}\mu(x) \right)^{\frac{1}{2}},$$

where T is the optimal transport map from μ to ν. This distance is called the **2-Wasserstein distance**, and can be defined more generally for metric measure spaces via an alternate formulation of the optimal transport problem.

By the work of McCann, for any two measures $\mu_0, \mu_1 \in \mathcal{P}_2^{ac}$ there is a unique Wasserstein geodesic $(\mu_t)_{0\leq t\leq 1}$ between them, so that $W_2(\mu_s, \mu_t) = |t-s| W_2(\mu_0, \mu_1)$ for all $0 \leq s, t \leq 1$. Moreover, there must exist $T\colon [0,1] \times M \to M$ given by

$$T(t, x) = \exp_x(t\nabla\varphi(x)),$$

so that for each $t \in [0,1]$, the map $T_t\colon M \to M$ defined by $T_t(x) = T(t,x)$ is the optimal transport map from μ_0 to μ_t. Therefore, we can write $\mu_t = (T_t)_\sharp \mu_0$, and for each t the Jacobian determinant of T_t at x is $\rho_0(x)/\rho_t(T_t(x))$ μ_0-almost everywhere. Observe that by properties of the exponential map, $d(x, T_1(x)) = |\nabla\varphi(x)|$ for all x, hence

$$W_2(\mu_0, \mu_1)^2 = \int_M |\nabla\varphi|^2 \mathrm{d}\mu.$$

Optimal Transport Maps and Ricci Curvature

The key connection between optimal transport on M and the Ricci curvature of M is that Ricci curvature features in a differential inequality for the Jacobian determinant of the map $T(t,x) = \exp_x(t\nabla\varphi(x))$.

Let $J_t(x)$ be the Jacobian of T_t at x, and let $\mathcal{J}_t(x) = \det J_t(x)$. We can write $\mathcal{J}_t(x)$ in terms of Jacobi fields along the geodesic $\gamma(t) = \exp_x(t\nabla\varphi(x))$. Then, via the Jacobi equation and the Cauchy-Schwarz inequality, we obtain the inequality

$$\frac{\mathcal{J}_t''}{\mathcal{J}_t} - \left(\frac{\mathcal{J}_t'}{\mathcal{J}_t}\right)^2 \leq \frac{\mathcal{J}_t''}{\mathcal{J}_t} - \left(1 - \frac{1}{n}\right)\left(\frac{\mathcal{J}_t'}{\mathcal{J}_t}\right)^2 \leq -\mathrm{Ric}(\nabla\varphi, \nabla\varphi).$$

Curvature Bounded Below by K

For $\mu \in \mathcal{P}_2^{ac}(M)$, define the **entropy** of μ by

$$H(\mu) = \int_M \rho \log \rho \,\mathrm{d}\,\mathrm{vol},$$

where $\mu = \rho \,\mathrm{vol}$. We say M has **curvature bounded below by K** if for any measures $\mu_0, \mu_1 \in \mathcal{P}_2^{ac}(M)$, the unique Wasserstein geodesic $(\mu_t)_{0\leq t\leq 1}$ satisfies

$$H(\mu_t) \leq (1-t)H(\mu_0) + tH(\mu_1) - K\frac{t(1-t)}{2} W_2(\mu_0, \mu_1)^2$$

for all $0 \leq t \leq 1$.

References

[1] Luigi Ambrosio and Nicola Gigli. *A User's Guide to Optimal Transport*. Springer, Berlin, Heidelberg, 2013.

[2] Xianzhe Dai and Guofang Wei. *Comparison geometry for ricci curvature*.

[3] Shin-ichi Ohta. Ricci curvature, entropy and optimal transport. 2014.

[4] Karl-Theodor Sturm. On the geometry of metric measure spaces. 2006.

[5] Cedric Villani. *Optimal Transport: Old and New*. Springer, Berlin, Heidelberg, 2008.

The Main Theorem

Theorem 1 (Equivalence of $\mathrm{Ric}_M \geq K$ and Curvature Bounded Below By K). *A compact Riemannian manifold M^n satisfies $\mathrm{Ric}_M \geq K$ if and only if it has curvature bounded below by K.*

Proof. First, suppose $\mathrm{Ric}_M \geq K$. Let $\mu_0, \mu_1 \in \mathcal{P}_2^{ac}(M)$ be arbitrary, and let $T(t,x) = \exp_x(t\nabla\varphi(x))$ the map associated with the unique Wasserstein geodesic $(\mu_t)_{0\leq t\leq 1}$. Then we have $\rho_t(T_t(x)) = \rho_0(x)/\mathcal{J}_t(x)$, and so by this change of variables we have

$$\frac{d}{dt}H(\mu_t) = \frac{d}{dt}\int_M \frac{\rho_0}{\mathcal{J}_t} \log\frac{\rho_0}{\mathcal{J}_t} \mathcal{J}_t \mathrm{d}\,\mathrm{vol} = -\int_M \frac{\mathcal{J}_t'}{\mathcal{J}_t}\rho_0 \mathrm{d}\,\mathrm{vol},$$
$$\frac{d^2}{dt^2}H(\mu_t) = -\int_M \left(\frac{\mathcal{J}_t''}{\mathcal{J}_t} - \left(\frac{\mathcal{J}_t'}{\mathcal{J}_t}\right)^2\right)\rho_0 \mathrm{d}\,\mathrm{vol}\,.$$

By the differential inequality for $\mathcal{J}$, we have

$$\frac{d^2}{dt^2}H(\mu_t) \geq \int_M \mathrm{Ric}(\nabla\varphi, \nabla\varphi)\mathrm{d}\,\mathrm{vol} \geq K\int_M |\nabla\varphi|^2 \mathrm{d}\,\mathrm{vol} = KW_2(\mu_0, \mu_1)^2.$$

Defining $f(t) = H(\mu_t) + K\frac{t(1-t)}{2}W_2(\mu_0, \mu_1)^2$, we see that $\frac{d^2}{dt^2}f(t) = \frac{d^2}{dt^2}H(\mu_t) - KW_2(\mu_0, \mu_1)^2 \geq 0$. Therefore f is convex, and for any $t \in [0,1]$ we have

$$H(\mu_t) + K\frac{t(1-t)}{2}W_2(\mu_0, \mu_1)^2 = f(t) \leq (1-t)f(0) + tf(1) = (1-t)H(\mu_0) + tH(\mu_1),$$

as desired.

The other direction is more complicated, so we only sketch an outline here. A more detailed account can be found in [5]. Suppose M has curvature bounded below by K. Let $x \in M$, $v \in T_xM$ be arbitrary. Take μ_0 to be the normalized volume measure on a small ball $B_\varepsilon(x)$, and take the transport map to be $T_t(x) = \exp_x(t\delta\nabla\varphi(x))$ for some small δ and suitable φ satisfying $\nabla(\varphi(x_0)) = v$. From this we obtain a Wasserstein geodesic $(\mu_t)_{0\leq t\leq 1}$ given by $\mu_t = (T_t)_\sharp\mu_0$, and choosing ε, δ, and φ carefully, the inequality for $H(\mu_t)$ gives us the desired Ricci curvature bound $\mathrm{Ric}(v,v) \geq K|v|^2$. □

Brunn-Minkowski

One geometric consequence of curvature $\geq K$ is the following Brunn-Minkowski inequality, generalizing the classic Brunn-Minkowski inequality in $\mathbb{R}^n$.

Theorem 2 (Generalized Brunn-Minkowski). *Suppose that M has curvature bounded below by K. For nonempty, compact $A_0, A_1 \subseteq M$ and $t \in (0,1)$, let A_t be the set of all points $\gamma(t)$, where γ runs over all unit-length geodesics with $\gamma(0) \in A_0, \gamma(1) \in A_1$. Then*

$$\ln \mathrm{Vol}(A_t) \geq (1-t)\ln \mathrm{Vol}(A_0) + t\ln \mathrm{Vol}(A_1) + K\frac{t(1-t)}{2}d(A_0, A_1)^2.$$

The following diagram depicts this inequality in the case $K > 0$, reflecting the fact that geodesics spread out and then return back together in positive curvature.

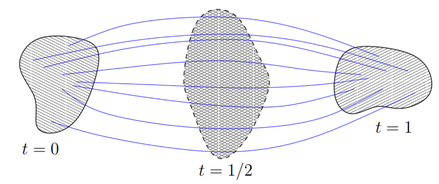


Figure 2. Brunn-Minkowski when $K > 0$. Image from [5].

This inequality can be improved if we introduce a $CD(K,N)$ condition, which encapsulates both "curvature bounded below by K" and "dimension bounded above by N", and in fact the validity of this improved inequality is equivalent to the $CD(K,N)$ condition.

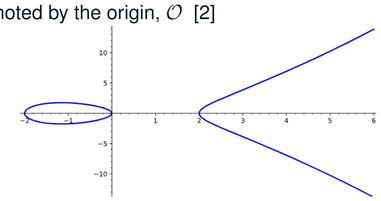
Exploring Rational Points on Elliptic Curves

Catherine Chen and Anna Li, mentored by Marcos Reyes

University of California, Santa Barbara

Introduction to Elliptic Curves

By definition, elliptic curves are smooth, projective, cubic curves with at least one rational point, denoted by the origin, $\mathcal{O}$ [2]



Throughout history, mathematicians have been interested in finding integer solutions to polynomial equations. However, there's no concrete algorithm for computing rational solutions of cubic equations with two variables. The main focus of this poster will be surrounding curves written in the normal Weierstrass form and their torsion points:

$$y^2 = x^3 + Ax^2 + Bx + C \text{ where A, B, and C are integers}$$

This form can be achieved from the general form through a change of variables. Our reading has been centered around investigating strategies to compute the rational points on ellipical curves by utilizing the properties of the finite torsion subgroup.

Basics of Groups

What is a group? A group is a set with a binary operation or a law of composition on a pair of elements (i.e. (5, 3) $\rightarrow$ (8)), where the group is defined by addition) that it satisfies the associative property, contains an identity element, and contains an inverse.[3]
Additionally, if a group satisfies the commutative property, then it is considered an *abelian group*.

Torsion Subgroup
We'll be looking at the torsion subgroup of an elliptic curve. It is defined as:

$$E(\mathbb{Q}) = \{P \in E(\mathbb{Q}) : n \in \mathbb{N}, \text{ s.t. } nP = \mathcal{O}\}$$

Simply put, a torsion subgroup is a collection of elements each with a finite order.

Order
The order of a point P on an elliptic curve is the smallest positive integer n such that nP = $\mathcal{O}$. ie: The curve $y^2 = x^3 - 15x + 22$ would have a torsion of $\mathbb{Z}/\mathbb{Z}6$

Cyclic Group
A Cyclic Group is an abelian group with a generator that can "generate" every point in the group.
Example: the group, 7$\mathbb{Z}$ = {..., -7, 0, 7, 14, ...} is generated by the element 7 since every number in the group is a multiple of 7. [3]

Fundamental Theorem of Finitely Generated Abelian Groups
Theorem 13.5: Every finitely generated abelian group G is isomorphic to a direct product of cyclic groups of the form

$$\mathbb{Z}_{p_1^{\alpha_1}} \times \mathbb{Z}_{p_2^{\alpha_2}} \times \cdots \times \mathbb{Z}_{p_n^{\alpha_n}} \times \mathbb{Z} \times \cdots \times \mathbb{Z}$$

where the $\mathbb{Z}_{p_i}$'s are primes.

Applying this to elliptic curves, the torsion points on the curve generate the entire torsion subgroup by repeated addition under the group operation. In this case, the inverse element will be the reflection of the point across the x-axis, denoted $-P$ and the identity would be the point at infinity $\mathcal{O}$. The order can be determined based on the largest order of the individual elements within the cyclic group.

Computing the Rational Points via Group Structure

One way to compute the rational points of an elliptic curve, E, is by acknowledging the group structure of an elliptic curve's rational points. Suppose $P, Q \in E$. Since E is a cubic function, then there must exist a third rational point on the secant line between P and Q (tangent if we're just working with P). Let R be the reflection of this third point over the x-axis where P + Q = R. Thus, it is possible to prove (E, +) is a finitely generated abelian group. [2]

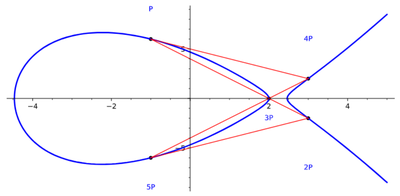
Example Let's look at the curve E: $y^2 = x^3 - 15x + 22$ and take the point P: (-1, 6) as our starting point. We want to compute P + P as defined above. To compute the tangent line of P on E, take the derivative $\frac{dy}{dx}$ to get:

$$\frac{dy}{dx} = \frac{(3x^2 - 15)}{2y}$$

Plugging in the torsion generator (-1, 6), we get a system of equations:

$$\begin{cases} 1. & y^2 = x^3 - 15x + 22 \\ 2. & y = -x + 5 \end{cases}$$

By substituting equation 1. into 2., P + P = 2P = (3, -2). Then, to compute 3P, we solve for P + 2P by finding the secant line between the two points and finding the third point. This process ends once the point -P is reached. For this curve, the order of this cyclic group is 6, since 5P = -P, and 6P = $\mathcal{O}$



Relevant Theorems

Some important theorems related to finding the torsion of elliptic curves are as follows:

Mordell-Weil Theorem
$E(\mathbb{Q})$ is a finitely generated abelian group, meaning that a finite set of points can generate all other rational points on the elliptic curve over the rationals. [1] Thus, it follows that:

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{torsion}} \oplus \mathbb{Z}^{r_E}$$

$\mathbb{Z}^{r_E}$ is the points generated by the points of infinite order. We will only be focused on the first part of this equation, the elements of torsion groups since it's finite.

Nagell-Lutz Theorem
Let $E(\mathbb{Q})$ be a Weierstrass elliptic curve. Then, every torsion point $P \neq \mathcal{O}$ of E satisfies:

1. The coordinates of P are integers
2. If y = 0, then it is a point of order 2.
3. If $y \neq 0$, then y divides D.

Mazur's Theorem
Let $E(\mathbb{Q})$ be a non-singular rational cubic curve containing a point of finite order m. Then:

$$1 \leq m \leq 10 \text{ or } m = 12, \text{ or}$$
$$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2M\mathbb{Z} \text{ where } 1 \leq M \leq 4$$

The Consequence of Nagell-Lutz

Because of the Nagell-Lutz Theorem, there is a way to determine a starting point with which we can use to generate the other rational points using the discriminant of the curve.

The discriminant is defined as follows:

$$D = 4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2$$

Suppose integer D is the discriminant of the elliptic curve E with the polynomial f(x). Then, find the finite amount of integers, y, such that y^2 divides D. Afterwards, take these y values and substitute them into the equation $y^2 = f(x)$ to solve for integer roots. If $D \neq 0$, then the roots of f(x) are distinct and the curve is smooth.

This is useful for figuring out the *possible* torsion points of a given elliptic curve, providing a reliable way to calculate a starting point P when we don't know any rational points off the top of our head. [1]

Example: Let's look at the elliptic curve E: $y^2 = x^3 + 1$. Using the formula from above, we see that the discriminant is -27. The multiples of 27 are: 1, 3, 9, and 27. The only numbers that satisfy our criteria above are ± 1 and ± 3. Then, we substitute these y-values into E to see what yields us an integer solution for x. Thus, the possible torsion points of this elliptic curve are $(-1, 0)$ and $(2, 3)$. Subsequently, we can generate a list of rational points and check the order for these points to determine the torsion.

To calculate 2P, the following equation, called the duplication formula, can be used to determine the slope of the tangent line:

$$\lambda = \frac{3x^2 + 2Ax + B}{2y}$$

Calculating P + nP can subsequently be done by the same method of calculating the slope between the two points and solving for the systems of equations. We have created an algorithm modeled after this strategy to compute the torsion group of elliptic curves. Scan the QR code below to access the website.

Acknowledgements

We would like to thank the Directed Reading Program at UCSB for providing us with the opportunity to work on this project, and our mentor Marcos Reyes for the continued guidance and support.

References

[1] Lorenzo-Robledo Alvaro. *Elliptical Curves, Modular Forms, and Their L-functions*. American Mathematical Society, 2011.

[2] Silverman Joseph H. and John T. Tate. *Rational Points on Elliptical Curves*. 2nd edition. Springer Cham, 2015.

[3] Judson Thomas W. and Robert A. Beezer. *Abstract Algebra: Theory and Applications*. Annual edition 2022. Orthagonal Publishing L3c, 2022.

Generating Functions and Percolation on Graphs

Vidushi Mittal Greta Glueck Mentor: Sawyer Dobson

University of California - Santa Barbara, Department of Mathematics - Directed Reading Program 2024

Introduction to Generating Functions

An ordinary generating function is a formal power series $f(x) = \sum_{k=0}^{\infty} a_k x^k$ whose coefficients correspond to terms of a sequence $\{a_k\}$. Since generating functions are *formal* power series, they can be added, multiplied, divided, etc. without considering issues of convergence. One of the main uses of generating functions is solving recurrence equations. Their power stems from the fact that they contain all the information of a sequence in a single series, whereas recurrence relations can only relate a few terms.

Interesting Examples of Generating Functions

We can use generating functions to find a ***closed form expression*** for the Fibonacci numbers, which are defined by $F_0 = 0, F_1 = 1$, and $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$. Let $f(x) = \sum_{k=0}^{\infty} F_k x^k$ be the generating function for the Fibonacci numbers. We multiply both sides of the recurrence relation by x^k and sum over k to obtain:

$$\sum_{k=0}^{\infty} F_{k+2}x^k = \sum_{k=0}^{\infty} F_{k+1}x^k + \sum_{k=0}^{\infty} F_k x^k$$

$$\frac{1}{x^2}(f(x) - x) = \frac{1}{x}f(x) + f(x)$$

Solving this equation yields a closed form expression for the generating function:

$$f(x) = \tfrac{x}{1-x-x^2}.$$

By computing the formal power series expansion of $f(x)$ about $x = 0$, we find

$$F_k = [x^k]f(x) = [x^k]\frac{x}{1-x-x^2} = \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^k - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^k$$

(Here we use the coefficient operator $[x^k]$, which extracts the coefficient of x^k in a formal power series $A(x)$ so that $[x^k]A(x) = a_k$.)

As another application of generating functions, we will count the number of binary trees with $k+1$ leaves, denoting this number by C_k. From this definition, we obtain the following recurrence relation:

$$C_0 = 1 \quad \text{and} \quad C_{k+1} = \sum_{n=0}^{k} C_n C_{k-n}$$

The numbers solving this recurrence are the ***Catalan numbers***, and they appear in many different counting problems. We can also obtain a closed form expression for the Catalan Numbers by using generating functions.

Theorem: If C_k is the k^{th} Catalan number, then

$$\sum_{k=0}^{\infty} C_k x^k = \frac{1-\sqrt{1-4x}}{2x} \quad \text{and} \quad C_k = \frac{1}{k}\binom{2k}{k-1}$$

Key Definitions and Notation

Having familiarized ourselves with generating functions, we will apply this knowledge to a problem in graph theory. We'll derive a recurrence relation, and use a generating function to find a closed-form expression for the solution. To state the problem, we need some definitions:

- ***P-percolation:*** Given a graph G with vertex set $V(G)$, a ***p-percolation*** is a random graph on $V(G)$, where each edge in $E(G)$ has a probability p of being ***open***, or included in the p-percolation, and a probability $1-p$ of being ***closed***, or deleted.
- ***Rooted Graph:*** A rooted graph is a graph with a specified vertex $\boldsymbol{v}$ called the ***root***.
- ***Cluster size:*** The cluster size, denoted $\boldsymbol{K_p(G)}$, of a ***p-percolation*** on a rooted graph G is the random variable counting how many vertices remain in the same connected component of the root $\boldsymbol{v}$. This includes $\boldsymbol{v}$ itself.
- ***Mass spectrum:*** The mass spectrum is the probability distribution of $K_p(G)$.
- ***D-regular tree:*** A ***d-regular tree*** is a rooted infinite tree where every vertex is adjacent to d vertices. Up to isomorphism, there is only one d-regular tree, denoted $\boldsymbol{V_d}$.

Setup of the Main Problem

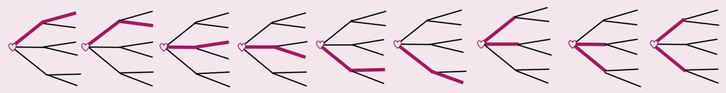
For the purpose of this poster, we consider d-regular trees, V_d. Our final goal is to compute the mass spectrum of a p-percolation on V_d. We will first construct a recurrence relation for the number of ways a cluster of a specific size can occur in our p-percolation, then use generating functions to obtain an explicit formula for the solution to the recurrence. We conclude, by using this formula, to compute the mass spectrum of a cluster size $K_p(G)$.

To construct the recurrence we need to establish some notation. Let $d, k \in \mathbb{N}$ and define the numbers $C_{d,k}$ and $D_{d,k}$ as follows:

$\boldsymbol{C_{d,k}}$: Fixing an arbitrary neighboring vertex v' of the root v, let $C_{d,k}$ be the number of connected subtrees of V_d with k vertices containing v but not containing v'. Note that the value of $C_{d,k}$ is independent of the choice of v' due to the symmetry of V_d.

Observe $C_{3,3} = 5$. Root vertex v denoted by ♡

$\boldsymbol{D_{d,k}}$: Let this be the number of connected subtrees of V_d with k vertices containing v

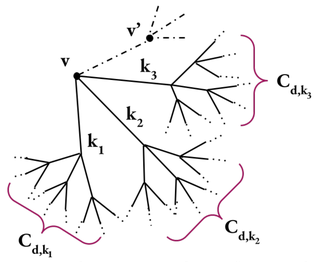


Observe that $D_{3,3} = 9$. Root vertex v denoted by ♡

Counting Subtrees

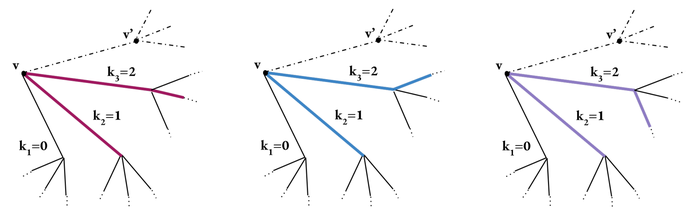
We will construct a recurrence relation for the numbers $C_{d,k}$. As an example, we'll start by considering V_4 and show how to write $C_{4,4}$ in terms of $C_{4,0}$, $C_{4,1}$, $C_{4,2}$, and $C_{4,3}$. Consider the set of all possible $(d-1) = 3$-tuples summing to $k - 1 = 3$:

$$K_{d,k} = K_{4,4} = \{(k_1, k_2, k_3) \in \mathbb{N}^3 : k_1 + k_2 + k_3 = 3\}$$



Essentially, the tuples tell us what size subtree we want from each immediate branch of the root, which we can visualize with k_1, k_2, k_3 in the diagram to the right. Since, within each branch, there are C_{d,k_i} different subtrees with k_i vertices, there are $C_{d,k_1} * C_{d,k_2} * C_{d,k_3}$ ways to choose one subtree from each branch (that satisfy the constraints of the tuple) and glue them together.

For example, for the tuple $(0, 1, 2)$, we want zero vertices from the k_1 branch, one vertex from the k_2 branch, and two vertices from the k_3 branch. There are $C_{4,0} * C_{4,1} * C_{4,2} = 1 * 1 * 3 = 3$ ways to construct such a subtree, pictured below:



Summing over the tuples,

$$C_{4,4} = \sum_{(k_1,k_2,k_3)\in K_{4,4}} C_{d,k_1} * C_{d,k_2} * C_{d,k_3} = \sum_{(k_1,k_2,k_3)\in K_{4,4}} \prod_{i=1}^{d-1=3} C_{d,k_i}$$

Counting Subtrees (Continued)

To generalize this example, to create a subtree of size k, we need $k-1$ additional vertices. We consider elements, $\vec{k}$, of $K_{d,k} = \{(k_1, ..., k_{d-1}) \in \mathbb{N}^{d-1} : \sum k_i = k - 1\}$. Now for each branch, there are C_{d,k_i} ways to choose a subtree of size k_i within each branch. Using the same reasoning we followed for V_4, we conclude:

Theorem: For $k \geq 1$, $C_{d,k}$ is recursively given by:

$$C_{d,k} = \sum_{\vec{k}\in K_{d,k}} \prod_{i=1}^{d-1} C_{d,k_i}$$

Now that we have obtained a recursive formula for $C_{d,k}$, we aim to use this to reach an explicit one. Let's setup the generating function $C_d(t) = \sum_{k\geq 0} C_{d,k}t^k$ and observe:

$$C_d(t) = \sum_{k=0}^{\infty} C_{d,k}t^k = 1 + t\sum_{k=1}^{\infty} t^{k-1}\left(\sum_{\vec{k}\in K_{d,k}} \prod_{i=1}^{d-1} C_{d,k_i}\right) = 1 + tC_d(t)^{d-1}$$

To extract the coefficients of $C_d(t)$ from this equation, we need the following theorem:

Lagrange Inversion Theorem

Suppose $Z_d(t)$ and $G(x)$ are power series, and $G(0) = 1$. If If $t\,G(Z_d(t)) = Z_d(t)$, then

$$[t^k]Z_d(t) = \left(\frac{1}{k}\right)[x^{k-1}]G(x)^k$$

Let $Z_d(t) = C_d(t) - 1$ and $G(x) = (x+1)^{d-1}$, so that

$$Z_d(t) = t(Z_d(t) + 1)^{d-1} = tG(Z_d(t))$$

Now, we may apply Lagrange Inversion theorem to obtain that for $k \geq 1$,

$$C_{d,k} = [t^k]Z_d(t) = \left(\frac{1}{k}\right)[x^{k-1}]G(x)^k = \frac{1}{k}\binom{k(d-1)}{k-1}$$

Here, we can note that for 3-regular trees, $C_{3,k}$ is the k^{th} Catalan number!

Computing the Mass Spectrum

Now we can use the same logic that we used earlier to find the recurrence for $C_{d,k}$ in order to get a recurrence for $D_{d,k}$ in terms of $C_{d,k}$ and then simplify to obtain the following:

$$D_{d,k} = \sum_{\vec{k}\in K_{d,k}} \prod_{i=1}^{d} C_{d,k_i} = \sum_{s=0}^{k-1} \frac{\binom{s(d-1)}{s-1}\binom{(k-s)(d-1)}{k-s-1}}{s(k-s)}$$

Lastly, we compute the mass spectrum $\mathbb{P}(K_p(V_d) = k)$. Well, this is the sum of probabilities of occurrence for each tree counted by $D_{d,k}$. Each of these trees has $k-1$ edges we want open in the percolation, and $dk - 2(k-1)$ edges we want closed. Moreover, each tree has an equal chance of occurring, so we multiply $D_{d,k}$ by the probability of these edges being open or closed and conclude:

$$\boldsymbol{\mathbb{P}(K_p(V_d) = k) = (D_{d,k})p^{k-1}(1-p)^{dk-2(k-1)}}$$

References

Bonnano, Leo, Sawyer Dobson, Juyon Lee, and Titus Sharman. Percolation on Graphs. Apr. 2024.

West, Douglas Brent. Combinatorial Mathematics. Cambridge University Press, 2021.

Special thanks to Alexander Urena for his support, dedication, and friendship, and to Sawyer Dobson for his time and mentorship.

Hopf Algebra and Representation Theory

Peihang Lin, Kriteen Shrestha, Ziqian Zhao (Mentor: Quinn Kolt)

Department of Mathematics, University of California, Santa Barbara

Abstract

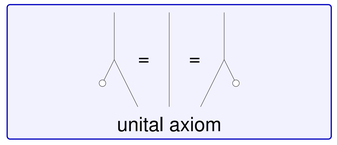
This poster will explore various algebraic structures and, in particular, Hopf algebras, a special algebraic structure that plays an important role in a variety of fields due to its algebraic properties. We will also discuss its representation theory, and study these fields with examples. A website containing more information and complementary code can be accessed through the QR code at the end of the poster.

Unital Associative Algebra

We start with **unital associative algebra**, a mathematical structure where elements interact with a binary operation, *multiplication*, and contains a multiplicative identity element, the *unit*. Multiplication is an operation used in many other mathematical structures that takes two elements to construct a new element in the algebra.

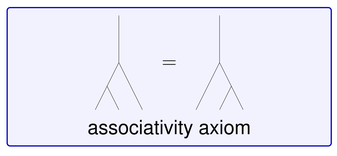
We can write these algebraic axioms diagrammatically, as seen in the right column. This diagrammatic notation will follow us throughout the poster, and **all operations are assumed to be linear**. Studying such algebraic structures can serve as powerful tools to uncover underlying patterns and abstracted relationships in various mathematical systems.

multiplication	unit
$m : H \otimes H \to H$	$1 : \mathbb{C} \to H$



unital axiom

$$1 \cdot h = h = h \cdot 1$$



associativity axiom

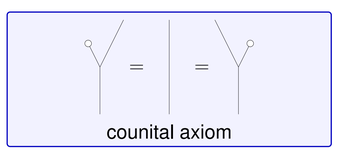
$$(g \cdot h) \cdot k = g \cdot (h \cdot k)$$

Counital Coassociative Coalgebra

Another interesting structure is the dual of unital associative algebras, **counital coassociative coalgebra**. The "co-" prefix indicates swapping the domain and codomain of our operations, namely the multiplication and unit of the algebra.

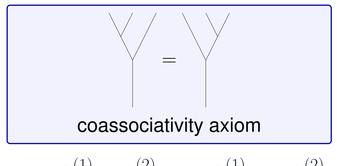
To motivate the concept of coalgebra, note that defining a tensor product of modules requires a method to distribute scalar multiplication over each tensor component. Comultiplication, defined as $\Delta(h) = \sum_i h_i^{(1)} \otimes h_i^{(2)}$, provides a natural mechanism for achieving such distribution. This will also be further discussed in the representation theory section.

comultiplication	counit
$\Delta : H \to H \otimes H$	$\varepsilon : H \to \mathbb{C}$



counital axiom

$$\sum_i \varepsilon(h_i^{(1)}) h_i^{(2)} = h = \sum_i h_i^{(1)} \varepsilon(h_i^{(2)})$$

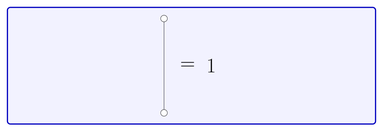


coassociativity axiom

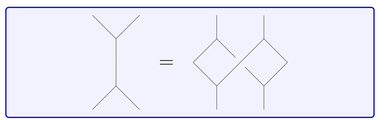
$$\sum_i \Delta(h_i^{(1)}) \otimes h_i^{(2)} = \sum_i h_i^{(1)} \otimes \Delta(h_i^{(2)})$$

Bialgebra

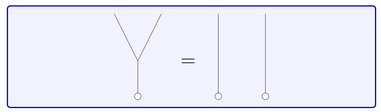
Combining these definitions of two important algebraic building blocks, we can create a **bialgebra**, a mathematical structure that is a unital associative algebra and a counital coassociative coalgebra with compatible operations. This compatability is encoded in the diagrams below.



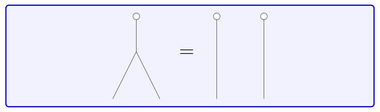
$\varepsilon(1) = 1$



$\Delta(hk) = \Delta(h)\Delta(k)$ Braid:



$\Delta(1) = 1 \otimes 1$

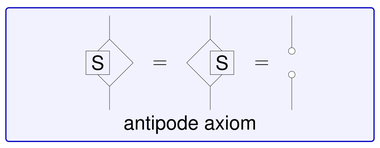


$\varepsilon(hk) = \varepsilon(h)\varepsilon(k)$

Hopf Algebras

We now have built a foundation to define Hopf algebras. A **Hopf algebra** is a structure that is a bialgebra, and it contains an additional linear operation – the *antipode* $S : H \to H$.

The antipode operation acts similar to the inverse operation of groups, except that it is linear and adapts to the counit and comultiplication operations.



antipode axiom

$$\sum_i h_i^{(1)} S(h_i^{(2)}) = \sum_i S(h_i^{(1)}) h_i^{(2)} = \varepsilon(h)1$$

Given a group G, we can define a **group Hopf algebra** $\mathbb{C}[G]$, a complex vector space with basis G and multiplication induced by the group multiplication. $\mathbb{C}[G]$ is always *cocommutative*. Cocommutativity is analagous to commutivity, but with comultiplication instead. On group elements g, comultiplication Δ is defined as $\Delta(g) = g \otimes g$, counit ε is defined as $\varepsilon(g) = 1$, and antipode S is defined as $S(g) = g^{-1}$. These operations can be computed on arbitrary elements by linearity.

Sweedler's Hopf algebra, denoted as H_4, is the smallest Hopf algebra that is both *noncommutative* and *noncocommutative* - meaning that elements neither commute nor cocommute. For example, $\Delta(\theta) \neq \Delta^{op}(\theta)$. Studying Hopf algebras gives a deeper understanding for how such properties behave and gain insight to representations of quantum groups and other important structures.

Let $G = D_3$, then $\mathbb{C}[D_3]$ has basis $\{1_G, s, r, r^2, sr, sr^2\}$. The multiplication is the same as the group operation:

$$s^2 = r^3 = (sr)^3 = 1_G,$$

and comultiplication, counit, and antipode are defined as follows:

$$\Delta(r) = r \otimes r, \quad \varepsilon(r) = 1,$$
$$S(r) = r^{-1}.$$

The *cocommutativity* of $\mathbb{C}[G]$ can be shown as

$$\Delta(r) = r \otimes r = \Delta^{op}(r).$$

The basis of H_4 consists of $\{1, K, \theta, K\theta\}$. With this basis, multiplication is subject to the following constraints:

$$K^2 = 1, \qquad K\theta = -\theta K, \qquad \theta^2 = 0.$$

We define comultiplication, counit, and the antipode identity as follows (note these are all linear maps):

$$\Delta(K) = K \otimes K, \quad \Delta(\theta) = K \otimes \theta + \theta \otimes 1,$$
$$\epsilon(K) = 1, \quad \epsilon(\theta) = 0,$$
$$S(K) = K, \quad S(\theta) = \theta K.$$

Representation Theory

Representation theory studies **abstract algebraic structures** by representing their **elements** as **linear transformations** of vector spaces. Motivating with representations of groups, we will see how the representations of Hopf algebras are defined analogously.

Definitions

A **representation** for a group G is equivalent to a G-module. A **G-module** M is a vector space with a map defined as

$$\rho : G \times M \to M,$$
$$(g, m) \mapsto g \cdot m,$$

such that as ρ is linear in M and

$$(g_1 \cdot g_2) \cdot m = g_1 \cdot (g_2 \cdot m),$$
$$1_G \cdot m = m,$$

for $g_1, g_2 \in G$, and $m \in M$.

A **representation** for a Hopf algebra H is equivalent to an H-module. An **H-module** M is a vector space with a linear map defined as

$$t : H \otimes M \to M,$$
$$h \otimes m \mapsto h \cdot m,$$

such that

$$(h \cdot k) \cdot m = h \cdot (k \cdot m),$$
$$1_H \cdot m = m,$$

for $h, k \in H$, and $m \in M$.

Example

An example of a 2-D D_3-module is the vector space $\mathbb{C}^2_{rot}$ with multiplication

$$s \cdot v = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \cdot v,$$
$$r \cdot v = \begin{bmatrix} \cos\frac{2\pi}{3} & \sin\frac{2\pi}{3} \\ -\sin\frac{2\pi}{3} & \cos\frac{2\pi}{3} \end{bmatrix} \cdot v,$$

for $v \in \mathbb{C}^2_{rot}$.

An example of a 1-D D_3-module is the vector space $\mathbb{C}_{sign}$ with multiplication

$$s \cdot \lambda = -\lambda, \qquad r \cdot \lambda = \lambda,$$

for $\lambda \in \mathbb{C}_{sign}$.

An example of a 1-D H_4-module is the vector space $\mathbb{C}_\varepsilon$ with multiplication

$$K \cdot \lambda = \varepsilon(K)\lambda = \lambda,$$
$$\theta \cdot \lambda = \varepsilon(\theta)\lambda = 0,$$

for $\lambda \in \mathbb{C}_\varepsilon$.

Another example of a 1-D H_4-module is the vector space $\mathbb{C}_k$ with multiplication

$$1 \cdot \lambda = \lambda, \qquad K \cdot \lambda = -\lambda,$$
$$\theta \cdot \lambda = 0, \qquad K\theta \cdot \lambda = 0,$$

for $\lambda \in \mathbb{C}_K$.

Tensor Product

Given M, N are G-modules, we can make $M \otimes N$ a G-module by

$$g \cdot (m \otimes n) = (gm) \otimes (gn).$$

Given M, N are H-modules, we can make $M \otimes N$ an H-module by

$$h \cdot (m \otimes n) = \Delta(h)(m \otimes n).$$

Example

Using the D_3-module examples above, $\mathbb{C}_{sign} \otimes \mathbb{C}^2_{rot} \cong \mathbb{C}^2_{rot}$ as D_3-modules.

Using the H_4-module examples above, $\mathbb{C}_K \otimes \mathbb{C}_K \cong \mathbb{C}_\varepsilon$ as H_4-modules.

Acknowledgements and References

We would like to thank our mentor Quinn Kolt for their support and inspiration for this project, as well as the UCSB Directed Reading Program for this invaluable experience.

[1] https://ncatlab.org/nlab/show/coalgebra
[2] https://en.wikipedia.org/wiki/Hopf_algebra
[3] https://en.wikipedia.org/wiki/Representation_theory

Kirby's Dream 3-Manifold

Miranda Jiang [1] Mentored by Sanjay Kumar [1]

[1]University of California - Santa Barbara
Department of Mathematics - Directed Reading Program 2024

Objective

Welcome to Kirby's dream land. Your goal is to distinguish 3-manifolds through the Kirby Calculus!

Background

The sphere S^3 can be obtained by gluing together two solid tori along the homeomorphism on their boundaries that interchanges longitudes and meridians.

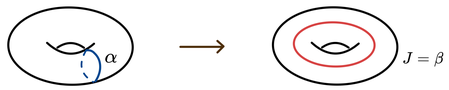


Figure 1. Demonstration of S^3

In fact, any orientable 3-manifold M^3 may be obtained by cutting out some solid tori from the 3-sphere S^3 and then pasting them back in, but along different homeomorphisms of their boundaries. This process is called a **surgery** on S^3.

We can describe the resulting 3-manifold entirely by the image of the meridian α under the attaching homeomorphism of the boundary torus $S^1 \times S^1$. Suppose that the meridian is sent to the curve $J = p\alpha + q\beta$, then J is the closed curve that winds around the boundary torus p times around the meridian and q times around the longitude.

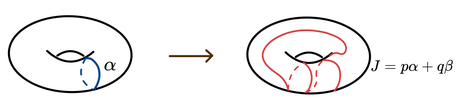


Figure 2. The meridian α is being sent to the curve J

The **framing** of the trivial knot is a rational number $r = p/q$ that determines the surgery of the 3-sphere.

We will consider the case of **integer surgery**, in which we choose q to be 1 so that the framing r is an integer. In fact, any knot diagrams has a natural framing coming from the number of times the normal vector turns when we draw the knot, called the **blackboard framing**. You can think of it as adding twists or kinks to the knot.

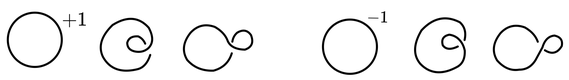


Figure 3. trivial knot with framing +1 and -1, denoted U_+ and U_-, respectively

Theorem *(Dehn-Lickorish): Any compact orientable 3-manifold without boundary can be obtained from the sphere S^3 by integer surgery on a framed link.*

Acknowledgements

Huge thanks to my mentor, Sanjay Kumar, for his continuous support during times of ups (the Eureka moments!) and downs (the painful struggles). His enthusiasm deeply inspired me. Thanks to the DRP committee for making this possible!

Thanks to Robion Kirby for all of his good math especially the Kirby Calculus and thank you Nintendo for developing the epic game "Kirby's Dream Land" which made this little poster possible.

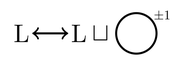
Happy problem-solving!

The Kirby Calculus

Manifolds of the same type can have extremely varied presentations by framed links. We want to find a systematic way to modify framed links so that they represent the same manifold. Here are your allowed moves:

1. The Kirby Moves:

- **The First Kirby Move**: Adding (or deleting) an unknotted circle with framing +- 1 that is unlinked with the other component of the given framed link $L \in S^3$.



- **The Second Kirby Move *("handle-slide")***: In a given framed link diagram with two distinguished unlinked components C and K, with framing indices n and k, respectively, we can slide the first curve C over K so that it encircles K and picking up the framing of K, while leaving the other components unchanged.

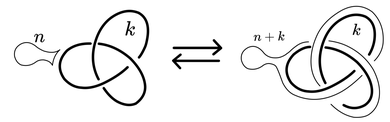


Figure 4. Kirby II

2. The Reidemeister Moves:

- **The Modified First Reidemeister Move:** The original first reidemeister move for knot diagrams involves resolving kinks. In the context of framed links, resolving kinks will result in a change of framing by ± 1. But we can, on the other hand, resolving two kinks with opposite orientations.
- **The Second and Third Reidemeister Move:**

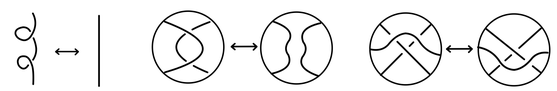


Figure 5. The double twist move Ω_1', Ω_2, and Ω_3

3. **Planer Isotopies:** The action of "smoothing out" the curves.

Theorem *(The Kirby Calculus): Two framed links in S^3 produce the same 3-manifold if and only if they can be obtained from each other by a finite sequence of the Kirby moves, the moves Ω_1', Ω_2, and Ω_3 and planer isotopies.*

Example

The following framed link diagrams represent the same 3-manifolds. They can be obtained from each other by a finite sequence of the Kirby moves and the Reidemeister moves.

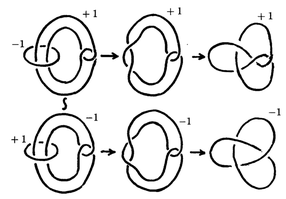


Figure 6. Two modifications of the Whitehead link

Applications

Analogous to building knot invariants with the Jones polynomial, we can build 3-manifold invariants with the Kirby calculus. An example would be the complex-valued 3-manifold invariant $I(D)$ that arises from the **Temperley-Lieb algebra** TL_n. The idea is as follows: given a framed link presentation of a 3-manifold, we can compute a polynomial by means similar to the Jones polynomial, and with a certain choice of coefficients (primitive $4rth$ root of unity), we obtain a complex number that stays invariant under both the Kirby moves and the Reidemeister moves.

We first define the Jones-Wenzl idempotent $f^{(n)}$ by a recursive formula:

$$f^{(n+1)} = f_1^{(n)} - \frac{\Delta_{(n-1)}}{\Delta_n}(f_1^n e_n f_1^{(n)})$$

where f_1^n is $f^{(n)}$ with a strand added on top, and e_n is the generator of TL_n

We then define the element $\omega = \sum_{n=0}^{r-2} \Delta_n S_n(\alpha)$, where $S_n(\alpha)$ is the image of the closure of $f^{(n)}$ under the map $TL_n \to S(S^1 \times I)$, and Δ_n is a complex number obtained from computing the polynomial of the closure of $f^{(n)}$ under the map $TL_n \to S(\mathbb{R}^2)$

Theorem: *Suppose that a closed oriented 3-manifold M is obtained by surgery on a framed link represented by a planar diagram D. Let b_+ be the number of positive eigenvalues and b_- be the number of negative eigenvalues of the linking matrix of this link. Suppose $r > 3$ and that A is a primitive $4rth$ root of unity. Then*

$$I(D) = \langle \omega, \dots \omega \rangle_D \langle \omega \rangle_{U_+}^{-b_+} \langle \omega \rangle_{U_-}^{-b_-}$$

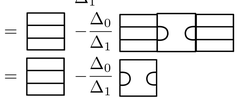
is a well-defined invariant of M.

Let's do an example.

Example Let $r = 4$, and let D to be this diagram:

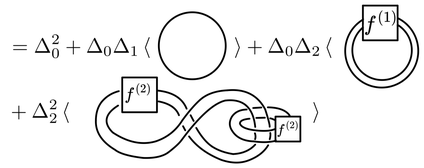
We will compute $f^{(2)}$ first:

$$f^{(2)} = f_1^{(1)} - \frac{\Delta_0}{\Delta_1} f_1^{(1)} e_n f_1^{(1)}$$



Then $\omega = \sum_{n=0}^{2} \Delta_n S_n(\alpha) = \Delta_0\alpha_0 + \Delta_1\alpha_1 + \Delta_2\alpha_2$

$$\langle \omega, \omega \rangle_D = \langle \Delta_0\alpha_0 + \Delta_1\alpha_1 + \Delta_2\alpha_2, \Delta_0\alpha_0 + \Delta_1\alpha_1 + \Delta_2\alpha_2 \rangle_D$$
$$= \Delta_0^2 + \Delta_0\Delta_1 \langle \alpha_0, \alpha_1 \rangle_D + \Delta_0\Delta_2 \langle \alpha_0, \alpha_2 \rangle_D + \dots + \Delta_2^2 \langle \alpha_2, \alpha_2 \rangle_D$$



Plug in $f^{(n)}$ into the square, and do the same operation to $\langle \omega \rangle_{U_+}^{-b_+} \langle \omega \rangle_{U_-}^{-b_-}$, we obtain the desired 3-manifold invariant $I(D)$.

References

[1] A. B. Sossinsky V. V. Prasolov. *Knots, Links, Braids and 3-Manifolds: An Introduction to the New Invariants in Low-Dimensional Topology*. American Mathematical Society, 1997.

[2] W. B. Raymond Lickorish. *An Introduction to Knot Theory*. Springer New York, NY, 1997.

Near-group Fusion Categories

Connor Lindquist-Carrillo
University of California, Santa Barbara

Fusion Categories

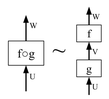
Definition [2]. A **fusion category** is a category $\mathcal{C}$ that is

- monoidal: $(\otimes, \mathbf{1}, \alpha, \lambda, \rho)$
- semisimple: $X \cong \bigoplus_{X_i \in \mathrm{Irr}(\mathcal{C})} m_i X_i$
- $\mathbb{C}$-linear: $\mathrm{Hom}(X, Y) \in \mathrm{Vect}_{\mathbb{C}}$
- rigid: $X^* \otimes X \xrightarrow{\mathrm{ev}_X} \mathbf{1} \xrightarrow{\mathrm{coev}_X} X \otimes X^*$
- finite rank: $|\mathrm{Irr}(\mathcal{C})| < \infty$
- $\mathbf{1}$ is simple

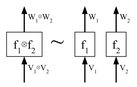
Graphical Language for Fusion Categories

Let $\mathcal{C}$ be a strict monoidal category. We have graphical representations [1] for morphisms in $\mathcal{C}$ and their relations, where $\sim$ denotes equivalence.

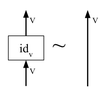
- $f \circ g$, where $f : V \to W$ and $g : U \to V$.



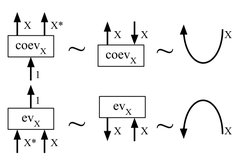
- $f_1 \otimes f_2$



- id_V



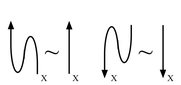
- ev_X and coev_X:



- The rigidity axioms require the following maps to be the identity on X and X^*

$$X \xrightarrow{\mathrm{coev}_X \otimes \mathrm{id}_X} X \otimes X^* \otimes X \xrightarrow{\mathrm{id}_X \otimes \mathrm{ev}_X} X$$
$$X^* \xrightarrow{\mathrm{id}_{X^*} \otimes \mathrm{coev}_X} X^* \otimes X \otimes X^* \xrightarrow{\mathrm{ev}_X \otimes \mathrm{id}_{X^*}} X^*$$

with graphical form:



Examples of Fusion Categories

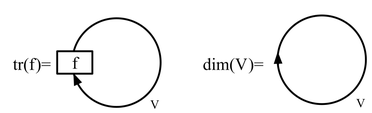
Let G be a finite group.

1. $\mathrm{Rep}(G)$: Category of finite-dimensional representations of G
   - Objects: finite-dimensional representations of G over $\mathbb{C}$
   - Morphisms: interwiners
   - Tensor product: tensor product of representations
   $$\rho_{V \otimes W}(g) := \rho_V(g) \otimes \rho_W(g)$$
2. Vec_G^{ω}: Category of G-graded Vector Spaces
   - Objects: G-graded finite dimensional vector spaces $V = \bigoplus_{g \in G} V_g$.
   - Morphisms: linear maps which preserve the grading.
   - Tensor product: $(\bigoplus_{g \in G} V_g) \otimes (\bigoplus_{h \in G} W_h) = \bigoplus_{\substack{k \in G \\ gh=k}} (V_g \otimes W_h)$
   - $\mathbf{1} = \mathbb{C}_e$, and for $g, h, k \in G, v_g \in V_g, w_h \in W_h, z_k \in Z_k$, associativity
   $$\alpha_{V,W,Z} : (v_g \otimes w_h) \otimes z_k \mapsto \omega(g, h, k) v_g \otimes (w_h \otimes z_k),$$
   where ω is a 3-cocycle: $\omega : G \times G \times G \to \mathbb{C}^\times$ such that
   $$\omega(ab, c, d)\omega(a, b, cd) = \omega(a, b, c)\omega(a, bc, d)\omega(b, c, d)$$
3. Fusion categories of rank 2 with objects $\mathbf{1}$, X, and $X \otimes X = \mathbf{1} \oplus nX$. The only possible values for n are 0 and 1. There are 4 fusion categories of rank 2 [5].

Invariants of Fusion Categories

Let $\mathcal{C}$ be a fusion category.

1. **Fusion ring and fusion coefficients**
   - The **fusion ring** $K(\mathcal{C})$ is a unital $\mathbb{Z}_+$-ring whose elements are isomorphism classes of objects in $\mathcal{C}$. The addition is defined by $[X]+[Y] = [X \oplus Y]$, and the multiplication is defined by $[X] \cdot [Y] = [X \otimes Y]$.
   - **Fusion rule** : $X \otimes Y = \oplus N_{XY}^Z Z$
   - **Fusion coefficients** : $N_{X,Y}^Z = \dim \mathrm{Hom}(X \otimes Y, Z)$
2. **Frobenius-Perron Dimension**
   - **Fusion matrix**: Let N_X be the matrix with (Y, Z)-entry N_{XY}^Z for simple objects, which is called the fusion matrix. N_X is a square matrix of nonnegative integers.
   - **Frobenius-Perron dimension:** Let $\mathrm{FPdim}(X)$ be the maximal eigenvalue of the fusion matrix $N_X, X \in \mathrm{Irr}(\mathcal{C})$.
   - Frobenius-Perron dimensions give a **character of the fusion ring** since
   $$\mathrm{FPdim}(V \otimes W) = \mathrm{FPdim}(V) \cdot \mathrm{FPdim}(W),$$
   $$\mathrm{FPdim}(V \oplus W) = \mathrm{FPdim}(V) + \mathrm{FPdim}(W), \quad \mathrm{FPdim}(\mathbf{1}) = 1$$
3. **Quantum Dimension**: Let $\mathcal{C}$ be a spherical category and $f : V \to V$ be a morphism.
   - The **quantum trace** of f is defined as:



   - The **quantum dimension** of V is $\dim V := \mathrm{tr}(\mathrm{id}_V)$.
   - Quantum dimensions also give a **character of the fusion ring.**

Near-group Categories

An object X in $\mathcal{C}$ is **invertible** if ev_X and coev_X are isomorphisms. The invertible objects in a fusion category $\mathcal{C}$ form a subcategory and their fusion rule can be described by a finite group G.

Definition. A **near-group category** is a fusion category $\mathcal{C}$ in which all but one simple object is invertible.

- Simple objects: $G \cup \{\rho\}$
- Fusion rules: $g\rho = \rho g = \rho$, $\forall g \in G$, and
$$\rho \otimes \rho = n'\rho + \sum_{g \in G} g.$$
- $\dim(\rho) = \frac{n' + \sqrt{n'^2 + 4n}}{2}$, where $n = |G|$.

Denote such a near-group category by $G + n'$.

- Given a near-group category of type $G + n'$, the only possible values for n' are $0, n-1$, or $n' \in n\mathbb{Z}$ [3], with $n' = 0$ completely classified by Tambara and Yamagami [6].

The Case $n' = |G|$

Near-group categories of type $G + n$ are determined by $(G, a, b, c, \langle\ ,\ \rangle)$, where $c \in \mathbb{T}, a : G \to \mathbb{T}, b : G \to \mathbb{C}$ and $\langle\ ,\ \rangle$ is a non-degenerate symmetric bicharcter satisfying

$$a(0) = 1, \quad a(x) = a(-x), \quad a(x+y)\langle x, y\rangle = a(x)a(y), \quad \sum_{x \in G} a(x) = \sqrt{n}c^{-3},$$
$$b(0) = -\frac{1}{d}, \quad \sum_y \overline{\langle x, y\rangle} b(y) = \sqrt{n}c\overline{b(x)}, \quad a(x)b(-x) = \overline{b(x)},$$
$$\sum_x b(x+y)\overline{b(x)} = \delta_{y,0} - \frac{1}{d}, \quad \sum_x b(x+y)b(x+z)\overline{b(x)} = \overline{\langle y, z\rangle} b(y)b(z) - \frac{c}{d\sqrt{n}},$$

where $d = \frac{n + \sqrt{n^2+4n}}{2}$ [3, 4].

Acknowledgements

I would like to thank my family and friends for their support, UCSB's DRP program, and my DRP mentor, Qing Zhang, whose kindness and dedication towards my growth as a mathematician has been genuinely incredible.

References

[1] Bojko Bakalov and Alexander A. Kirillov. "Lectures on tensor categories and modular functors". In: 2000. URL: https://api.semanticscholar.org/CorpusID:52201867.

[2] Pavel Etingof et al. *Tensor categories*. Vol. 205. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2015.

[3] David E. Evans and Terry Gannon. "Near-group fusion categories and their doubles". In: *Adv. Math.* 255 (2014), pp. 586–640.

[4] Masaki Izumi. "The structure of sectors associated with Longo-Rehren inclusions. II. Examples". In: *Rev. Math. Phys.* 13.5 (2001), pp. 603–674.

[5] Viktor Ostrik. *Fusion categories of rank 2*. 2002. arXiv: math/0203255 [math.QA].

[6] Daisuke Tambara and Shigeru Yamagami. "Tensor Categories with Fusion Rules of Self-Duality for Finite Abelian Groups". In: *Journal of Algebra* 209 (1998), pp. 692–707. URL: https://api.semanticscholar.org/CorpusID:121300911.

On constructing the weak solution to 2D Euler equation with non-smooth initial conditions

Alan Yin, Katelyn Matchett, Ryoma Kozaki, Kieran Salib

University of California Santa Barbara

Introduction

Euler and Navier-Stokes equations govern the motion of fluid with or without viscosity. They give rise to complex phenomena and intricate structures. The Navier-Stokes equations consists of a time-dependent continuity equation for conservation of mass, three time-dependent conservation of momentum equations and a time-dependent conservation of energy equation.

Navier-Stokes Equation

$$\frac{\partial u}{\partial t} + (u \cdot \nabla u) = -\nabla p + v\Delta u + f(x,t)$$
$$div(u) = 0$$

In the study of fluid dynamics, there is an important dimensionless quantity named Reynold's number that helps predict the flow pattern by measuring the ratio between inertial and viscous forces. A high R_e of a fluid system indicates a persistent presence of turbulence. Turbulent flows are much more difficult to describe than the laminar ones. They are, nevertheless, important to gain insights from to develop a deep understanding of physical systems like jet engine and tornado. In this poster, we discuss the rigorous construction of solution to Euler equation in 2D when subject to these non-smooth conditions.

Preliminaries

Sobolev Spaces

Sobolev spaces consist of functions whose weak derivatives belong to L^p. These spaces provide one of the most useful settings for the analysis of PDEs.
Suppose that Ω is an open set in $\mathbb{R}^n$, $k \in \mathbb{N}$, and $1 \leq p \leq \infty$. The Sobolev space $W^{k,p}(\Omega)$ consists of all locally integrable functions $f : \Omega \to \mathbb{R}$ such that

$$\partial^\alpha f \in L^p(\Omega) \qquad \textit{for} 0 \leq |\alpha| \leq k$$

We write $W^{k,2}(\Omega) = H^k(\Omega)$. The Sobolev space is a Banach space when equipped with the norm:

$$||f||_{W^{k,p}(\Omega)} = \left(\sum_{\alpha \leq |k|} \int_\Omega |\partial^\alpha f|^p dx \right)^{1/p}$$

for $1 \leq p < \infty$ and $p = \infty$,

$$||f||_{W^{k,p}(\Omega)} = max_{|\alpha| \leq k} sup_\Omega |\partial^\alpha f|.$$

Weak Derivatives

Let $u \in L^p_{loc}(\Omega)$ with open $\Omega \subset R^n$. We will call $\frac{du}{dx_j}$ the **weak derivative** of u if every smooth compactly supported function $\phi \in C_c^\infty(\Omega)$ gives the equality

$$\int_\Omega \frac{du}{dx_j} \phi dx = -\int_\Omega u \frac{d\phi}{dx_j} dx$$

Weak derivatives are unique up to almost everywhere equivalence. In our construction, we will encounter functions that seem differentiable except on sets of zero measure. Weak derivative enables us to construct derivatives for functions as such.

Weak* Convergence

Let (u_n) be a bounded sequence of $L^\infty(\Omega)$; then, from the sequence (u_n), we can extract a subsequence which is weakly-convergent; that is

$$\exists (u_{n_k})_k, \exists u \in L^\infty(\Omega), \lim_{k \to \infty} \int_\Omega u_{n_k} \phi \, dx = \int_\Omega u\phi \, dx, \quad \forall \phi \in L^1(\Omega).$$

Reformulation of Euler Equations for $v \in L^1$ and $\omega \in L^p$

Reality is not so ideal, and in many scenarios we have to work with highly unstable fluid structure where instantaneous local dynamics occurs in all time.

Weak Solution in Primitive-Variable Form

(i) A velocity field $u(x,t)$ with initial data v_0 such that $v \in L^1([0,T]) \times B_R$ for any $R > 0, B_R = x \in \mathbb{R}^2, x \leq R$
(ii) $v \otimes v = (v_i v_j) \in L^1([0,T] \times B_R$
(iii) div $v = 0$
(iv) for any $\Phi = (\Phi_1, \phi_2) \in C^1[0,T], C^1(\mathbb{R}^2)$ with div $\Phi = 0$,

$$\int \Phi(x,T) \cdot v(x,T) dx - \int \Phi(x,0) \cdot v_0(x) dx = \int_0^T \int (\Phi_t \cdot v + \nabla\Phi : v \otimes v) dx dt,$$

$$\nabla\Phi = \left(\frac{\partial}{\partial x_i} \Phi_j \right), A : B = \sum_{i,j=1}^{2} A_{ij} B_{ij}$$

Now, following this weak formulation, we are able to construct and approximate solution sequence for the Euler equation with respect to conditions of interests.

Lions-Aubin Lemma

Let $B_0 \subset B_1 \subset B_2$ be three Banach spaces. We assume that the embedding of B_1 in B_2 is continuous and that the embedding of B_0 in B_1 is compact. Let p, r such that $1 \leq p, r \leq +\infty$. For $T > 0$, we define

$$E_{p,r} = \left\{ v \in L^p(]0,T[, B_0) \,\middle|\, \frac{dv}{dt} \in L^r(]0,T[, B_2) \right\}.$$

1. If $p < +\infty$, the embedding of $E_{p,r}$ in $L^p(]0,T[, B_1)$ is compact.
2. If $p = +\infty$ and if $r > 1$, the embedding of $E_{p,r}$ in $C^0([0,T], B_1)$ is compact.

A more useful version of the lemma would be its reformulation as follows: let $\{f^\epsilon(t)\}$ be a sequence in $C\{[0,T], H^s(\mathbb{R}^N)\}$ such that:

1. $max_{0 \leq t \leq T} ||f^\epsilon(t)||_s \leq C$.
2. for any $\rho \in C_0^\infty(\mathbb{R}^N)$, $\{\rho f^\epsilon\}$ is uniformly in Lip$\{[0,T], H^M(\mathbb{R}^N)\}$, i.e.,
$$||\rho f^\epsilon(t_1) - \rho f^\epsilon(t_2)||_M \leq C_M |t_1 - t_2|, \quad 0 \leq t_1, \quad t_2 \leq T$$
for some constant C_M; then there exists a subsequence $\{f^{\epsilon_j}\}$ and $f \in C\{[0,T] H^s(\mathbb{R}^N)\}$ such that for all $R \in (M,s)$ and $\rho \in C_0^\infty(\mathbb{R}^N)$:
$$max_{0 \leq t \leq T} ||\rho f^{\epsilon_j}(t) - \rho f(t)||_R \to 0 \quad as \quad j \to \infty$$

This version of lemma completes the picture of formalism of approximate-solution sequence by allowing us to construct the last necessary condition.

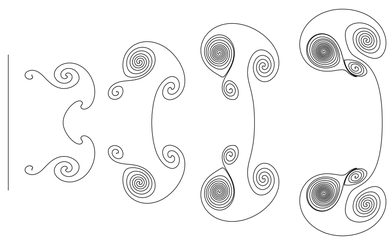


Figure 1: An Evolving Vortex Sheet

Convergence Results in 2D

Reformulated, we can make sense of solutions where the vorticity is unbounded:

$$\omega_0 \in L^p(\mathbb{R}^2) \cap L^1(\mathbb{R}^2)$$

Non-smooth initial conditions like this would create large-scale coherent structures with incredible small-scale complexity, and we refer to them as vortex sheets. Vortex sheets suggest a natural framework within which to build a mathematical theory, since, despite the singularity in vorticity, they still retain the physically significant feature of finite kinetic energy:

$$\int_\Omega |v^\epsilon|^2 dx \leq C(\Omega) \quad \forall \in \Omega, \quad v^\epsilon \Rightarrow v \in L^2(\Omega)$$

The uniform bound suggests there's a function v and a subsequence, $\{v^\epsilon\}$, that weakly converges to it in L^2_{loc}; hence, it is natural to ask: is v still a solution to the Euler's equation, and what kind of defect develop in this limit? A sounding plan of attack would be developing a rigorous framework of approximate-solution sequences of 2D Euler equation.

Theorem 1. *A sequence of function $v^\epsilon \in C\{[0,T], L^2_{loc}(R)^2\}$ is an approximate solution sequence for the 2D Euler equation if*

- *for all R, T > 0, max$_{0 \leq t \leq T} \int_{|x| \leq R} |v^\epsilon(x,t)|^2 dx \leq C(R)$*
  $div(v) = 0$ in the sense of distributions

- *(weak consistency with the Euler equation),*

$$lim_{\epsilon \to 0} \int_0^T \int_{(R)^2} (v^\epsilon \cdot \Phi_t + \nabla\Phi : v^\epsilon \otimes v^\epsilon) dx dt = 0$$

In addition to the conditions above, we also require the following control for $p = 1$ and $p > 1$ by assumption.
Seeking to use the Lions-Aubin lemma, we show that, for an proper approximate-solution sequence, it should also satisfy, for some constant $C > 0$:

$$||\rho v^\epsilon(t_1) - \rho v^\epsilon(t_2)|| - L \leq C|t_1 - t_2| \quad t_1, t_2 \in [0,T] \quad \forall L > 0 \quad \forall \rho \in C_0^\infty(\mathbb{R}^N)$$

i.e., $\{v^\epsilon\}$ is uniformly bounded in $Lip\{[0,T], H^{-L}_{loc}(\mathbb{R}^N)\}$.

With these tools at hand, it is possible to prove the following theorem:

Theorem 2. *Given vorticity in $L^1 \cap L^p$, for an approximate-solution sequence that satisfies conditions (i)-(vi), there exists $v \in L^2(\Omega)T)$, $\Omega_T = [0,T] \times B_R$, such that*

$$||v^\epsilon - v||_{L^2(\Omega_T)} \to 0 \quad as \quad \epsilon \to 0$$

and v is a weak solution for the 2D Euler equation. Additionally, we are also able to make the substantial claim that, out of the two limiting behaviors, only concentration is possible in 2D.

Acknowledgements

Thank you to the DRP program and to our mentor Pranav for his guidance!

References

Boyer, Franck, and Pierre Fabrie. Mathematical Tools for the Study of the Incompressible Navier-Stokes Equations and Related Models. Springer, 2013.
Majda, Andrew, and Andrea L. Bertozzi. Vorticity and Incompressible Flow. Cambridge University Press, 2002.

Partitions and Representations of the Symmetric Group

Sogol Cyrusian

University of California, Santa Barbara - Directed Reading Program 2024

Introduction

- A **partition** of n is any $\lambda = (\lambda_1, \lambda_2, ..., \lambda_\ell)$ such that $\sum_{i=1}^{\ell} \lambda_i = n$ and $\lambda_i \geq \lambda_{i+1}$.
- **Example.** Let $\lambda = (5, 3, 3, 2, 1)$, then the Ferrers diagram of λ is

$$\begin{array}{l} \bullet\ \bullet\ \bullet\ \bullet\ \bullet \\ \bullet\ \bullet\ \bullet \\ \bullet\ \bullet\ \bullet \\ \bullet\ \bullet \\ \bullet \end{array}$$

- The **symmetric group** on a set of n elements, denoted S_n, consists of the bijections between that set and with the group operation of composition.
- Let $\lambda = (\lambda_1, \lambda_2, ..., \lambda_\ell)$ be a partition of n. The **Young subgroup** of S_n corresponding to λ is

$$\begin{aligned} S_\lambda &= S_{\{1,2,...\lambda_1\}} \times S_{\{\lambda_1+1,\lambda_1+2,...,\lambda_1+\lambda_2\}} \times \cdots \times S_{\{n-\lambda_\ell+1,n-\lambda_\ell+2,...,n\}} \\ &\cong S_{\lambda_1} \times S_{\lambda_2} \times \cdots \times S_{\lambda_\ell}. \end{aligned}$$

- **Example.** Let $\lambda = (5, 3, 3, 2, 1)$, then

$$\begin{aligned} S_{(5,3,3,2,1)} &= S_{\{1,2,3,4,5\}} \times S_{\{6,7,8\}} \times S_{\{9,10,11\}} \times S_{\{12,13\}} \times S_{\{14\}} \\ &\cong S_5 \times S_3 \times S_3 \times S_2 \times S_1 \end{aligned}$$

- A **Young tableau** of shape λ, where λ is a partition of n, is any array obtained by bijectively replacing the dots in the Ferrers diagram of λ with $1, 2, ..., n$.
- **Example.** Let $\lambda = (2, 1)$, then the Ferrers diagram of λ is

$$\begin{array}{l} \bullet\ \bullet \\ \bullet \end{array}$$

And so, all the Young tableaux of shape λ are

$$\begin{array}{ll} 1\ 2 \\ 3 \end{array}, \quad \begin{array}{ll} 2\ 1 \\ 3 \end{array}, \quad \begin{array}{ll} 3\ 2 \\ 1 \end{array}, \quad \begin{array}{ll} 2\ 3 \\ 1 \end{array}, \quad \begin{array}{ll} 1\ 3 \\ 2 \end{array}, \quad \begin{array}{ll} 3\ 1 \\ 2 \end{array}.$$

- **Remark.** There are exactly $n!$ Young tableaux of shape λ, where λ is any partition of n, because the number of distinct Young tableaux is in bijection with the number of ways to arrange $1, ..., n$ in n spaces.

Representations

- A matrix representation of a group is a way of viewing its elements as matrices.
- **Example.** The homomorphism ρ defined to take all elements of S_n to 1 is called the trivial representation.
- Let $\mathrm{GL}(V)$ denote the group of invertible $m \times m$ matrices. A **representation** of S_n is (V, φ), where V is a vector space and $\varphi : S_n \to \mathrm{GL}(V)$ is a homomorphism.
- **Example.** An element of S_n is either odd or even. In particular we can write every element in S_n as the composition of 2-cycles, which are odd. So, we can determine whether any element in S_n is even or odd.
  Consider the representation (φ, V) where $\varphi : S_n \to \mathrm{GL}(V)$ is defined by $\varphi(x) \mapsto \mathrm{sgn}(x)$.

An Algorithm

- **Theorem.** There is a one-to-one correspondence between elements in S_n and pairs of Young tableuax of the same shape with increasing rows and columns.
- A **partial tableau** is an array with distinct entries whose rows and columns are increasing.
- Let P be a partial tableau. Take $x \notin P$, to insert x in P we use the following algorithm:
  1. Let $R :=$ the first row of P.
  2. **While** x is less than some element of row R, **do**
     a. Let y be the smallest element in row R that is strictly greater than x. Replace y by x in row R.
     b. Set $x := y$ and $R = R + 1$.
  3. When x is greater than every element in row R, put x at the rightmost end of row R.
- **Example.** Let $x = 3$,

$$P = \begin{array}{llll} 1 & 2 & 5 & 8 \\ 4 & 7 & & \\ 6 & & & \\ 9 & & & \end{array}$$

We first insert 3 into the first row, then insert 5 in the second row, and finally we insert 7 in the third row. The partial tableaux resulting from each step are, respectively

$$\begin{array}{llll} 1 & 2 & 3 & 8 \\ 4 & 7 & & \\ 6 & & & \\ 9 & & & \end{array}, \quad \begin{array}{llll} 1 & 2 & 3 & 8 \\ 4 & 5 & & \\ 6 & & & \\ 9 & & & \end{array}, \quad \begin{array}{llll} 1 & 2 & 3 & 8 \\ 4 & 5 & & \\ 6 & 7 & & \\ 9 & & & \end{array}.$$

- Using the algorithm above, we can construct a Young tableau from any permutation in S_n. To construct the pair of Young tableaux as in the theorem statement, we must take the permutation $\sigma = (\sigma_1 \sigma_2 ... \sigma_k) \in S_n$, then
  1. Let P, Q be the empty tableaux.
  2. For i increasing from 1 to n, insert σ_i in P, and set the position where σ_i was inserted to i in Q.

  We end with a pair of Young tableaux with increasing rows and columns.
- **Example.** Let $\pi = (4236517) \in S_n$. Then, the tableaux constructed using the algorithm above are

$$\emptyset, \quad \begin{array}{l} 4, \\ \end{array} \begin{array}{ll} 2 \\ 4 \end{array}, \begin{array}{ll} 2\ 3 \\ 4 \end{array}, \begin{array}{lll} 2\ 3\ 6 \\ 4 \end{array}, \begin{array}{lll} 2\ 3\ 5 \\ 4\ 6 \end{array}, \begin{array}{lll} 1\ 3\ 5 \\ 2\ 6 \\ 4 \end{array}, \begin{array}{lll} 1\ 3\ 5\ 7 \\ 2\ 6 \\ 4 \end{array} = P$$

$$\emptyset, \quad \begin{array}{l} 1, \\ \end{array} \begin{array}{ll} 1 \\ 2 \end{array}, \begin{array}{ll} 1\ 3 \\ 2 \end{array}, \begin{array}{lll} 1\ 3\ 4 \\ 2 \end{array}, \begin{array}{lll} 1\ 3\ 4 \\ 2\ 5 \end{array}, \begin{array}{lll} 1\ 3\ 4 \\ 2\ 5 \\ 6 \end{array}, \begin{array}{lll} 1\ 3\ 4\ 7 \\ 2\ 5 \\ 6 \end{array} = Q$$

- To prove the other direction of the theorem, we basically follow the algorithm in reverse. We take the last element we inserted to construct Q, then we know that the position of that element corresponds to the position of the last element inserted to construct P. Now we can delete that element from P by retracing the steps in the algorithm above.

Specht Modules

- Two λ-tableaux are **row equivalent**, $t_1 \sim t_2$, if for all i, the ith row of t_1 contains the same elements as the ith row of t_2.
- **Example.** For $\lambda = (2, 1)$, the following Young tableaux are row equivalent,

$$\begin{array}{ll} 1\ 2 \\ 3 \end{array} \sim \begin{array}{ll} 2\ 1 \\ 3 \end{array}$$

- A **λ-tabloid** is $\{t\} = \{t_1 : t_1 \sim t\}$, where t is a Young tableau of shape λ.
- **Example.** Let $t = \begin{array}{ll} 1\ 2 \\ 3 \end{array}$, then $\{t\} = \left\{ \begin{array}{ll} 1\ 2 \\ 3 \end{array}, \begin{array}{ll} 2\ 1 \\ 3 \end{array} \right\}$.
- Let λ partition n, and let $\{t_1\}, ..., \{t_k\}$ be a complete list of distinct λ-tabloids. Then, the **permutation module corresponding to λ**, M^λ, is the vector space over $\mathbb{C}$ with basis $\{\{t_1\}, ..., \{t_k\}\}$.
- **Example.** $M^{(n)}$ is a vector space over $\mathbb{C}$ with basis $\{\{1\ 2\ 3 ... n\}\}$ as all tableaux of shape (n) are row equivalent.
- Let t be a Young tableau with rows $R_1, .., R_\ell$ and columns $C_1, .., C_k$. Then,

$$R_t = S_{R_1} \times S_{R_2} \times \cdots \times S_{R_\ell}$$

and

$$C_t = S_{C_1} \times S_{C_2} \times \cdots \times S_{C_k}$$

are the **row-stabilizer** and **column-stabilizer**, respectively.
- **Remark.** The row-stabilizer permutes elements within each row of the Young tableau, and the column-stabilizer permutes elements within each of its columns.
- **Example.** Consider the Young tableau $t = \begin{array}{lll} 4\ 1\ 2 \\ 3\ 5 \end{array}$. The row-stabilizer is $S_{\{1,2,4\}} \times S_{\{3,5\}}$, and the column-stabilizer is $S_{\{3,4\}} \times S_{\{1,5\}} \times S_{\{2\}}$.
- The **polytabloid** associated with a tabloid $\{t\}$ is

$$e_t = \sum_{\pi \in C_t} \mathrm{sgn}(\pi) \pi \{t\}.$$

In particular, we can obtain e_t by adding together all tabloids obtained by column permutations of t.
- **Example.** Let $t = \begin{array}{lll} 1\ 3\ 5 \\ 4\ 2 \end{array}$, then

$$e_t = \left\{ \begin{array}{lll} 1\ 3\ 5 \\ 4\ 2 \end{array} \right\} - \left\{ \begin{array}{lll} 4\ 3\ 5 \\ 1\ 2 \end{array} \right\} - \left\{ \begin{array}{lll} 1\ 2\ 5 \\ 4\ 3 \end{array} \right\} + \left\{ \begin{array}{lll} 4\ 2\ 5 \\ 1\ 3 \end{array} \right\}.$$

- For any partition of n, λ, the corresponding **Specht module**, S^λ, is the submodule of M^λ spanned by the polytabloids e_t, where t is of shape λ.
- **Theorem.** Let λ partition n. The Specht modules S^λ are all of the irreducible representations of S_n over $\mathbb{C}$.

References & Acknowledgments

Thanks to my DRP mentor Andres Barei and the DRP Organizing Committee.

Bruce E. Sagan. *The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions*. Springer Science & Business Media, 2001.

Prime Numbers in RSA Cryptosystem

Sam Ream and Weimo Zhu, mentored by Charles Kulick

University of California, Santa Barbara - Directed Reading Program (DRP) 2024

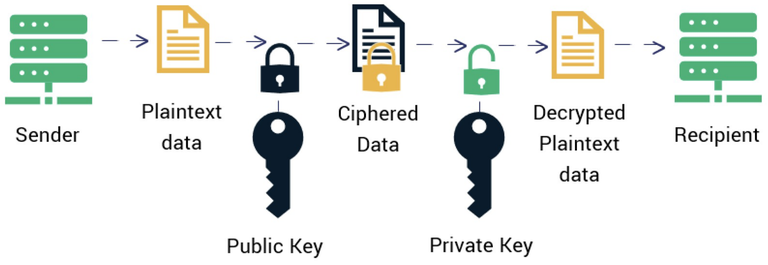
Introduction

Throughout history, the need for secure communication has always been an important issue in various areas, such as private communication during war or credit card encryption. Prior to 1970s, symmetric-key cryptosystems were mainly implemented. Such cryptosystems required the sender and the receiver to agree on a private key, which led to the difficulty of finding a secure line and exchanging keys without being intercepted. Later, public-key cryptosystems (asymmetric cryptosystems) were invented, where the sender and receiver could publicly agree on the public key and set their own private keys. Without the need to send private keys, public-key cryptosystems are much less vulnerable. The RSA cryptosystem is one of the most famous public-key cryptosystems.

RSA Algorithm

1. Receiver: Choose two distinct large prime numbers p and q.
2. Receiver: Compute $n = pq$ and $\phi(n) = (p-1)(q-1)$.
3. Receiver: Choose an integer e such that $gcd(e, \phi(n)) = 1$.
4. Receiver: Compute d such that $de \equiv 1 \pmod{\phi(n)}$.
5. Receiver: Send n, e publicly.
6. Sender: Send $c \equiv m^e \pmod{n}$.
7. Receiver: Decrypt $m \equiv c^d \pmod{n}$.

In general, d is the private key, (n, e) are the public keys.



Significance of Prime Numbers

To attack RSA, the most straightforward way is to factor n. With $n = pq$, the observer can compute $\phi(n) = (p-1)(q-1)$, and thus get the private key d. In modern implementations of the RSA cipher, the prime numbers p and q chosen to compute the encryption key n should be at least hundreds of digits long to ensure security. However, up until now, there was no known factoring method for the product of two hundreds-digit-length primes that could be done in a feasible time frame. Therefore, the high security level of RSA cryptosystems relies on the difficulty of factoring n to get p and q. In the following sections, we will introduce methods to find large primes needed for decryption and also possible attacks on factoring the product of two large primes.

References

[1] Evan Chen.
An Infinitely Large Napkin.
Evan Chen, 2023.

[2] Simon Rubinstein-Salzedo.
Cryptography.
Springer, 2018.

[3] Lawrence C. Washington Wade Trappe.
Introduction to Cryptography with Coding Theory.
Pearson, 2006.

Primality Testing

To choose our p and q, we need to find large, unique numbers and ensure their primality. Instead of trying to factorize an integer x, we can use primality testing to be more efficient.

In the following, we will introduce two probabilistic primality tests, which determine how likely it is that a given integer is prime.

Fermat Primality Test

Recall Fermat's Little Theorem:
If p is a prime number and a is not divisible by p, then $a^{p-1} \equiv 1 \pmod{p}$.

Implementing the contrapositive of Fermat's Little Theorem, Fermat "Primality Test" is actually a "Composite Test". It concludes x is composite if there exists an integer a such that $gcd(x, a) = 1$ and $a^{x-1} \not\equiv 1 \pmod{x}$. While if $a^{x-1} \equiv 1 \pmod{x}$, x is probably prime.

Given that this is a probabilistic primality test, there are infinitely many cases where this test fails. They are called Carmichael numbers or absolute pseudoprimes. For example, the smallest Carmichael number is 561.

Miller-Rabin Primality Test

Compared to Fermat Primality Test, Miller-Rabin Primality Test is stronger and has a lower probability in concluding a composite number as prime. It is currently used in many RSA implementations.

The main idea:

1. Given n, an odd integer.
2. Write $n - 1 = 2^k m$, where m is odd now.
3. Choose a random positive integer a such that $1 < a < n - 1$.
4. Compute $b_0 \equiv a^m \pmod{n}$.
5. If $b_0 \equiv \pm 1 \pmod{n}$, then stop the test and n is probably prime.
   If not, continue the test and compute $b_i \equiv b_{i-1}^2 \pmod{n}$.
6. If $b_i \equiv 1 \pmod{n}$, then n is composite.
   If $b_i \equiv -1 \pmod{n}$, then n is probably prime.
7. Iterate until stopping or reaching b_{k-1}.
   If $b_{k-1} \not\equiv -1 \pmod{n}$, then n is composite.
   If not, then n is probably prime.

Scan this QR code for the Python implementation of Miller-Rabin Primality Test.

For a given integer n and a choice of a, the probability of Miller-Rabin Test failing and wrongly declaring that a composite n is prime is at most $\frac{1}{4}$. Thus, if this test is repeated for k times, the probability of failing is at most $(\frac{1}{4})^k$. Repeating 5 times with 5 different a, we can reduce the probability of error to below 0.1%, which is usually accurate enough.

Reconsider the Carmichael number $n = 561$. Then $n - 1 = 560 = 2^4 \times 35$. Let $a = 2$. Then:

$$b_0 \equiv 2^{35} \equiv 263 \pmod{561}$$
$$b_1 \equiv {b_0}^2 \equiv 166 \pmod{561}$$
$$b_2 \equiv {b_1}^2 \equiv 67 \pmod{561}$$
$$b_3 \equiv {b_2}^2 \equiv 1 \pmod{561}$$

Since $b_3 \equiv 1 \pmod{561}$, 561 is correctly declared to be composite.

Factoring Attacks

The Birthday Attack

The motivation behind the birthday attack is the idea that, for example, if there are 23 people in a room, then there is about a 50% chance of two people sharing a birthday and additionally that probability increases to about 70% with 30 people in the room. In general, if there were n unique birthdays, it would take about $\sqrt{2n\log(2)}$ people before we would expect a match.

The birthday attack uses this idea to find factors of n. It will take us about $\sqrt{2n\log(2)}$ random numbers before we find a factor of n. The following algorithm allows us to do this efficiently:

1. Choose a random polynomial f(x) that maps values $\mathbf{Z}/n\mathbf{Z} \Rightarrow \mathbf{Z}/n\mathbf{Z}$.
2. Let $x = 2$ and $y = 2$ (standard convention).
3. Replace x with f(x) and y with f(f(y)).
4. Compute $d = gcd(|x - y|, n)$.
5. If $d = 1$, return to step 3. If $d = n$, then the algorithm fails, so we must restart at step 1 and pick a new function f(x). Otherwise, if $d \neq 1$ and $d \neq n$, then d is our factor of n.

Quadratic Sieve

In this factoring method, if we want to factor some number n, we must find integers x and y such that $x^2 \equiv y^2 (mod\, n)$, but $x \not\equiv y (mod\, n)$. In this case, n is composite and gcd(x-y,n) gives us our nontrivial factor of n.

In order to find our integers x and y, we must produce squares that are slightly larger than a multiple of n using $\lfloor in + j \rfloor$ for various values of i and small j.

Once we find our x integers, we must write them as products of primes less than 20, which will comprise our factor base. Each of our squares will represent a row of a matrix with the entries being the exponents of the primes. For example:

9398	0	0	5	0	0	0	0	1
19095	2	0	1	0	1	1	0	1
1964	0	2	0	0	0	3	0	0
17078	6	2	0	0	1	0	0	0
8077	1	0	0	0	0	0	0	1
3397	5	0	1	0	0	2	0	0
14262	0	0	2	2	0	1	0	0

Now, if we have a linear dependency mod 2 among the rows, the product of the numbers yields another square, our y^2. If $x \not\equiv \pm y (mod\, n)$, then gcd(x-y,n) gives us our factor of n.

The $p - 1$ Factoring Algorithm

1. Choose an integer $a > 1$ ($a = 2$ is common).
2. Choose a bound B. The size of B will depend on the situation, but if B is too small, the chance of success is small and if B is too big, then the algorithm is very slow.
3. Compute $b \equiv a^{B!} (mod\, n)$ where
   a. $b_1 \equiv gcd(b - 1, n)$.
   b. $b_j \equiv b_{j-1}^j (mod\, n)$.
   c. Then $b_B \equiv (mod\, n)$.
4. Let $d = gcd(b - 1, n)$. If $1 < d < n$, we have our nontrivial factor of n.

Rest A-Shor-ed: Quantum Computing Revolutionizes Cryptographic Security

Samuel Caruthers, Ljosha Kremliovsky, and Daric Zhou, mentored by Kyle Hansen

University of California, Santa Barbara

Abstract

Shor's Algorithm is a quantum algorithm used to efficiently factor large numbers. We investigate quantum computing and examine the idea of leveraging this algorithm as a quantum attack against the classical RSA cryptosystem.

Cryptography and RSA

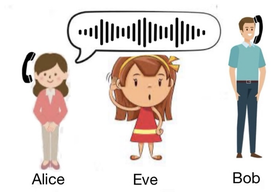


Fig. 1: Our main characters

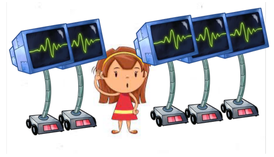


Fig. 2: Eve with her quantum computers

Euler's φ Function

Let $m \in \mathbb{N}$. The value $\varphi(m)$ is

$$\#\{k \in \mathbb{N} \mid \gcd(k, m) = 1,\ k \leq m\}$$

If $m = pq$ for p, q prime, then

$$\varphi(m) = (p-1)(q-1).$$

Euler's Theorem

For $x \in \mathbb{Z}$ with $gcd(x, n) = 1$, we have

$$x^{\varphi(n)} \equiv 1 \pmod{n}.$$

The RSA Algorithm

Bob's Private Knowledge

$p = 17$ $\quad q = 5$

$n = pq = 85$ $\quad \varphi(n) = (p-1)(q-1) = 64$

$e = 7$ $\quad d \equiv e^{-1} \pmod{\varphi(n)} = 55$

Bob's Public Key

$n = 85 \quad e = 7$

Bob Checks

p and q are prime

$\gcd(e, \varphi(n)) = 1$

Alice Computes

She encodes her message m as the number $m = 11$.

Computes $c \equiv m^e \pmod{n}$ or

$71 \equiv 11^7 \pmod{85}$

and sends c to Bob

Bob Decrypts

He computes $c^d \pmod{n}$ and gets

$$71^{55} \equiv 11 \pmod{85}$$

because, by Fermat's Little Theorem,

$$c^d \equiv m^{ed} \equiv m^{1+\beta\varphi(n)} \equiv m \pmod{n}.$$

Eve Wants to Know

She knows $85 = pq$ for *some* primes p and q.

If she knew p and q, then she could compute $\varphi(n) = (p-1)(q-1)$.

Then she could compute $d \equiv e^{-1} \pmod{\varphi(n)}$, and decrypt any messages sent to Bob with this public key.

If only she could factor n...

Eve Interferes and accesses c, but...

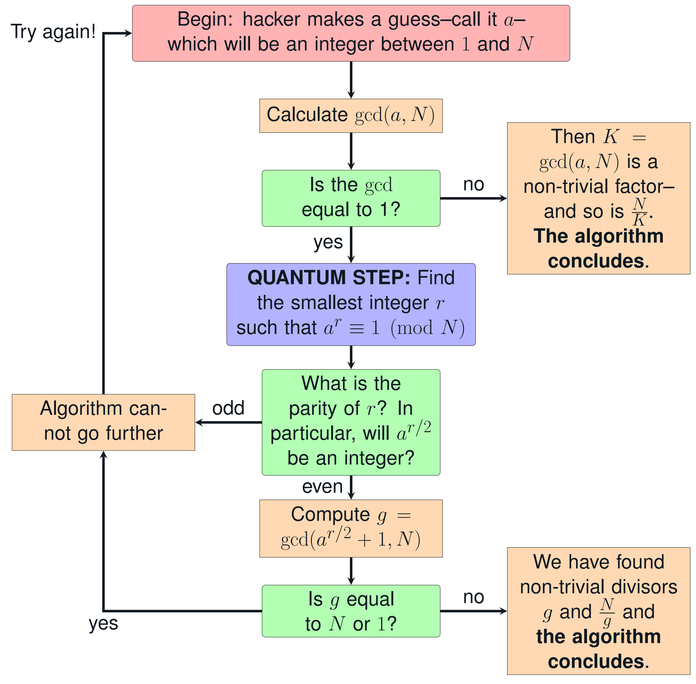
She only knows c, n, and e. She knows that

$$c \equiv m^e \pmod{n},$$

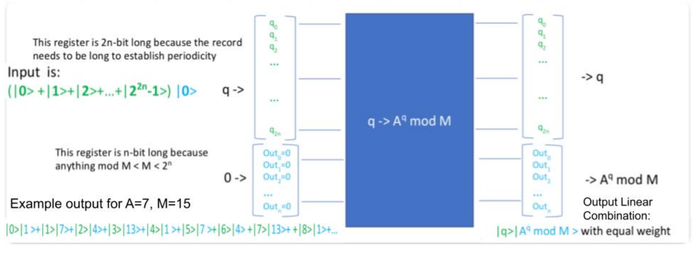
but can't figure out what m is without factoring n.

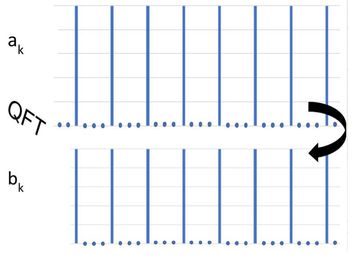
Shor's Algorithm

Herein lies our ultimate tool for decrypting the once-impenetrable fortress of RSA encryption: Shor's algorithm, a quantum breakthrough poised to shatter the very foundation of digital security. We have a simple goal: **find the factors of some integer** N. Let's go through the process to do so:



0	1	2	3	4	5	6	7	q
1	7	4	13	1	7	4	13	$7^q \bmod 15$





Shor's Algorithm used to find the period of

$$f(q) = A^q \pmod{M}$$

where $M = 15$, and $A = 7$. In this case, it is easy to see that $p = 4$. Images adapted from [5].

The Quantum Step

Without regard for the quantum step, Shor's algorithm is a rather straightforward way to find the factors of large integers and destroy cybersecurity. But the quantum step is crucial–how does it go? Let's discuss:

In quantum mechanics, information is encoded by **qubits** which can exist simultaneously (until measured). These qubits can be superpositioned or entangled using **quantum gates** in ways that determine their outcome when measuring them–how likely certain qubits are measured are determined by their **probability amplitude**. Here is the process:

$$\begin{array}{l} |x\rangle \ \text{—}\boxed{H}\text{—}\bullet\text{—} \\ |a^x \pmod N\rangle \ \text{———}\oplus\text{—} \end{array} \quad \frac{1}{\sqrt{Q}} \sum_{x=0}^{Q-1} |x\rangle |a^x \pmod N\rangle \quad (*)$$

where $Q > N^2$. We measure $(*)$, and we obtain a quantum interference that collapses the first register containing $|x\rangle$ into a singular $y = a^{x_0} \pmod N$. But properties of modular arithmetic tell us that $x_0 + kr$ satisfy the equation for all k and a singular r, so we take a superposition of x_0 like so:

$$|x_0\rangle \ \text{—}\boxed{H}\text{—} \ \frac{1}{\sqrt{r}} \sum_{k=0}^{r-1} |x_0 + kr\rangle \quad (**)$$

The output reveals the periodicity of the function, precisely captured by r. And in order to extract this value from our periodic function, we will apply a **quantum Fourier transform**, which, when applied to $(**)$, gives us

$$\frac{1}{\sqrt{Q}} \sum_{c=0}^{Q-1} e^{\frac{2\pi i x_0 c}{Q}} \left(\sum_{k=0}^{r-1} e^{\frac{2\pi i k r c}{Q}} \right) |c\rangle$$

This looks quite messy, but what's important to us is that this QFT will **constructively interfere** at multiples of r–take a look at the series in the parenthesis. It is a geometric series whose value will be large when $r\frac{c}{Q}$ is close to an integer. We can measure the c that gives us an integer multiple of r by using a classical post-processing algorithm such as continued fractions to obtain r–and once we do, the process concludes, and the algorithm continues.

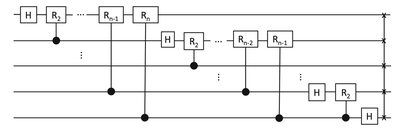


Fig. 5: Quantum Fourier Transform

Image from [4]

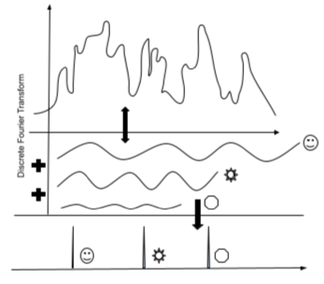


Fig. 6: Fourier Transform

Acknowledgements & References

We would like to thank the DRP team for organizing the program this year. We would also like to thank our mentor Kyle for his patient guidance and mentorship during this whole process.

References

[1] *Bloch Sphere*. https://en.wikipedia.org/wiki/Bloch_sphere. Accessed: 1 May 2024.

[2] David J. Hunter. *Cryptography and Coding Theory*. URL: https://djhunter.github.io/cryptography/.

[3] *Karen Plankton*. URL: https://spongebob.fandom.com/wiki/Karen_Plankton.

[4] Ray LaPierre. *Introduction to quantum computing*. Springer Nature, 2021.

[5] Michael Pushkarsky. "Seminar - quantum computers". Unpublished.

[6] Wade Trappe. *Introduction to cryptography with coding theory*. Pearson Education India, 2006.

Self Consistent Field Theory: The numerical magic behind polymer engineering

Yibo Liang and Weishu Tan
University of California, Santa Barbara

Introduction to Polymer Physics

Suppose there are many linear homopolymer chains dancing in a spatial domain Ω with volume V. The **chemical potential** of such system can be represented as a smooth function $\mu : \Omega \to \mathbb{R}$.

Objective

Give a numerical algorithm which takes a function ρ as input and returns an optimal $\mu^* : \Omega \to \mathbb{R}$ to minimize the functional,

$$G[\mu] := \int_\Omega d\mathbf{r}\mu(\mathbf{r})\rho(\mathbf{r}) + n \ln Q[\mu] \tag{1}$$

where n is the number of identical homopolymer chains in the melt, $\rho : \Omega \to [0, +\infty)$ **prescribed monomer density profile**, a smooth function satisfying **periodic boundary conditions**, and Q is another functional which can be computed by:

$$Q[\mu] = \frac{1}{V}\int_\Omega d\mathbf{r} q(\mathbf{r}, 1; [\mu]) \tag{2}$$

where $q(\cdot, \cdot, [\mu]) : \Omega \times [0, 1] \to \mathbb{R}$ the **chain propagator** corresponds to μ, satisfying Fokker-Planck Equations,

$$\frac{\partial}{\partial s}q(\mathbf{r}, s, [\mu]) = \nabla^2 q(\mathbf{r}, s, [\mu]) - \mu(\mathbf{r})q(\mathbf{r}, s, [\mu]), \quad q(\mathbf{r}, 0, [\mu]) = 1 \tag{3}$$

In this problem, the G is convex and thus any local minimum is also global minimum. Moreover, its gradient is given by,

$$\frac{\delta G[\mu]}{\delta \mu(\mathbf{r})} = \rho(\mathbf{r}) - \tilde{\rho}(\mathbf{r}, [\mu]) \tag{4}$$

where $\tilde{\rho}(\cdot, [\mu]) : \Omega \to \mathbb{R}$ is the **monomer density** generated by μ,

$$\tilde{\rho}(\mathbf{r}, [\mu]) = \frac{\rho_0}{Q[\mu]}\int_0^1 ds\, q(\mathbf{r}, s, [\mu])q(\mathbf{r}, 1-s, [\mu]) \tag{5}$$

where $\rho_0 := \frac{1}{V}\int d\mathbf{r}\rho(\mathbf{r})$ is called the volume-averaged monomer density of the system.

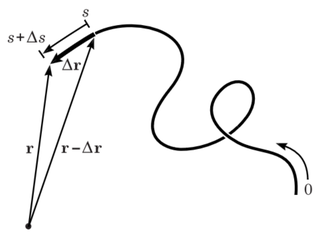


Fig 1 (Gaussian chain model [2] pg. 42)

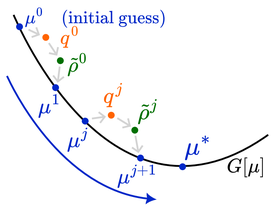


Fig 2 (Gradient descent on μ)

Self Consistent Field Theory

The system above can be solved using **Self Consistent Field Theory (SCFT)**. Since $G[\mu]$ is convex, a natural way to find $\mu^*(\mathbf{r})$ would be an iterative process where we traverse down the gradient of G — a **gradient descent** (See Fig 2). Here, we introduce a fictitious time variable to denote our iteration step. Upon each iteration, the $\mu^j \to \mu^{j+1}$ update is given by,

$$\mu^{j+1/2}(\mathbf{r}) = \mu^j(\mathbf{r}) - \Delta t\left(\frac{\delta G[\mu]}{\delta\mu(\mathbf{r})}\bigg|_{\mu=\mu^j}\right) \tag{6}$$

$$\mu^{j+1}(\mathbf{r}) = \mu^{j+1/2}(\mathbf{r}) - \frac{1}{V}\int d\mathbf{r}\,\mu^{j+1/2}(\mathbf{r}) \tag{7}$$

Equation (6) is the **forward Euler formula** applied at the half-step, but equation (7) is a correction to keep the mean 0 and lift the degeneracy of the solution. Additionally, the integral in equation (7) can be evaluated by **composite trapezoidal rule**, which is highly efficient over periodic boundary conditions. Note that this scheme is **conditionally stable**, so Δt must be chosen appropriately.

Solving 1-D Fokker-Planck Equations

Discretizing the space: Let $\Omega = [0, L]$ 1-D spatial domain, $0 = x_0 < x_1 < \cdots < x_{N_x-1} = L$ equispaced partition of Ω where N_x spatial resolution, $\Delta x := L/(N_x - 1)$.
Discretizing the chains: Let $0 = s_0 < s_1 < \cdots < s_{N_s-1} = 1$ equispaced partition of $[0, 1]$ where N_s chain resolution, $\Delta s := 1/(N_s - 1)$.

To solve eqn. (3) efficiently, we use the so-called **pseudo-spectral method**. Fix a μ^j and rewrite eqn. (3) as,

$$\frac{\partial}{\partial s}q^j(x, s) = \mathcal{L}q^j(x, s)$$

where differential operators $\mathcal{L} := \mathcal{L}_D + \mathcal{L}_W$, $\mathcal{L}_D := d^2/dx^2$, and $\mathcal{L}_W(q) := -\mu^j q$. The exact solution of this differential equation is $q^j(x, s) = \exp(s\mathcal{L})q^j(x, 0)$ with initial condition $q^j(x, 0) = 1$ for all x.

Proposition 1: Strang Splitting

Suppose $\mathcal{L}$, $\mathcal{L}_A$, and $\mathcal{L}_B$ are differential operators with $\mathcal{L} = \mathcal{L}_A + \mathcal{L}_B$. Then:

$$\exp(\Delta s\mathcal{L}) = \exp(\Delta s\mathcal{L}_A/2)\exp(\Delta s\mathcal{L}_B)\exp(\Delta s\mathcal{L}_A/2) + O(\Delta s^3) \text{ as } \Delta s \to 0.$$

From this proposition, the $q^j(x, s) \to q^j(x, s + \Delta s)$ update is given by:

$$q^j(x, s + \Delta s) \approx \exp[-\Delta s\mu^j(x)/2]\exp[\Delta s\mathcal{L}_D]\exp[-\Delta s\mu^j(x)/2]q(x, s) \tag{8}$$

While the first and third terms $\exp[-\Delta s\mu^j(x)/2]$ are **diagonalizable**, the middle term $\exp[\Delta s\mathcal{L}_D]$ is not. However, the middle term is diagonalizable in Fourier space, that is:

Proposition 2: In Fourier space, $\exp(\Delta s\mathcal{L}_D)$ behaves like function multiplication

Let $f(x)$ be a smooth L-periodic function with domain $[0, L]$. Discretize f as $\{f(x_l) : l = 0, \ldots, N_x - 1\}$, and let $\hat{f}_k$ denote the k-th entry of the FFT of f; then, for each $k = 0, 1, \ldots, N_x - 1$:

$$\widehat{\exp(\Delta s\mathcal{L}_D)f}_k = \begin{cases}\exp\left[-4\pi^2k^2\Delta s/L^2\right]\cdot\hat{f}_k & \text{if } k \le N_x/2\\ \exp\left[-4\pi^2(N_x-k)^2\Delta s/L^2\right]\cdot\hat{f}_k & \text{if } k > N_x/2\end{cases} \tag{9}$$

The **central result** of this section is an algorithm which takes a chemical potential $\mu^j : \Omega \to \mathbb{R}$ as input, and produce a corresponding chain propagator $q^j(x, s) : \Omega \times [0, 1] \to \mathbb{R}$:

Discretized Pseudo-Spectral Algorithm, with time complexity $O(N_sN_x \log N_x)$

Input: $\{\mu^j(x_l) : l = 0, \ldots, N_x - 1\}$ - The chemical potential at j-th iteration step.
Steps:

- 1: Set $q^j(x_l, s_0) = 1$ for each l
- 2: For each $n = 1, \ldots, N_s - 1$ do:
  - 2a: Let $q'_l = \exp[-\Delta s\mu(x_l)/2]\cdot q^j(x_l, s_{n-1})$ for each l
  - 2b: Let $q'' =$ the FFT of q'
  - 2c: Let $q'''_k = \exp(-4\pi^2k^2\Delta s/L^2)\cdot q''_k$ for each $k \le N_x/2$
  - 2d: Let $q'''_k = \exp[-4\pi^2(N_x-k)^2\Delta s/L^2]\cdot q''_k$ for each $k > N_x/2$
  - 2e: Let $q'''' =$ the inverse FFT of q'''
  - 2f: Set $q^j(x_l, s_n) = \exp[-\Delta s\mu(x_l)/2]\cdot q''''_l$ for each l

Output:$\{q^j(x_l, s_n) : l = 0, \ldots, N_x - 1, n = 0, \ldots, N_s - 1\}$ - The chain propagator corresponds to μ^j.

Once we have q^j, we can put it into eqn. (2) first to compute $Q[\mu^j]$ and put q^j and $Q[\mu^j]$ together into eqn. (5) to get $\tilde{\rho}^j$. Then, from eqn. (4), the new gradient of G is given by $\rho - \tilde{\rho}^j$. Therefore, **SCFT is a complete algorithm** for solving this minimization problem.

Concrete Example: 1-D Case

Consider the target density profile,

$$\rho(x) = \rho_0[1 + \tanh(\eta\cos(2\pi L^{-1}x))], \quad 0 \le x \le L \tag{10}$$

Set $L = 10$, $\rho_0 = 0.5$, $N_x = N_s = 128$, $\Delta t = 2.5$, and run the algorithm for both $\eta = 2$ and $\eta = 5$. The result is shown in Fig 3.

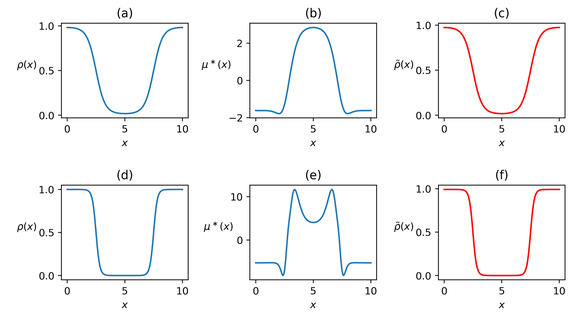


Fig 3 (Sample inputs and outputs of the algorithm) (a) Input ρ specified by eqn. (10) with $\eta = 2$. (b) Output μ^*. (c) $\tilde{\rho}$ generated by μ^*, where this $\tilde{\rho}$ should be identical to (a). (d), (e), (f) are given by the same routine with $\eta = 5$.

Using Neural Network to Optimize

With that being said, the Big-O for the Discretized Pseudo-Spectral Algorithm is $O(N_sN_x \log N_x)$; **for large systems, this is still expensive**. Therefore, we propose a **machine learning approach** where we train a Feed-Forward Neural Network (FNN) to learn μ^* from a given target density ρ.

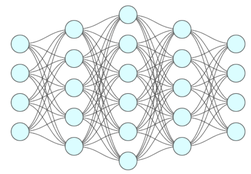


Fig 4 (FNN Architecture) Not representative of actual layer widths (128-256-512-256-128); compile with optimizer **ADAM** and loss function **MSE**.

Taking $N_x = 128$, $\eta = 5$, and using a periodic Gaussian Process to generate 1000 training data, we attempt to learn the optimal μ^* (d) of Fig 3.

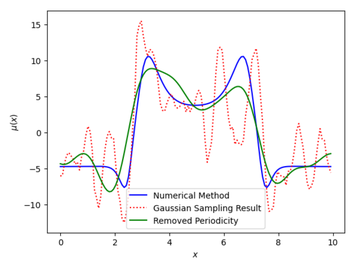


Fig 5 (Learned μ^* with FNN) Model output along with post-processing that removes output periodicity.

The model successfully learned the global behavior; however, we suspect that the model picked up on the periodic trends within the training data. To confirm this suspicion, we removed high-frequency Fourier components from the FNN predictions and observed an improved fit.

Acknowledgments and References

We thank Will Sheppard for his guidance as well as the UCSB Directed Reading Program for the opportunity to work on this project.

[1] Ceniceros, Hector D., and Fredrickson, Glenn H. "Numerical solution of polymer self-consistent field theory." *Multiscale Modeling & Simulation* 2, no. 3 (2004): 452-474.
[2] Fredrickson, Glenn H. *The equilibrium theory of inhomogeneous polymers*. No. 134. Oxford University Press, 2006.