Exploring the Jump From Pre to Post Quantum Cryptography

Fei Du and Carly Greutert, advised by Joel E. Pion
UC Santa Barbara Directed Reading Program (DRP)

Traditional Cryptography

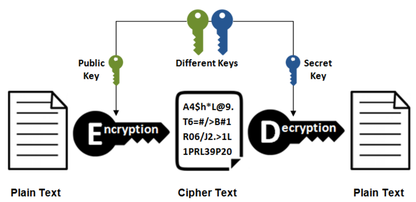
Symmetric encryption

1. Caesar Shift (50 BCE)
2. Vigenère (1553)
3. Enigma Machine (1920)

Asymmetric encryption

1. Rivest–Shamir–Adleman (1977)
2. Elliptic-Curve Cryptography (1985)

Asymmetric Encryption



Hidden Subgroup Problem (HSP)

Suppose there is a known group G and a function $f : G \to S$ where S is some finite set.
Suppose f has the property that there exists a subgroup $H \leq G$ such that f is constant within each coset, and distinct on different cosets: $f(g) = f(g') \iff gH = g'H$. This condition says f is well-defined on the set of left cosets G/H. Since H may be large, "finding H" typically means finding a **generating set** for H.

Discrete logarithm

Given a generator γ (people often use a prime number) of a cyclic multiplicative group C of size N.
This means that $C = \{\gamma^a | a \in \{0, ..., N-1\}\}$, and $A \in C$, can we find the unique $a \in \{0, 1, ..., N-1\}$ such that $\gamma^a = A$?

Classical computers need a lot of time to compute a from A (need time roughly exponential in $\log N$). Take $G = \mathbb{Z}_N \times \mathbb{Z}_N$ and define $f : G \to C$ by $f(x, y) = \gamma^x A^{-y}$. For group elements $g_1 = (x_1, y_1), g_2 = (x_2, y_2) \in G$ we have

$$f(g_1) = f(g_2) \iff \gamma^{x_1 - ay_1} = \gamma^{x_2 - ay_2} \iff (x_1 - x_2) = a(y_1 - y_2) \iff g_1 - g_2 \in \langle (a, 1) \rangle$$

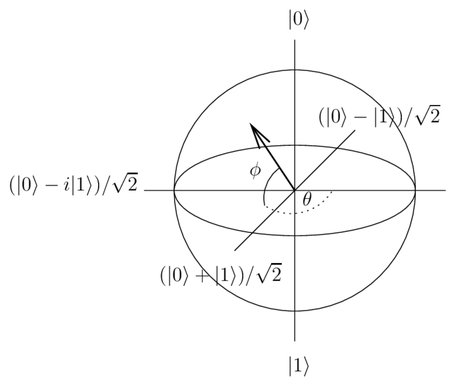
Let H be the subgroup of G generated by the element $(a, 1)$, then finding the generator of the hidden subgroup H gives us a.

The Abelian Case

If G is Abelian, the QFT operator shown on the right helps compute a generating set $\mathcal{L}$ for the period lattice

$$L = \{(x_1, ..., x_n) | \sum_{i=1}^{n} g_i^{x_i} \in H\}$$

Bloch Sphere Illustration



A geometric representation of a qubit

Quantum Fourier Transformation (QFT)

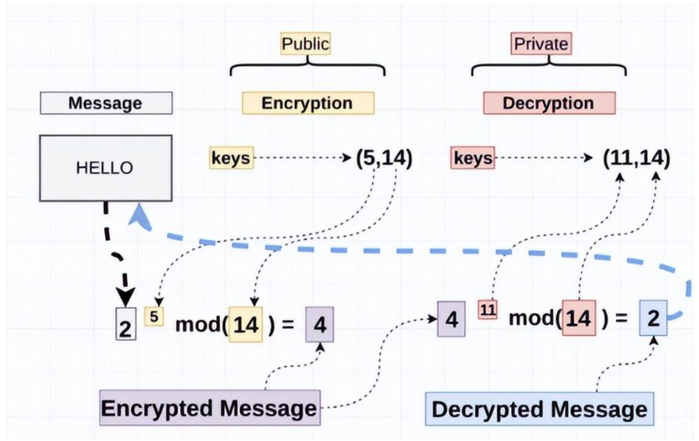
Suppose we have a group G, a set generating G, $\{g_1, ..., g_n\}$, a periodic function f on $\mathbb{Z}^n$ where there exists a normal subgroup H of G ($GHG^{-1} = H$), and an injective function g on the quotient group G/H such that

$$f(x_1, ..., x_n) = g(\sum_{i=1}^{n} g_i^{x_i} \bmod H)$$

The HSP then asks us to present a generating set of the largest such H and the relations between its elements. Define a 2^n dimensional Hilbert space as follows: $\mathcal{H}_n = \mathcal{H} \otimes \cdots \otimes \mathcal{H} = \mathbb{C} \oplus \mathbb{C} \otimes \cdots \otimes \mathbb{C} \oplus \mathbb{C}$. The QFT operator is then defined on an interval of length $N = 2^n$ below:

$$QFT_n : \mathcal{H}_n \longrightarrow \mathcal{H}_n : |x\rangle \longrightarrow 2^{\frac{-N}{2}} \sum_{y=0}^{N-1} e^{\frac{2\pi i x y}{N}} |y\rangle$$

RSA Illustration



1. Pick two prime numbers $p = 2$ and $q = 7$ and multiply them to get the modulus 14
2. Compute $L = \text{LCM}(p - 1, q - 1) = 6$ and choose the integer (public key) $e = 5$ such that $1 < e < L$ and $\text{GCD}(e, L) = 1$
3. Solve the private key $d = 11$ such that $d \cdot e = 1 (\text{MOD } L)$

RSA Algorithm

Step One (Key Generation): Choose two secret prime numbers, **p** and **q** (typically, p and q are very large to ensure your message is secure). Then, multiply them together to obtain **n**, the modulus for encryption/decryption. n is a part of the publicly available key.

Then, compute **L(n)**=lcm(L(p), L(q))=lcm(p-1, q-1), and keep L(n) a secret. We then choose a number **e** such that $1 < e < L(n)$ and gcd(e, L(n))=1 (i.e. e and L(n) are **relatively prime**). The integer e is then released as part of the public key. Note the size and length of e will determine how fast and secure the encryption is.

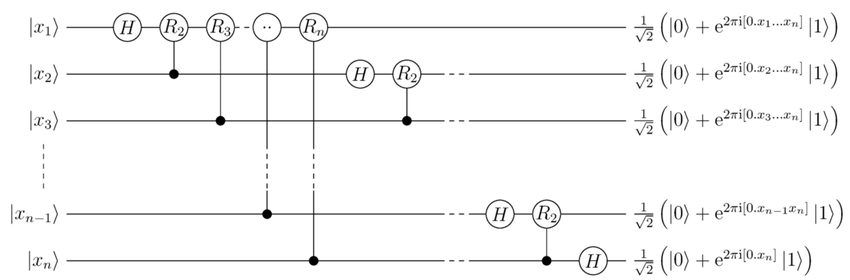
Finally, solve for **d** (the **modular multiplicative inverse** of e modulo L(n) in $d \equiv e^{-1} (\text{MOD } L(n))$. We know such an inverse exists since e and L(n) are coprime. This d will work as our **private key** component.

Step Two (Key Distribution): Suppose Alice is sending a message to Bob. Alice must know Bob's public key (n, e) to encrypt the message, and Bob must use his private key (d) to decrypt the message.

Step Three (Encryption): After Alice obtains Bob's public key, she can send a message M by converting it into an integer from **plain text** such that $(0 \leq M \leq n)$, $M \in \mathbb{Z}$. She computes the **cipher text (c)** by $c \equiv M^e \text{ MOD } n$. Alice then sends c to Bob.

Step Four (Decryption): Once Bob receives the cipher text, he can compute Alice's message M by solving $c^d \equiv (M^e)^d \equiv M^{ed} \equiv M^{k(L(n))+1} \equiv M(M^{k(L(n))}) \equiv M(1) \equiv M \text{ MOD } n$.

QFT Illustration



Implementation of the discrete Fourier transform on $\mathbf{2^n}$ amplitudes into a quantum circuit consisting of only $\frac{\mathbf{n(n+1)}}{\mathbf{2}}$ Hadamard gates **H** (the gate that creates an equal superposition of the two basis states: $|0\rangle \longrightarrow \frac{|0\rangle + |1\rangle}{\sqrt{2}}, |1\rangle \longrightarrow \frac{|0\rangle - |1\rangle}{\sqrt{2}}$, so $H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$) and controlled phase shift gates, $\boldsymbol{R_m} = \begin{bmatrix} 1 & 0 \\ 0 & e^{\frac{2\pi i}{2^m}} \end{bmatrix}$, that modify the phase of the quantum state. Note **n** is the number of qubits.

Related Topics

- Elliptic Curve Cryptography (Pollard's p-1 and Lenstra's Factorization Algorithms)
- Classical Cryptosystems Not Yet Broken by the Quantum Algorithm (McEliece, NTRU, and Lattice-Based public key encryptions)
- Special Cases of the HSP (Pell's Equation, Non-abelian Groups, etc)
- Extended Euclidean Algorithm & Bezout's Identity

Acknowledgements and References

We would like to thank our wonderful advisor, Joel Pion, for his support and guidance through the intricate world of cryptography.
We would also like to thank the 2022 DRP program for giving us this opportunity to further our studies in a supportive environment.

[1] Bernstein, Daniel J., et al., editors. Post-Quantum Cryptography. Springer-Verlag , 2009.

[2] Ronald de Wolf. "Quantum Computing Lecture Notes, Extra Chapter".

The Bernstein Problem

Xingzhe Li, Graduate Mentor: Junrong Yan

Introduction

In minimal surface theory, the celebrated Bernstein problem is as follows: if the graph of a function on $\mathbb{R}^{n-1}$ is a minimal surface in $\mathbb{R}^n$, does this imply that the function is linear? This is proven to be true in dimensions at most 8 but false in dimensions at least 9. Bernstein solved $n = 3$ case at the beginning of 20th century. In 1962, Fleming gave a new proof by deducing it from the fact that all area-minimizing hypercones in $\mathbb{R}^3$ are flat. A few years later, De Giorgi solved $n = 4$ case and Almgren solved $n = 5$ case. In 1968, Simons showed that all area-minimizing hypercones in $\mathbb{R}^7$ are flat, thus extending the Bernstein theorem to dimension 8. Moreover, he gave examples of locally stable cones in $\mathbb{R}^8$, which were proven to be area-minimizing by Bombieri, De Giorgi, and Giusti in 1969. They also showed that there exists complete minimal graphs that are not hyperplanes for $n \geq 9$. Combined with the result of Simons, this gives a complete solution to Bernstein problem in $\mathbb{R}^n$.

Minimal Submanifolds

Let (M^n, g) be a Riemannian manifold with Levi-Civita connection ∇ and let Σ be a k-dimensional submanifold of M. If $X \in \mathfrak{X}(\Sigma)$, then let X^T and X^N denote the tangential and normal components, respectively. For $X, Y \in T_x\Sigma$, the vector-valued bilinear form A on Σ is given by

$$A(X, Y) = (\nabla_X Y)^N.$$

In literature, A is called the second fundamental form, and the trace of A at x is the mean curvature vector

$$H = \sum_{i=1}^{k} A(E_i, E_i),$$

where E_i is an orthonormal basis for $T_x\Sigma$. The normed squared of the second fundamental form at x is

$$|A|^2 = \sum_{i,j=1}^{k} |A(E_i, E_j)|^2.$$

An immersed submanifold $\Sigma^k \subset M^n$ is said to be minimal if the mean curvature H vanishes everywhere. This is equivalent to Σ^k being the critical point for the area functional. In particular, if $\Sigma \subset \mathbb{R}^3$ is a graph of C^2 function $u : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$, then Σ satisfies the minimal surface equation

$$\operatorname{div}\left(\frac{\nabla u}{\sqrt{1 + |\nabla u|^2}}\right) = (1 + u_y^2)u_{xx} + (1 + u_x^2)u_{yy} - 2u_x u_y u_{xy} = 0.$$

Examples of embedded minimal surfaces in $\mathbb{R}^3$ include the helicoid $H = \{(t\cos s, t\sin s, s) | t, s \in \mathbb{R}\}$ and the catenoid $C = \{(x_1, x_2, x_3) | x_1^2 + x_2^2 = (\cosh x_3)^2\}$. By viewing H as the graph of function $u(x, y) = \arctan(y/x)$, one can check that the minimal surface equation holds.

Figure 1: The Helicoid

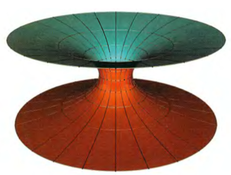


Figure 2: The Catenoid

References

[1] Tobius Holck Colding and William P. Minicozzi II. *A Course in Minimal Surfaces*. Volume 121 of Graduate studies in mathematics. American Mathematical Society, 2011.

[2] Xin Zhou. Lecture notes on minimal surfaces, 2021.

[3] Wendell H. Fleming. On the oriented plateau problem. *Rendiconti del Circolo Matematico di Palermo*, 11:69–90, 1962.

[4] James Simons. Minimal varieties in riemannian manifolds. *Annals of Mathematics*, 88:62–105, 1968.

[5] E. Giusti E. Bombieri, E. De Giorgi. Minimal cones and the bernstein problem. *Inventiones mathematicae*, 7:243–268, 1969.

The Second Variation Formula

Let $\Sigma^k \subset M^n$ be a minimal submanifold and let $F : \Sigma \times (-\epsilon, \epsilon) \to M$ be a variation of Σ with compact support and fixed boundary. In terms of local coordinates, we have the pullback metric $g_{ij}(t) = g(F_{x_i}, F_{x_j})$, the measure $\nu(t) = \sqrt{\det(g_{ij}(t))} \cdot \sqrt{\det(g^{ij}(0))}$, and the area formula

$$\operatorname{Vol}(F(\Sigma, t)) = \int \nu(t)\sqrt{\det(g_{ij}(0))}.$$

Since

$$\left.\frac{d^2}{dt^2}\right|_{t=0} \operatorname{Vol}(F(\Sigma, t)) = \int \nu''(0)\sqrt{\det(g_{ij}(0))},$$

it suffices to derive a formula for $\nu''(0)$ at some $x \in \Sigma$. Choose the normal coordinate system at x. By differentiating the first variation formula $2\nu'(t) = \operatorname{Tr}(g'_{ij}(t)g^{lm}(t))\nu(t)$, we obtain that

$$\begin{aligned} 2\nu''(0) &= \left.\frac{d}{dt}\right|_{t=0} (\operatorname{Tr}(g'_{ij}(t)g^{lm}(t))\nu(t)) \\ &= \left.\frac{d}{dt}\right|_{t=0} (\operatorname{Tr}(g'_{ij}(t)g^{lm}(t))) + \operatorname{Tr}(g'_{ij}(0)g^{lm}(0)) \cdot \frac{1}{2}\operatorname{Tr}(g'_{ij}(0)g^{lm}(0)) \\ &= \frac{1}{2}[\operatorname{Tr}(g'_{ij}(0))]^2 + \operatorname{Tr}(g''_{ij}(0)) - \operatorname{Tr}(g'_{ij}(0)g'_{lm}(0)). \end{aligned}$$

At the point x, we have

$$\begin{aligned} &|g'(0)|^2 = 4|\langle A(\cdot,\cdot), F_t\rangle|^2; \\ &\operatorname{Tr}(g''(0)) = 2|\langle A(\cdot,\cdot), F_t\rangle|^2 + 2|\nabla^N_\Sigma F_t|^2 + 2\operatorname{Tr}\langle R_M(\cdot, F_t)F_t, \cdot\rangle + 2\operatorname{div}_\Sigma(F_{tt}). \end{aligned}$$

Substituting $|g'(0)|^2$ and $\operatorname{Tr}(g''(0))$ inside yields that

$$\nu''(0) = -|\langle A(\cdot,\cdot), F_t\rangle|^2 + |\nabla^N_\Sigma F_t|^2 - \operatorname{Tr}_\Sigma\langle R_M(\cdot, F_t)\cdot, F_t\rangle + \operatorname{div}_\Sigma(F_{tt}) \text{ and}$$

$$\begin{aligned} \left.\frac{d^2}{dt^2}\right|_{t=0} \operatorname{Vol}(F(\Sigma, t)) &= -\int_\Sigma |\langle A(\cdot,\cdot), F_t\rangle|^2 + \int_\Sigma |\nabla^N_\Sigma F_t|^2 - \int_\Sigma \operatorname{Tr}_\Sigma\langle R_M(\cdot, F_t)\cdot, F_t\rangle \\ &= -\int_\Sigma \langle F_t, LF_t\rangle, \end{aligned}$$

where L is the stability operator introduced below.

The Stability Inequality

Suppose that Σ has a trivial normal bundle. By identifying a normal vector field $X = \eta N$ with η, we define the stability operator L as

$$L\eta = \Delta_\Sigma\eta + |A|^2\eta + \operatorname{Ric}_M(N, N)\eta.$$

In particular, if $M = \mathbb{R}^n$, then the Ricci tensor vanishes everywhere and

$$L\eta = \Delta_\Sigma\eta + |A|^2\eta.$$

We say that a minimal submanifold $\Sigma^k \subset M^n$ is stable if for all variations F with boundary fixed,

$$\left.\frac{d}{dt}\right|_{t=0} \operatorname{Vol}(F(\Sigma, t)) = -\int_\Sigma \langle F_t, LF_t\rangle \geq 0.$$

Intuitively, being stable means that the second derivative is positive and the graph is convex. Substituting the formula for L inside and applying the divergence theorem yield the stability inequality

$$\int_\Sigma (\inf_M \operatorname{Ric}_M + |A|^2)\eta^2 \leq \int_\Sigma |\nabla_\Sigma\eta|^2,$$

where $\Sigma^{n-1} \subset M^n$ is a stable minimal hypersurface with trivial normal bundle. In particular, if $M = \mathbb{R}^n$, then the stability inequality reduces to

$$\int_\Sigma |A|^2\eta^2 \leq \int_\Sigma |\nabla_\Sigma\eta|^2.$$

Acknowledgement

I would like to thank my mentor Junrong Yan for helping me develop the intuition for concepts in minimal surface theory and answering my numerous questions about proofs. I would also like to thank the organizer of 2022 UCSB DRP for running this fantastic program.

The Bernstein Theorems

The Bernstein theorem says that if $u : \mathbb{R}^{n-1} \to \mathbb{R}$ is an entire solution to the minimal surface equation and $n \leq 8$, then u is a linear function.

To see why it's true for $n \leq 6$, we make use of the L^p bound of $|A|^2$ for stable hypersurfaces along with the area bound. Let $\Sigma^{n-1} \subset \mathbb{R}^n$ be an orientable stable minimal hypersurface. For all $p \in [2, 2 + \sqrt{2/(n-1)}]$ and every nonnegative Lipschitz function ϕ with compact support, we have the estimate

$$\int_\Sigma |A|^{2p}\phi^{2p} \leq C(n, p)\int_\Sigma |\nabla\phi|^{2p}.$$

The proof is just a computation involving the stability inequality, the Cauchy-Schwarz inequality, the absorbing inequality $2xy \leq \epsilon x^2 + y^2/\epsilon$, and the Simons' inequality

$$|A|\Delta|A| + |A|^4 \geq \frac{2}{n-1}|\nabla|A||^2.$$

Suppose in addition that Σ is complete and

$$\sup_{R>0} \frac{\operatorname{Vol}(B_R \cap \Sigma)}{R^{n-1}} \leq V$$

for some $V < \infty$.

If we consider $2p = 4 + \sqrt{7/5} < 4 + \sqrt{8/(n-1)}$, then the above L^p bound of $|A|^2$ for the cutoff function

$$\phi(x) = \begin{cases} 1, & \text{if } |x| \leq r \\ 0, & \text{if } |x| \geq 2r \\ -\frac{1}{r}|x| + 2, & \text{otherwise} \end{cases}$$

implies that

$$\begin{aligned} \int_{B_r\cap\Sigma} |A|^{4+\sqrt{7/5}} &\leq C(n, p)r^{-4-\sqrt{7/5}} \operatorname{Vol}(B_{2r} \cap \Sigma) \\ &\leq C(n, p)2^{n-1}Vr^{n-5-\sqrt{7/5}} \to 0 \text{ as } r \to \infty. \end{aligned}$$

It follows that $|A|^2$ vanishes everywhere and Σ is flat.

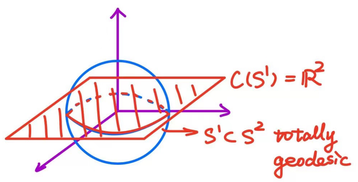
Now, if $u : \mathbb{R}^{n-1} \to \mathbb{R}$ solves the minimal surface equation entirely, then the area-minimizing property of a minimal graph (deduced from the monotonicity formula along with the calibration argument) gives an area bound and the previous argument shows that Σ is a hyperplane. Hence, u is a linear function and the Bernstein theorem holds for $n \leq 6$. For $6 < n \leq 8$, the proof relies on the fact that the hyperplanes are the only area-minimizing hypercones in $\mathbb{R}^n$ for $3 \leq n \leq 7$, which will be explained below.

Minimal Cones

Let N^{k-1} be a submanifold of $S^{n-1} \subset \mathbb{R}^n$. The cone over N is a smooth k-dimensional submanifold away from the origin

$$C(N) = \{x \in \mathbb{R}^n | x/|x| \in N\}.$$

It's immediate from definition that a cone is invariant under dilations about the origin. An example is given by the cone over the equator of S^2, which is just the horizontal plane. More generally, if S^{k-1} is a totally geodesic $(k-1)$-sphere in S^{n-1}, then $C(S^{k-1})$ is a k-dimensional plane through the origin in $\mathbb{R}^n$.



We mention two consequences of $N^{k-1} \subset S^{n-1}$ being a minimal submanifold. Let $\Delta x = (\Delta x_1, \ldots, \Delta x_n)$ denote the metric Laplacian on N. Since a submanifold $N^{k-1} \subset S^{n-1}$ is minimal if and only if Δx is normal to $S^{n-1} \subset \mathbb{R}^n$, we have $\Delta x = xf$ for some function f. As $|x|^2 = 1$, a simple calculation yields that

$$0 = \Delta|x|^2 = 2\langle x, \Delta x\rangle + 2|\nabla x|^2 = 2f + 2(k-1).$$

Hence, $f = 1 - k$ and the coordinate functions are eigenfunctions with eigenvalue $k - 1$. To obtain the other consequence, we observe that Δ_N and $\Delta_{C(N)}$ are related by the formula at $x \neq 0$:

$$\Delta_{C(N)}u = \frac{1}{r^2}\Delta_N u(\frac{1}{r}x) + (k-1)\frac{1}{r}\frac{\partial}{\partial r}u + \frac{\partial^2}{\partial r^2}u,$$

where $r = |x|$. Given x_i a coordinate function on $C(N)$, we may write it as $x_i = ru_i$ with x_i and u_i agreeing on $N \subset S^n$. By the chain rule, we know that

$$\begin{aligned} \Delta_{C(N)}x_i &= \frac{1}{r}\Delta_N u_i + u_i(k-1)\frac{1}{r}\frac{\partial}{\partial r}r + u_i\frac{\partial^2}{\partial r^2}r \\ &= -(k-1)\frac{1}{r}u_i + (k-1)\frac{1}{r}u_i = 0. \end{aligned}$$

Hence, every coordinate function is harmonic on $C(N)$ and $C(N) \subset \mathbb{R}^n$ is minimal.

Now, consider the Bernstein theorem for $n \leq 8$. Let Σ_u be the minimal graph of u and assume $x_0 \in \Sigma_u$. The monotonicity formula at x_0 yields that

$$\frac{\operatorname{Vol}(B_R(x_0) \cap \Sigma_u)}{R^{n-1}} - \frac{\operatorname{Vol}(B_r(x_0) \cap \Sigma_u)}{r^{n-1}} = \int_{(B_R(x_0)\setminus B_r(x_0))\cap\Sigma_u} \frac{|(x - x_0)^N|^2}{|x - x_0|^{n+1}}.$$

Let the density at infinity be

$$\Theta_\infty(x_0) = \lim_{r\to\infty} \Theta_r(x_0) = \lim_{r\to\infty} \frac{\operatorname{Vol}(B_r(x_0) \cap \Sigma_u)}{\omega_{n-1}r^{n-1}},$$

whose existence is guaranteed by the nondecreasing of $\Theta_r(x_0)$ as $r \to \infty$. Moreover, since $\Theta_0(x_0) \geq 1$, we have $\Theta_\infty(x_0) \geq 1$. If $\Theta_\infty(x_0) = 1$, then $\Theta_0(x_0) = 1$ and the monotonicity formula implies that

$$\lim_{\substack{r\to 0 \\ R\to\infty}} \int_{(B_R(x_0)\setminus B_r(x_0))\cap\Sigma_u} \frac{|(x - x_0)^N|^2}{|x - x_0|^{n+1}} = \Theta_\infty(x_0) - \Theta_0(x_0) = 1 - 1 = 0.$$

Hence, $x \in T_{x_0}\Sigma_u$ for all $x \in \Sigma_u$ and we obtain Σ_u as a cone. Blowing up Σ_u at x_0 gives

$$\Sigma_u = \lim_{r_n\to 0} \frac{1}{r_n}\Sigma_u = T_{x_0}\Sigma_u,$$

which implies that Σ_u is flat.

To prove the Bernstein theorem, we suppose that Σ_u is not a hyperplane. Then we have $\Theta_\infty > 1$. Blowing down Σ_u at 0 gives

$$\Sigma_\infty = \lim_{r_n\to 0} r_n\Sigma_u.$$

Based on the fact that

$$\Theta_r^{\Sigma_\infty}(x_0) = \lim_{n\to\infty} \frac{\operatorname{Vol}(r_n\Sigma_u \cap B_r(x_0))}{\omega_{n-1}r^{n-1}} = \lim_{n\to\infty} \frac{\operatorname{Vol}(\Sigma_u \cap B_{r/r_n}(x_0))}{\omega_{n-1}(r/r_n)^{n-1}} = \Theta_\infty(x_0),$$

we have $\Theta_r^{\Sigma_\infty} = \Theta_\infty$ for every $r > 0$. Since $r_j\Sigma_j$ is area-minimizing for every j, by stationary varifold theory or integral current theory, we deduce that Σ_∞ is also area-minimizing. As Σ_∞ has constant density, the monotonicity formula implies that Σ_∞ is a cone. However, by the splitting theorem of De Giorgi, there exists an area-minimizing hypercone with an isolated singularity, which contradicts with the fact that all area-minimizing hypercones in $\mathbb{R}^n$ for $3 \leq n \leq 7$ are flat.

The Bernstein problem is false for $n \geq 9$. In fact, there exists a singular area-minimizing cone in $\mathbb{R}^8$,

$$C_4 = \{(x_1, \ldots, x_8) | x_1^2 + \cdots + x_4^2 = x_5^2 + \cdots + x_8^2\}.$$

More generally, for $m \geq 4$, the cones

$$C_{2m} = \{(x_1, \ldots, x_{2m}) | x_1^2 + \cdots + x_m^2 = x_{m+1}^2 + \cdots + x_{2m}^2\}$$

are singular area-minimizing. Then, Fleming's result implies that there exists a nonlinear entire solution to the minimal surface equation.

Block Chain Mining and Game Theory

Eric Liu, Ryan Stofer, advised by Andre Martins Rodrigues

University of California, Santa Barbara

What is Block Chain?

Block chain is a public data structure used by crypto-currency networks, such as Bitcoin, to perform peer-to-peer transactions and decentralized governance. The block chain has the following four characteristics: it is a decentralized network, a tamperproof ledger, displays transparent transactions, and is trustless but has secure trading [1].

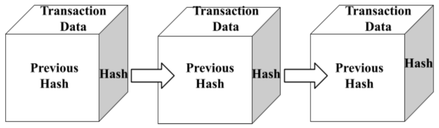


Fig. 1: Simple visualization of Bitcoin blockchain

Since any individual can join the public system and participate in the block chain, popular crypto-currencies, such as Bitcoin, employ a **proof of work** mechanism in order to secure all transactions and avoid potential attacks from hackers. Proof of work is performed by analyzing the exerting computing power made by participants when utilizing the block chain. When an individual (also known as a **miner**) has proven that they have exerted enough resources to the chain, they are then allowed to create a new **block** and are compensated with newly minted crypto-currency.

What are Mining Pools?

As for single miners, it may take an extremely long period to generate a new block because of the difficulty of the proof of work's protocol. Therefore, to decrease the uncertainty and make the revenue more predictable, miners form **mining pools** where all miners mine concurrently and share rewards with the whole mining pool when someone generates a new block.

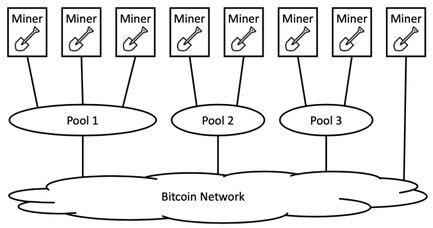


Fig. 2: Example of a Bitcoin system with 3 pools and 1 solo miner [1]

Mining pools are typically implemented as a **pool manager** and a group of miners, and the pool manager representing the whole pool joins the proof of work as a single miner[1]. The pool manager estimates the miners' power by accepting **partial proof of work** and allocates the revenue to miners according to the power they provided.

Game Theory and Nash Equilibrium

Game theory is a study of mathematical models of strategic interactions among rational players. In block chain, we can consider mining pools as rational players, since they always want to maximize their revenue. Through the game, each player will perform their own optimal strategy to maximize the profit, and strategies for mining pools will be discussed in the next section. A **non-cooperative game** is formed if players cannot collaborate or form alliance voluntarily.

One important concept of game theory is **Nash equilibrium** where the optimal outcome of a game is where there is no incentive to deviate from the initial strategy of each player. Therefore, Nash equilibrium is the most common way to define the solution of a non-cooperative game.

Pool Game

One of the classical attacks between mining pools is **pool block withholding attack**, where the attacking pool infiltrates other pools with attacking miners. Registered as miners in attacked pools, attacking miners only send partial proof of work and discards the full proof of work. Thus, the attacking miners can share the revenue obtained by other honest miners without contributing, which reduces the total revenue of the attacked pool. In addition, the total effective mining power in the block chain system will be reduced, decreasing the difficulty of the proof of work protocol.

Since a pool can potentially increase its revenue by attacking other pools, the strategies for each mining pool are to either attack other pools or mine honestly. The interaction between pools give rise to the **pool game** [1]. Since mining pools do not cooperate with other pools in our pool game model, the pool game is considered a non-cooperative game.

Attack between Two Pools

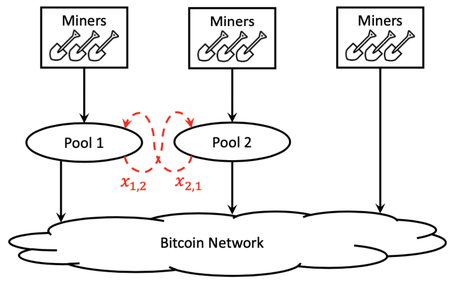


Fig. 3: Two pools attacking each other by infiltrating attacking miners [1]

We will begin the analysis with the case of two pools, pool 1 and 2. Let m_1 and m_2 denote the number of miners inside each pool and $x_{1,2}$ and $x_{2,1}$ denote the number of miners used by pool 1 to infiltrate pool 2 and the number of miners used by pool 2 to infiltrate pool 1 respectively. Then, the direct mining power of each pool is $m_1 - x_{1,2}$ and $m_2 - x_{2,1}$, and the effective mining power of the whole block chain is $m - x_{1,2} - x_{2,1}$ where m denotes all miners of the block chain.

Now, we define R_i as the direct mining rate of pool i which is the ratio between the direct mining power of pool i and the total effective mining power of the block chain. Therefore, the direct mining rate of two pools are:

$$R_1 = \frac{m_1 - x_{1,2}}{m - x_{1,2} - x_{2,1}}, \quad R_2 = \frac{m_2 - x_{2,1}}{m - x_{1,2} - x_{2,1}}$$

Then, we define r_i as the revenue density [2] of pool i which indicates the average revenue a miner can obtain inside pool i. We can obtain r_1 and r_2, based on the infiltration rate, by dividing the pool's revenue among all miners inside the pool:

$$r_1(x_{1,2}, x_{2,1}) = \frac{m_2 R_1 + x_{1,2}(R_1 + R_2)}{m_1 m_2 + m_1 x_{1,2} + m_2 x_{2,1}}, \quad r_2(x_{2,1}, x_{1,1}) = \frac{m_1 R_2 + x_{2,1}(R_1 + R_2)}{m_1 m_2 + m_1 x_{1,2} + m_2 x_{2,1}}$$

Since each pool will choose the optimal infiltration rate $x_{1,2}$ and $x_{2,1}$ that maximizes its revenue density, r_1 and r_2 will be maximized at single points in the range $0 \le x_{1,2} \le m_1$ and $0 \le x_{2,1} \le m_2$. We denote the optimal infiltration rate by $\bar{x}_{i,j} = \arg\max_{x_{i,j}} r_i$ and the corresponding revenue density $\bar{r}_i$ [1], where $i \neq j, i, j \in \{1, 2\}$ in this case.

Therefore, equilibrium can be achieved by finding pairs $x'_{1,2}$ and $x'_{2,1}$ such that

$$\begin{cases} \arg\max_{x_{1,2}} r_1(x_{1,2}, x'_{2,1}) = x'_{1,2} \\ \arg\max_{x_{2,1}} r_2(x'_{1,2}, x_{2,1}) = x'_{2,1} \end{cases}$$

under the constraints $0 < x'_{1,2} < m_1$ and $0 < x'_{2,1} < m_2$.

Two Pools Numerical Analysis and Equilibrium

Nash Equilibrium exists for $x_{1,2}, x_{2,1}$ when

$$\begin{cases} \frac{\delta r_1(x_{1,2}, x_{2,1})}{\delta x_{1,2}} = 0 \\ \frac{\delta r_2(x_{2,1}, x_{1,2})}{\delta x_{2,1}} = 0 \end{cases}$$

which is shown in the figure [1] below:

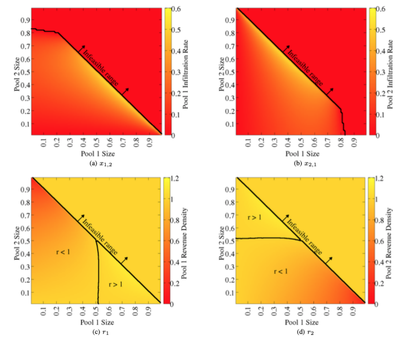


Fig. 4: Infiltration Rate and Revenue Graphs for 2 pools

We observe that only in extreme cases a pool does not attack its counterpart. Specifically, at equilibrium, a pool will refrain from attacking only if the other pool is larger than around 80% of the total mining power. Furthermore, we observe that a pool improves its revenue compared to the no-pool-attacks scenario only when it controls a strict majority of the total mining power. Thus, we see that the dominant strategy is to attack, regardless of what the other pool decides. The table below shows the Prisoner's Dilemma [1] for the Two Pools:

Pool 1 \ Pool 2	No Attack	Attack
No Attack	$(r_1 = 1, r_2 = 1)$	$(r_1 > 1, r_2 = \tilde{r}_2 < 1)$
Attack	$(r_1 = \tilde{r}_1 < 1, r_2 > 1)$	$(\tilde{r}_1 < r_1 < 1, \tilde{r}_2 < r_2 < 1)$

Practicalities

Although the model presented is simplistic, there are many factors that can perturb our model due to the assumptions we have made. For instance, we assume that the infiltrating miners are loyal to the attacker. However, some of the pool's members may be disloyal infiltrators. To avoid such a risk, a pool needs a sufficient number of verified miners — miners that it knows to be loyal. In general, the optimal infiltration rate may be as high as 60% of the pool size, but this is only in extreme cases when pools are large [1]. For practical pool sizes, a pool may need up to 25% of its mining power for infiltration [1].

Furthermore, a pool may engage in an attack against another pool not to increase its absolute revenue, but to attract miners by temporarily increasing its revenue relative to a competing pool. Such sabotage attack does not transfer revenue from victim to attacker, and migrating miners will switch to less attacked pools, changing pool sizes and hence revenues until convergence. Thus, many requirements must be satisfied for our model to be accurate in practice.

References

[1] I. Eyal. "The Miner's dilemma". In: *2015 IEEE Symposium on Security and Privacy* (May 2015), pp. 89–103.

[2] C. Grunspan and R. Pérez-Marco. "On profitability of stubborn mining". In: *2010 Mathematics Subject Classification* (2010), pp. 1–16.

Algorithms to Generate Random Gentle Algebras

Brian Fan Max Heneghan Jose Landa

University of California, Santa Barbara

Representation Theory

Representation Theory is a branch of mathematics that allows us to take intricate objects and "represent" them with simpler objects. Moreover, these simpler objects correspond and link to elements of Linear Algebra and Abstract Algebra. This area of math studies these algebraic structures, specifically finite dimensional algebra. One of the most recognized class of algebras is the gentle algebras. To understand this class of algebras, it would be helpful to become familiar with several definitions.

Definitions

Throughout, K is a field. We will first talk about quivers and then move towards algebras. Quivers are important because every finite dimensional algebra can be associated with a quiver and quivers give us a visual way of representing complex aspects of algebras.

Quiver - A quiver $Q = (Q_0, Q_1, s, t)$ is a quadruple consisting of two sets: Q_0 (whose elements are called vertices) and Q_1 (whose elements are called arrows), and two maps s, t : $Q_1 \to Q_0$ which associate to each arrow $\alpha \in Q_1$ its source $s(\alpha) \in Q_0$ and its target $t(\alpha) \in Q_0$, respectively.

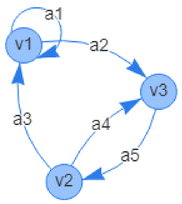


Figure 1. Example quiver Q, with $Q_0 = \{v1, v2, v3\}$ and $Q_1 = \{a1, a2, a3, a4, a5\}$

Adjacency Matrix - A square $n \times n$ matrix M which represents a quiver of n elements. The entry M_{ij} represents the number of arrows from vertex i to vertex j. If $M_{ij} = 0$, then there are no arrows from vertex i to vertex j. A non-zero entry on the diagonal of the matrix M (ie. $M_{ii} > 0$) represents an arrow(s) from vertex i to itself, and this is called a **loop**.

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Figure 2. Adjacency matrix for the quiver Q

K-Algebra - A K-algebra is a ring with identity, A, such that A has a K-vector space structure compatible with the multiplication of the ring. We say A is finite dimensional if the dimension of the K-vector space A is finite.

Path Algebra - Let Q be a quiver. The path algebra KQ of Q is the K-algebra whose underlying K-vector space has as its basis the set of all paths of length $l \geq 0$ in Q and such that the product of two paths $\alpha_1 ... \alpha_l$ and $\beta_1 ... \beta_k$ is equal to zero if $t(\alpha_l) \neq s(\beta_1)$ and is equal to the composed path $\alpha_1 ... \alpha_l \beta_1 ... \beta_k$ if $t(\alpha_l) = s(\beta_1)$.

Relations - Let Q be a quiver. A relation in Q with coefficients in K is a K-linear combination of paths of length at least two having the same source and target. Given a set of relations, let I be the ideal generated by these. Then KQ/I is the algebra bound by these relations.

Goal

A ***gentle algebra*** is a finite dimensional algebra

$$A = KQ/I$$

where Q is a quiver, KQ is a path algebra, and I is an ideal generated by paths of length 2 and satisfies:

1) At most 2 arrows enter and 2 leave each vertex of the quiver Q.

2) For each arrow $\beta \in Q_1$, there is at most one arrow $\gamma \in Q_1$ and at most one arrow $\alpha \in Q_1$ such that $\gamma\beta$ and $\beta\alpha$ are relations contained in I and at most one arrow $\gamma' \in Q_1$ and at most one arrow $\alpha' \in Q_1$ such that $\gamma'\beta$ and $\beta\alpha'$ are relations not contained in I.

The goal for our team was to develop an algorithmic code that enables the generation of gentle algebras. Adhering to the conditions for this class algebra were hard and generating them while maintaining their relations was even harder. Gentle algebras in Representation Theory are an interesting, pretty well-known type of algebra, and were considered to be a good challenge to test out examples.

Introducing GAP and QPA

Evidently, translating theoretical models into a coding language can prove to be a difficult feat, especially since for our project data inputs and outputs were both considered and desired respectfully. Thus, we used the programming language Groups, Algorithms, Programming (GAP) to construct our random generator for gentle algebras. Moreover, GAP has large data libraries that house many packages that contain functions implementing algebraic models written into the preceding programming language. The most frequented package used in GAP for our project was the Quivers and Path Algebras (QPA) package, which contains data structures for quivers and finite dimensional algebras.

Algorithms Mind-Map

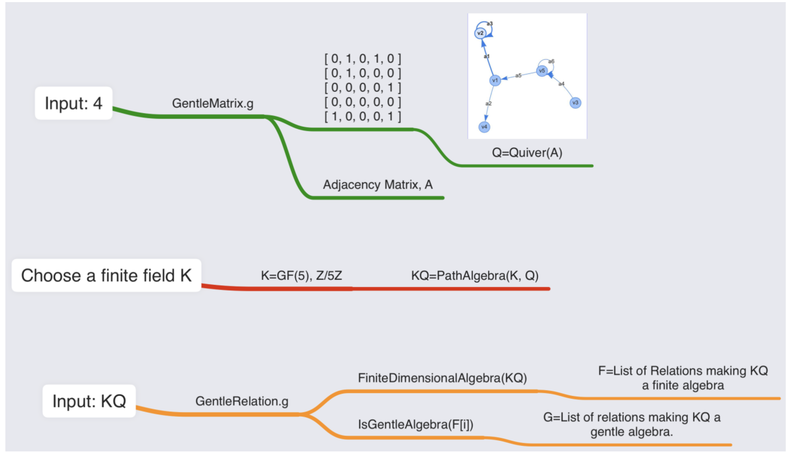


Figure 3. Example Algorithm for a Random 4 Vertices Quiver

Algorithm Explanations

Our approach to generate gentle algebra relations consists of three functions, allocated in "GentleMatrix.g", "FiniteDimensionalAlgebraRelation.g", and "GentleRelation.g". Due to the limitation of space, we have included our actual code works in the QR code below. Please scan it for detailed information.

(a) (b) (c) (d)

Figure 4. (a)-(b) GentleMatrix.g (c) FiniteDimensionalAlgebraRelation.g (d) GentleRelation.g

GentleMatrix.g

In this file, we are trying to generate an adjacency matrix that can be used to construct a gentle quiver which has at most two arrows entering and 2 leaving each vertex. The primary function in this file will take an integer n, which represents the number of vertices, as its input. To make the resulted adjacency matrix as random as possible, we utilize the built-in function **Random()** in GAP to create randomness. Moreover, to ensure our matrix can generate a special biserial quiver, we control the sum of entries in each row and column with an upper limit of 2.

FiniteDimensionalAlgebraRelation.g

In this file, we created a function called **FiniteDimensionalAlgebraRelation()** that takes a path algebra created by the adjacency matrix generated by the **GentleMatrix()** as its input and outputs a list of relations that can make this inputted path algebra finite dimensional. As said by the definition, every gentle algebra is finite dimensional, so the existence of this function helps us filter out all the relations that wouldn't make our path algebra finite. This function takes advantage of a built-in function called **IsFiniteDimensional()** in GAP's QPA package.

GentleRelation.g

In this file, we created a function called **GentleRelation()** that takes the path algebra created by the adjacency matrix generated by the **GentleMatrix()** as its input and outputs a list of relation that can make this inputted path algebra finite dimensional. In our algorithm, we employ the **FiniteDimensionalAlgebraRelation()** defined above and the built-in function **IsGentleAlgebra()** to filter out relations that can make the path algebra gentle.

Acknowledgement

We would like to extend our gratitude towards our advisor and mentor Andres Barei, whose research pertains in Representation Theory of Algebras. Also, it goes without saying, that we would like to thank the Directed Reading Program for allowing us to take part in this amazing opportunity.

References

[1] Assem, I., Simson, D., &; Skowroński Andrzej. (2006). Elements of the Representation Theory of Associative Algebras, Volume 1. Cambridge: Cambridge University Press.

[2] Geiß, C., &; Reiten, I. (2005). Gentle algebras are Gorenstein. Representations of Algebras and Related Topics, 129–133. doi:10.1090/fic/045/09

RSA Encryption

Anna Maximova and James McNeice

2022 Mathematics Directed Reading Program. Department of Mathematics, University of California, Santa Barbara

Why Public Key Cryptographic systems?

Consider the following situation: A message needs to be sent to someone over a public channel. As the channel is not secure, anyone can look at whatever you send to the other party. The question is, how do you send a message to the other party that without compromising the information contained in the message. This is the crux of the field of cryptography. Public Key systems come into play when there is no way to transmit a key safely. The solution is creating a system where all the necessary information to encrypt a message is available publicly, but decrypting the message is very difficult without some sort of key.

An Introduction To Number Theory

Modular Arithmetic

One of the most basic ideas in number theory is modular arithmetic, which is a system of arithmetic that centers around the remainder after repeated subtraction.

Consider a 12-hour clock. Suppose the hour hand points at 12 when no time has elapsed. When 3 hours pass, the hour hand will point at the 3. When 19 hours elapse, the hand will point at the 7 because the hand will cycle through the 12 hours and then restart its cycle to reach 7. Similarly when 25 hours elapse, the hand will point at the 1. We can represent this using the symbol for congruence $\equiv$ as follows:

$3 \equiv 3 \bmod 12$ $\qquad$ $19 \equiv 7 \bmod 12$ $\qquad$ $25 \equiv 1 \bmod 12$

So $a \equiv b \bmod (n)$ if and only if there exists an integer k such that $(a-b) = nk$.[3]

Modular Exponentiation

Suppose you were asked to find the smallest x such that $x \equiv 3^{173} \bmod 11$. We could multiply 3 by itself 173 times and then subtract 11 until we were left with a remainder between 0 and 11 but that would take too long and we're feeling a little lazy today. Luckily for us there is a very simple process we could employ to help us that relies on modular arithmetic: modular exponentiation. The process is relatively simple. First we find a few other congruences.

$3^1 \equiv 3 \bmod 11$ $\qquad$ $3^{16} \equiv 5^2 \equiv 3 \bmod 11$

$3^2 \equiv 9 \equiv 9 \bmod 11$ $\qquad$ $3^{32} \equiv 3^2 \equiv 9 \bmod 11$

$3^4 \equiv 9^2 \equiv 4 \bmod 11$ $\qquad$ $3^{64} \equiv 9^2 \equiv 4 \bmod 11$

$3^8 \equiv 4^2 \equiv 5 \bmod 11$ $\qquad$ $3^{128} \equiv 4^2 \equiv 5 \bmod 11$

Now we can use smaller powers to get to larger powers based on the fact that $x^m \cdot x^n = x^{m+n}$. So, $3^{173} \equiv 3^{(1+2+2+8+32+128)} \equiv 3 \cdot 9 \cdot 9 \cdot 9 \cdot 5 \equiv 1 \bmod 11$.

Chinese Remainder Theorem

Suppose gcd(m, n) = 1. Given integers a and b, there exists exactly one solution x mod (mn) to the simultaneous congruences: $x \equiv a \bmod (m)$ and $x \equiv b \bmod (n)$. [3]

Fermat's Little Theorem

If p is a prime and p does not divide a, then $a^{p-1} \equiv 1 \bmod (p)$. [3]

Intuition for RSA

Using the classic example in cryptography, suppose Bob wants to send a secret message over an unsecure channel to Alice such that if Eve (the eavesdropper) who is listening in on the channel isn't able to understand the message. Alice would create a lock and a key that only she possesses. She would send the unlocked lock to Bob who would use it to lock his message and send it back to Alice. Finally Alice would unlock the lock with her private key and read the message. Eve would only have information about the unlocked and locked lock and therefore theoretically would not be able to read the message.

RSA

RSA works by first choosing two large prime numbers, p and q, then multiplying them to make N, that is:

$$pq = N$$

This is the value that will serve as the modulus for encryption and decryption. At this point a message can be given a numeric representation, M, such that $0 \leq M \leq N-1$. We now choose some e with the following property

$$\gcd(e, (p-1)(q-1)) = 1$$

We now choose value d such that $ed \equiv 1 \mod (p-1)(q-1)$. The setup is now complete, and (n, e) are released as the public key. A message is encrypted by taking $a \equiv M^e \mod N$, and decrypted by taking $M = a^d \mod N$. [3].

Example [2]

Alice:

1. Chooses two primes: $p = 7$ and $q = 19$.
2. Calculates the product: $N = 7 \cdot 19 = 133$.
3. Calculates the totient: $\phi(N) = (p-1)(q-1) = 6 \cdot 18 = 108$.
4. Selects a public key: $e = 29$
5. Selects a private key: $d = 41$
6. Sends the public key: $(N, e) = (133, 29)$

Bob:

1. Chooses a message: $m_o = 99$.
2. Encrypts the message: $m_e = 99^{29} \bmod 133 = 92$.
3. Sends the encrypted message.

Alice:

1. Decrypts the message: $m_o = 92^{41} \bmod 133 = 99$

Note: The efficiency of RSA lies in the fact that it is significantly faster to multiply two numbers than it is to factor a number of the same size as their product. This means that even if an eavesdropper is able to read a message in its encrypted state, they are unable to understand its content because finding the value of d is difficult. Factoring can be made arbitrarily difficult by choosing sufficiently large numbers. For simplicity's sake, we used very small numbers in our example. However to make the encryption feasible and secure, the primes used are typically 1024 to 2048 bits long, approximately 300 to 600 digits long.

Attacks On RSA

Timing Attack

It was demonstrated in 1995 that by timing the process of decrypting multiple messages a malicious party is able to determine the key. This attack is worth mentioning because it does not attack the fundamental process of encryption[3]. Its more akin to having a storefront tightly locked up, and instead of picking the locks a thief throws a rock through the front window [3].

Fermat Attack

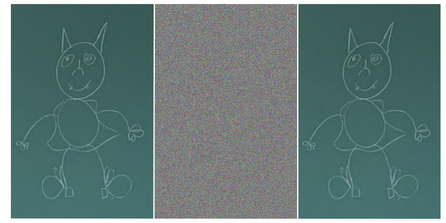
If the primes chosen for encryption are too close to each other then it has been demonstrated that an algorithm can factor N very efficiently. Using the fact that $N = a^2 - b^2 = (a-b)(a+b)$ we can tell that if we find a, b then $(a-b) = p$ and $(a+b) = q$. This is accomplished by taking $\lceil \sqrt{N} \rceil = a$, and determining if $b^2 = a^2 - N$ is an integer. If not, then increment a by one and try again until either a value of b is found, or until 100 or so values of a have been tried [1].

Shor's Algorithm

Shor's Algorithm is a quantum computing algorithm that shatters the security of RSA. It does this taking a 'bad' guess for two numbers that factor some given integer, and spitting out a 'good' guess [3].

Additional Applications of RSA

By coming up with a clever way to express some message many different forms of media can be transmitted via RSA, for instance:



Original Image | Encrypted Image | Decrypted Image

Acknowledgements

We would like to thank the organizers of the of the 2022 Directed Reading Program for the opportunity to learn about cryptography. We would also like to thank our mentor Charles Kulick for his incredible support and mentorship throughout this program.

References

[1] Hanno Böck. *Fermat Attack on RSA*. 2022. URL: https://fermatattack.secvuln.info/.

[2] Ed Harmoush. *RSA Example*. 2021. URL: https://www.practicalnetworking.net/series/cryptography/rsa-example/.

[3] W. Trappe and L.C. Washington. *Introduction to Cryptography: With Coding Theory*. Prentice Hall, 2002. ISBN: 9780130618146. URL: https://books.google.com/books?id=kVU%5C_AQAAIAAJ.

Algebraic Geometry and Moduli Spaces

Kai Yan, Graduate Mentor: Daniel Hamrast

Algebraic Varieties

An **affine algebraic variety** is the common zero set of a collection $\{F_i\}_{i\in I}$ of complex polynomials. In particular, the zero sets of homogeneous polynomials can be viewed as a **projective variety** in a quotient of $\mathbb{C}^{n+1}$ known as the projective space $\mathbb{P}^n$. These varieties form **Zariski topology**, where the open sets are the complement of the varieties. These varieties are completely determined by their **coordinate rings**, defined as $\mathbb{C}[V] = \mathbb{C}[x_1, ..., x_n]/\mathbb{I}(V)$, and conversey every reduced, finite type $\mathbb{C}$-algebra gives an affine/projective variety. The passage from a $\mathbb{C}$-algebra to its variety is denoted by $Spec$, which consists of all the prime ideals of the algebra.

Veronese Map

One useful relationship between projective spaces is the following: All homogeneous degree d polynomial in the polynomial ring $\mathbb{C}[x_0, ..., x_n]$ form a finite dimensional $\mathbb{C}-$vector space with the basis consisting of $\binom{d+n}{d}$ monomials: $x_0^{d_0}...x_n^{d_n}$ with $\sum d_i = d$. This motivates the **Veronese embedding** of the projective space $\mathbb{P}^n$ into $\mathbb{P}^m (m = \binom{d+n}{d} - 1)$, which is the morphism:

$$[x_0 : ... : x_n] \xrightarrow{\nu_d} [x_0^d : x_0^{d-1}x_1 : ... : x_n^d]$$

Five Points Determine a Conic

A conic in projective space $\mathbb{P}^2$ is the zero set of the polynomial:

$$F(x, y, z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz$$

where the coefficients are not all 0. Hence each line through $\mathbb{C}^6$, denoted by $[a : b : c : d : e : f]$ uniquely determines a conic. Therefore we can identify sets of conics in $\mathbb{P}^2$ with points in $\mathbb{P}^5$, and we say that $\mathbb{P}^5$ parameterizes conics in $\mathbb{P}^2$. This is an example of a solution to a **moduli problem**, which I will talk about later.
Now consider a fixed point $[x_0 : y_0 : z_0]$ in $\mathbb{P}^2$, $F(x_0, y_0, z_0) = 0$ now defines a linear equation satisfied by a, b, c, d, e, f. Hence each point in $\mathbb{P}^2$ defines a hyperplane in $\mathbb{P}^5$ through F! Therefore five points(we require there can be no more than three collinear points) $p_1, p_2, p_3, p_4, p_5 \in \mathbb{P}^2$ determines five hyperplanes $H_1, ..., H_5 \subset \mathbb{P}^5$. The intersection of five linearly independent hyperplanes is nothing but a point in $\mathbb{P}^5$, since intersecting once reduce the dimension by one. So there is exactly one conic passing through five fixed point.

References

[1] Ravi Vakil. *THE RISING SEA, Foundations of Algebraic Geometry*. 2017.

[2] Pekka Kekalainen Karen E. Smith, Lauri Kahanpaa and William Traves. An invitation to algebraic geometry. 2000.

[3] Robin Hartshorne. *Algebraic geometry*. Springer-Verlag, New York, 1977. Graduate Texts in Mathematics, No. 52.

Hilbert Polynomial

A **graded ring** is a ring that decomposes into direct sum of its subrings. The **Hilbert function** is defined on a graded ring $R = R_0 \oplus R_1 \oplus R_2 ... \oplus R_m$ by:

$$m \longrightarrow \dim R_m$$

Let $V \subset \mathbb{P}^n$ be a projective variety, whose coordinate ring is clearly a graded ring. In this way we can define the Hilbert function of a projective variety.
For large m Hilbert function agrees with a polynomial, called the **Hilbert polynomial**:

$$P(m) = e_0 m^d + ... + e_{d-1}m + e_d$$

with degree $d = \dim V$ and $e_0 = \frac{degV}{d!}$. $degV$ is the **degree** of V, which is defined to be the largest possible number of intersections between V and a codimension $\dim V$ linear subvariety of $\mathbb{P}^n$.

The Hilbert Scheme

Fixing an arbitrary polynomial P, the set of all subvarieties with P as its Hilbert polynomial naturally forms a variety, or more precicely, a **scheme**(a generalization of a variety) in its own right. We call this the **Hilbert scheme**. To contruct the Hilbert scheme, note that any projective variety $V \in \mathbb{P}^n$ is uniquely defined by a homogeneous radical ideal $I = \mathbb{I}(V) \subset \mathbb{C}[x_0, ..., x_n]$. Grothendieck showed that for any P, there exists a positive integer r(depending on P) such that for all ideals I defining a variety with Hilbert polynomial P, I is the radical of the subideal generated by its elements of degree r. Hence, to every Hilbert Polynomial P, one can associate a vector subspace $I_r \subset S_r$, where S_r is the vector space of all homogeneous polynomials of degree r. One can compute the dimension of the vector subspace I_r by:

$$d_r = \dim I_r = \dim S_r - \dim S_r/I_r = \binom{r+n}{r} - P(r)$$

In this way, a Hilbert polynomial, together with r, uniquely specifies a **Grassmannian** $G(\binom{r+n}{r}, d_r)$, which consists of all the d_r-dimensional vector subspaces of a $\binom{r+n}{r}$-dimensional vector space S_r. And a variety uniquely determines a single point in the Grassmannian. Therefore, the Hilbert scheme is a very good way to classify and parameterize subvarieties(or more generally, subschemes) of projective space.

Acknowledgement

I would like to thank my mentor Daniel Halmrast for helping me develop the intuition for concepts used in this poster as well as answering my questions. I would also like to thank the organizer of UCSB DRP for running this fantastic program.

Category, Natural Transformation and the Yoneda Lemma

If F, G are functors between categories $\mathcal{A}, \mathcal{B}$, then a **natural transformation** $\eta : F \implies G$ is a set of morphisms that satisfies:

- The natural transformation must associate a morphism $\eta_A : F(A) \longrightarrow G(A)$ to every object $A \in \mathcal{A}$. This morphism is called a **component** of A.
- For every morphism $f : A_1 \longrightarrow A_2$, we have:

  $$\eta_{A_2} \circ F(f) = G(f) \circ \eta_{A_1}$$

  In other words, the following diagram must commute:

$$\begin{array}{ccc} F(A_1) & \xrightarrow{\eta_{A_1}} & G(A_1) \\ \downarrow{\scriptstyle F(f)} & & \downarrow{\scriptstyle G(f)} \\ F(A_2) & \xrightarrow{\eta_{A_1}} & G(A_2) \end{array}$$

In category theory, one of the most important results regarding natural transformation is called the **Yoneda lemma**. Given a fixed category $\mathcal{A}$, each object $X \in \mathcal{A}$ naturally gives a functor h_X defined by:

$$h_X = Hom(-, X)$$

Hence for any objects $Y \in \mathcal{A}$, $h_X(Y) = Hom(Y, X)$, which is the set of all morphisms from Y to X. The Yoneda lemma states that the set of natural transformation between h_X and h_Y is isomorphic to the set of morphisms from Y to X. In other words:

$$Hom(h_x, h_y) \cong Hom(Y, X)$$

The Yoneda lemma allows us to completely determine any object by looking at the morphisms that maps into it. This is very powerful in the context of moduli problem, where the structure of the moduli space is not obvious.

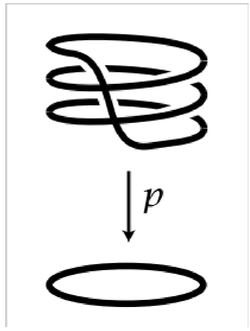
Hilbert functor and Moduli Space

In a more categorical term, the Hilbert scheme is a representation of a functor that sends topological spaces to sets. **Hilbert Functor** It can be defined as:

$$Hilb_X^d : Top \longrightarrow Sets$$

$$Hilb_X^d(Y) = \left\{ Z \subset X \times Y : \begin{array}{c} Z \xrightarrow{\pi_Y} Y \text{is finite and} \\ \text{locally free of rank } d \end{array} \right\}$$

In particular, π_y is analogous to a finite and locally free covering map:



To give a more concrete example, let X be a topological space. Consider the Hilbert functor $Hilb_X^1$. It sends any topological space Y to the set where the elements are topological subspaces $Z \subset X \times Y$ such that the projection from the Z to Y is a homeomorphism. Interestingly enough, $Hilb_X^1$ is in fact naturally isomorphic to the functor h_X and the components of the natural transformation map the set Z to the function $\pi_X \circ \pi_Y^{-1}$. If you let X be $\mathbb{R}$, then $Hilb_X^1$ simply sends any function $g : \mathbb{R} \longrightarrow \mathbb{R}$ to its graph.
Another example concerns the previously mentioned process of five points determining a conic.The Hilbert functor corresponding to a Hilbert polynomial $P(m)$ and subvarieties of $\mathbb{P}^n$ is denoted by: $Hilb_{\mathbb{P}^n}^{P(m)}$. Conics in $\mathbb{P}^2$ has Hilbert polynomial $2m + 1$ and degree 2. The veronese map of degree d associates the set of all degree d hypersurfaces in $\mathbb{P}^n$ with the set of all linear hyperplanes in $\mathbb{P}^M (M = \binom{d+n}{d} - 1)$, which is isomorphic to $\mathbb{P}^M$. Hence we have:

$$Hilb_{\mathbb{P}^2}^{2m+1} \cong h_{\mathbb{P}^5}$$

We say $\mathbb{P}^5$ **represents** the Hilbert functor $Hilb_{\mathbb{P}^n}^{2m+1}$.
These are the simple examples of **moduli spaces**, whose points represent algebraic subvarieties, or more generally subschemes, up to isomorphisms. In the language of moduli spaces, one can parameterize different classes of interesting geometric objects. More often than not, the moduli spaces themselves can have interesting structures beyond merely being a set of points representing classes of objects. And the Yoneda lemma is presicely the tool to study abstract objects like moduli spaces: one can probe the structure of moduli spaces by looking at how other topological spaces map into them. For instance, the map from $\mathbb{A}^1$ into any moduli spaces can give us information about their path connectedness. Hence, solving moduli problems not only helps one classify interesting objects, but give insight into how these classes relate to each other. This makes the study of moduli spaces a very active area in mathematics and physics.

Čech Cohomology!

Lawrence Wu Mentor: **Danning Lu**

University of California, Santa Barbara | 2022 Directed Reading Program

Abstract

In essence, this poster is a brief exploration into ideas from Sheaf Theory, specifically focusing on the Čech Cohomology. Firstly, we introduce the definition of a sheaf, as well as ways to construct sheaves. We then explore the Čech Cohomology, a powerful tool centered around intersections and open covers of a Topological Space.

What is a sheaf

We first must introduce some technical machinery before we discuss further topics.

Sheaves!

Let X be a topological space. A *sheaf of abelian groups on X* consists of:

(a) *a function $x \to \mathcal{F}_x$, assigning each $x \in X$ to some Abelian group $\mathcal{F}_x$.*

(b) *a topology on the set $\mathcal{F}$, the sum of the sets $\mathcal{F}_x$.*

If f is an element of $\mathcal{F}_x$, we put $\pi(f) = x$; we call the mapping of π the *projection* of $\mathcal{F}$ onto X; the family in $\mathcal{F} \times \mathcal{F}$ consisting of pairs (f, g) such that $\pi(f) = \pi(g)$ is denoted by $\mathcal{F} + \mathcal{F}$.

Furthermore, we impose two axioms on **(a)** and **(b)**.
(I) *for all $f \in \mathcal{F}$ there exist open neighborhoods of V of f and U of $\pi(f)$ to V is a homeomorphism of V and U.*

(II) *the mapping $f \to -f$ is a continuous mapping from $\mathcal{F}$ to $\mathcal{F}$, and the mapping $(f, g) \to f + g$ is a continuous mapping from $\mathcal{F} + \mathcal{F}$ to $\mathcal{F}$.*

Sections!

Let $\mathcal{F}$ be a sheaf, and let open $U \subseteq X$. We define a section of $\mathcal{F}$ over U as a continuous mapping $s : U \to \mathcal{F}$ such that $\pi \circ s$ coincides with the identity on U. The set of sections of $\mathcal{F}$ over U is denoted as $\Gamma(U, \mathcal{F})$ and is an abelian group.

Construction of Sheaves!

Suppose for all open $U \subset X$, we have an abelian group $\mathcal{F}_U$, and for all pairs of open subsets $U \subseteq V$ a homomorphism $\phi_U^V : \mathcal{F}_V \to \mathcal{F}_U$, satisfying the transitivity condition $\phi_U^V \circ \phi_U^W = \phi_U^W$. With these conditions, we can define $\mathcal{F}_x = \lim \mathcal{F}_U$ as the inductive limit of the system of open neighborhoods of x. Furthermore, let $t \in \mathcal{F}_U$ and denote $[t, U]$ as the set of $\phi_x^U(t)$ for x running over U. Now, we give $\mathcal{F}$ the topology generated by $[t, U]$. This guarantees that the system $(\Gamma(U, \mathcal{F}), \rho_U^V)$ is a sheaf, but it doesn't guarantee that it is isomorphic to $\mathcal{F}$.

Observe that $x \to \phi_x^U(t)$ is a section of $\mathcal{F}$ over U, which allows us to define the canonical morphism $\iota : \mathcal{F}_U \to \Gamma(U, \mathcal{F})$.

Proposition 1: $\iota : \mathcal{F}_V \to \Gamma(U, \mathcal{F})$ is injective if and only if the following condition holds:

If an element $t \in \mathcal{F}_U$ is such that there exists an open covering $\{U_i\}$ of U with $\phi_{U_i}^U = 0$.

Proposition 2: Let U be an open subset of X, and let $\iota : \mathcal{F}_V \to \Gamma(U, \mathcal{F})$ be injective for all open $V \subset U$. Then ι is surjective if and only if the following condition is satisfied:

For all open coverings $\{U_i\}$ of U, and all systems $\{t_i\}$, $t_i \in \mathcal{F}_{U_i}$ such that $\phi_{U_i \cap U_j}^{U_i}(t_i) = \phi_{U_i \cap U_j}^{U_j}(t_j)$ for all pairs (i, j), there exists a $t \in \mathcal{F}_U$ with $\phi_{U_i}^U(t) = t_i$ for all i.

Proposition 3: If $\mathcal{F}$ is a sheaf of abelian groups on X, the sheaf defined by the system $(\Gamma(U, \mathcal{F}), \rho_U^V)$ (with propositions 1,2) is canonically isomorphic with $\mathcal{F}$.

Some examples of sheaves

The definition of sheaves is undoubtedly daunting, but there are several examples that are relatively easy to grasp. Consider the following examples,

- Let X be some topological space. Let G be an abelian group, and set $\mathcal{F}_x = G$ for all $x \in X$. Now, our sheaf $\mathcal{F}$ can be identified as the $X \times G$ with the product topology of X and G, equipped with the discrete topology. This construction can be verified to be a sheaf, and is known as the *constant sheaf*.
- Let X be some topological space. Let $x \in X$, and let G be some abelian group. Let U be an open subset of X, we define $\mathcal{F}(U)$ as

$$\mathcal{F}(U) := \begin{cases} G & \text{if } x \in U \\ 0 & \text{if } x \notin U \end{cases}$$

  Indeed, we can construct a sheaf from $\mathcal{F}(U)$ and is known as the *skyscraper sheaf*.
- There are also more concrete examples we can talk about! For instance, we consider the topological space $\mathbb{C}$. Let $U \subset \mathbb{C}$ be an open subset of $\mathbb{C}$. We associate each U with the set of holomorphic functions $\mathcal{F}(U) := \mathcal{C}(U)$. Under a system of inclusion maps, it's easy to see that we in fact do yield a rather visual sheaf!

The Čech Cohomology

With some tools in our inventory, we can begin to talk about the Čech Cohomology! The full construction of the Čech Cohomology is quite long and technical, and the curious reader should turn their attention to Coherent Algebraic Sheaves.

Let $\mathcal{U} = \{U_i\}_{i \in I}$ be an open cover of X. If $s = (i_0, \ldots, i_p)$ is a finite sequence of elements in I, we put $U_s = U_{i_0} \cap \ldots U_{i_p}$. A p-cochain of $\mathcal{U}$ is a function f assigning every sequence s of $p+1$ elements of I to a section of $\mathcal{F}$ over U_s. Note that the p-cochains form an abelian group, denoted by $C^p(\mathcal{U}, \mathcal{F})$.

Let $S(I)$ be the simplex with I as its vertices. Let $K_p(I)$ be the free group with the set of simplexes of dimension p of $S(I)$ as its base. Now, we are beginning to delve into familiar territory. We define our boundary map $\partial : K_{p+1}(I) \to K_p(I)$ in the usual way,

$$\partial(i_0, \ldots, i_{p+1}) = \sum_{j=0}^{p+1} (-1)^j (i_0, \ldots, \hat{i}_j, \ldots, i_{p+1}).$$

Now, we define the coboundary operator ${}^t\partial : C^{p+1}(\mathcal{U}, \mathcal{F}) \to C^p(\mathcal{U}, \mathcal{F})$ as

$$({}^t\partial f)_{(i_0, \ldots, i_{p+1})} = \sum_{j=0}^{p+1} (-1)^j \rho_j (f_{i_0, \ldots, \hat{i}_j, \ldots, i_{p+1}}).$$

where $\rho_j : \Gamma(U_{i_0, \ldots, \hat{i}_j, \ldots, i_{p+1}}, \mathcal{F}) \to \Gamma(U_{i_0, \ldots, i_{p+1}}, \mathcal{F})$ denotes the restriction homomorphism.

With this, we can finally define the q-th cohomology group of the complex $C(\mathcal{U}, \mathcal{F})$ as $H^q(\mathcal{U}, \mathcal{F}) := \text{Ker}\,({}^t\partial_q)/\text{Im}\,({}^t\partial_{q-1})$. However, this is not enough to define the Čech Cohomology on X as our cohomology groups generally depend on our choice of $\mathcal{U}$. To combat this issue, we consider finer open covers of X.

A cover $\mathcal{U}$ is said to be finer than $\mathcal{V}$ if there exists a mapping $\tau : I \to J$, such that $U_i \subset V_{\tau(i)}$ for all $i \in I$. If $\mathcal{U}$ is finer than $\mathcal{V}$, there exists a canonical mapping $\sigma(\mathcal{U}, \mathcal{V})$ from $H^q(\mathcal{V}, \mathcal{F})$ to $H^q(\mathcal{U}, \mathcal{F})$.

Finally, we are ready to define the Čech Cohomology on X. Under refinement, the covers of X form a directed set, which allows us to set $H^q(X, \mathcal{F}) := \lim H^q(\mathcal{V}, \mathcal{F})$.

Čech Cohomology Isomorphic?

The construction of the Čech Cohomology is quite undeniably complicated. This begs the question, why exactly do we care about the Čech Cohomology? What exactly does Čech Cohomology bring to the table?

Firstly, the Čech Cohomology has many applications in Algebraic Geometry, which is a beautiful field in its own right. The curious reader should once again turn their attention towards the reference section.

In our construction of the Čech Cohomology, we are reminded of the construction of other Cohomologies. In some sense, the Čech Cohomology can be thought of as a generalization of both the Singular Cohomology and the de Rham Cohomology. While in general, the Čech Cohomology groups for an arbitrary space X is not isomorphic to either Cohomology groups, we can impose certain conditions such that they always coincide.

Proposition 4.

Let X be a paracompact topological space, and $\mathcal{F} = A$ a constant sheaf. Then the following is true,

$$\check{H}(X, \mathcal{F}) \cong H_{\text{Sing.}}(X, A).$$

Furthermore, since CW-complexes are paracompact, if X is homotopic equivalent to a CW-complex, then our two cohomology groups coincide.

Proposition 5.

Let X be a differential manifold, and $\mathcal{F} = \mathbb{R}$. Then the following holds,

$$\check{H}(X, \mathcal{F}) \cong H_{\text{de Rham}}(X, \mathbb{R}).$$

The proofs can be found in the references [2][3] respectively.

Acknowledgements

I would like to thank the Directed Reading Program for giving me this amazing opportunity to explore mathematics in such a hands-on fashion. I would also like to extend my gratitude to the many graduate students in the Mathematics Lab who helped me understand the material in a deeper level. Lastly, I would to like to give special thanks to my mentor, Danning Lu, who has been a terrific guide through this project. Without his boundless patience and wisdom, this project quite literally could not be possible.

References

[1] *Achinger, (1965). Faisceaux Algebriques Coherents [Coherent Algebraic Sheaves].*
[2] *Spanier, E. H. (1966). Algebraic topology. Springer.*
[3] *Bott, Raoul, and Loring W. Tu. Differential Forms in Algebraic Topology. Springer, 1995.*

UC SANTA BARBARA

An Introduction to Optimal Transport

Ayesha Usmani †

†Department of Mathematics, University of California, Santa Barbara

Introduction

As humans, we naturally aim to find the most efficient way to do things. Optimal Transport, as can be deduced by its name, is an area of study dedicated to finding the most efficient way to transport units from one location to another. Its origins can be traced back to French mathematician Gaspard Monge who, in the 1780s, considered a simple problem whereby a worker moves one pile of sand to create a new pile of a specific shape in another location. To do this while also using the least amount of effort, one must consider the local cost of moving each grain of sand from the original pile to the targeted pile and use this to find the minimum global cost. For the sake of simplicity, we will consider discrete optimal transport problems.

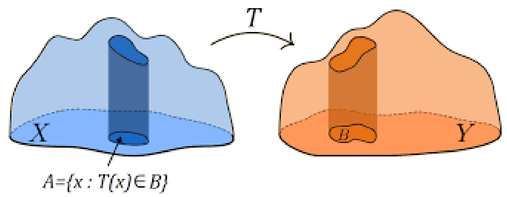


Figure 1. An example of the Monge problem - sand being moved from one location to another. Source: Matthew Thorpe, University of Cambridge.

Assignment Problem

A special case of the Monge Problem is the assignment problem.
Suppose we have equal masses contained in locations x_i that we wish to transfer to locations y_j. In this case, suppose that all the mass from any x_i must be transported to only one y_j.

$$\begin{array}{ccc} x_1 \bullet & & \bullet\ y_1 \\ x_2 \bullet & \xrightarrow[\text{permutation}]{\sigma} & \bullet\ y_2 \\ \vdots & & \vdots \\ x_n \bullet & & \bullet\ y_n \end{array}$$

The cost of transporting something from x_k to y_j is $C_{k,\sigma(k)}$. The optimal transport cost is

$$\min_{\sigma\in\text{Perm}(n)} \sum_{k=1}^{n} C_{k,\sigma(k)}.$$

There are two points of concern in this problem. Firstly, it does not allow the splitting of mass to transport multiple locations. Secondly, it does not have a unique solution, as you can see in Figure 1.

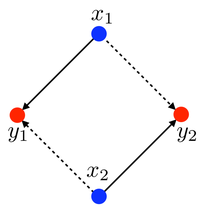


Figure 2. Notice that the cost of transporting mass from x_1 to y_1 and x_1 to y_2 is the same. Similarly, the cost of transporting mass from x_2 to y_1 and x_2 to y_2 is the same. Source: Gabriel Peyré.

Kantorovich Problem

In the 1940s, Soviet mathematician, Leonid Kantorovich, revisited Monge's problem but allowed for the splitting of mass, and admits a dual formulation. This problem also does not have a unique solution. Similar to the Monge problem, we wish to transfer mass from location x to location y.

$$\begin{array}{ccc} x_1 \bullet & & \bullet\ y_1 \\ x_2 \bullet & \xrightarrow{\sigma} & \bullet\ y_2 \\ \vdots & & \vdots \\ x_n \bullet & & \bullet\ y_m \end{array}$$

Here, it is possible for $m \neq n$, since it possible to send mass from one location to multiple destinations. Let $\sum_n$ denote a collection of n nonnegative numbers that add up to 1. We define a set of matrices

$$\mathbf{U}(\mathbf{a},\mathbf{b}) \stackrel{\text{def}}{=} \{\mathrm{P} \in \mathbb{R}_+^{n,m} : \mathrm{P}\mathbb{1}_m = a \text{ and } \mathrm{P}^T\mathbb{1}_n = b\}$$

where

$$\mathrm{P}\mathbb{1}_m = \sum_j (\mathrm{P}_{i,j})_i \in \mathbb{R}^n \text{ and } \mathrm{P}\mathbb{1}_n = \sum_i (\mathrm{P}_{i,j})_j \in \mathbb{R}^m.$$

Then the most efficient cost of transport in this case is

$$\mathbf{L}_{\mathbf{C}(\mathbf{a},\mathbf{b})} \stackrel{\text{def}}{=} \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{C}, \mathbf{P}\rangle \stackrel{\text{def}}{=} \sum_{i,j} \mathbf{C}_{i,j}\mathbf{P}_{i,j},$$

which is attained for transport plan P^*.

Lemma. The Kantorovich problem is more efficient than the Monge problem.

$$\mathbf{L}_{\mathbf{C}}(\mathbb{1}_n/n, \mathbb{1}_n/n) \leq \min_{\sigma\in\text{Perm(n)}} \langle \mathbf{C}, \mathbf{P}_\sigma\rangle$$

Theorem. If $m = n$ and $a = b = \frac{\mathbb{1}_n}{n}$, then there exists an optimal solution for the Kantorovich problem $\mathbb{P}_{\sigma^*}$, which is a permutation matrix associated to an optimal permutation $\sigma \in \text{Perm}(n)$.

Coffee Break!

To illustrate the Kantorovich problem, it helps to think of n warehouses that store coffee beans required by m coffee shops. Suppose each warehouse is indexed with an integer i and contains a_i units of the resource, while the coffee shops are indexed with integer j and require b_j units of the resource. To transport the raw materials, the warehouse manager can hire a transportation company that charges $\mathbf{C}_{i,j}$ to transport one unit from i to j. In order to get the most ideal deal, the manager must solve the Kantorovich problem to obtain a transportation plan $\mathbf{P}^*$. The total amount they would have to pay the transportation company would then be $\langle \mathbf{P}^*, \mathbf{C}\rangle$.

Kantorovich Dual Problem

Theorem. The Kantorovich problem admits the dual

$$\mathbf{L}_{\mathbf{C}}(\mathbf{a},\mathbf{b}) = \max_{(\mathbf{f},\mathbf{g})\in\mathbf{R}(\mathbf{C})} \langle \mathbf{f},\mathbf{a}\rangle + \langle \mathbf{g},\mathbf{b}\rangle$$

where the set of admissible dual variables is

$$\mathbf{R}(\mathbf{C}) \stackrel{\text{def}}{=} \{(\mathbf{f},\mathbf{g}) \in \mathbb{R}^n \times \mathbb{R}^m : \forall (i,j) \in [n]\times[m], \mathbf{f}\oplus\mathbf{g} \leq \mathbf{C}\}.$$

The Kantorovich problem is a linear minimization problem with convex constraints. Therefore, it admits a dual problem. Looking at the same example of warehouses and coffee, suppose the manager outsources the problem of solving for the ideal transportation plan to a third party. The third party vendor will suggest a price of

$$\langle f, g\rangle + \langle a, b\rangle$$

where fi is the cost to collect a unit of resource at each warehouse i, g_j is the cost to deliver a unit of resource to factory j. a_i is the total number of units at warehouse i and b_j is units required factory j. The vendor will try to make f and g as high as possible. The warehouse manager should check the recommended price by checking if $f_i + g_j \leq \mathbf{C}_{i,j}$. If this inequality fails, then the manager should reject the vendor's offer. The manager's own transport plans would be too expensive

$$\sum_{i,j} \mathbf{P}_{i,j}\mathbf{C}_{i,j} \geq \sum_{i,j} \mathbf{P}_{i,j}(\mathbf{f}_i + \mathbf{g}_j) = (\sum_i \mathbf{f}_i \sum_j \mathbf{P}_{i,j}) + (\sum_j \mathbf{g}_j \sum_i \mathbf{P}_{i,j}) = \langle \mathbf{f}, \mathbf{a}+\mathbf{g}, \mathbf{b}\rangle.$$

So, the manager should accept the vendor's offer while the vendor should seek prices $\mathbf{f}, \mathbf{g}$ that maximize $\langle \mathbf{f},\mathbf{a}\rangle + \langle \mathbf{g},\mathbf{b}\rangle$ but also satisfy $\mathbf{f}_i + \mathbf{g}_j \leq \mathbf{C}_{i,j}$.

The Auction Algorithm

One algorithm to solve the optimal assignment problem is the auction algorithm. Suppose you have an equal number of buyers and goods. The algorithm consists of distributing the goods in a way such that the maximum amount of satisfaction is reached by the buyers. Here, individual satisfaction isn't the goal. We instead aim to find the maximum satisfaction of the group as a whole.
Let a_{ij} denote the "happiness" person i receives from good j, let $j = \sigma(i)$ denote the good, where σ is some permutation of the goods among all of the buyers, and let p_j be the price of good j. All buyers are content with their purchases if the following condition is satisfied:

$$a_{i\sigma(i)} - p_{\sigma(i)} = \max_{j=1,\ldots,N}\{a_{ij} - p_j\}.$$

The way the algorithm works is we begin with a random injective map of buyers and goods. Then, we select a specific buyer such that the above condition does not hold, and we exchange the good they have with a good that brings them more satisfaction. Continue this process until we reach a point where buyers are indifferent with the good they possess and the second best option. The auction algorithm can be extended to solve optimal transport problems. It applies mostly to linear optimal transport problems such as network optimization, shortest path and max-flow problems.

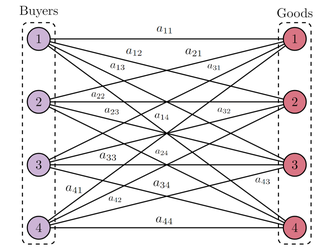


Figure 3. The Auction Algorithm. a_{ij} represents the satisfaction buyer i receives from good j.

Acknowledgements

I would like to thank the UCSB Directed Reading Program for giving me the opportunity to work on this project. I'm especially thankful to my mentor, Đorđe Nikolić, for his time and guidance throughout the past few months.

References

[1] Leonid V. Kantorovich. On translation of mass (in russian). Proceedings of the USSR Academy of Sciences, pages 199—201, 1942.

[2] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. De l'Imprimerie Royale, 1781.

[3] Gabriel Peyré and Marco Cuturi. Computational optimal transport. Foundations and Trends in Machine Learning, 11(5-6):355–607, 2019.

[4] Filippo Santambrogio. Optimal Transport for Applied Mathematicians, volume 87. Springer International Publishing, 2015.

Representation Theory (for Finite Groups)

Daniel Zhang [1]

[1]*University of California, Santa Barbara, Santa Barbara, CA*
UCSB Directed Reading Program 2022

Introduction

Representation theory is the study of how groups can act on vector spaces.

Definitions

A **representation** of a finite group G on a finite-dimensional vector space V is a homomorphism $\rho : G \to GL(V)$. We often refer to V as the representation and omit ρ.

A G**-linear map** between two representations V and W is a map $\varphi : V \to W$ where for all $g \in G$, $\rho_W(g) \circ \varphi = \varphi \circ \rho_V(g)$

Two representations are isomorphic if they are isomorphic as vector spaces by a G-linear isomorphism.

A **subrepresentation** W of V is a subspace W of V that is invariant under G.

An **irreducible** representation is a representation with no proper nonzero subspace invariant under G.

The **direct sum** $V \oplus W$ of two representations is a representation formed by taking the direct sums of their vector spaces with a group action defined by

$$g \cdot (v \oplus w) = (g \cdot v) \oplus (g \cdot w)$$

The **tensor product** $V \otimes W$ of two representations is a representation formed by taking tensor products of their vector spaces with a group action defined by

$$g \cdot (v \otimes w) = (g \cdot v, g \cdot w)$$

Properties of vector spaces such as $U \otimes (V \oplus W) = (U \otimes V) \oplus (U \otimes W)$ are also true for representations.

The **permutation representation** associated to a left action of G on a finite set X is a vector space with basis $\{e_g : g \in G\}$ with the left action

$$g \cdot \sum_{x \in G} a_x e_x = \sum_{x \in G} a_x e_{gx}$$

The **regular representation** of G is the permutation associated to the left action of G on itself.

Representations of S_3

Example (Trivial representation)

The trivial representation is the one-dimensional representation where the action of any group element is the identity.

Example (Alternating representation)

The alternating representation is the one-dimensional representation where the action of any permutation of even parity is the identity and the action of any permutation of odd parity is negation.

Example (Permutation representation on $\{1, 2, 3\}$)

This is the representation on a three dimensional vector space where the left action of $g \in S_3$ on the vector (z_1, z_2, z_3) is $g \cdot (z_1, z_2, z_3) = (z_{g^{-1}(1)}, z_{g^{-1}(2)}, z_{g^{-1}(3)})$

Example (Standard representation)

The permutation representation of S_3 acting on $\{1, 2, 3\}$ is not irreducible since it has an invariant subspace $\{(z_1, z_2, z_3) \in \mathbb{C}^3 : z_1 + z_2 + z_3 = 0\}$ The representation on this subspace is called the standard representation of S_3, and is irreducible.

The permutation representation on $\{1, 2, 3\}$ is the direct sum of the trivial representation and the standard representation. The trivial representation and standard representation are subrepresentations of the permutation representation on $\{1, 2, 3\}$.

Example (Regular representation of S_3)

This is the representation on a six-dimensional vector space with basis $\{e_h : h \in S_3\}$ where the left action of $g \in S_3$ is

$$g \sum_{h \in S_3} a_h e_h = \sum_{h \in G} a_h e_{gh}$$

Complete reducibility

If W is a subrepresentation of V, then there is a subspace W' of V invariant under G such that $V = W \oplus W'$. This can be shown by taking any projection onto W, averaging over all group elements, and looking at the kernel.

This means any representation can be recursively decomposed into a direct sum of irreducible representations. This means that if we find all irreducible representations of G, then any other representation can be written as a direct sum of those.

This decomposition is in some sense unique. Every representation V of G has a unique factorization $V = \oplus_i V_i^{\oplus a_i}$ where the V_i are irreducible representations of G. Each $V_i^{\oplus a_i}$ is uniquely determined, but the decomposition of $V_i^{\oplus a_i}$ into copies of V_i is not. A simple counterexample would be decomposing a 2-dimensional representation of the trivial group into 2 copies of the trivial representation.

Showing uniqueness requires the following lemma.

Theorem (Schur's lemma)

If $\varphi : V \to W$ is a G-module homomorphism between irreducible representations,

- *Either φ is an isomorphism or $\varphi = 0$*
- *If $V = W$, then φ is scalar multiplication by a constant.*

A consequence of Schur's lemma is that all irreducible representations of abelian groups are 1-dimensional.

Characters

A **character** of a representation V is a map $\chi_V : G \to \mathbb{C}$ defined by $\chi_V(g) = Tr(\rho(g))$.

χ_V is a class function, i.e. it is constant on conjugacy classes.

The character has the following nice properties, which make it convenient for computations:

$$\chi_{V \oplus W} = \chi_V + \chi_W$$
$$\chi_{V \otimes W} = \chi_V \cdot \chi_W$$
$$\chi_{V^*} = \overline{\chi_V}$$

We can define an inner product on characters

$$\langle \alpha, \beta \rangle = \frac{1}{|G|} \sum_{g \in G} \overline{\alpha(g)} \beta(g)$$

The characters of irreducible representations are orthonormal under this inner product.

The number of irreducible representations equals the number of conjugacy classes of G.

In other words, the characters of irreducible representations form an orthonormal basis on the space of class functions.

An irreducible representation V_i appears in V $\langle \chi_V, \chi_{V_i} \rangle$ times.

Since a representation is determined up to isomorphism by the number of copies of each irreducible representations it contains, this means a representation is determined up to isomorphism by its character.

If we know the character of a representation, it is very easy to check if it is irreducible: a representation is irreducible iff $\langle \chi_V, \chi_V \rangle = 1$.

Character tables

If we know all of a group's irreducible representations, we can decompose any representation with a known character into irreducible representations by taking inner products. It is convenient to summarize all this information about the group into a **character table**.

Since characters are constant on conjugacy classes, we only need its value on each conjugacy class. We also label each conjugacy class with how many elements it contains, since this is needed during inner product calculations. Character tables always have the same number of rows as columns.

Example (Character table for S_3)

	1	3	2
S_3	1	(1 2)	(1 2 3)
trivial U	1	1	1
alternating U'	1	-1	0
standard V	2	0	-1

We can check that these representations are irreducible because their inner product with themselves is 1. We know there are no other irreducible representations because S_3 only has three conjugacy classes.

Fixed-point formula

If V is a permutation representation of G acting on X, then $\chi_V(g)$ is the number of elements of X fixed by g. This is because if we write $\rho(g)$ as a matrix, the only nonzero diagonal entries are 1s where g fixes an element of X.

Example (Permutation representation of S_3 on $\{1, 2, 3\}$)

Let W be the permutation representation of S_3 on $\{1, 2, 3\}$.
The identity leaves all 3 elements fixed. A cycle of length 2 leaves 1 element fixed. A cycle of length 3 leaves no elements fixed.

$$\chi_W(1) = 3$$
$$\chi_W((1\,2)) = 1$$
$$\chi_W((1\,2\,3)) = 0$$

Now that we know the character, we can take inner products with irreducible representations to determine how many times each one occurs in W.

$$\langle \chi_W, \chi_U \rangle = \frac{1}{6}(1(3)(1) + 3(1)(1) + 2(0)(1)) = 1$$
$$\langle \chi_W, \chi_{U'} \rangle = \frac{1}{6}(1(3)(1) + 3(1)(-1) + 2(0)(1)) = 0$$
$$\langle \chi_W, \chi_V \rangle = \frac{1}{6}(1(3)(2) + 3(1)(0) + 2(0)(-1)) = 1$$

Hence, $W = U \oplus V$.

Irreducible representations of S_4

We can use the properties of characters to find all irreducible representations of S_4. Since S_4 has 5 conjugacy classes, it must have 5 irreducible represenations.

First, like S_3, S_4 has the trivial representation U, alternating representation U', and standard representation V. If we tensor the standard representation with the alternating representation, we get a distinct irreducible representation V'. We can verify this by computing its character by multiplying the characters for U' and V, and checking it is irreducible by taking its inner product with itself. The character of the remaining irreducible representation W must be orthogonal to all the other ones, and have inner product with itself equal to 1. The sign is determined since $\chi_W(1) = \dim W > 0$.

	1	6	8	6	3
S_4	1	(12)	(123)	(1234)	(12)(34)
U	1	1	1	1	1
U'	1	-1	1	-1	1
V	3	1	0	-1	-1
V'	3	-1	0	1	-1
W	2	0	-1	0	2

Acknowledgements

Thanks to the UCSB Directed Reading Program, and to Edward Chen for being my mentor.

References

[1] William Fulton; Joe Harris. *Representation Theory*. Springer, New York, NY, 2004.

Introduction to Quantum Computing

Sebastian Nunez and Shawn (Yeming) Xiao - Mentored by Greg McGrath

2022 Mathematics Directed Reading Program, University of California Santa Barbara

Introduction

Quantum computing is an emerging field and technology that uses the properties and behaviors of quantum mechanics to create more efficient computers. The main differentiation between classical and quantum computing is the possibility that quantum computing can challenge the weak church turing thesis. In classical computing all information exists in a simple "off" or "on" state, such as 0 or 1, state called "bits". Qubits do not have these restrictions and can exist in any probabilty of being 1 or 0. This means that the amount of information a systems can hold stored grows exponentially with additional qubits being added.

Mathematical Primer

- **Hilbert Spaces -** Hilbert spaces are special vector spaces denoted $\mathcal{H}$ that are necessary for the formulation of the notation of quantum computing. The Hilbert spaces that are relevant are finite-dimensional complex and will typically have a dimension 2^n.
- **Dirac Notation -** A quantum mechanic systems of vectors to represent the state of a qubit. Row vectors are named *kets* and are represented as $|\psi\rangle$ with ψ being the identifier of the ket. In a similar system column vectors are called *bras* and are represented as $\langle\phi|$.

$$\langle\psi| = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \qquad |\phi\rangle = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

- **Orthonormal Basis -** Consider a Hilbert space $\mathcal{H}$ of dimension 2^n. A set of 2^n vectors $B = \{|b_m\rangle\} \subseteq \mathcal{H}$ is called an orthonormal basis for $\mathcal{H}$ if

$$\langle b_n|b_m\rangle = \delta_{n,m} \qquad \forall b_m, b_n \in B$$

and every $|\psi\rangle \in \mathcal{H}$ can be written as

$$|\psi\rangle = \sum_{b_n \in B} \psi_n |b_n\rangle \quad \text{for some } \psi_n \in \mathbb{C}$$

The set $\{\langle b_n|\}$ is the orthonormal basis for $\mathcal{H}^*$ called the dual space.

- **Operators -** An operator on a vector space $\mathcal{H}$ is a linear transformation $\mathbf{T} : \mathcal{H} \to \mathcal{H}$. It is useful to note that by constructing an orthonormal basis $B = \{|b_m\rangle\}$ for a vector space $\mathcal{H}$. Then every linear operator $\mathbf{T}$ on $\mathcal{H}$ can be written as

$$\mathbf{T} = \sum_{b_n, bm \in B} \mathbf{T}_{n,m} |b_n\rangle \langle b_m|$$

where $\mathbf{T}_{n,m} = \langle b_n| \mathbf{T} |b_m\rangle$ are matrix elements. Additionally, $|b_n\rangle \langle b_m|$ is the outer product.

- **Tensor Products -** The tensor product is a way of combining spaces, vectors, or operators together. Suppose $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces of dimension *n* and *m* respectively. Then the tensor product space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is a new, larger Hilbert space of dimension *n*×*m*. Very often the $\otimes$ symbol is left out of the tensor product notation and $|\psi\rangle \otimes |\phi\rangle$ becomes $|\psi\rangle |\phi\rangle$ or $|\psi\phi\rangle$.

Unitary - An operator $\mathbf{U}$ is unitary if

$$\mathbf{U}^\dagger = \mathbf{U}^{-1}$$

Hermitean - An operator $\mathbf{T}$ is Hermitean (or self-adjoint) if

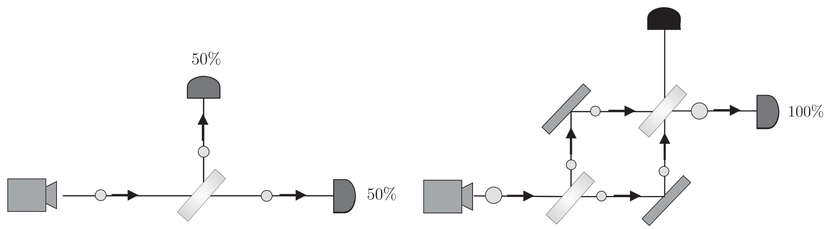
$$\mathbf{T}^\dagger = \mathbf{T}$$

Schmidt Decomposition Theorem

If $|\psi\rangle$ is a vector in a tensor product space $\mathcal{H}_A \otimes \mathcal{H}_B$, then there exists an orthonormal basis $\{|\phi_i^A\rangle\}$ for $\mathcal{H}_A$, and an orthonormal basis $\{|\phi_i^B\rangle\}$ for $\mathcal{H}_B$, and non-negative real numbers $\{p_i\}$ so that

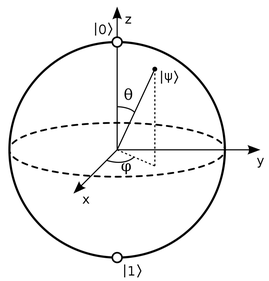
$$|\phi\rangle = \sum_i \sqrt{p_i} |\phi_i^A\rangle |\phi_i^B\rangle$$

Quantum Physics



Imagine that a beam of light is shot through a polished glass and two photon detectors are placed in the path of the reflect photons. After running the experiment on the left we observe that 50% of photons land in the path above and 50% of photons travel through to the right. This result is easily explained by classical mechanics as the polished glass randomly with a coin-flip to transmit or reflect the photons. On the right is the same setup with a few modifications to allow for an additional polished glass. Using our previous analysis we should expect that both photon sensors receive an equal distribution of photons. However, when performed the experiment on the right shows that 100% of photons travel to the right sensor. This non-intuitive behaviour occurs because of a unique property of quantum mechanics called *superposition*.

Superposition



Quantum bits exist in a superposition of states associated with weighted probabilities corresponding to the root of the likelihood of being observed in that state. A useful example quantum computing chooses is complex unit vector $|\Psi\rangle$ in a 2-dimensional Hilbert space.

$$|\Psi\rangle = \cos(\frac{\theta}{2}) |0\rangle + e^{i\Phi} \sin(\frac{\theta}{2}) |1\rangle$$

Where $e^{i\Phi}$ is a global phase factor and the kets $|0\rangle$ and $|1\rangle$ are the basis. When the qubit is measured such as the photon sensors in the experiment the superposition will "collapse" into the state $|0\rangle$ or $|1\rangle$. The state which the qubit collapses is determined by the state probabilities of $|0\rangle$ and $|1\rangle$.

Composite Systems and Measurements

The state space of the combined physical system is the tensor product space $\mathcal{H}_1 \otimes \mathcal{H}_2$ of the state spaces of the component subsystems. If the first system is in the state $|\psi_1\rangle$ and the second system is in the state $|\psi_2\rangle$, then the state of the combined system is

$$|\psi_1\rangle \otimes |\psi_2\rangle \,.$$

Importantly, qubits that can not be written as such are referred to as *entangled*.

For a given orthonormal basis $B = \{|\varphi_i\rangle\}$ of a state space $\mathcal{H}_A$ for a system *A*, it is possible to perform a *Von Neumman measurement* on system $\mathcal{H}_A$ with respect to the basis *B*, given a state

$$|\psi\rangle = \sum_i \alpha_i |\varphi_i\rangle \,,$$

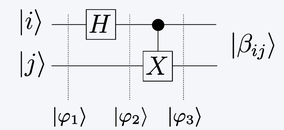
outputs a label *i* with a probability $|\alpha_i|^2$ and leaves the system in state $|\varphi_i\rangle$. Furthermore, given a state $|\psi\rangle = \sum_i \alpha_i |\varphi_i\rangle |\gamma_i\rangle$ fromm a bipartite state space $\mathcal{H}_A \otimes \mathcal{H}_B$ (the φ_i are orthonormal; the γ_i have a un it norm but are not necessarily orthogonal), then performing a Von Neumann measurement on system *A* will yield outcome *i* with a probability $|\alpha_i|^2$ and leave the bipartite system in state $|\varphi_i\rangle |\gamma_i\rangle$.

Quantum Circuits and Bell Basis

Quantum computing is performed on circuits that apply operators on a set of input qubits. The operators are called "gates" and are represented with a box spanning the qubits the operator acts on.

$$\mathbf{X} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \quad \mathbf{Z} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \mathbf{H} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

The $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ gates are fundamental operators that correspond to a rotation of a qubit on one of the axis. Using them in combination can translate a qubit to any point on the Bloch sphere so they are referred to as a set of *universal gates*. The $\mathbf{H}$ gate is useful because it can take any qubit already collapsed and turn it into a superposition.



The Bell Basis is a constructed 2-qubit set of superpositions $\{|\beta_{00}\rangle, |\beta_{01}\rangle, |\beta_{10}\rangle, |\beta_{11}\rangle\}$ that are necessary to generate many of the significant applications of quantum computing.

$$|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \qquad |\beta_{01}\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$$

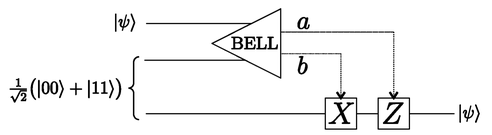
$$|\beta_{10}\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle) \qquad |\beta_{11}\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

Superdense Coding and Quantum Teleportation

Superdense coding allows for a qubit to send two classical pieces of information through a channel with only one qubit. The setup required is both ends of the channel to have the same initial $|\beta_{00}\rangle$ state. This is done through applying a combination of $\mathbf{Z}$ and $\mathbf{X}$ gates.

To send	Transformation
00	$\mathbf{I} \otimes \mathbf{I}$: $\|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle) \mapsto \frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle) = \|\beta_{00}\rangle$
01	$\mathbf{X} \otimes \mathbf{I}$: $\|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle) \mapsto \frac{1}{\sqrt{2}}(\|01\rangle + \|10\rangle) = \|\beta_{01}\rangle$
10	$\mathbf{Z} \otimes \mathbf{I}$: $\|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle) \mapsto \frac{1}{\sqrt{2}}(\|00\rangle - \|11\rangle) = \|\beta_{10}\rangle$
11	$\mathbf{Z} \cdot \mathbf{X} \otimes \mathbf{I}$: $\|\beta_{00}\rangle = \frac{1}{\sqrt{2}}(\|00\rangle + \|11\rangle) \mapsto \frac{1}{\sqrt{2}}(\|01\rangle - \|10\rangle) = \|\beta_{11}\rangle$

Quantum teleportation allows the abilty to send one qubit of information using only two *bits* of information.



Crucially, this is possible with neither party of the exchange knowing their own state.

References

[1] Michele Mosca Philip Kaye, Raymond Laflamme. An introduction to quantum computing. *Oxford University Press*, pages 1–270, 2007.

[2] Thomas G. Wong. Introduction to classical and quantum computing. *www.thomaswong.net*, pages 1–95, 2022.

Hyperbolic Geometry and the Farey Tessellation

Misha Kulshresta
University of California, Santa Barbara

What is Hyperbolic Geometry?

Hyperbolic geometry is a geometry in which Euclid's parallel postulate is rejected. Two-dimensional hyperbolic geometry can be modeled in two ways, the open half-plane and the disk model. We can define the open half-plane as follows:

$$\mathbb{H}^2 = \{(x, y) \in \mathbb{R}^2; y > 0\} = \{z \in C : Im(z) > 0\}$$

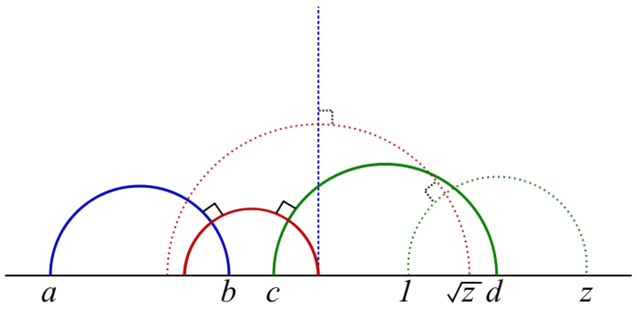


Fig. 1: An example of the half-plane model.

The vertical line here represents the imaginary axis, z. Each arc of Fig. 1 represents a *geodesic*, which can be defined as the curve which is the shortest distance between two points.

The *disk model* is represented by a disk of radius 1. It has a circle at infinity, such that as you get closer to the outer circle, you approach infinity. This model is additionally popular in mathematically-inspired art, as can be seen below in artist M.C. Escher's piece *Circle Limit 1*.

Fig. 2: M.C. Escher's Circle Limit 1.

The below definition will be useful in the following columns:
An *isometry* is a distance-preserving transformation between metric spaces (which includes both the euclidean and hyperbolic planes).

Tessellating

A *tessellation* of a surface (such as the euclidean or hyperbolic plane) is a family of tiles $X_n, n \in \mathbb{N}$, such that:
1. each tile X_m is a connected polygon on the surface.
2. any two X_m, X_n are isometric.
3. the X_m cover the whole surface, in the sense that their union is equal to this space.
4. the intersection of any two distinct tiles X_m and X_n consists only of vertices and edges of X_m, which are also vertices and edges of X_n.

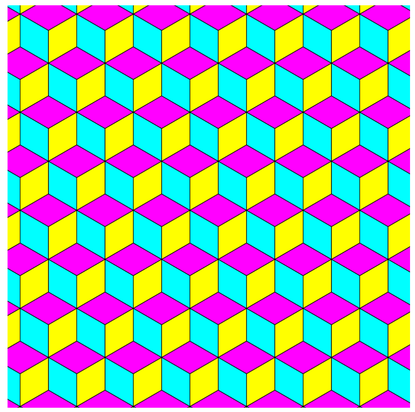


Fig. 3: A tessellation of the euclidean plane with isometric quadrilaterals.

The above diagram is classified as a p6m tessellation (primitive cell, 6-fold rotation, and mirrored), which is one of 17 wallpaper groups.

Tessellations of the hyperbolic plane follow the same rules. Isometries of the hyperbolic plane are not as immediately visible, but the diagram below does in fact satisfy point 2 from above.

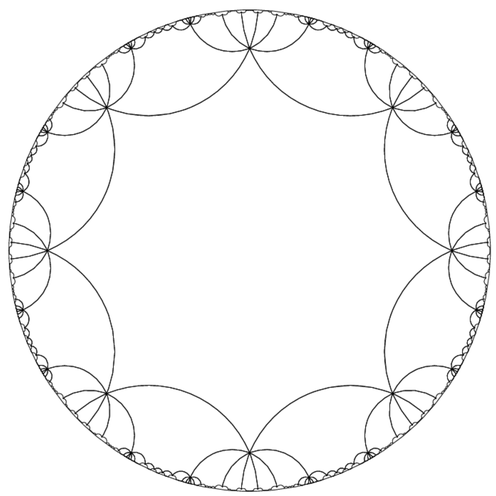


Fig. 4: An order-8 octagon tiling of the hyperbolic plane, resulting in a tessellation of the surface.

Farey Circle Packing

For every rational number $\frac{p}{q} \in \mathbb{Q}$ with p, q coprime and $q > 0$, draw in the plane $\mathbb{R}^2$ and the circle $C_{\frac{p}{q}}$ of diameter $\frac{1}{q^2}$ that is tangent to the x-axis at $(\frac{p}{q}, 0)$ and lies above this axis. These circles $C_{\frac{p}{q}}$ fit together to form a pattern of tangent circles with disjoint interiors as seen below.

$$\infty = \tfrac{1}{0}$$

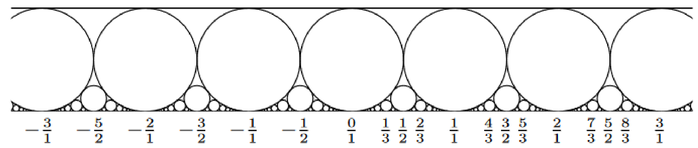


Fig. 5: Farey circle packing.

The Farey Tessellation

Suppose we erase the circles $C_{\frac{p}{q}}$ from the diagram, and instead connect the points $(\frac{p}{q}, 0)$ and $(\frac{p'}{q'}, 0)$ with a semi-circle centered on the x-axis where the circles $C_{\frac{p}{q}}$ and $C_{\frac{p'}{q'}}$ are tangent. The resulting set of hyperbolic geodesics form the *Farey Tessellation*.

$$\infty = \tfrac{1}{0}$$

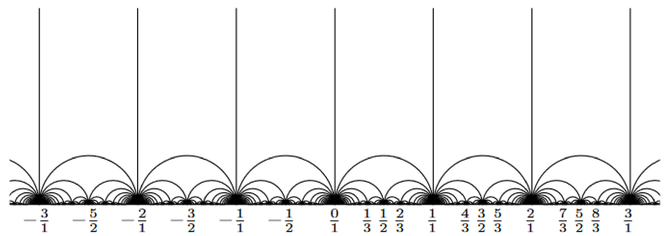


Fig. 6: The Farey tessellation of the hyperbolic plane.

Acknowledgements

A big thank you to Jaime Vandeveer for guiding me through this exciting adventure into hyperbolic geometry, and for making my participation in the Directed Reading Program possible!

References

Bonahon, Francis. *Low-Dimensional Geometry: From Euclidean Surfaces to Hyperbolic Knots.* American Mathematical Society, 2009.

Category Theory and Homological Algebra

Isaiah Martinez, Joseph Sullivan. Mentor: Jeremy Brightbill

2022 Mathematics Directed Reading Program.

Categories

A **category** $\mathcal{C}$ consists of the following data:

- a collection $\mathrm{Ob}(\mathcal{C})$ called the **objects** of $\mathcal{C}$,
- for each $A, B \in \mathrm{Ob}(\mathcal{C})$, a collection $\mathrm{Hom}_{\mathcal{C}}(A, B)$, called the **morphisms** from A to B,
- a morphism $1_A \in \mathrm{Hom}_{\mathcal{C}}(A, A)$ for each $A \in \mathrm{Ob}(A)$
- for each $A, B, C \in \mathrm{Ob}(\mathcal{C})$, a function called **composition**,

$$\begin{aligned} \circ : \mathrm{Hom}_{\mathcal{C}}(A, B) \times \mathrm{Hom}_{\mathcal{C}}(B, C) &\longrightarrow \mathrm{Hom}_{\mathcal{C}}(A, C) \\ f, g &\longmapsto g \circ f. \end{aligned}$$

with the axioms that composition is associative, and for each morphism $f \in \mathrm{Hom}_{\mathcal{C}}(A, B)$, $\mathrm{id}_B \circ f = f = f \circ \mathrm{id}_A$.

Example. We give three examples of categories.

- "Set" has sets as objects and set functions as morphisms.
- Let R be a ring. "$R-$Mod" has left R-modules as objects and R-module homomorphisms as morphisms.
- The category "Top" has topological spaces as objects and continuous maps as morphisms. "Haus" is the subcategory of "Top" consisting of Hausdorff spaces.

We give two more definitions—special types of morphism that generalize injective/surjective functions respectively.

- A **monomorphism** (mono) is a morphism $f : B \hookrightarrow C$ such that whenever morphisms $g, h : A \to B$ satisfy $f \circ g = f \circ h$, then $g = h$.
- An **epimorphism** (epi) is a morphism $f : A \twoheadrightarrow B$ such that whenever morphisms $g, h : B \to C$ satisfy $g \circ f = h \circ f$, then $g = h$.

Example. In the categories "Set" and "$R-$Mod", monos and epis are exactly injections and surjections. In "Haus", monos are still injections, but epis need not be surjective—they only need to have dense image.

Universal Properties

All objects live inside a category $\mathcal{C}$.

- An **initial object** is an object I such that for any object C, there exists a unique morphism $f : I \to C$.
- A **terminal object** is an object T such that for any object C, there exists a unique morphism $g : C \to T$.
- A **zero object** is an object 0 which is both initial and terminal.
- A **product** of a family $\{A_\lambda\}_{\lambda\in\Lambda}$ is a pair $(A, \{\pi_\lambda\}_{\lambda\in\Lambda})$ of an object A and morphisms $\pi_\lambda : A \to A_\lambda$ (called *projection morphisms*) such that for any object T and morphisms $\{f_\lambda : T \to A_\lambda\}_{\lambda\in\Lambda}$, there exists a unique $f : T \to A$ such that Figure 1 commutes (for each $\lambda \in \Lambda$).
- A **coproduct** of a family $\{B_\lambda\}_{\lambda\in\Lambda}$ is a pair $(B, \{\iota_\lambda\}_{\lambda\in\Lambda})$ of an object B and morphisms $\iota_\lambda : B_\lambda \to B$ (called *inclusion morphisms*) such that for any object T and morphisms $\{g_\lambda : B_\lambda \to T\}_{\lambda\in\Lambda}$, there exists a unique $g : B \to T$ such that Figure 2 commutes (for each $\lambda \in \Lambda$).

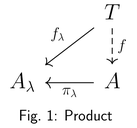


Fig. 1: Product

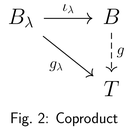


Fig. 2: Coproduct

Remark. In an arbitrary category objects satisfying these universal properties may not exist, but when they do exist, they are unique up to unique isomorphism.

Example. In "Set" ("Top" is similar),

- $\varnothing$ is initial.
- $\{*\}$ is terminal.
- There is no zero object.
- The product of $\{A_\lambda\}_{\lambda\in\Lambda}$ is the usual Cartesian product $\prod_{\lambda\in\Lambda} A_\lambda$ with componentwise projections.
- The coproduct of $\{B_\lambda\}_{\lambda\in\Lambda}$ is the *disjoint union* $\bigsqcup_{\lambda\in\Lambda} B_\lambda$ with inclusions.

Example. In "$R-$Mod",

- 0 is zero.
- The (finite) product and coproduct of $A_1, \ldots, A_n$ is $A_1 \oplus \cdots \oplus A_n$ with componentwise projections and inclusions.

Kernels and Cokernels

Let $\mathcal{C}$ be a category with zero object 0. We say a morphism $g : A \to B$ is 0 if g is the composition of the unique morphisms $A \to 0$ and $0 \to B$.

- A **kernel** of a morphism $f : A \to B$ is a pair (K, k) of an object K and a morphism $k : K \to A$ such that $f \circ k = 0$ and for any object K' and morphism $k' : K' \to A$ such that $f \circ k' = 0$, there exists a unique morphism $u : K' \to K$ such that Figure 3 commutes.
- A **cokernel** of a morphism $f : B \to A$ is a pair (Q, q) of an object Q and a morphism $q : A \to Q$ such that $q \circ f = 0$ and for any object Q' and morphism $q' : A \to Q'$ such that $q' \circ f = 0$, there exists a unique morphism $u : Q \to Q'$ such that Figure 4 commutes.

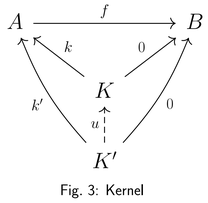


Fig. 3: Kernel

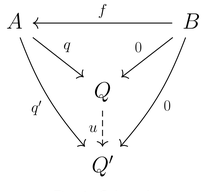


Fig. 4: Cokernel

- An **image** of a morphism $f : A \to B$ is a kernel of the cokernel of f.
- A **coimage** of a morphism $f : A \to B$ is a cokernel of the kernel of f.

Example. In the category $R-$Mod, a kernel of $f : A \to B$ is just $(\ker f, \iota)$, where ι is the inclusion of $\ker f \hookrightarrow A$. A cokernel of $f : A \to B$ is just $(B/\operatorname{im} f, \pi)$, where π is the quotient map $B \to B/\operatorname{im} f$.

Abelian Categories

- A **preadditive category** is a category $\mathcal{C}$ where all collections $\mathrm{Hom}_{\mathcal{C}}(A, B)$ are endowed with an abelian group structure, and all morphisms $f, g, h \in \mathrm{Hom}_{\mathcal{C}}(A, B)$ satisfy
  - $f \circ (g + h) = (f \circ g) + (f \circ h)$
  - $(f + g) \circ h = (f \circ h) + (g \circ h)$.

  i.e., $\circ$ is biadditive.
- An **additive category** is a preadditive category that has all finite products and a zero object. In an additive category, all finite products are also coproducts.
- An **abelian category** is an additive category where
  - every morphism $f : A \to B$ has a kernel and cokernel.
  - every monomorphism is a kernel, and every epimorphism is a cokernel.
- An **exact sequence** is a sequence

$$\cdots \to X_{i-1} \xrightarrow{f_{i-1}} X_i \xrightarrow{f_i} X_{i+1} \to \cdots$$

of objects and morphisms in an abelian category $\mathcal{A}$ such that $\operatorname{im} f_i = \ker f_{i+1}$ for all i. In particular, $f_{i+1} \circ f_i = 0$.

Example. "$R-$Mod" is an abelian category. "Set" and "Top" are not abelian (they do not even have a zero object).

Proposition. Let $f : A \to B$ in an abelian category $\mathcal{A}$. Then, f is a mono iff $\ker f = 0$, and f is an epi iff $\operatorname{coker} f = 0$.

The Five Lemma

Theorem. (The Five Lemma) Let $\mathcal{A}$ be an abelian category. Consider the commutative diagram (Figure 5) of objects and morphisms in $\mathcal{A}$.

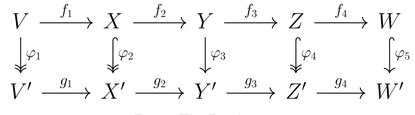


Fig. 5: The Five Lemma

Suppose both rows of the diagram are exact sequences. Additionally, suppose that φ_2 and φ_4 are isomorphisms (both an epimorphism and a monomorphism), φ_1 is an epimorphism, and φ_5 is a monomorphism. Then φ_3 must be an isomorphism.

Proof. For simplicity, we shall assume $\mathcal{A}$ is the category of R-modules. To prove that φ_3 is an isomorphism, it suffices to show that φ_3 is both injective and surjective.

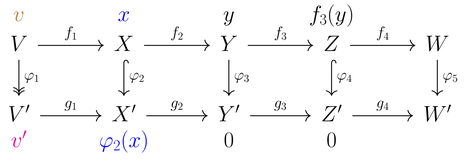


Fig. 6: Proof of Injectivity

To show injectivity of φ_3, we prove that $\ker(\varphi_3) = \{0\}$. Let $y \in \ker(\varphi_3)$, so that $\varphi_3(y) = 0 \in Y'$. Then $g_3(\varphi_3(y)) = g_3(0) = 0$. Since the diagram commutes, $\varphi_4(f_3(y)) = g_3(\varphi_3(y)) = 0$. Since φ_4 is injective, $\ker(\varphi_4) = \{0\}$. Thus $f_3(y) = 0$, so $y \in \ker(f_3)$. Exactness of the first row tells us that $\operatorname{im}(f_2) = \ker(f_3)$, so there exists an $x \in X$ such that $f_2(x) = y$. Now by commutativity of the diagram, we have

$$g_2(\varphi_2(x)) = \varphi_3(f_2(x)) = \varphi_3(y) = 0.$$

Thus $\varphi_2(x) \in \ker(g_2) = \operatorname{im}(g_1)$, so there exists $v' \in V'$ such that $g_1(v') = \varphi_2(x)$. Surjectivity of φ_1 implies that there exists a $v \in V$ such that $\varphi_1(v) = v'$. Then,

$$\varphi_2(f_1(v)) = g_1(\varphi_1(v)) = g_1(v') = \varphi_2(x).$$

By injectivity of φ_2, $f_1(v) = x$, so $f_2(f_1(v)) = f(x) = y$. By exactness, $f_2 \circ f_1$ is zero, so $y = 0$. Thus $\ker(\varphi_3) = \{0\}$, so φ_3 is injective.

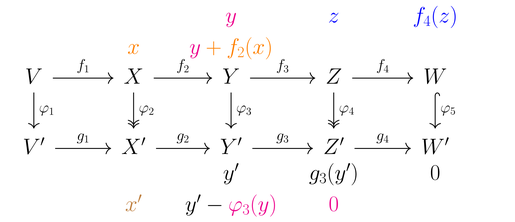


Fig. 7: Proof of Surjectivity

To prove φ_3 is surjective, let $y' \in Y'$. By surjectivity of φ_4 there exists $z \in Z$ such that $\varphi_4(z) = g_3(y')$. By exactness of the bottom row, $g_4 \circ g_3$ is zero. Then, by commutativity of the diagram,

$$0 = g_4(g_3(y')) = g_4(\varphi_4(z)) = \varphi_5(f_4(z)).$$

Thus $f_4(z) \in \ker(\varphi_5)$. Since φ_5 is injective, $f_4(z) = 0$. Thus, $z \in \ker(f_4) = \operatorname{im}(f_3)$, so there exists $y \in Y$ such that $f_3(y) = z$. Then,

$$g_3(\varphi_3(y)) = \varphi_4(f_3(y)) = \varphi_4(z) = g_3(y').$$

Thus $g_3(y' - \varphi_3(y)) = 0$ because g_3 is a homomorphism. Then $y' - \varphi_3(y) \in \ker(g_3) = \operatorname{im}(g_2)$ so there exists $x' \in X'$ such that $g_2(x') = y' - \varphi_3(y)$. Since φ_2 is surjective, there exists $x \in X$ such that $\varphi_2(x) = x'$. Then $\varphi_3(f_2(x)) = g_2(\varphi_2(x)) = g_2(x') = y' - \varphi_3(y)$. Therefore, $\varphi_3(f_2(x) + y) = \varphi_3(f_2(x)) + \varphi_3(y) = y' - \varphi_3(y) + \varphi_3(y) = y'$. Thus, every element of Y' has a nonempty preimage in Y, and therefore φ_3 is surjective.

□

References

(1) Francis Borceux. The Handbook of Categorical Algebra I: Basic Category Theory. Cambridge University Press, 1994.

(2) Thomas W. Hungerford. Algebra. Springer Science & Business Media, 2012.

(3) Samuel Petterson. Additive, abelian, and exact categories. 2016.