An Introduction to Information Geometry

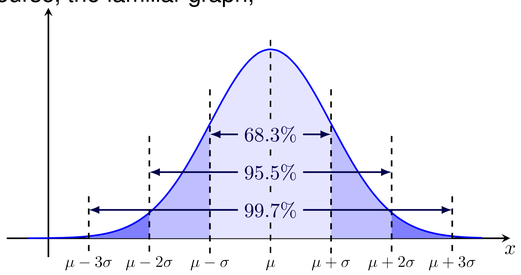
Daniel Badilla Jr, Ran Ji, James McNeice
University of California - Santa Barbara

The Normal Distribution

The standard normal distribution, defined by parameters μ, the mean of a population for some statistic and σ, the standard deviation. Letting $\theta = (\mu, \sigma)$ where $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$ then the corresponding probability density function (PDF)

$$p_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{\sigma^2}\right] \quad (1)$$

and of course, the familiar graph,



From (1) we notice that there is a one to one correspondence between PDFs and ordered pairs, so if we take $p_\theta(x)$ as a particular PDF, we can identify $S := \{p_\theta(x) \mid \theta \in \mathbb{H}^2\}$ which leaves the relation $\mathbb{H}^2 \cong S$.

Information Geometry

Information Geometry looks to use differential geometry to garner a deeper understanding of statistics. In particular, we can apply various techniques to make optimal decisions given a set of parameters (Nielsen). Statistical Inference, or the problem of choosing a model given sample data is one common application.

It turns out S naturally becomes a statistical manifold, a special case of **Riemannian manifolds**. Generally, Riemannian manifolds possess a **Riemannian metric** which we can think of as the collection of inner products on the tangent spaces of all points on the manifold. We would like to define a special Riemannian metric g which is called the **Fisher Information Metric**. We let $\ell_\theta(x) = \log(p_\theta(x))$ (the so called log-likelihood), then

$$g_{ij}(\theta) = \mathsf{E}\left[\frac{\partial \ell_\theta}{\partial \theta_i}\frac{\partial \ell_\theta}{\partial \theta_j}\right] \quad (2)$$

which results in

$$g_S = \frac{1}{\sigma^2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \quad (3)$$

The pair (S, g) is then a **statistical manifold**, that is, a Riemannian manifold the set of which represents a statistical family. Furthermore, (S, g) is isometric to the Poincaré upper half plane $\mathbb{H}^2$.

The Upper Half Plane $\mathbb{H}^2$ (Hyperbolic Space)

$\mathbb{H}^2$ refers to the hyperbolic space, which is a two-dimensional manifold with constant negative curvature. Defined as

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

- Since $\mathbb{H}^2$ is a manifold (locally Euclidean), there exists a mapping, commonly known as a patch. A **patch** provides a way to locally parametrize a portion of the manifold using coordinates from Euclidean space.
- Given a smooth surface S, $\forall p \in S$, there exists a regular patch $\sigma : U(\subseteq \mathbb{R}^2) \to W(\subseteq \mathbb{R}^3)$ such that $\sigma(U) = S \cap W, \sigma : U \to S \cap W$ is a homeomorphism.
- Poincaré Metric of $\mathbb{H}^2$: $g_{\mathbb{H}^2} = \frac{dx^2+dy^2}{y^2} = \frac{1}{y^2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$
- The Poincaré metric and the Fisher Information metric coincide, sharing the same mathematical form

$\mathbb{H}^2$ Isometric to Poincaré Disk $\mathbb{D}$

- Poincaré Disc : $\mathbb{D} = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 < 1\} = \{z = x + iy \in \mathbb{C} | z\overline{z} < 1\}$ whose Poincaré metric is $g_{\mathbb{D}} = ds^2 = \frac{dx^2+dy^2}{(1-x^2-y^2)^2}$
- Given two smooth surfaces S_1 and S_2, a smooth map $f : S_1 \to S_2$ is called an isometry if it takes any curve in S_1 to a curve of the same length in S_2. In other words, isometry preserves geometric structures of surfaces.
- The **Cayley Transform**

$$c : \mathbb{H}^2 \to \mathbb{D} \qquad \text{defined as} \qquad z \longmapsto \frac{z-i}{z+i}$$

is an isometry from $\mathbb{H}^2$ to $\mathbb{D}$.

Space of Normal Distributions Is Isometric to Hyperbolic Space

Gauss's Theorema Egregium states that the Gaussian curvautre of a surface depends only on its Riemannian metric (The first fundamental form).

- $\mathbb{H}^2$ and $\mathbb{D}$ are isometric and their Gaussian curvatures are negative constant.
- By Uniformization theorem, there is only one possible simply-connected smooth surface whose Gaussian curvature are negative constant up to isometries.
- Consequently, $\mathbb{H}^2$, $\mathbb{D}$ with the Poincaré metric, and the space of one-dimensional normal distributions with the Fisher Information metric are all regarded as a same smooth surface.

Implication of Differential Geometry in Statistics

Cramér–Rao Inequality

- A fundamental result in mathematical statistics that provides a lower bound for the variance of any unbiased estimator of a parameter.
- Let *X* be an observation taking values in measurable space (*X*,*B*) whose distribution P_θ depends on an s-variate parameter $\theta \in \Theta$ and is given by density $p(x;\theta) = p(x;\theta_1, ..., \theta_s)$ with respect to a measure μ on (*X*,*B*)

$$P_\theta(B) = \int_B p(x;\theta)d\mu(x), B \in \boldsymbol{B}$$

Suppose that $I_X(\theta)$ is positive definite so that $I_X^{-1}(\theta)$ exists. The Cramér-Rao inequality claims that the covariance matrix $V_{\hat{\theta}}$ of any unbiased estimator $\theta(X)$ of θ is bound from below by $I_X^{-1}(\theta)$:

$$V_{\hat{\theta}}(\theta) \geq I_X^{-1}(\theta),$$

Information Geometry Interpretation

- We can show that this inequality is a direct corollary of the following two facts:
  1. Monotonicity of $I(\theta$. If S=S(X) is a statistic, then
  $$I_S(\theta) \leq I_X(\theta), (1)$$
  2. In a linear model, if Y is an n-dimensional random vector with
  $$E_\theta(Y) = A\theta, V_Y(\theta) = E_\theta(Y - A\theta)(Y - A\theta)', (2)$$
  Then we have
  $$I_Y(\theta) \geq A' V_Y^{-1} A, (3)$$
  From (1) and (3), we can derive the **Cramér-Rao Inequality.**
- The above theorem indicates that the Cramér-Rao Inequality can be interpreted from a differential geometry perspective as the comparison between two Fisher Information Metric.

Acknowledgements

Reference Material:

- "Another Look at the Cramér-Rao Inequality by Abram Kagan
- "An Elementary Introduction to Information Geometry" by Frank Nielsen
- "The Geometry of Asymptotic Inference" by Robert E.Kass
- "Riemannian Geometry" by Manfredo Perdigão do Carmo

Thank you to the UCSB Directed Reading Program and to our mentor **Gunhee Cho** for making this project possible.

Basic Computations of Mapping Class Groups

Ajay Manneth

University of California - Santa Barbara

Basic terminology and notation of surfaces

A surface, at its most general, is a two-dimensional manifold possibly with boundary, and an orientable surface is one where the property of chirality is preserved for figures embedded in the surface. The following fundamental theorem, as often attributed to Möbius, gives a classification of all compact, orientable, and connected surfaces:

Theorem: Any closed, connected, orientable surface is homeomorphic to the connect sum of a 2-dimensional sphere with $g \geq 0$ tori. Any compact, connected, orientable surface is obtained from a closed surface by removing $b \geq 0$ open disks with disjoint closures. The set of homeomorphism types of compact surfaces is in bijective correspondence with the set $\{(g, b) : g, b \geq 0\}$

g is known as the genus of our surface, and b gives us the number of boundary components of the surface. Furthermore, we can create noncompact, orientable surfaces by introducing "punctures" into these surfaces by removing points from the interior of the surface. If we were to remove $n > 0$ points from the interior of one such surface, we would say the resulting surface had n punctures.
We then denote a surface of genus g with 0 boundary components and n punctures as $S^b_{g,n}$. For a surface with no boundary components, we may simply write $S_{g,n}$.

Definition of a Mapping Class Group

Let S be a surface of the form $S^b_{g,n}$ as defined above. Let $\mathrm{Homeo}^+(S, \partial S)$ be the group of all homeomorphisms of S to itself that restrict to the identity on ∂S, the boundary of S. We define the mapping class group of S, $\mathrm{Mod}(S)$, to be

$$\mathrm{Mod}(S) = \pi_0(\mathrm{Homeo}^+(S, \partial S)).$$

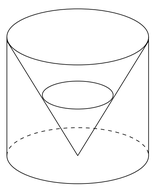
In other words, the mapping class group is the group of homeomorphisms of S to itself, up to isotopies, or continuous deformations, between homeomorphisms, that fix the boundary pointwise.
We notice that by the above definitions, punctures in a surface must be sent to punctures by both the defined homeomorphisms and isotopies, but are allowed to be permuted with each other. This is distinct from the behavior of boundary components, which must be fixed pointwise. However, like boundary components, we can see that the requirement of isotopy must leave punctures fixed, as well.

Alexander's Lemma

Let us begin by computing one particularly easy, yet useful, example of a mapping class group. Consider the surface $D^2 \simeq S^1_0$, which is just the closed disk, and let $\phi : D^2 \to D^2$ be a homeomorphism such that the restriction $\phi|_{\partial D^2}$ is the identity.
Let us identify D^2 with the closed unit disk centered at the origin in $\mathbb{R}^2$. Now, define the following function $F : D^2 \times [0, 1] \to D^2$ for $x \in D^2$ and $0 \leq t \leq 1$:

$$F(x,t) = \begin{cases} (1-t)\phi(\frac{x}{1-t}) & 0 \leq |x| < 1-t \\ x & 1-t \leq |x| \leq 1 \end{cases}$$

We can verify that $F(x, 0) = \phi(x)$, and that $F(x, 1) = x$ is the identity. Finally, we show that $F(x, t)$ is an isotopy between the two given homeomorphisms. To see why, we can consider F to itself be a map of the cylinder $D^2 \times [0, 1]$ to itself, as seen below:



Every lateral slice represents a different value of t, going downwards from 0 to 1, and the section of the cone in each slice represents the $(1-t)\phi(\frac{x}{1-t})$ term, which we can think of as "shrinking" the homomorphism ϕ down to the origin. Since ϕ is fixed on the boundary ∂D^2, we can see that every point on the boundary of the cone agrees with both the identity map and the scaled copy of ϕ, and so every lateral slice, corresponding to some value of t, is itself a homeomorphism. Since the slices continuously vary from $t = 0$ to $t = 1$, we have that F must therefore be a homotopy from ϕ to the identity on D^2.
In particular, we have shown that every member of $\mathrm{Homeo}^+(D^2, \partial D^2)$ is isotopic to the identity map, and therefore the group $\mathrm{Mod}(D^2)$ is trivial.
In fact, we can do a similar calculation with a once punctured-disk; by identifying the punctured disk with the closed unit disk in $\mathbb{R}^2$ punctured at the origin, we are able to use the same function $F(x, t)$, and a similar argument, to show that the mapping class group of the punctured disk is also trivial.

This proof is known as Alexander's lemma, and it ends up being an invaluable tool in the computation of many mapping class groups, as we shall soon see.

Mapping Class Group of $S_{0,3}$

We can use Alexander's Lemma to compute the mapping class group of the thrice-punctured sphere, $S_{0,3}$.

Theorem: The natural map $\mathrm{Mod}(S_{0,3}) \to \Sigma_3$ given by the action of elements of the mapping class group on the surface is an isomorphism, where Σ_3 is the symmetric group on 3 elements.

Proof: We can clearly see that this map is well-defined and surjective, and that the kernel of this map is the elements of the mapping class group which fix all 3 punctures individually.
Now, take any homeomorphism $\phi \in \mathrm{Homeo}^+(S_{0,3}, \partial S_{0,3})$, such that ϕ fixes all 3 punctures. Treating the punctures, which we label p, q, r, as marked points, we let α be a simple arc connecting p and q. We can see that $\phi(\alpha)$ is another simple arc, and since the punctures remain fixed pointwise, we are able to say that $\phi(\alpha)$ is another simple arc connecting p, q. We then introduce a the following proposition:

Proposition: Given ϕ and α defined above, if $\phi(\alpha)$ and α connect the same pair of punctures, then there exists a homeomorphism of $S_{0,3}$ isotopic to ϕ, which we label ϕ', which fixes α pointwise.

Assuming this proposition, we need only show that ϕ' is isotopic to the identity. To complete the proof, we need only "cut open" the surface $S_{0,3}$ along α. This action produces a once-punctured disk, with a lone puncture r and a boundary consisting of two copies of α, and the punctures p, q. ϕ' then induces a homeomorphism on the disk, which we denote $\bar{\phi}'$, that fixes the boundary and r. Therefore, we can see that $\bar{\phi}'$ is isotopic to the identity using Alexander's lemma. This induces an isotopy from ϕ' to the identity on $S_{0,3}$, and so we can see that ϕ must therefore be isotopic to the identity.
Therefore, we have that any homeomorphism from $S_{0,3}$ to itself that fixes the punctures is isotopic to the identity. We can thus determine that this constitutes all the elements that have mapping class in the kernel of our map $\mathrm{Mod}(S_{0,3}) \to \Sigma_3$, and so we determine that $\mathrm{Mod}(S_{0,3}) \simeq \Sigma_3$.

Acknowledgements

Reference Material: "A Primer on Mapping Class Groups" by Benson Farb and Dan Margalit
Thank you to the UCSB Directed Reading Program, and to my DRP mentor Michael Zshornack, for making this study possible.

Categorical Thinking: Category Theory and the Yoneda Lemma

by Alexander Urena and Jacob Hembree

University of California, Santa Barbara

Philosophy Of Categories

Mathematician Michael Atiyah once described mathematics as the "science of analogy." In the same vein, the objective of category theory then is *mathematical analogy*.

Thus, rather than characterize mathematical objects directly, the categorical approach emphasizes the transformations between objects of the same general type. A fundamental lemma in category theory implies that any mathematical object can be characterized by its representations of morphisms to or from other objects of a similar form.

The Recipe For a Category

The contents of a **Category** are as follows:

- A collection of **Objects** $X, Y, Z, \ldots$
- A collection of **Morphisms** $f, g, h, \ldots$

so that

- Each morphism has specified **domain** and **codomain** objects; e.g. $f : X \to Y$ is a morphism with domain X and codomain Y.
- Each object has an **identity morphism** $1_X : X \to X$.
- Any pair of morphisms f, g with the codomain of f equal to the domain of g, there exists a specified **composite morphism** gf whose domain is equal to the domain f and whose codomain is equal to the codomain of g i.e.:

$$f : X \to Y, \quad g : Y \to Z \qquad \rightsquigarrow \qquad gf : X \to Z.$$

Where two axioms must hold:

- For any $f : X \to Y$, the composites $1_Y f$ and $f 1_X$ are both equal to f.
- For any composable triple of morphisms f, g, h, the composites $h(gf)$ and $(hg)f$ are equal (thus denoted hgf).

$$f : X \to Y, \quad g : Y \to Z, \quad h : Z \to W \qquad \rightsquigarrow \qquad hgf : X \to W.$$

Types of Categories

Categories assemble various mathematical objects which can be either *concrete* or *abstract*. Concrete categories have underlying sets as objects and their morphisms as functions between those underlying sets. Examples include:

- **Set** has sets as its objects and functions, with specified domain and codomain, as its morphisms.
- **Top** has topological spaces as its objects and continuous functions as its morphisms.
- **Graph** has graphs as objects and graph morphisms as morphisms.

Other Types of Categories

By contrast, abstract categories do not restrict their morphisms to simply being functions e.g.:

- $\mathbf{Mat}_R$, for a unital ring R, is the category whose objects are positive integers and in the set of morphisms from n to m is the set of $m \times n$ matrices with values in R.
- A poset $(\mathbf{P}, \leq)$ can be treated as a category. The elements of $\mathbf{P}$ are the objects and there exists a unique morphism $x \to y$ if and only if $x \leq y$.

Duality

For any category $\mathcal{C}$, there is the opposite category $\mathcal{C}^{op}$ which has

- the same objects as in $\mathcal{C}$, and
- a morphism f^{op} in $\mathcal{C}^{op}$ for each morphism f in $\mathcal{C}$ where the domain of f^{op} is defined to be the codomain of f and the codomain of f^{op} is defined to be the domain of f i.e.,

$$f^{op} : X \to Y \quad \in \mathcal{C}^{op} \quad \rightsquigarrow \quad f : Y \to X \quad \in \mathcal{C}$$

Any theorem quantifying over "all categories $\mathcal{C}$" then necessarily applies to the opposite category. So any proof in category theory simultaneously proves two theorems, the original statement and its dual.

Example: Below is the category $\mathcal{C}$ with elements A,B,C and morphisms $f : A \to B$ and $g : B \to C$ and its dual $\mathcal{C}^{op}$.

$$\mathcal{C}: \quad A \xrightarrow{f} B \xrightarrow{g} C \qquad\qquad \mathcal{C}^{op}: \quad A \xleftarrow{f^{op}} B \xleftarrow{g^{op}} C$$

Functors and Natural Transformations

A **functor** $F : \mathcal{C} \to \mathcal{D}$, between categories $\mathcal{C}$ and $\mathcal{D}$, consists of:

- An object $Fc \in \mathcal{D}$, for each object $c \in \mathcal{C}$.
- A morphism $Ff : Fc \to Fc' \in D$, for each morphism $f : c \to c' \in \mathcal{C}$, so that the domain and codomain of Ff are equal to F applied to the domain or codomain of f.

A **natural transformation** $\alpha : F \Rightarrow G$ between two functors $F, G : \mathcal{C} \to \mathcal{D}$ is a family of morphisms $\alpha_A : F(A) \to G(A)$, where A is an object in $\mathcal{C}$, such that for every morphism $f : A \to B$ in $\mathcal{C}$, the following diagram commutes:

$$\begin{array}{ccc} F(A) & \xrightarrow{F(f)} & F(B) \\ \downarrow{\scriptstyle \alpha_A} & & \downarrow{\scriptstyle \alpha_B} \\ G(A) & \xrightarrow[G(f)]{} & G(B) \end{array}$$

The Yoneda Lemma

Lemma: Let $\mathcal{C}$ be a locally small category, and let A be an object in $\mathcal{C}$. For any functor $F : \mathcal{C} \to \mathbf{Set}$, there is a natural bijection between the set of all natural transformations $\mathrm{Nat}(\mathrm{Hom}(A, -), F)$ and the set $F(A)$. In other words, we have:

$$\mathrm{Nat}(\mathrm{Hom}(A, -), F) \cong F(A)$$

where $\mathrm{Nat}(\mathrm{Hom}(A, -), F)$ represents the collection of all natural transformations from the covariant hom functor $\mathrm{Hom}(A, -)$ to the functor F.

Applications of Yoneda Lemma

The Yoneda Lemma is a fundamental result in category theory that establishes a powerful connection between representable functors and natural transformations. Here are some significant applications:

- **Use in Category Theory:** Provides a way to characterize the structure of categories and to establish isomorphisms between different categories
- **Functorial Semantics:** Allows us to represent types and operations in a programming language as functors and natural transformations, allowing us to study their behavior and interactions
- **Universal Algebra:** Helps in characterizing and understanding the universal properties of algebraic objects, such as groups, rings, and modules
- **Data Analysis and Machine Learning:** Provides a theoretical foundation for understanding the relationship between data points, features, and transformations
- **Type Theory and Programming Languages:** Helps in understanding the relationship between types and programs, and in the design of expressive and powerful type systems
- **Quantum Information Theory:** Establishes connections between quantum channels and their representations as exclusively positive trace-preserving maps

Acknowledgements

We would like to thank our graduate mentor, Andres Barei, and the DRP community for the great opportunity to learn and delve deeper into some new math! Andres was very helpful through this process and made the experience super enjoyable.

References

[1] Emily Riehl.
Category Theory In Context.
Dover Publications Inc., 2017.

Category Theory and the Snake Lemma

Theodore Bafrali, Jackson Weidmann, Adrian Weitzer. Mentor: Jeremy Brightbill

2023 Mathematics Directed Reading Program.

Categories

A **category** $\mathscr{C}$ is the following data:

(i) A collection $\mathrm{Ob}(\mathscr{C})$, called **objects** of $\mathscr{C}$.

(ii) For any two objects A, B of $\mathscr{C}$, a collection $\mathrm{Hom}_{\mathscr{C}}(A, B)$ of **morphisms/arrows** f from A to B, denoted $A \to B$, $f : A \to B$, or $A \xrightarrow{f} B$.

(iii) A **composition law** $\circ : \mathrm{Hom}_{\mathscr{C}}(A, B) \times \mathrm{Hom}_{\mathscr{C}}(B, C) \to \mathrm{Hom}_{\mathscr{C}}(A, C)$, written as $\circ(f, g) = g \circ f$ or gf. Additionally $\circ$ must satisfy

$$h \circ (g \circ f) = (h \circ g) \circ f$$

whenever all the compositions are defined.

(iv) For every object A of $\mathscr{C}$ there exists a morphism $1_A \in \mathrm{Hom}_{\mathscr{C}}(A, A)$ called the **identity (on A)** satisfying:

$$1_A \circ f = f, \qquad g \circ 1_A = g$$

for all morphisms $f \in \mathrm{Hom}_{\mathscr{C}}(B, A)$ and $g \in \mathrm{Hom}_{\mathscr{C}}(A, C)$.

A collection of morphisms, e.g., $f : A \to B$, $g : A \to C$, $h : C \to D$, and $k : B \to D$, may be arranged in a diagram:

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle k} \\ C & \xrightarrow{h} & D \end{array}$$

Fig. 1: Diagram

We say this diagram **commutes** if $k \circ f = h \circ g$.

Some example of categories:

(i) **Top** The category of topological spaces and continuous maps.

(ii) R-**Mod** The category of left R modules and R-homomorphisms.

(ii) Define a category $\mathscr{C}$ with objects the positive integers and

$$\mathrm{Hom}_{\mathscr{C}}(n, m) = \begin{cases} \{*\} & \text{if } n \mid m \\ \varnothing & \text{otherwise} \end{cases}$$

(Co)Limits

(i) Given an indexing set I and a family of objects $(A_i)_{i\in I}$ in $\mathscr{C}$, A **product** of $(A_i)_{i\in I}$ is a pair $(\prod A_i, (\pi_i)_{i\in I})$, where $\prod A_i \in \mathrm{Ob}(\mathscr{C})$ and $\pi_i \in \mathrm{Hom}_{\mathscr{C}}(\prod A_i, A_i)$ for every $i \in I$, and such that if $(T, (t_i)_{i\in I})$ is another pair also satisfying these conditions, there exists a unique morphism $r : T \to \prod A_i$ making Figure 2 commute for every $i \in I$.

(ii) Given the diagram

$$Y \xrightarrow{g} Z \xleftarrow{f} X$$

A pullback is an object P and morphisms $\tilde{f} : P \to Z, \tilde{g} : P \to Y$ such that $f \circ \tilde{f} = g \circ \tilde{g}$, and for any T and f_t, g_t with the same properties there is a unique $\varphi \in \mathrm{Hom}_{\mathscr{C}}(T, P)$ such that Figure 4 commutes.

These are special cases of "categorical limits" and they recover important constructions in many disciplines. In **Top** the product is the product space. In the category $\mathscr{C}$, the product of the integers $\{n_i\}$ is

$$\prod_{i\in I} n_i = \mathrm{GCD}(n_i)$$

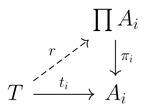


Fig. 2: Product

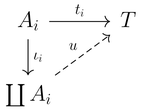


Fig. 3: Coproduct

We can dualize(reverse all the arrows) in the above definitions and get examples of "categorical colimits":

(i) A **coproduct** of $(A_i)_{i\in I}$ is a pair $(\coprod A_i, (\iota_i)_{i\in I})$ with $\coprod A_i$ an object and $\iota_i : A_i \to \coprod A_i$ morphisms such that for any other pair $(T, (t_i)_{i\in I})$ of an object T and morphisms $t_i : A_i \to T$, there exists a unique morphism $u : C \to T$ making Figure 3 commute for every $i \in I$.

(ii) Given the diagram

$$Y \xleftarrow{g} Z \xrightarrow{f} X$$

A pushout is an object P and morphisms $\tilde{f} : P \to Z, \tilde{g} : P \to Y$ such that $\tilde{f} \circ f = \tilde{g} \circ g$, and for any T and f_t, g_t with the same properties there is a unique $\varphi \in \mathrm{Hom}_{\mathscr{C}}(P, T)$ such that Figure 5 commutes.

Going back to our examples in **Top**, the coproduct is the disjoint union with the disjoint union topology, and in the category $\mathscr{C}$,

$$\coprod_{i\in I} n_i = \mathrm{LCM}(n_i)$$

When a (co)limit exists, it is unique up to unique isomorphism.

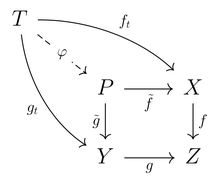


Fig. 4: Pullback

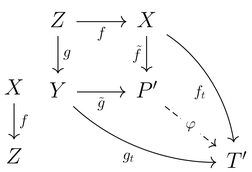


Fig. 5: Pushout

Abelian Categories

We can also consider categories that have more structure, with an addition defined on the morphism sets and (co)kernels with nice properties. First define

(i) A **zero object** for $\mathscr{C}$, denoted by 0, is an object such that for every object X of $\mathscr{C}$ there are unique morphisms $X \to 0$ and $0 \to X$. Similarly, the **zero morphism** $0_{XY} : X \to Y$ is the unique morphism $X \to 0 \to Y$.

(ii) A **kernel** of a morphism $f : X \to Y$ is a pullback of $X \xrightarrow{f} Y \xleftarrow{0} 0$

(iii) A **cokernel** of a morphism $f : X \to Y$ is a pushout of $Y \xleftarrow{f} X \xrightarrow{0} 0$

(iv) An **image** of a morphism $f : A \to B$ is a kernel of the cokernel of f.

(v) A **coimage** of a morphism $f : A \to B$ is a cokernel of the kernel of f.

An **Abelian category** is defined by requiring a category satisfy a sequence of conditions:

(i) A **preadditive category** is a category $\mathscr{C}$ such that the morphism sets are abelian groups under an operation $+$ and for all morphisms $f, f_1 \in \mathrm{Hom}_{\mathscr{C}}(A, B)$, $h, h_1 \in \mathrm{Hom}_{\mathscr{C}}(C, A)$

(a) $(f + f_1) \circ h = (f \circ h) + (f_1 \circ h)$

(b) $f \circ (h + h_1) = (f \circ h) + (f \circ h_1)$.

(ii) An **additive category** is a preadditive category with a zero object in which every nonempty, finite collection of objects admits a product.

(iii) An **abelian category** is an additive category where every morphism f has a kernel and cokernel and there is a unique isomorphism θ such that $\mathrm{im}(f) \circ \theta \circ \mathrm{coim}(f) = f$

In an abelian category, a morphism f is called an epimorphism/epi(resp. monomorphism/mono) if $\mathrm{coker}\, f = 0$ (resp., $\ker(f) = 0$. Epis/Monos are denoted by $\twoheadrightarrow/\hookrightarrow$.

Within an abelian category $\mathscr{A}$ consider the sequence of objects:

$$0 \longrightarrow A \xrightarrow{f} B \xrightarrow{g} C \longrightarrow 0$$

Fig. 6: Exact Sequence

This sequence is called **exact at B** if $\ker(g) = \mathrm{im}(f)$. It is called a **short exact sequence** if additionally f is a monomorphism and g is an epimorphism.

A longer diagram is called **exact** if it is exact at every position.

The Snake Lemma

Theorem.(The Snake Lemma) In an abelian category $\mathscr{A}$ if the diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A_1 & \xhookrightarrow{f_1} & B_1 & \xrightarrow{g_1} & C_1 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \\ 0 & \longrightarrow & A_2 & \xhookrightarrow{f_2} & B_2 & \xrightarrow{g_2} & C_2 & \longrightarrow & 0 \end{array}$$

commutes and has exact rows, there is a connecting morphism $\delta : \ker(\gamma) \to \mathrm{coker}(\beta)$ giving the following exact sequence

$$0 \longrightarrow \ker(\alpha) \longrightarrow \ker(\beta) \longrightarrow \ker(\gamma) \xrightarrow{\delta} \mathrm{coker}(\alpha) \longrightarrow \mathrm{coker}(\beta) \longrightarrow \mathrm{coker}(\gamma) \longrightarrow 0$$

Schanuel's Lemma. The snake lemma provides a quick proof of many different diagram lemmas.

First we introduce the concept of a **projective object** in an abelian category $\mathscr{A}$. There are two equivalent definitions for projective P. Let N, X be any objects in $\mathscr{A}$.

- If we have an epi, h, there exists f' such that Figure 9 commutes
- If we have an exact sequence as in Figure 10, then $N \simeq X \oplus P$ ($\oplus$ denotes product/coproduct, which coincide in abelian categories).

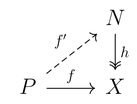


Fig. 9:

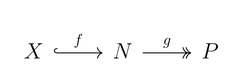


Fig. 10:

Schanuel's lemma states that if we have two exact sequences like below, then $K \oplus P' \simeq K' \oplus P$.

$$K' \xhookrightarrow{f'} P' \xrightarrow{g'} \!\!\!\!\rightarrow M, \qquad K \xhookrightarrow{f} P \xrightarrow{g} \!\!\!\!\rightarrow M$$

We can construct the second row of this diagram to be exact using the properties of the product. We can then find the map $K \oplus P' \to P$ using the properties of the product and projective modules. The snake lemma connects the top and bottom rows to form an exact sequence which gives that $\ker(h) = K$ and $\mathrm{coker}\, h = 0$.

$$\begin{array}{ccccccccc} & & 0 & \longrightarrow & \ker(h) & \longrightarrow & K & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & K & \xrightarrow{\iota_K} & K \oplus P' & \xrightarrow{\pi_P} & P' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle 1} & & \downarrow{\scriptstyle h} & \nearrow{\scriptstyle \tilde{g}} & \downarrow{\scriptstyle g'} & & \\ 0 & \longrightarrow & K & \xrightarrow[f]{} & P & \xrightarrow[g]{} \!\!\!\!\rightarrow & M & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & 0 & \longrightarrow & \mathrm{coker}(h) & \longrightarrow & 0 & & \end{array}$$

Third Isomorphism Theorem This result is often seen in group theory, but it also applies more generally to any abelian category. The theorem states that given groups $T \subset S \subset M$ with T, S normal in M, we get an isomorphism $(M/T)/(S/T) \simeq (M/S)$.

In an abelian category $\mathscr{A}$, suppose we have M, S, T and $\sigma \circ \eta = \tau$ (note $\tau = \tau \circ 1$), from the diagram below. As mentioned earlier we can create the quotients M/T and S/T which makes the middle two rows of the diagram exact. The properties of cokernels give us a map $S/T \to M/T$, and we have the setup for the snake lemma. The exact sequence given by the snake lemma shows that $K = 0$ and $M/S \simeq (M/T)/(S/T)$.

$$\begin{array}{ccccccccc} & & 0 & \longrightarrow & 0 & \longrightarrow & K & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & T & \xhookrightarrow[\eta]{} & S & \xrightarrow{} \!\!\!\!\rightarrow & S/T & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle 1} & & \downarrow{\scriptstyle \sigma} & & \downarrow & & \\ 0 & \longrightarrow & T & \xhookrightarrow[\tau]{} & M & \xrightarrow{} \!\!\!\!\rightarrow & M/T & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & 0 & \longrightarrow & M/S & \longrightarrow & \frac{(M/T)}{(S/T)} & \longrightarrow & 0 \end{array}$$

References

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

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

CONGRUENCE RELATIONS AND FERMAT'S LITTLE THEOREM

Tingyu Fu, Chang Long
University of California Santa Barbara

Linear Equations

Given two whole numbers a and b, we are going to look at all the possible numbers we can get by adding a multiple of a to a multiple of b. In other words, we will consider all numbers obtained from the formula:

$$ax + by$$

Every number of the form $ax + by$ is divisible by $gcd(a, b)$, since both a and b are divisible by $gcd(a, b)$. In Conclusion, the smallest positive value of $ax+by$ is equal to $gcd(a, b)$.
We will use the **Euclidean algorithm** to find integers x and y that are solutions to the equation

$$ax + by = gcd(a, b)$$

We are going to start by illustrating the Euclidean algorithm method for solving $ax + by = gcd(a, b)$ with a particular example:

$$22x + 60y = gcd(22, 60)$$

$$\begin{aligned} 60 &= 2 \times 22 + 16 \\ 22 &= 1 \times 16 + 6 \\ 16 &= 2 \times 6 + 4 \\ 6 &= 1 \times 4 + 2 \\ 4 &= 2 \times 2 + 0 \end{aligned}$$

We can summarize the above computation in the following efficient tabular form. Note that the left-hand equations are the Euclidean algorithm, and the righthand equations compute the solution to $ax + by = gcd(a, b)$.

$$\begin{aligned} a &= 2 \times b + 16 & 16 &= a - 2b \\ b &= 1 \times 16 + 6 & 6 &= b - 1 \times 16 \\ & & &= b - 1 \times (a - 2b) \\ & & &= -a + 3b \\ 16 &= 2 \times 6 + 4 & 4 &= 16 - 2 \times 6 \\ & & &= (a - 2b) - 2 \times (-a + 3b) \\ & & &= 3a - 8b \\ 6 &= 1 \times 4 + 2 & 2 &= 6 - 1 \times 4 \\ & & &= (-a + 3b) - 1 \times (3a - 8b) \\ & & &= -4a + 11b \\ 4 &= 2 \times 2 + 0 \end{aligned}$$

Eventually, we get down to the last nonzero remainder, which we know is equal to $gcd(a, b)$, and this gives the desired solution to $gcd(a, b) = ax + by$. This process is summarized in the **Linear Equation Theorem**. For nonzero integers a and b, and let $g = gcd(a, m)$. The equation

$$ax + by = g$$

always has a solution (x_0, y_0) in integers that can be found by using the **Euclidean Algorithm** .

Extension to Linear Equation Theorem

Given the Linear Equation Theorem, we may want to figure out that, under what conditions on a,b,c, the equation

$$ax + by + cz = 1$$

has a solution. Suppose that (x_0, y_0) is a solution to

$$ax + by = gcd(a, b)$$

Then by the Linear Equation Theorem, there exist a solution (w_0, z_0) to

$$gcd(a, b)w + cz = gcd(gcd(a, b), c) = gcd(a, b, c)$$

Hence, there exist a solution (x_0, y_0, z_0) to

$$axw + byw + cz = gcd(a, b, c)$$

So the equation

$$ax + by + cz = 1$$

always has a solution if $gcd(a, b, c) = 1$.

Fermat's Little Theorem

Fermat's Little Theorem
Let p be a prime number, and let a be any number with $a \not\equiv 0 (mod\, p)$. Then

$$a^{p-1} \equiv 1 (mod\, p).$$

Before giving the proof of Fermat's Little Theorem, we want to indicate its power and show how it can be used to simplify computations. As a particular example, consider the congruence

$$6^{22} \equiv 1 (mod\, 23).$$

This says that the number 6^{22-1} is a multiple of 23. If we wanted to check this fact without using Fermat's Little Theorem, we would have to multiply out 6^{22}, subtract 1, and divide by 23. Here's what we get:

$$6^{22-1} \equiv 23 \times 5722682775750745$$

Now we are ready to prove the theorem. First, we may want to introduce a lemma that helps us with the proof. However, we will skip the verification of the lemma due to space constraint.
Lemma *Let p be a prime number and let a be a number with a ≢ 0 (mod p). Then the numbers*

$$a, 2a, 3a, ..., (p-1)a \ (mod\, p)$$

are same as the numbers

$$1, 2, 3, ..., (p-1) \ (mod\, p),$$

although they may be in a different order.
Using the lemma, it is easy to finish the proof of Fermat's Little Theorem. By the lemma, we know that $a, 2a, 3a, ..., (p-1)a \ (mod\, p)$ and $1, 2, 3, ..., (p-1) \ (mod\, p)$, are the same. Thus,

$$a \cdot (2a) \cdot (3a) \cdots ((p-1)a) \equiv 1 \cdot 2 \cdot 3 \cdots (p-1) (mod\, p).$$

Now if we factor our $p - 1$ copies of a from the left-hand side of the equation, we will get

$$a^{p-1} \cdot (p-1)! \equiv (p-1)! (mod\, p).$$

Since $(p-1)!$ is relatively prime to p, we are able to cancel it from both sides of the equation to obtain Fermat's Little Theorem,

$$a^{p-1} \equiv 1 (mod\, p).$$

Congruences

We say that a is congruent to b modulo m, and we write $a \equiv b (mod\, m)$. Before giving the general theory, let's try an example. We will solve the congruence

$$18x \equiv 8 (mod\, 22)$$

This means we need to find a value of x with 22 dividing $18x - 8$, so we have to find a value of x with $18x - 8 = 22y$ for some y. In other words, we need to solve the linear equation

$$18x - 22y = 8$$

Using the **Linear Equation Theorem** we can solve the equation

$$18u - 22v = gcd(18, 22) = 2,$$

and indeed we easily find the solution $u = 5$ and $v = 4$. But we really want the right-hand side to equal 8, so we multiply by 4 to get

$$18 \times (5 \times 4) - 22 \times (4 \times 4) = 8$$

Thus, $18 \times 20 \equiv 8 (mod\, 22)$, so $x \equiv 20 (mod\, 22)$ is a solution to the original congruence. We will soon see that this congruence has two different solutions modulo 22; the other one turns out to be $x \equiv 9 (mod\, 22)$. Suppose now that we are asked to solve an arbitrary congruence of the form

$$ax \equiv c (mod\, m)$$

Linear Congruence Theorem
Now we generalize the solution from the example above to solve an arbitrary congruence of the form

$$ax \equiv c (mod\, m)$$

where a,c m are integers with $m \geqslant 1$, and let $g = gcd(a, m)$. Rearranging this equation, we find that $ax \equiv c (mod\, m)$ if and only if the linear equation $ax - my = c$ has a solution. By the Linear Equation Theorem, we know that there is always a solution to the equation

$$au + mv = g$$

Suppose g divides c, if we multiply this equation by the integer $\frac{c}{g}$ to obtain

$$a\left(\frac{cu}{g}\right) + m\left(\frac{cv}{g}\right) = c$$

Thus, $x_0 \equiv \frac{cu}{g} (mod\, m)$ is a solution to the congruence $ax \equiv c (mod\, m)$ and

$$x \equiv x_0 + k \cdot \frac{m}{g}$$

where $k = 0, 1, ..., g - 1$.
If g does not divide c, however, then the congruence $ax \equiv c (mod\, m)$ has no solutions.

Acknowledgements

We would like to thank Mychelle Parker for her guidance as well as the UCSB Directed Reading Program for the opportunity to work on this project.[1]

References

[1] Joseph H. Silverman. *A Friendly Introduction to Number Theory*. Providence, RI: Pearson Education, Inc., 2012.

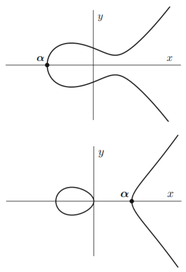
FOUR TIMES ELLIPTIC CURVES

Bonny Wang, Yaoyu Li, and Tagore Zhao
University of California Santa Barbara

Introduction

A (rational) elliptic curve is a curve given by points (x, y) in the plane satisfying an equation

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

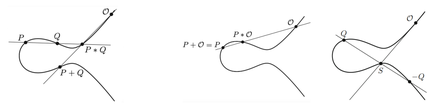


that is typically in one of the two shapes. The study of elliptic curves has garnered significant attention for its involvement in crucial theorems, notably including the proof of Fermat's Last Theorem, a renowned problem unsolved within 350 years since proposed. Additionally, the inherent fascinating properties of elliptic curves make them an intriguing subject for investigation in their own right.

In this expository writing, we focus on the unique features that make elliptic curves stand out-the **group structure** on elliptic curves. Simply put, we would like to find a reasonable way to 'add' two points on an elliptic curve to get another, in such a way that this addition satisfies all desired properties such as associativity. Let C denote an elliptic curve.

A geometric approach

Fix a point O on C. Starting with two points P and Q, let P be intersection of line $\overline{PQ}$ and C. Define $P+Q = O*(P*Q)$. If $P = Q$, we can take the tangent line of C at P in place of $\overline{PQ}$.



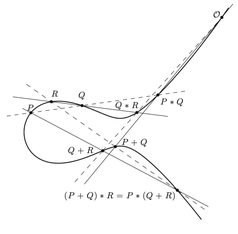
Now, we show the above addition satisfies the following properties.

Identity element: Identity element is O, with $P + O = P$ for all P.

Inverse element: For any Q on C, let $-Q = Q*(O*O)$. Then, $Q+(-Q) = O$.

Commutativity: $P + Q = Q + P$ since line $\overline{PQ}$ = $\overline{QP}$.

Associativity: Let P, Q, and R be three points on the curve. We want to prove that $(P + Q) + R = P + (Q + R)$. By definition, we only need to prove $(P + Q) * R = P * (Q + R)$. In the figure, there are 10 points O, P, Q, R, $P * Q$, $P + Q$, $Q * R$, $Q + R$, $P * (Q + R)$ and $(P + Q) * R$, each of the first 8 lying on one of the dashed lines and one of the solid lines by construction, and our goal is to show the 9-th point agrees with the 10-th. We can regard C_1, the union of the three dashed lines, and C_2, the union of the three solid lines, both as cubic curves since their equations are given by products of 3 linear equations. Now, the cubic curve C intersects both C_1 and C_2 at the 8 points $O, P, Q, R, P * Q, P + Q, Q * R, Q + R$. A theorem of Bèzout then says that the 9-th intersection of C and C_1 must agree with the 9-th intersection of C and C_2, that is, $(P + Q) * R = P * (Q + R)$.

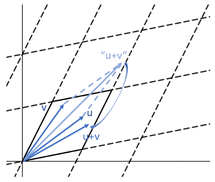


An analytic approach

The goal is to define the addition on the ellipse curve. Instead of looking at the curve directly, we first look at the addition on a parallelogram at the origin.

Addition on a parallelogram

For points u and v inside a parallelogram, we have the usual vectors addition. But if $u + v$ is out of this parallelogram, we need to find a new $u + v$ inside this parallelogram. We consider all parallelograms in this plane as equavelent by allowing translation. In this way we can find a new $u + v$ in the origin parallelograms which is a shift of the "$u + v$" outside the parallelogram.

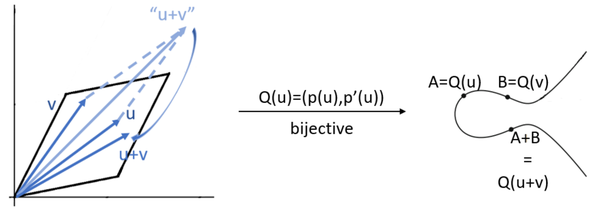


Mapping parallelogram to elliptic curve

Consider the function: $\wp : \mathbb{C} \longrightarrow \mathbb{C}$ by $\wp(u) = \frac{1}{u^2} + \sum_{\omega \in L}(\frac{1}{(u-\omega)^2} - \frac{1}{\omega^2})$, where L is the set of endpoints of all parallelograms except the origin. Note that if we replace u by a correpsonding point in a different parallelogram, we will get the same summation(just shifting the summands)! Thus the values of $\wp$ on all parallelograms agree and we can consider it as a function from one parallelogram to $\mathbb{C}$. Computation yields an differential equation: $\wp'(u)^2 = 4\wp(u)^3 - g_2\wp(u) - g_3$ for any $u \in \mathbb{C}$. Thus every complex number u in the parallelograms has a corrosponding point $Q(u) = (\wp(u), \wp'(u))$ on a curve $y^2 = 4x^3 - g_2x - g_3$ which is essentially an elliptic curve.

Defining the addition on elliptic curve

It turns out that the map Q is onto and one to one. There is then a natural addition on ellipse curve P coming from the one on a parallelogram: For points $A = Q(u), B = Q(v)$ correpsonding to u, v, simply define $A + B$ to be $Q(u + v)$!



An algebraic approach

As before, we will first consider a seemingly irrelevant group that will lead to a solution eventually. Note that all we need is to find a bijection from the points on C to something that is already a group!

The divisor group Div

We define a divisor group $DivC$ consisting of elements of the form

$$D = a_1P_1 + ... + a_mP_m$$

where $P_i \in C$ with integer coefficients a_i. Obviously, divisors can be added and subtracted. Let $D, D' \in DivC, D' = a_1'P_1 + ... + a_m'P_m$. We can easily conclude that

$$D + D' = (a_1 + a_1')P_1 + ... + (a_m + a_m')P_m = D' + D$$

The group Pic^0

However, $DivC$ is too big for our problem-solving; we need a smaller group. We identify D and D' if they differ by "plus or minus zeros counted with multiplicities" of a polynomial. For example, consider the polynomial $f = (x - y)^2(x - 2y)$. Let P_1, P_2, P_3 be points on C satisfies $x - y = 0$, Q_1, Q_2, Q_3 be points on C satisfies $x - 2y = 0$. If D and D' differ with $2(P_1 + P_2 + P_3) + Q_1 + Q_2 + Q_3$, then they are equivalent. It turns out that the equivalent classes of $D = a_1P_1 + ... + a_mP_m$ satisfying $a_1 + ... + a_n = 0$ is the correct thing. We call it Pic^0C. Let $P_0 \in C$ be a point. Then the map

$$C \to Pic^0C, \quad P \mapsto P - P_0$$

can be shown to be a bijection.

Rational points on elliptic curves

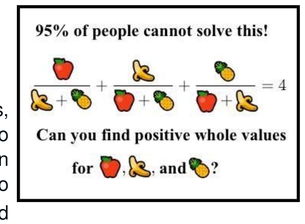
If a point has rational x-coordinate and y-coordinate, we call it a **rational** point. **We mention here that all three approaches before essentially define the same addition on elliptic curves.** In what follows we take the geometric approach. If we start from P and Q that are both rational, to find the x-coordinate of $P * Q$, we need to solve a cubic equation with rational coefficients, obtained by solving the system of equations for line $\overline{PQ}$ and C(both with rational coefficients). Since P and Q has rational x-coordinates, by Vieta theorem, $P * Q$ must have rational x-coordinate. Since $P * Q$ is on a rational line, it must have a rational y-coordinate, too. Thus $P * Q$ is rational. Similarly, $P + Q = (P * Q) * O$ must be rational .

Upshot: if P,Q are rational, so is $P + Q$.

An arithmetic application

Can you find positive integers a, b, c such that

$$\frac{a}{b+c} + \frac{b}{c+a} + \frac{c}{a+b} = 4?$$



It is not complex to find a solution which are integers, such as $a = 4, b = -1, c = 11$. However, it is hard to find a positive integer solution by simple calculation or even using computers. In this case, we need to use some other methods. Let $x = \frac{-28(a+b+2c)}{6a+6b-c}$ and $y = \frac{364(a-b)}{6a+6b-c}$.

We can show that x and y satisfy an equation $y^2 = x^3 + 109x^2 + 224x$, which is an elliptic curve. In fact, we can also recover (a, b, c) from (x, y):

$$a = \frac{56 - x + y}{56 - 14x}, b = \frac{56 - x - y}{56 - 14x}, c = \frac{-28 - 6x}{28 - 7x}$$

Through the equation above, we can see that there exists a bijection between the set of rational points on the curve and the (rational) solution set $\{a, b, c\}$. By observation, we can find a solution $a = 4, b = -1, c = 11$ which are integers, but they are not all positive. From this solution, we can find a rational point P on the elliptic curve. Now by addition defined in the previous sections, we can find other rational points $2P, 3P$, and so on with their corresponding a, b, c. In some cases, the a, b, c might be rational but not integers. In these cases, we can multiply a, b, c with $(56 - 14x)$. Then, they will be integers. Now for $2P, 3P, ...$, we compute their corresponding a, b, c and ask if they are all positive with the help of computer programs, until we eventually find the a, b, c corresponding to $9P$ that are all positive: $(a, b, c) =$

$(154476802108746166441951315019919837485664325669565431700026634898253202035277999,$
$368751317941299998271978115652254748254929799689719709962831374716372246340555 79,$
$437361267792869725786125260237139015281653755816161361862143799337842346777 2036)$

Acknowledgements

We extend our heartfelt gratitude to our mentor, Mulun Yin, for his invaluable guidance and steadfast support throughout our project. Additionally, we want to thanks to the UCSB DRP program for giving us this opportunity to learn and showcase our work.

References

Joseph H. Silverman, John T Tate(2005), Rational Points on Elliptic Curves-Springer
Andreas Gathmann(2002/2003), Algebraic Geometry -University of Kaiserslautern
https://www.quora.com/How-do-you-find-the-positive-integer-solutions-to-frac-x-y+z-+-frac-y-z+x-+-frac-z-x+y-4/answer/Alon-Amit

Geometry and Lie Theory of PSL(2,$\mathbb{R}$)

Ethan Mader, Clayton Ellis
University of California Santa Barbara

What is PSL(2, $\mathbb{R}$)?

Let us start by defining $SL(2,\mathbb{R})$, the matrix group representing all real 2x2 matrices with determinant 1. Note that the determinant of a matrix product is the product of the determinants, satisfying closure.

$$SL(2,\mathbb{R}) \coloneqq \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a,b,c,d \in \mathbb{R}, ad-bc=1 \right\}.$$

$PSL(2,\mathbb{R})$ is simply $SL(2,\mathbb{R})/\{\pm I\}$. In this way, $SL(2,\mathbb{R})$ can be thought of as two copies of $PSL(2,\mathbb{R})$, so we can learn about $PSL(2,\mathbb{R})$ by studying $SL(2,\mathbb{R})$. Consider the characteristic polynomial of $A \in SL(2,\mathbb{R})$:
$det(A-\lambda I) = (a-\lambda)(d-\lambda) - bc = \lambda^2 - (a+d)\lambda + ad - bc = \lambda^2 - tr(A)\lambda + 1$, which has roots

$$\lambda = \frac{tr(A) \pm \sqrt{tr(A)^2 - 4}}{2}.$$

Note that the discriminant is imaginary when $|tr(A)| < 2$, 0 when $|tr(A)| = 2$, and positive real when $|tr(A)| > 2$. These cases describe a classification of $SL(2,\mathbb{R})$ as transformations on $\mathbb{R}^2$. Quotienting $SL(2,\mathbb{R})$ by $\{\pm I\}$ will give us a convenient relationship between $PSL(2,\mathbb{R})$ and isometries of the hyperbolic plane $\mathbb{H}^2$.

Relation to Hyperbolic Geometry

Let us use Poincaré's upper half-plane model to represent $\mathbb{H}^2$, which has space $\mathbb{U}$ and transformation group $\mathcal{U}$. $\mathbb{U} = \{z \in \mathbb{C} | \operatorname{Im}(z) > 0\}$, and this has metric $ds^2 = (dx^2+dy^2)/y^2$ with ideal points on the x-axis and at ∞. The group $\mathcal{U}$ is all Möbius transformations T such that $T(\mathbb{U}) = \mathbb{U}$, called an isometry. We can associate every $A \in PSL(2,\mathbb{R})$ with an orientation-preserving Möbius transformation:

$$T_A(z) \mapsto \frac{az+b}{cz+d} \in \mathcal{U}.$$

To classify these, let us consider the following subsets of $SL(2,\mathbb{R})$:

$$\left\{ \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \middle| \theta \in \mathbb{R} \right\}, \left\{ \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \middle| t \in \mathbb{R} \right\}, \left\{ \begin{bmatrix} r & 0 \\ 0 & \frac{1}{r} \end{bmatrix} \middle| r \in \mathbb{R}, r > 0 \right\}.$$

$SL(2,\mathbb{R})$ can actually be expressed as a product of these three sets. Disregarding the identity, these subsets are distinct and correspond to the three types of isometries of $\mathbb{U}$, which are elliptic, parabolic, and hyperbolic, respectively.
An element from each is visualized to the right, with bolded lines denoting orbits of the corresponding transformation on $\mathbb{U}$.
Elliptic Möbius transformations are characterized by $tr < 2$ and have one interior fixed point. In Fig 1, this point is i, and the orbits are circular.
Parabolic Möbius transformations are characterized by $tr = 2$ and have one boundary fixed point. In Fig 2, this point is ∞, and the orbits are horizontal.
Hyperbolic Möbius transformations are characterized by $tr > 2$ and have two boundary fixed points. In Fig 3, these points are 0 and ∞, and the orbits are radial.
It turns out that every element of $PSL(2,\mathbb{R})$ is conjugate to a member of one of the three sets. $PSL(2,\mathbb{R})$ is a useful tool for studying $\mathbb{H}^2$ as it represents precisely its set of isometries.

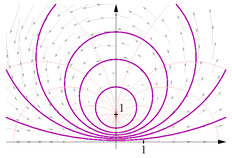


Figure 1: Elliptic

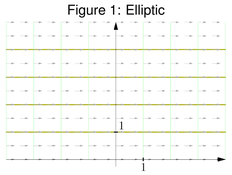


Figure 2: Parabolic

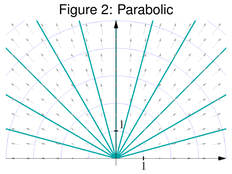


Figure 3: Hyperbolic

Lie Algebra of PSL(2, $\mathbb{R}$)

We can also learn about $PSL(2,\mathbb{R})$ by investigating the Lie algebra of $SL(2,\mathbb{R})$.
Let $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$. The Lie algebra of a matrix group $G \subseteq GL_n(\mathbb{K})$, where $GL_n(\mathbb{K})$ is the group of invertible n by n matrices, is the tangent space to G at I, denoted

$$\mathfrak{g} \coloneqq \mathfrak{g}(G) \coloneqq T_I(G) \coloneqq \{\gamma'(0) | \gamma : (-\epsilon,\epsilon) \to G \text{ differentiable and } \gamma(0) = I\}.$$

We can think of a matrix as living in $\mathbb{K}^{n^2}$, where the tangent vector at any point is also matrix. In this way, we can view matrix groups as manifolds, in which the dimension of the Lie algebra defines the dimension of the manifold which is equivalent to that of the matrix group. In the case of $SL(2,\mathbb{R})$, the Lie algebra is denoted $sl(2,\mathbb{R})$ and is the set of all traceless 2x2 matrices. To see why, we can prove that for any differentiable $\gamma : (-\epsilon,\epsilon) \to M_n(\mathbb{K})$ where $\gamma(0) = I$, we have

$$\left.\frac{d}{dt}\right|_{t=0} det(\gamma(t)) = tr(\gamma'(0)).$$

For any $\gamma'(0) \in sl(2,\mathbb{R})$, we know $\gamma(t) \in SL(2,\mathbb{R})$ which means $det(\gamma(t)) = 1$ and subsequently has derivative 0, so $tr(\gamma'(0)) = 0$. Now we have

$sl(2,\mathbb{R}) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \middle| a,b,c,d \in \mathbb{R}, a+d=0 \right\}$, which has basis $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\}$.

Thus we can think of $SL(2,\mathbb{R})$ as a 3-dimensional manifold which is embedded in $\mathbb{R}^4$.
Another useful way to explore the relationship between matrix groups and their Lie algebra is through matrix exponentiation, which is defined for all $A \in M_n(\mathbb{K})$ by

$$e^A = I + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \ldots = \sum_{k=0}^{\infty} \frac{1}{k!}A^k.$$

It turns out that when applied to an element of a Lie algebra, this exponential map gives us an element of the corresponding matrix group. This also solidifies the connection between addition in a Lie algebra and multiplication in the matrix group, since as one might expect $e^{A+B} = e^A e^B$ for all $A, B \in M_n(\mathbb{K})$.
One can also prove that $det(e^A) = e^{tr(A)}$ for all $A \in M_n(\mathbb{K})$, verifying that the exponential map sends elements $A \in sl(2,\mathbb{R})$, which have $tr(A) = 0$, to elements $e^A \in SL(2,\mathbb{R})$ with unit determinant.
A Lie algebra $\mathfrak{g}$ also comes with an additional operation to the vector space structure. This operation is called the Lie bracket which is an $\mathbb{R}$-bilinear form denoted

$$[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}, \text{ where } [A,B] = AB - BA \text{ for all} A, B \in \mathfrak{g}.$$

Note that $tr[A,B] = tr(AB-BA) = tr(AB) - tr(BA) = 0$, which verifies that the Lie bracket sends two elements of $sl(2,\mathbb{R})$ to $sl(2,\mathbb{R})$.
Moving from $SL(2,\mathbb{R})$ to $PSL(2,\mathbb{R})$, notice that we have an immersive inclusion mapping f between $PSL(2,\mathbb{R})$ and $SL(2,\mathbb{R})$ so for any $\gamma : (-\epsilon,\epsilon) \to SL(2,\mathbb{R})$ satisfying $\gamma(0) = I$, we can generate $psl(2,\mathbb{R})$ by considering the elements $(f \circ \gamma)'(0)$ to generate $psl(2,\mathbb{R})$. Through the same means as to determine the structure of $sl(2,\mathbb{R})$, we can find that composing γ and f does not impact the contents of the lie algebra described so we find $sl(2,\mathbb{R}) = psl(2,\mathbb{R})$ so in general, we just consider $sl(2,\mathbb{R})$.

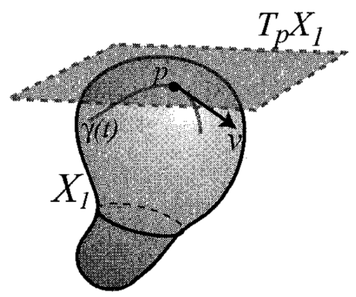
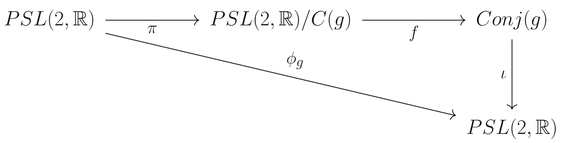


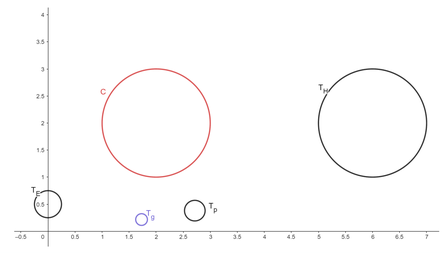
Figure 4: Visual of a manifold

Conjugacy Class Structure

In order to describe the conjugacy classes of $PSL(2,\mathbb{R})$, we first must define a few maps. First, we pick any $g \in PSL(2,\mathbb{R})$ to be the group element whose conjugacy classes we are looking at. We define the first map to be $\pi : PSL(2,\mathbb{R}) \to PSL(2,\mathbb{R})/C(g)$, where $\pi(A) = A \cdot C(g)$ to be the typical coset mapping of $PSL(2,\mathbb{R})$ to its quotient group. Next, we note that we have a bijection between our quotient group and the set of the conjugacy classes of g, $Conj(g)$. In order to make sense of what this means, we first note that as $C(g)$ is closed, the Closed-Subgroup Theorem tells us that $PSL(2,\mathbb{R})/C(g)$ is in fact a manifold as well so due to the existence of this bijection, we also have a diffeomorphism f between these two spaces and so $Conj(g)$ is both a manifold and has the same dimension as our quotient group. It is important to note though that when looking at the images of f and ϕ_g, we will need to choose representatives of each of these cosets in the quotient group in order to ensure our maps are well-defined.



Now, we investigate the differential map $d\phi_g : sl(2,\mathbb{R}) \to sl(2,\mathbb{R}) \cdot g$. The kernel of this map will tell us exactly how many fewer dimensions the image of ϕ_g is than its domain. As it turns out, this kernel is exactly the centralizer of g which we already know to be one dimensional so then the image of $d\phi_g$, and thus that of ϕ_g, is two dimensional so ϕ_g is a submersion into a two dimensional submanifold of $PSL(2,\mathbb{R})$. As f is known to not be a subersion and ι is an immersion, it follows that π must be a submersion thus $PSL(2,\mathbb{R})/C(g) \cong Conj(g)$ must be two-dimensional as ϕ_g is just the composition of our other three maps.
As an example, we can look at the picture below to see how these maps transform the circle C centered at $2+2i \in \mathbb{U}$ first by the Möbius transformation that can be represented by the matrix $g = \begin{bmatrix} 2 & 0 \\ 1 & \frac{1}{2} \end{bmatrix}$ which has its image represented by the curve T_g. Furthermore, we also have the images of C under the conjugation of the elliptic element $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ denoted T_E, the hyperbolic element $\begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$ denoted T_H, and finally the parabolic element $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ with image denoted T_P.



Acknowledgements

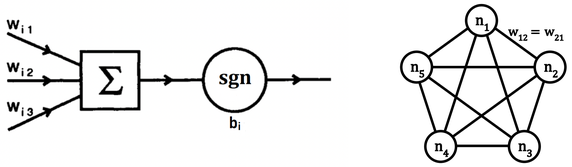
Reference Material: "Geometry with an Introduction to Cosmic Topology" by Michael Hitchman, "Matrix Groups for Undergraduates" by Kristopher Tapp, and "Introduction to Smooth Manifolds" by John M. Lee, and "Geometry of Möbius Transformations Color Illustrations" by Vladimir Kisil.
Big thank you to Paige Hillen for being a great mentor and guiding us through the reading materials!

Hopfield Models and Modern Implementations

Pariya Akhiani, Daniel Naylor
University of California Santa Barbara

Introduction

An **artificial neural network** is a computational model that is based on organic neural networks in brains. A human brain is robust, efficient and flexible at achieving many tasks, including being able to associate unrelated concepts with each other. One of the earliest model neurons that was designed to emulate **associative memory** is the **McCulloch-Pitts model**.



Each model neuron takes a value of -1 or 1, based on a weighted sum of the values of all other neurons. Note that all neurons are pairwise connected in this model. The value is determined by the following update rule:

$$n_i(t+1) = \operatorname{sgn}\left(\sum_{j=1}^{N}\left[w_{ij}n_j(t)\right] + b_i\right) = \operatorname{sgn}(h_i) \tag{1}$$

Where:

- n_i = ith neuron, which takes on one of two values $\{-1, 1\}$.
- t = current time iteration/step. We will always start with $t = 0$.
- w_{ij} = weight describing how n_i is affected by n_j.
- b_i = bias term for n_i. This is a real-valued constant that can change the activation threshold for each neuron. For simplicity's sake, we can let $b_i = 0$.
- N = length of patterns. In this case, it is also the number of neurons.
- $h_i = \sum_{j=1}^{N}\left[w_{ij}n_j(t)\right] + b_i$, which is a shorthand for the expression inside of the sign function.

The neurons can update **synchronously**, which means all at the same time, or **asynchronously**, meaning that only one neuron is updated at each time step. These choices are equivalent.

The Discrete Hopfield Model

The **Hopfield Model** is a neural network designed to model associative memory using McCulloch-Pitts neurons with predetermined weights. Fundamentally, the associative memory problem states that given a set of p patterns $\{\xi^1, \dots, \xi^p\} \subseteq \{-1, 1\}^N$ to "remember," when a new pattern ζ is inputted into the network, it should output one of the stored patterns that best resembles ζ. We define the weights

$$w_{ij} = \frac{1}{N} \cdot \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$$

which are chosen in accordance with **Hebb's Law of Association**. At each time step, the state of the network can be described by the following real-valued function H, which is referred to as an **energy function**:

$$H(t) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} \cdot n_i(t)n_j(t) \tag{2}$$

An application of the triangle inequality shows that the energy function is bounded. It is simple to show that the energy function decreases as the system updates, which means that (2) is a **Lyapunov Function**. Therefore, the system always converges to a local minimum of the energy function, or the **attractors** of the network, which is the final output.

By construction, the stored patterns $\{\xi^1, \dots, \xi^p\}$ are local minima of H, but they are not the only attractors. It is clear from the definition of (2) that $-\xi^\mu$ and ξ^μ have the same energy, so the reverse of a stored pattern is also an attractor.

As it turns out, linear combinations of an odd number of stored patterns are also local minima, referred to as **mixture states**. Attractors that are not stored patterns are referred to as **spurious states**. These patterns have relatively small basins of attraction compared to the stored patterns (as shown in Figure 1), but the existence of these states imply that this model does not work perfectly. One can reduce the basins of attraction for spurious states by changing the bias terms b_i of (1), or by introducing stochasticity.

Figure 1: Basins of Attraction over $\{-1, 1\}^{10}$

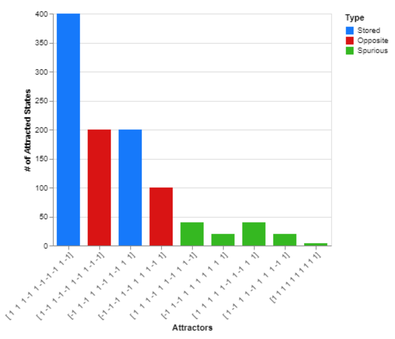


Figure 1 shows each attractor of a model trained on 2 stored patterns (in blue). The model computed the stable state of every element in $\{-1, 1\}^{10}$.

Figure 2: Image Samples

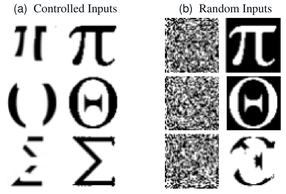


Figure 2 shows a model trained on 3 black and white 50×50 ($\{-1, 1\}^{2500}$) images. (a) shows the result of controlled inputs, giving us our 3 learned patterns. (b) shows the result of 3 random inputs. The first two outputs are opposite attractors, while the third is a spurious attractor.

The Continuous Case

In most scenarios, binary sequences do not accurately represent real-world patterns. Thus, it useful to generalize our pattern space from $\{-1, 1\}^N$ to $[-1, 1]^N$, thereby allowing our neurons to take on a continuum of values.

A useful continuous variant of (1) we will use is

$$\tau_i \frac{dn_i}{dt} = -n_i(t) + \sigma\left(\sum_{j=1}^{N} w_{ij} \cdot n_j(t)\right) \tag{3}$$

Where τ_i is a time constant, chosen such that (3) admits a stable system, and where σ is a strictly increasing, differentiable, and bounded function between -1 and 1 (e.g. $\tanh$). For a discussion on how to obtain suitable τ_i, refer to Hertz and Krogh (1991).

As a result, we can obtain a new energy function similar to (2), given by

$$H_c(t) = H(t) + \sum_{i=1}^{N}\int_0^{n_i(t)} \sigma^{-1}(x)dx \tag{4}$$

Note that even though (3) and (4) suggests a continuous time domain, in practice we will still operate with discrete values of t. One can obtain an update rule by applying numerical methods.

Figure 3: Sample Energy Function

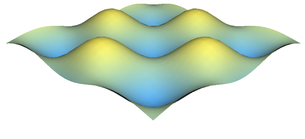


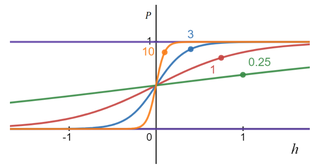
Figure 3 shows an example of an energy function defined over $[-1, 1]^2$. The minima of the graph are attractors of the system. The blue areas are basins of attraction for each of these attractors; patterns that are equidistant to these basins of attraction are spurious states.

Stochasticity

To minimize the appearance of spurious states, we can introduce a notion of randomness. We consider a new parameter $\beta \in [0, \infty]$, inspired by statistical mechanics. The update rule for each neuron n_i becomes the following:

$$\operatorname{Prob}(n_i = \pm 1) = f_\beta(\pm h_i) = \frac{1}{1 + \exp(\mp 2\beta h_i)}$$

Each spurious state has a "critical point" such that sufficiently chosen β less than the critical value ceases it from functioning as an attractor. However, making β too small will increase the amount of noise (and thus, unwanted attractors).

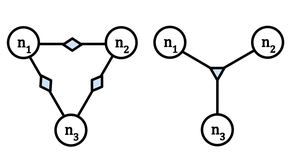


The graph above displays f_β with different parameters. As $\beta \to \infty$, f_β becomes deterministic. An optimal β can be found through numerical methods.

Dense Associative Memories

The maximum number of patterns that can be recalled within a certain error threshold, referred to as the **storage capacity**, is relatively small in the standard Hopfield Model. In the following table, P_{error} denotes the probability of error per neuron, p_{max} denotes the storage capacity and N is the number of neurons.

P_{error}	p_{max}/N
0.001	0.105
0.0036	0.138
0.01	0.185
0.05	0.37
0.1	0.61



Note that we use the Gauss error function to compute P_{error}. After some computations, we get that $p_{max} = N/2\log N$. A **Dense Associative Memory** is a generalization of the Hopfield Model that significantly increases the storage capacity by increasing the number of neurons that are connected to each other per connection. The energy function is generalized to the following:

$$E = -\sum_{\mu=1}^{p} F\left(\sum_i \xi_i^\mu n_i\right)$$

where F is a function depending on the interaction between the neurons. In the image above, a cubic interaction term $F(x) = x^3$ is implemented next to the standard Hopfield Model with 3 neurons. The energy function implicitly uses Hebb's Law to compute weights.

Acknowledgements and Citations

We would like to thank our mentor, William Sheppard, and the DRP Committee. All material presented on this poster are from the following:

(1) John Hertz, Anders Krogh, Richard G. Palmer (1991) *Introduction to the Theory of Neural Computation*

(2) Dmitry Krotov, John Hopfield (2021) *Large Associative Memory Problem In Neuro-Biology and Machine Learning*

Figures 1 and 2 can be found from github.com/DanLeEpicMan/HopfieldModel

Hyperbolic Isometries and Related Topological Groups

Connor Ding, Joshua Howard, Jaime Vandeveer

University of California - Santa Barbara

Background: Hyperbolic Geometry and Hyperbolic Isometeries

The **Hyperbolic Plane** is the metric space consisting of the open half-plane

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

Let $P, Q \in \mathbb{H}^2$, we define the **distance** between P, Q to be

$$d_{\mathrm{hyp}}(P, Q) = \inf\{\ell_{\mathrm{hyp}}(\gamma); \gamma \text{ goes from } P \text{ to } Q\}$$

with the explicit formula

$$D(P, Q) = d_{\mathrm{hyp}}(P, Q) = \log \frac{|P - \bar{Q}| + |P - Q|}{|P - \bar{Q}| - |P - Q|}$$

An important idea in geometry is the notion of a distance-preserving function, also called **isometry**. Let $(X, d), (X', d')$ be two metric spaces. Let $\varphi : X \to X'$ be a bijection. Then φ is an isometry if

$$d'(\varphi(P), \varphi(Q)) = d(P, Q)$$

For all $P, Q \in X$. Notably, all hyperbolic isometries can be written in the form of a **fractional linear map**: Let $z \in \mathbb{H}^2$, then a fractional linear map is the function

$$z \mapsto \frac{az + b}{cz + d}, \; z \mapsto \frac{c\bar{z} + d}{a\bar{z} + b}$$

where $a, b, c, d \in \mathbb{R}$ and $ad - bc = 1$.

Topological Groups

A **group** is the data $(G, \cdot)$ of a set G and a binary operation $\cdot$ on G, satisfying these axioms:

1. $\cdot$ is associative: $\forall a, b, c, (a \cdot b) \cdot c = a \cdot (b \cdot c)$.
2. Identity: there exists $e \in G$ such that for all $a \in G$, $a \cdot e = a = e \cdot a$.
3. Inverse: for all $a \in G$, there exists $a^{-1} \in G$ such that $a \cdot a^{-1} = e = a^{-1} \cdot a$.

A **topological space** is a pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a collection of subsets of X such that

1. $\varnothing, X \in \mathcal{T}$.
2. Let $X_1, X_2 \in \mathcal{T}$. Then $X_1 \cap X_2 \in \mathcal{T}$.
3. Let $\{X_i\}$ be a collection of elements in $\mathcal{T}$. Then

$$\bigcup_i X_i \in \mathcal{T}$$

We say that $\mathcal{T}$ is a **topology** on $\mathbb{X}$. Elements of $\mathcal{T}$ are called **open sets**. A subset $A \subseteq X$ is **closed** if its complement $X - A$ is open. A **topological group** is a group G that is also a topological space such that the group operation $(g, h) \mapsto gh$ and inversion $g \mapsto g^{-1}$ are continuous functions.

General Linear Group, Special Linear Group, and Projective Linear Group

- The **general linear group**, denoted by $\mathrm{GL}(n, F)$, is the set of all invertible $n \times n$ invertible complex matrices.
- The **special linear group**, denoted by $\mathrm{SL}(n, F)$, is the set of $n \times n$ matrices with determinant 1.

We can replace F with $\mathbb{C}$, $\mathbb{R}$, etc. We will consider the case of real numbers and $n = 2$. Define the **projective linear group**

$$\mathrm{PSL}(2; \mathbb{R}) = \mathrm{SL}(2, \mathbb{R})/(\pm I)$$

Let V be a real vector space of dimension n. Then $GL(V) \cong GL(n, F)$.
We have the following diagram of short exact sequences that illustrates the relationships of these groups:

$$\begin{array}{ccccc}
\mathrm{SZ}(V) & \hookrightarrow & \mathrm{Z}(V) \cong F^* & \xrightarrow{\det \cong z^n} & (F^*)^n \\
\downarrow & & \downarrow & & \downarrow \\
\mathrm{SL}(V) & \hookrightarrow & \mathrm{GL}(V) & \xrightarrow{\det} & F^* \\
\downarrow & & \downarrow & & \downarrow \\
\mathrm{PSL}(V) & \hookrightarrow & \mathrm{PGL}(V) & \longrightarrow & F^*/(F^*)^n
\end{array}$$

These Groups have the following relations to geometric spaces:

- $\mathrm{SL}(2, \mathbb{R})$ is the group of all linear transformations that preserve oriented area in $\mathbb{R}^2$.
- $\mathrm{PSL}(2, \mathbb{R})$ is the group of orientation-preserving isometries in $\mathbb{H}^2$

Group Actions on Hyperbolic Space

Let G be a group. Let A be a set. A **group action** of G on A is an operation $\cdot : G \times A \to A, (g, a) \mapsto g \cdot a$ that satisfy two axioms:

1. Associative: $\forall g_1, g_2 \in G, \forall a \in A, g_1 \cdot (g_2 \cdot a) = (g_1 \cdot g_2) \cdot a$.
2. Unitary: $\forall a \in A, e \cdot a = a$.

We say "G **acts on** A".

The **orbit** of $a \in A$ is given by the set $O(a) = \{g \cdot a \mid g \in G\}$.

In the context of hyperbolic space, we have $\mathrm{PSL}(2; \mathbb{R})$ acting on $\mathbb{H}^2$ by fractional linear transformations:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot z = \frac{az + b}{cz + d}$$

This gives us a way to describe all hyperbolic isometries in terms of groups.

Fuchsian Groups

To discuss Fuchsian Group, we introduce the following concepts: A subgroup of a topological group Γ is called a **discrete subgroup** if it contains no limit points.
Examples:

- $\mathbb{Z}$ is a discrete subgroup of $\mathbb{R}$.
- $\mathbb{Q}$ is **not** a discrete subgroup of $\mathbb{R}$.

A group action is **properly discontinuous** if:
$\forall x \in X$, there is a neighborhood V of x such that there are only finitely many $\gamma \in \Gamma$ where $\gamma V \cap V \neq \varnothing$.

Or equivalently, $\forall x \in X$, $O(x)$ under the action of Γ is locally finite.

With the necessary background information, we provide two equivalent characterizations of a **Fuchsian group**:

- A Fuchsian group is a discrete subgroup of $\mathrm{PSL}(2; \mathbb{R})$
- A Fuchsian group is a group that acts properly discontinuously on $\mathbb{H}^2$.

Applications of Hyperbolic Space/Groups to Neuroscience

Despite the abstract nature of topological groups, computational neuroscientists are able to use related concepts to apply them to the study of the brain. Here are some applications:

- Wang et. al. (2023) showed that the neurons in a particular part of the mouse brain (CA1 region of hippocampus) that facilitate spatial perception represent spatial information according to hyperbolic geometry.
- Manifold Gaussian process latent variable model (mGPLVM) introduced by Jensen et. al. (2020) is a novel method to understand neural representations of visual features and it does not assume Euclidean feature spaces but to identify the related manifolds to which the features belong.

Acknowledgements

Reference Material: "Low Dimensional Geometry" by Francis Bonahon
"Fuchsian Groups" by Svetlana Katok
"Foundations of Hyperbolic Manifolds" by John G. Ratcliffe
Math 111A Lecture Notes By Jeremy Brightbill, Math 240A Lecture Notes by John Douglas Moore.
Thank you to the UCSB Directed Reading Program and to our mentor Jaime Vandeveer for making this project possible.

UC **SANTA BARBARA**
Department of Mathematics

Introduction to Orbifolds

Yang Hu, Ethan Palomares, Leslie Mavrakis
University of California, Santa Barbara

Manifolds

While our main topic of discussion is orbifolds, some preliminary properties of manifolds will help clarify our description of an orbifold. A topological space M is a **topological manifold of dimension** n, which we will shorten to manifold, if it satisfies the following three properties:

- M is a **Hausdorff space**
- M is **second countable**
- M is **locally Euclidean** of dimension n

We will omit a thorough description of these properties, but the important part for the orbifold comparison is that manifolds are locally Euclidean; every point in the manifold has a neighborhood that is diffeomorphic to an open subset of $\mathbb{R}^n$.

Classification of 2-Dimensional Manifolds

Classification Theorem for Closed Surfaces: Every compact connected surface is homeomorphic to a sphere, a connected sum of tori, or a connected sum of projective planes.

Here we will introduce the classification of orientable closed 2-manifolds.

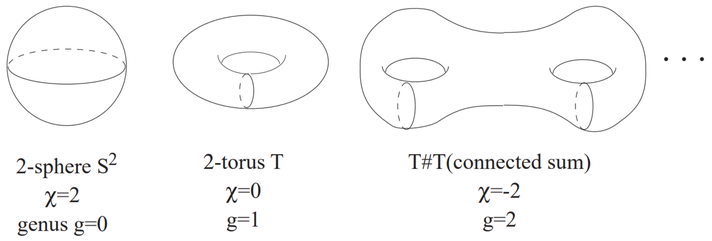


Figure 1. Orientable Closed 2-manifolds

Euler Characteristics for Manifolds

The Euler characteristic χ offers a way to categorize the manifolds. If a surface M has a cell decomposition with V vertices, E edges and F faces, its Euler Characteristic is calculated by:

$$\chi = V - E + F \tag{1}$$

For orientable surfaces, it's calculated as:

$$\chi = 2 - 2g, \tag{2}$$

where g is the genus.

It follows that if $\tilde{M}$ is a d-fold covering of M then $\chi(\tilde{M}) = d\chi(M)$ since $\tilde{M}$ has a cell decomposition obtained by lifting cells from M, and each cell of M has d lifts to M.

Acknowledgements

We would like to thank our Graduate mentor, Leslie Mavrakis, for all the help and guidance she has given us throughout the project.

Orbifolds

Orbifolds can be viewed as generalizations of manifolds. They retain some key features of manifolds, while also introducing new concepts. An orbifold Q is essentially a topological space that locally resembles "quotients" of Euclidean space, which we'll clarify. It follows these three properties:

- Q is a **Hausdorff space**.
- Q is **second countable**.
- Q is **locally modeled** on the **quotient spaces of Euclidean space** by finite group actions.

What distinguishes orbifolds from manifolds is the third property. Rather than being locally Euclidean, each point in an orbifold has a neighborhood that is diffeomorphic to the quotient of an open subset of $\mathbb{R}^n$ under the action of a finite group. For example, a point may have a neighborhood that has been "folded" or "creased" in a certain way. Generally, most points will have the trivial group as the associated action, and the neighborhood around these points will be diffeomorphic to a subset of $\mathbb{R}^n$

Singular locus:

The set of points with a nontrivial associated group is referred to as the "singular locus." In dimension 2, the singular locus has:

- Cone Points - Cyclic Rotation $\mathbb{Z}_n$
- Mirror Points - Reflection
- Corner Points - Dihedral Group

An orbifold with an empty singular locus is just the underlying manifold.

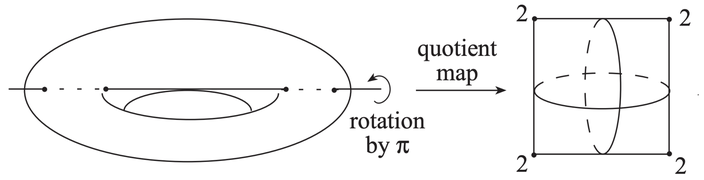


Figure 2. Pillow Case Example: 4 singular cone points where the local group of symmetries is $\mathbb{Z}_2$.

Euler Characteristics for Orbifolds

We would like to extend the Euler Characteristics formula to orbifolds. Considering the universal cover$\tilde{D} \to D^2(n)$. This is an n-fold cyclic branched cover of the disc, branching over the cone point p in $D^2(n)$ where p accounts for only $1/n$ of a vertex. The orbifold covering transformation group Z_n of $\tilde{D}$ fixes a point , projecting to p. Hence, we might perceive p as having n "separate lifts", each taking up $1/n$ of the point . This thought process leads to the following definition.

The orbifold Euler characteristic of Q is

$$\chi(Q) = \sum_{\sigma \in Q} \frac{(-1)^{\dim(\sigma)}}{|\Gamma(\sigma)|}, \tag{3}$$

where σ ranges over (open) cells in X_Q and $\Gamma(\sigma)$ is the local group assigned to points in σ.

Classification of $SO(3)$

The key observation is that any finite subgroup G of $SO(3)$ naturally acts on the sphere S^2, and the quotient space S^2/G is a 2-orbifold. The Euler characteristic χ of an orientable 2-orbifold is given by the formula:

$$\chi = 2 - 2g - \sum(1 - \frac{1}{n_i}), \tag{4}$$

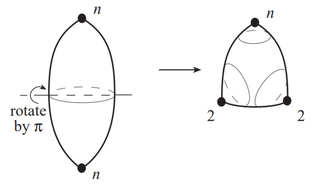
where g is the genus and would be 0 here.

For $\chi > 0$, the formula simplifies to $2 - \sum(1 - \frac{1}{n_i}) > 0$, which is equivalent to $\sum(1 - \frac{1}{n_i}) < 2$. Therefore, we can have:

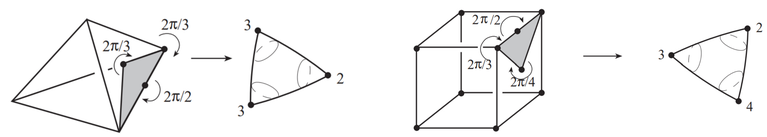
- The sphere S^2, which has $\chi = 2$ and no cone points.
- $S^2(n)$ for $n = 2, 3, 4, \ldots$, which is the sphere with one cone point of order n.
- $S^2(n, m)$ for $n, m = 2, 3, 4, \ldots$, which is the sphere with two cone points of order n and m.
- $S^2(2, 3, n)$ for $n = 3, 4, 5$ which is the sphere with three cone points of order 2, 3 and n.

$S^2(n)$ and $S^2(n, m)$ where $n \neq m$ are "bad orbifolds," meaning they are not a quotient of S^2, so the possible forms of G are:

- $S^2(n, n)$ where G is cyclic of order n.
- $S^2(2, 2, n)$ where G is dihedral of order $2n$. Since any two rotations about distinct axes in 3D space either generate a dihedral group or the whole group $SO(3)$ (if the axes are orthogonal), we see that G must be a dihedral group.



- The action of G on S^2 have more than two axes of rotation. The only way this can happen is if G is the symmetry group of a regular polyhedron since those are the only configurations of more than two axes of rotation in 3D space that are preserved by a group action. This corresponds to $S^2(2, 3, 3)$ if G are the symmetries of a regular tetrahedron, $S^2(2, 3, 4)$ if G are the symmetries of a cube or an octahedron, and $S^2(2, 3, 5)$ if G are the symmetries of an icosahedron or a dodecahedron which gives the symmetry groups of the Platonic solids.



References

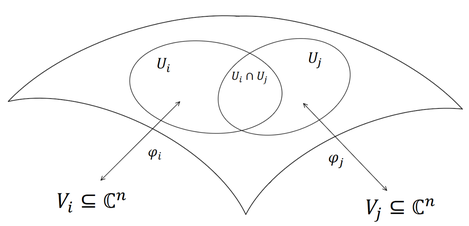
[1] Daryl Cooper, Craig D. Hodgson, and Steven P. Kerckhoff. Three-dimensional orbifolds and cone-manifolds. *Mathematical Society of Japan*, 5:178, 2000.

Manifolds, Cohomologies, and Hodge Structures

Hespos Goodman
University of California, Santa Barbara

COMPLEX MANIFOLDS

An n-dimensional **complex manifold** is a topological space that is locally isomorphic to $\mathbb{C}^n$. This means manifolds can take arbitrary, and often extremely complicated, forms on a global scale, but "zooming in" allows us to study their local properties with relative ease. This construction is defined by an **atlas** of open sets $(U_i)_{i\in M}$ that cover our manifold, each with a **chart** (φ) that that links it to $\mathbb{C}$



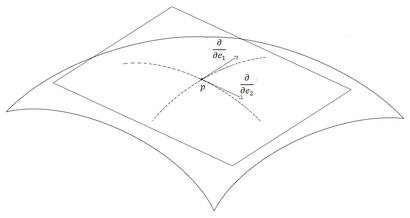
In order for this construction to be useful. It must guarantee continuity of functions on the surface of our manifold. This is achieved through requiring that our charts (φ_i) be *holomorphic (analytic) diffeomorphisms*, and requiring that, on the intersection, the composition $\varphi_i^{-1} \circ \varphi_j$ is a smooth map.

TANGENT BUNDLES

Now that we know what the surface of a manifold looks like, we can begin talking about what happens along that surface. At any particular point p we define T_pM, the **tangent space** at that point. This space in generated by the tangent vectors (at p) of every curve on our manifold that passes through p. As usual, this is equivalent to using the partial derivatives of our chart with respect to the basis vectors in V_i at the preimage of our point,

$$T_pM = \left\langle \frac{\partial}{\partial e_1}(\varphi_i)\Big|_{\varphi_i^{-1}(p)}, \frac{\partial}{\partial e_2}(\varphi_i)\Big|_{\varphi_i^{-1}(p)}, \cdots, \frac{\partial}{\partial e_{2n}}(\varphi_i)\Big|_{\varphi_i^{-1}(p)} \right\rangle = \left\langle \frac{\partial}{\partial e_1}, \frac{\partial}{\partial e_2}, \cdots, \frac{\partial}{\partial e_{2n}} \right\rangle$$

The notation on the right hand side is less formal, but is permissible in the local (U_i) frame. Note that this basis is isomorphic to $\mathbb{R}^{2n}$ under $\frac{\partial}{\partial e_1} \leftrightarrow e_1$. Which yields the usual understanding of a tangent space, depicted for a 2-(real)-dimensional manifold on the right.



The tangent space is specific to each individual point, because it relies on evaluating the partial derivative at the unique (restricted to U_i) preimage of p. In order to address the manifold at large, we can define a **tangent bundle** (TM) which is the set of all pairs of points (p), and vectors in that point's tangent space.

$$TM = \{(p, \vec{v}) \mid p \in M, \vec{v} \in T_pM\}$$

Naturally, there are a LOT of vectors in the tangent space of any particular point. The **(tangent) vector field** (ξ) provides us a method for selecting one of these vectors, given a particular point.

$$\begin{aligned} \xi := M &\mapsto TM \\ p &\mapsto (p, \vec{v}) \end{aligned}$$

For the purposes of integration, we want to remember which point each of these vectors comes from. This is why its essential for the vector field to map to the tangent bundle rather than a particular tangent space.

COTANGENT BUNDLES

Using our definitions of tangent spaces, bundles, and fields, we will define cotangent spaces, bundles, and fields. A covector (ω)(also called a 0-form, or a linear functional) is a function that takes in a vector and outputs a scalar.

$$\omega := \vec{v} \mapsto z$$

Naturally a **cotangent vector** is a covector who's domain is the tangent space (at a point), so we can be sure that it intakes tangent vectors. Applying what we know about tangent spaces, we can see that the **cotangent space** should be the space of all cotangent vectors.

$$T_p^*M = \{\omega | \omega : T_pM \mapsto \mathbb{C}\}$$

COTANGENT BUNDLES (CONT.)

Here we use the notation for the dual of the tangent space since thats exactly what the cotangent space is! It is the set of all maps(covectors) from the tangent space to the underlying field($\mathbb{C}$ in our case). In light of this, we can define a basis for the cotangent space, with the conventional linear functional basis of a dual space

$$T_p^*M = \langle de_1^p, de_2^p, \cdots, de_{2n}^p \rangle, \quad de_i^p\left(\frac{\partial}{\partial e_j}\Big|_p\right) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Similarly, the **cotangent bundle** is the set of all point-cotangent vector pairs

$$T^*M = \{(p, \omega) | p \in M, \omega \in T_p^*M\}$$

Again, this is the dual of the tangent bundle
Finally, a **covector field** is analogous to a vector field. It is a map that, given a point, provides a covector in the cotangent space of that point

$$\begin{aligned} \alpha := M &\mapsto T^*M \\ p &\mapsto (p, \omega) \end{aligned}$$

When we require this map to be smooth, we realize this "covector field" as a section of the cotangent bundle, or a **differential one form**

DIFFERENTIAL 1-FORMS AND EXTERIOR DERIVATIVES

Differential 1-forms are functions that are nearly equivalent to covector fields, the main difference is that we allow them to intake a point AND a vector (i.e. a vector field), so their output becomes a point-scalar pair . Differential forms are written

$$\alpha(p, \vec{v}) = \left(p, \sum_{i=1}^{2n} f_i(p) de_i^p(\vec{v})\right) = \left(p, f_1(p) de_1^p(\vec{v}) + f_2(p) de_2^p(\vec{v}) + \cdots + f_{2n}(p) de_{2n}^p(\vec{v})\right)$$

So, in the particular case where $\vec{v} = \frac{\partial}{\partial e_i}$ that we achieve

$$\alpha\left(p, \frac{\partial}{\partial e_i}\Big|_p\right) = 0 + \cdots + f_i(p) de_i^p\left(\frac{\partial}{\partial e_i}\Big|_p\right) + \cdots + 0 = (p, f_i(p))$$

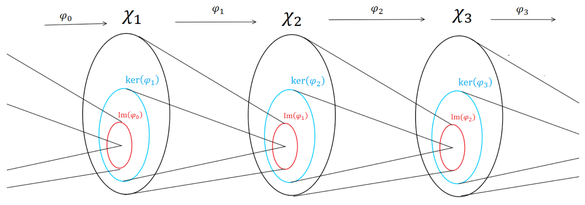
Inspecting the second term, evaluation of α at a point allows us to "measure" the value of α in the $\frac{\partial}{\partial e_i}|_p$ direction. So, summing α along a curve is equivalent to integrating f_i with respect to e_i. More generally, evaluating along some vector field, ξ, allows us to integrate along our entire manifold (with respect to ξ).
Thus, 1-forms are the tools we use in every one dimensional integral. We can use the **exterior derivative** to achieve 2-forms, which allow us to integrate area, and eventually m-forms, which measure m-dimensional oriented density.
The exterior derivative, d, asks us to differentiate each of our f_is with respect to each e_j^p and to note that differentiation in the result

$$d(\alpha) = \sum_{j=1}^{2n} \sum_{i=1}^{2n} \frac{\partial f_i}{\partial e_j} de_j^p \wedge de_i^p$$

The wedge product $(\wedge)$ here is a complicated algebraic structure that explicitly outlines how to evaluate the vector part of our input.

COHOMOLOGIES

In order to better understand the properties of a certian manifold, it can be helpful to understand how differential forms of change as we differentiate them. For this we will use a cohomology. **A cohomology** is a sequence of groups (and maps from one group to the next) with a useful property called *exactness*, that every element becomes 0 after being mapped twice.



Inspecting the quotient group $^{ker(\varphi_i)}/_{Im(\varphi_{i-1})}$ allows us to measure just how fast this process is happening.

THE DOLBEAULT COHOMOLOGY

Since we are working with a complex manifold. We can chose a convenient basis to address our tangent and cotangent spaces

$$\langle z_1, \overline{z_1}, \cdots, z_n, \overline{z_n}, \rangle$$

This choice of basis leads to a method for splitting the exterior derivative

$$d = \partial + \overline{\partial}$$

Where ∂ takes the partial derivatives with respect to the complex basis $\langle z_1, z_2, \cdots, z_n \rangle$, and $\overline{\partial}$ takes the partial derivatives with respect to the complex conjugate basis $\langle \overline{z_1}, \overline{z_2}, \cdots, \overline{z_n} \rangle$. Using these operators, we can construct a cohomology in two directions. Begining with $\Omega^{0,0}$ the space of 0-forms (covectors) we construct,

$$\begin{array}{ccccccc} \Omega^{0,0} & \xrightarrow{\partial_{0,0}} & \Omega^{1,0} & \xrightarrow{\partial_{1,0}} & \Omega^{2,0} & \xrightarrow{\partial_{2,0}} & \\ \downarrow{\scriptstyle \bar{\partial}_{0,0}} & & \downarrow{\scriptstyle \bar{\partial}_{1,0}} & & \downarrow{\scriptstyle \bar{\partial}_{2,0}} & & \\ \Omega^{0,1} & \xrightarrow{\partial_{0,1}} & \Omega^{1,1} & \xrightarrow{\partial_{1,1}} & \Omega^{2,1} & \xrightarrow{\partial_{2,1}} & \\ \downarrow{\scriptstyle \bar{\partial}_{0,1}} & & \downarrow{\scriptstyle \bar{\partial}_{1,1}} & & \downarrow{\scriptstyle \bar{\partial}_{2,1}} & & \\ \Omega^{0,2} & \xrightarrow{\partial_{0,2}} & \Omega^{1,2} & \xrightarrow{\partial_{1,2}} & \Omega^{2,2} & \xrightarrow{\partial_{2,2}} & \\ \downarrow{\scriptstyle \bar{\partial}_{0,2}} & & \downarrow{\scriptstyle \bar{\partial}_{1,2}} & & \downarrow{\scriptstyle \bar{\partial}_{2,2}} & & \end{array}$$

Note that $\partial \circ \overline{\partial} = \overline{\partial} \circ \partial$ and that $\Omega^{i,j} = \overline{\Omega^{j,i}}$. Inspecting $i = 1, j = 0$ reveals that $\Omega^{1,0}$ are the holomorphic 1-forms, and $\Omega^{0,1}$ are the antiholomorphic 1-forms, on our manifold.

THE HODGE DIAMOND

We now inspect the downward cohomology of the dobeault cohomology. We name the quotient groups that it creates

$$H^{i,j}(M) = ker(\overline{\partial}_{(i,j)}) \Big/ Im(\overline{\partial}_{(i,j-1)})$$

We call the dimension of these groups the **hodge numbers**, $h^{i,j} = |H^{i,j}(M)|$. Then we arrange these into **the Hodge diamond**

The hodge diamond is exceptionally useful in algebraic topology as a tool to classify manifolds. The row that each of these hodge numbers are in corresponds to the weight of the represented group. This weight is the order of the forms contained within the cosets that make up each individual $H^{i,j}(M)$.

$$\begin{array}{ccccc} & & h^{0,0} & & \\ & \iddots & \vdots & \ddots & \\ h^{0,n} & \cdots & h^{n/2,n/2} & \cdots & h^{n,0} \\ & \ddots & \vdots & \iddots & \\ & & h^{n,n} & & \end{array}$$

HODGE STRUCTURES

We call the direct sum of all cohomology groups of a particular weight(k), the **hodge structure of weight k**

$$H^k(M, \mathbb{C}) = \bigoplus_{i+j=k} H^{i,j}(M)$$

In the case of hodge structures of weight 1 we know

$$H^1(M, \mathbb{C}) = H^{1,0}(M) \oplus H^{1,0}(M) = H^{1,0}(M) \oplus \overline{H^{1,0}(M)}$$

So we conclude that $H^1(M, \mathbb{C})$ is of even dimension. This must also be true for the lattice subset $H^1(M, \mathbb{Z}) \subset H^1(M, \mathbb{C})$. Thus we identify a torus

$$T = H^{0,1}(M) \Big/ H^1(M, \mathbb{Z})$$

The map from a complex torus to the cohomology groups generated on that torus yields an inverse map and thus we establish a bijection between complex tori and hodge structrues of weight 1

$$T \leftrightarrow H^1(M, \mathbb{C})$$

ACKNOWLEDGEMENTS AND REFERENCES

Shout out to Chris Dare - pro sk8r & mathmatician - christopherdare.com
Thank you to the DRP committee for the opportunity to work on this project

Voisin, C. (2002). *Hodge Theory and Complex Algebraic Geometry, I.* Cambridge, UK. Cambridge University Press

Markov-Style State Machines

Nathaniel Hamovitz, Ron Kibel, Rianna Alers
University of California Santa Barbara

Reinforcement Learning

Reinforcement learning (RL) is a technique for training an agent to act favourably in an arbitrary environment. Environments consist of states, actions, observations, and rewards. The agent's job is to choose actions that maximize the total reward.
There are many approaches to solving this task. The agent can try to predict the long-term consequences of its actions and maximize its long-term benefits (value-based approach). It can also try to learn the environment itself (model-based approach). Here we will look at what is known as a **policy-based** approach to RL, which refers to optimizing the sequence of actions we take to maximize our current and future rewards.

Markov Models

These models come in different variants, with different tradeoffs. The following four classes of models are all "stochastic, discrete state, discrete time" [1] finite state machines with Markovian dynamics – that is, the next state depends only on the current state and the action. We restrict our discussion here to **MDPs (Markov Decision Processes)** and **POMDPs (Partially Observable Markov Decision Processes)**, but you may be familiar with other Markovian processes:

States Completely Observable?	Control over State Transitions?	
	NO	YES
YES	Markov Chain	MDPs
NO	Hidden Markov Model	POMDPs

MDPs and POMDPs

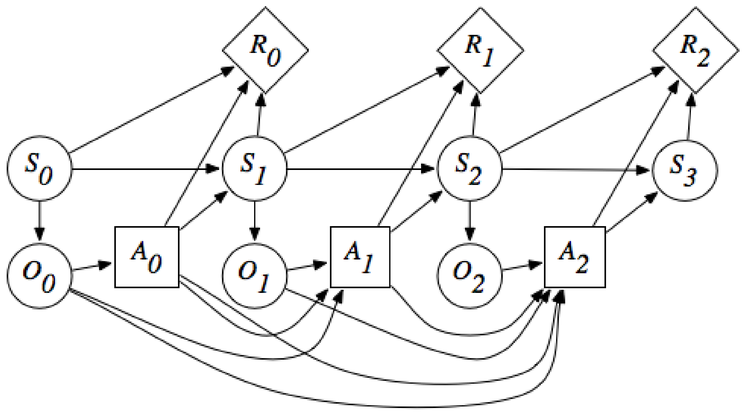
Both MDPs and POMDPs are especially helpful in the RL setting because they can represent the relationship between the agent and the environment.

- In an MDP, we assume the environment is fully observable. As such, the states fully capture all relevant information for decision-making (state transitions contain $P_a(s \mid s')$, the probability that taking action a at time t in state s will lead us to state s' at time $t+1$).
- We can generalize an MDP into a POMDP by assuming that the environment is only partially observable. This implies that states in our model only capture part of the true environment—observations give us information that are only partially informative, which makes it more challenging! We thus need to introduce what is known as the **belief state**, which is a probability distribution over all states encoding how closely a state resembles the true environment [3].

Formally, a POMDP is a tuple of the following items:

- S, a set of states
- A, a set of actions
- $P(s_t \mid a, s_{t-1})$, state-action transition probabilities
- $R : S \times A \to \mathbb{R}$, state-action reward function
- O, a set of observations
- $P(o \mid s)$, conditional observation probabilities
- $\gamma \in [0, 1)$, the discount factor

Diagram of a POMDP



Adaptive State Aggregation for MDPs

MDPs can be solved by iterating the operator

$$T_s(\mathbf{V}) = \min_{a \in \mathcal{A}} \left(r(s,a) + \gamma \cdot \mathbf{P}_{s,a}^{\top} \mathbf{V} \right).$$

Value iteration is simple and guarantees convergence, but is computationally expensive. For large state spaces solving an MDP with this technique becomes infeasible! We can use **state aggregation** to reduce these costs by dynamically grouping states with similar cost-to-go values [2].

This algorithm alternates between two phases: a global update phase and an aggregated update phase. The global update phase performs value iteration on S and the aggregated update phase groups together states with similar cost-to-go values.
We need both phases because the aggregated update phase will require updated knowledge of V* to perform aggregation. In the following algorithms, A_i reference our state-aggregation and B_i reference global iterations.

Value Iteration

We must use some form of value-iteration to obtain a value function for our aggregation. We observe an algorithm using a pre-specified aggregation where W is the value function generated by the mega-states and $\tilde{V}$ is the incuced value function. In the following algorithm α_t is the step size of the learning algorithm (α_t = 1 recovers the formula for value iteration). This has been proven to converge and our alterations maintain a similar convergence bound [2].

Algorithm 1 Random Value Iteration with Aggregation

Input: $\mathbf{P}, r, \gamma, \mathbf{\Phi}, \{\alpha_t\}_{t=1}^{\infty}$
Initialize $\mathbf{W}_0 = \mathbf{0}$
for $t = 1, \ldots, n$ **do**
  for $j = 1, \ldots, K$ **do**
    Sample state s uniformly from set S_j
    $W_{t+1}(j) = (1 - \alpha_t) W_t(j) + \alpha_t T_s \mathbf{V}(\mathbf{W}_t)$
  end for
end for
Output: $\tilde{\mathbf{V}}_n$

State Aggregation

Divide the state space S into K subsets and view these subsets as mega-states. The value function generated by each mega-state can be used to find the optimal value V^*. The algorithm is below:

Algorithm 2 Value-based Aggregation [2]

Input: $\varepsilon, V = (V(1), ..., V(|S|))^T$
$b_1 = \min_{s \in |S|} V(s), b_2 = \max_{s \in |S|} V(s), \Delta = (b_2 - b_1)/\varepsilon$
for $i = 1, ..., \Delta$ **do**
  $\hat{S}_i = \{s | V(s) \in [b_1 + (i-1)\epsilon, b_1 + i\epsilon)\}$
  $W(i) = b_i + (i - \frac{1}{2})\varepsilon$
end for
Return $\{S_i\}_{i=1}^{K}$ and W

Adaptive State Aggregation Algorithm

Combining the two techniques above, we get the following algorithm where A_i are our state-aggregation and B_i are our global iterations.
This method is separate from other aggregation techniques because it learns the cost-to-go values continuously, which aggregation methods need to generate mega-states.

Algorithm 3 Value Iteration with Adaptive Aggregation [2]

Input: $\mathbf{P}, r, \varepsilon, \gamma, \{\alpha_t\}_{t=1}^{\infty}, \{\mathcal{A}_i\}_{i=1}^{\infty}, \{\mathcal{B}_i\}_{t=1}^{\infty}$
Initialize $W_0 = \mathbf{0}, V_1 = \mathbf{0}, t_{sa} = 1$
for $t = 1, \ldots, n$ **do**
  if $t \in B_i$ **then**
    if $t = \min\{B_i\}$ **then**
      $\mathbf{V_{t-1}} = \tilde{\mathbf{V}}(\mathbf{W_{t-1}})$
    end if
    for $j = 1, \ldots, |S|$ **do**
      State $V_t(j) = T_j V_{t-1}$
    end for
  else
    Find current i s.t. $t \in \mathcal{A}_i$
    if $t = \min\{A_i\}$ **then**
      Run our Value-based Aggregation algorithm with input ε, V_{t-1}
      Set the $\{S_i\}_{i=1}^{K}$ and W_t to be the output of our Algorithm
    end if
    for $j = 1, ..., K$ **do**
      Sample state s uniformly from set S_j
      $W_t(j) = (1 - \alpha t_s a) W_{t-1}(j) + \alpha_{t_{sa}} T_s \mathbf{V}(\mathbf{W_{t-1}})$
    end for
    $t_{sa} = t_{sa} + 1$
  end if
end for
if $n \in \mathcal{B}_i$ **then** return V_n
end if
return $\tilde{\mathbf{V}}(\mathbf{W_n})$

Acknowledgements

We are grateful to Charles Kulick for his mentorship and guidance.

References

[1] Anthony R. Cassandra. *POMDP FAQ*. URL: https://pomdp.org/faq.html.

[2] Guanting Chen et al. "An Adaptive State Aggregation Algorithm for Markov Decision Processes". In: (2021). arXiv: 2107.11053 [cs.LG].

[3] Milos Hauskrecht. "Value-Function Approximations for Partially Observable Markov Decision Processes". In: *Journal of Artificial Intelligence Research* 13 (Aug. 2000).

Mathematics of General Relativity

Connor Lindsay

University of California Santa Barbara

Background

The theory of general relativity describes gravity and is based on coordinate invariance and the equivalence principle. In order satisfy these, we describe space as a smooth manifold equipped with a metric, so let us define what these are:

Smooth manifold

This is a space that may have curved geometry, but locally looks like $\mathbb{R}^n$. That is, for every point in the manifold, there exists a neighborhood around it where a diffeomorphism to $\mathbb{R}^n$ exists.

Metric

A metric in general allows for a notion of distance within a space. For us, the metric will be a map taking vectors to dual vectors.

$$g_{ab}V^a = V_b$$

Tensor

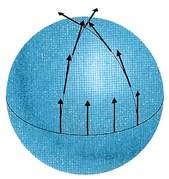
Coordinate invariant objects used to build the physics of general relativity. We will use Einstein summation notation to represent tensors, where repeated indices are summed over.

Levi-Civita Connection

In general relativity, our manifold can be curved, so tangent spaces at every point are independent from each other. We need the Levi-Civita connection to relate tangent spaces, given by,

$$\Gamma^a{}_{bc} = \frac{1}{2}g^{ad}\left(\partial_b g_{dc} + \partial_c g_{db} - \partial_d g_{bc}\right)$$

The connection is a coordinate dependent object, and so it is not tensorial. Instead the connection defines a notion of parallel transport in the manifold, as well as other useful properties [2]:



Covariant Derivative

The covariant derivative is a coordinate invariant form of the derivative.

$$\nabla_a V^b = \partial_a V^b + \Gamma^b{}_{ca}V^c$$

Geodesic Equation

The geodesic equation allows for you to calculate the path of an object in free fall.

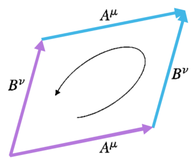
$$\frac{d^2x^a}{ds^2} + \Gamma^a{}_{bc}\frac{dx^b}{ds}\frac{dx^c}{ds} = 0$$

Riemann Curvature Tensor

The Riemann Curvature tensor, built out of the connection, fully describes the curvature of the manifold at every point.

$$R^a{}_{bcd}V^b = [\nabla_c, \nabla_d]V^a$$

This is based on the idea of parallel transporting a vector around a loop to detect a deviation from flat space.



By simplifying the above expression, we can also write this in terms of the Levi-Civita connection which is useful for calculation [2].

$$R^a{}_{bcd} = \partial_c\Gamma^a{}_{bd} - \partial_d\Gamma^a{}_{bd} + \Gamma^a{}_{ce}\Gamma^e{}_{bd} - \Gamma^a{}_{de}\Gamma^e{}_{bc}$$

Ricci tensor and scalar

The Ricci tensor, R_{bd}, and the Ricci scalar, R, are both built by contracting indices from the Riemann curvature tensor and are used in the Einstein field equations.

$$R_{bd} = R^a{}_{bad}, \qquad R = g^{bd}R^a{}_{bad}$$

Einstein-Hilbert Action

The curvature of our space is determined by the physical configuration of the system, so we need a way to solve for the metric based on this. To do this we can use the principle of least action and variational calculus:

Action

The action, denoted S, is a quantity obtained by integrating a scalar over the entire manifold. The scalar is chosen to embody the physical properties of the system studied.

Principle of Least Action

This states that any system will be at a minimum of its action. This means we can find the equations of motion by varying an action against its parameters.

In general relativity the Einstein-Hilbert action is used for our system, given below [2],

$$S = \int \sqrt{|g|}Rd^nx$$

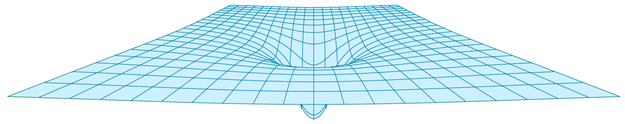
Here R is the Ricci scalar, and $|g|$ is the determinant of the metric. Together these encode the curvature of space, and so by finding the variation of this with respect to the metric we get the Einstein field equations.

Einstein Field Equations

For an n dimensional manifold, the Einstein field equations are a set of n^2 partial differential equations that solve for the metric.

$$R_{ab} - \frac{1}{2}Rg_{ab} = T_{ab}$$

Here R_{ab} is the Ricci tensor, R the Ricci scalar and T_{ab} is the energy momentum tensor, describing the energy present in a system.



In general, these are difficult equations to solve, but specific configurations have analytic solutions [2].

Einstein-Hilbert-Palatini Formalism

When choosing a connection we assumed it to be Levi-Civita since it has useful properties, but the Palatini formalism seeks to give a reason why this is the correct connection to use.

$$\Gamma^a{}_{bc} = \Gamma^a{}_{cb}$$

The Levi-Civita connection is torsion free, satisfying the above property, allowing for the simplification of the Riemann curvature tensor.

$$\nabla_a g_{bc} = 0$$

It is also metric compatible, allowing for simplifications when varying the Einstein-Hilbert action.

In the Palatini formalism we choose an arbitrary affine connection for our manifold as a second parameter for variation. This variation gives the usual Einstein field equations and an equation restricting our arbitrary connection. Solving these restrictions shows the only physical solutions to the equations of motion are given by the Levi-Civita connection [1].

Acknowledgements

I would like to thank Patrick Vecera for his guidance as well as the UCSB Directed Reading Program for the opportunity to work on this project.

References

[1] Antonio N. Bernal et al. “On the (non-)uniqueness of the Levi-Civita solution in the Einstein-Hilbert-Palatini formalism”. In: *Physics Letters B* 768 (May 2017), pp. 280–287.

[2] S.M. Carroll. *Spacetime and Geometry*. Cambridge University Press, 2019. ISBN: 9781108488396.

Modern Portfolio Theory: The Markowitz Model

Xinyue Li, Anna Maximova, Yujing Kang; Mentored by: Doris Padilla & Alex Bernstein

University of California, Santa Barbara

Abstract

Financial markets are notoriously complex and unpredictable. Therefore, investors must consider several factors and strategies in order to increase profits. **Modern portfolio theory**, or mean-variance analysis, is a mathematical framework which uses analysis to model the relationships between profits and losses in the portfolio selection process. Ultimately, the objective is to maximize expected investment return given a certain risk amount. In our project, using a dimension reduction procedure, we apply modern portfolio theory to a high-dimensional data set. The results include a sequence of optimal portfolios (efficient frontier).

Introduction

Let's generalize the description of the financial market with the following notation,

- **Assets:** there are N assets (securities) which are traded in our market. Let S_t^i represent the price of asset i at time t.
- **Returns:** The return of an asset i is given to be $R_i \equiv \frac{S_1^i - S_0^i}{S_0^i}$. The market with N assets is characterized by the `random vector` $R = (R_1, \ldots, R_N)$.
- **Portfolio Values:** The initial value of a portfolio equals $x = (x_1, \ldots, x_N)$; that is, at time zero, the investor **invests** amount x_i into asset i.
- **Reward vs Risk:** Our portfolio selection process is characterized through a trade-off between reward and risk. The variables **expectation** ($\mathbb{E}[R_x] \equiv m_x$) and **variance** ($\mathrm{Var}[R_x] \equiv \sigma_x^2$), respectively, capture this unique relationship. Suppose,
  - m represents the vector of expected returns, $m = (m_1, \ldots, m_N)$.
  - V is the *covariance matrix* of returns, that is: $V = (\sigma_{ij})$ where $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$. Note that, $W \equiv V^{-1}$.

Problem Statement: Consider the following scenario: an investor wishes to determine an optimal allocation of his/her assets according to an individual risk preference $\tau \geq 0$, also called the **risk parameter**. Applying Markowitz's result in [4] & [1], this problem can be solved as follows:

$$\max_{x \in \mathbb{R}^N} \{\tau m_x - \sigma_x^2\} \text{ subject to } \sum_{i=1}^N x_i = 1. \tag{1}$$

The trade-off lies in the conflicting parameters m_x and σ_x^2. Usually, a higher expected return is associated with more risk, and vice versa. Intuitively, the goal remains to *maximize* return while *minimizing* risk.

Principal Component Analysis: Principal component analysis (PCA, [2]) is a dimension reduction procedure, whose benefits include:

- Data transformation into a new space, with the same order, and orthogonal axes (i.e principal components).
- The newly formed axes are ordered, decreasingly, in terms of their explained variability (weight). That is, the first principal component explains more variance than the second one, and the pattern continues.

In our project, we apply PCA to our covariance matrix V for the following reasons:

- Understand the major sources of variance and the effect of those factors on each security in our data set.
- Optimal portfolios require the inversion of a covariance matrix (see Thm 2). Our sample V has a condition number on the order of 10^{20}. The inversion of such a large covariance matrix (dim: 5119×5119) is *numerically unstable*.

Theoretical Results

Assumptions:

- The law/distribution of the portfolio return is fully characterized by the mean $\mathbb{E}[R_x]$ and variance $\mathrm{Var}[R_x]$.
- The covariance matrix V is *positive definite*; that is, $\langle x, Vx\rangle > 0, \forall x \neq 0$.
- There are at least two assets i and j where $m_i \neq m_j$.

Theorem 1. The expected return & variance of a portfolio R_x equals,

$$\mathbb{E}[R_x] = \sum_{i=1}^N x_i m_i = \langle x, m\rangle \ , \ \mathrm{Var}[R_x] \equiv \sum_{i,j=1}^N x_i \sigma_{ij} x_j \equiv \langle x, Vx\rangle \tag{2}$$

Proof. Note that, $R_x \equiv \sum_{i=1}^N x_i R_i$ and $M_x = \sum_{i=1}^N x_i m_i$. As a result, by definition:

$$\mathbb{E}[R_x] = \mathbb{E}[\sum_{i=1}^N x_i R_i] = \sum_{i=1}^N x_i \mathbb{E}[R_i] = \sum_{i=1}^N x_i m_i \equiv \langle x, m\rangle$$
$$\mathrm{Var}[R_x] = \mathbb{E}[R_x - \mathbb{E}[R_x]]^2 = \mathbb{E}[R_x - M_x]^2$$

Continuing to expand the variance, we get the final answer:

$$\mathbb{E}[R_x - M_x]^2 = \mathbb{E}[\sum_{i=1}^N \sum_{j=1}^N x_i x_j (R_i - m_i)(R_j - m_j)] = \sum_{i,j=1}^N x_i x_j Cov(R_i, R_j) \equiv \langle x, Vx\rangle$$

□

Theorem 2. For each risk tolerance parameter $\tau \geq 0$ the Markowitz portfolio selection problem has a unique solution,

$$x_\tau^* = \frac{We}{\langle e, We\rangle} + \frac{\tau}{2}\left(Wm - \frac{\langle e, Wm\rangle}{\langle e, We\rangle} We\right), \text{with } e = (1, 1, \ldots, 1). \tag{3}$$

Proof. Suppose $L(x, \lambda)$ represents the Lagrange Multiplier i.e.

$$L(x, \lambda) = \underbrace{\tau\langle m, x\rangle - \langle x, Vx\rangle}_{\text{function to maximize}} + \underbrace{\lambda\big(\langle x, e\rangle - 1\big)}_{\text{constraint}}$$

The objective is to solve for the vector $x \equiv x_\tau^*$ such that i) $L_x = 0$ and ii) $m\tau - 2Vx + \lambda e = 0$. If $\tau = 0$, we see that:

$$-2Vx + \lambda e = 0 \Leftrightarrow -2x + \lambda We = 0; \quad \text{(applying } W \text{ both sides)}$$
$$\Leftrightarrow \lambda = \frac{2}{\langle e, We\rangle}; \quad \text{(since } \langle x, e\rangle = 1).$$

So the *minimum variance portfolio* becomes:

$$-2x + \frac{2}{\langle e, We\rangle} \times We = 0 \ \Rightarrow \ x^{\min} \equiv x_{\{\tau=0\}}^* = \frac{We}{\langle e, We\rangle}.$$

Assume $\tau > 0$. Define $z \equiv x - x_0^*$. Equiv, $x = z + \frac{We}{\langle e, We\rangle}$ Then, Equations $i)$ & $ii)$ can be re-written as such:

$$i^*) \ m\tau - 2V(z + x_0^*) + \lambda e \equiv m\tau - 2V\left(z + \frac{We}{\langle e, We\rangle}\right) + \lambda e = 0;$$
$$ii^*) \ \langle z, e\rangle = 0.$$

Applying W to Equation $i^*)$ and simplifying, we obtain: $\lambda = \frac{2}{\langle e, We\rangle} - \frac{\tau\langle e, Wm\rangle}{\langle e, We\rangle}$. In conclusion,

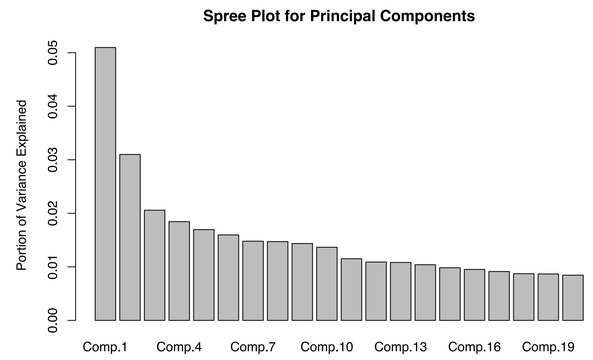
$$\tau Wm - 2z - \frac{\tau\langle e, Wm\rangle}{\langle e, We\rangle} We = 0 \quad \Rightarrow \quad z = \frac{\tau}{2} z^* \text{ where } z^* = Wm - \frac{\langle e, Wm\rangle}{\langle e, We\rangle} We.$$

□

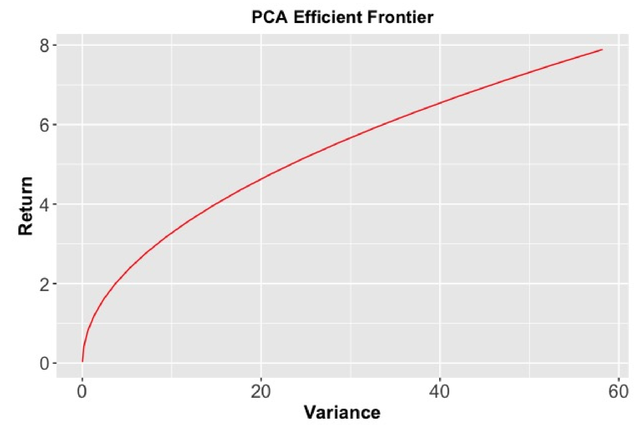
Coding Application

We collect data from [3], which provides daily stock returns for all US-based securities and ETFs trading on the NYSE, NASDAQ, and NYSE MKT. Stocks which contained missing values were excluded from the analysis, leaving a final sample of 5,119 assets across 218 data points (January 2nd, 2017-November 10th, 2017).

Dimension Selection: Let V represent the covariance matrix constructed from our estimated factors using the Principal Component method. Our analysis from Fig. 1 concludes that 3 eigenvectors sufficiently explain the *majority* of uncertainty in our dataset.



Results: Using this lower-dimensional matrix V, we applied Theorem 2 to find the optimal portfolio x_τ^*. The points (σ_x^2, m_x) on the *efficient frontier* represent the variances (x-axis) and expectations (y-axis) of the `returns` on x_τ^* with risk tolerance $\tau \geq 0$. *Intuition:* more money means higher risk.



References

[1] Igor V. Evstigneev. *Mathematical Financial Economics*. Springer International Publishing Switzerland, 2015. ISBN: 978-3-319-16570-7.

[2] Richard Johnson and Dean Wichern. *Applied Multivariate Statistical Analysis*. Pearson Education Limited, 2014. ISBN: 978-1-292-02494-3.

[3] Boris Marjanovic. *Huge Stock Market Dataset*. 2017. URL: https://www.kaggle.com/datasets/borismarjanovic/price-volume-data-for-all-us-stocks-etfs.

[4] Harry Markowitz. "Portfolio Selection". In: *Journal of Finance* 7.1 (1952), pp. 77–91.

A Brief Introduction to Lie Algebras

Lana Azar and Mychelle Parker

University of California Santa Barbara - Directed Reading Program 2023

What are Lie Algebras?

We define a Lie Algebra over a field $\mathbb{F}$ as an $\mathbb{F}$-vector space L, together with a bilinear map, which we call the Lie bracket. The Lie bracket is defined as

$$L \times L \Rightarrow L, \quad (x, y) \mapsto [x, y]$$

satisfying

$$[x, x] = 0, \ \forall x \in L,$$
$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0, \ \forall x, y, z \in L.$$

This second condition is known as the **Jacobi identity.** Here is a consequence of those two conditions:

$$0 = [x + y, x + y] = [x, x] + [x, y] + [y, x] + [y, y] = [x, y] + [y, x]$$
$$\implies [x, y] = -[y, x]$$

We note that the Lie bracket $[x, y]$ is often referred to as the commutator of x and y.

Examples of Lie Algebras

Here we highlight three examples of Lie Algebras.

(1) Suppose V is a finite-dimensional vector space over field $\mathbb{F}$ and let $gl(V)$ be the set of all linear maps from V to V. $gl(V)$ is also a vector-space over $\mathbb{F}$ and is a Lie algebra, known as the **general linear algebra**, defined by the Lie bracket

$$[x, y] := x \circ y - y \circ x, \ \forall x, y \in gl(V)$$

where $\circ$ denotes the composition of maps. We can check that the Jacobi identity holds for this Lie algebra:

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0, \ \forall x, y, z \in gl(V)$$
$$\implies (x \circ [y, z] - [y, z] \circ x) + (y \circ [z, x] - [z, x] \circ y) + (z \circ [x, y] - [x, y] \circ z) = 0$$
$$\implies x \circ (y \circ z - z \circ y) - (y \circ z - z \circ y) \circ x + y \circ (z \circ x - x \circ z) - (z \circ x - x \circ z) \circ y$$
$$+ z \circ (x \circ y - y \circ x) - (x \circ y - y \circ x) \circ z) = 0$$
$$\implies 0 = 0$$

since the composition of linear maps is commutative.

(2) Let $g(n, \mathbb{F})$ be the vector space of all $n \times n$ matrices over $\mathbb{F}$ with the Lie bracket defined by

$$[x, y] := xy - yx,$$

where xy is the usual product of matrices x and y. Since $gl(n, \mathbb{F})$ is a vector space, it has a basis consisting of the matrix units e_{ij} for $1 \leq i, \ j \leq n$, where e_{ij} is the $n \times n$ matrix that has a 1 in the ij-th position and 0 in all other positions. Letting δ be the Kronecker delta, defined by $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise, we have

$$[e_{ij}, e_{kl}] = \delta_{jk} e_{il} - \delta il e_{kj},$$

which can be verified using the Lie bracket.

(3) Let $sl(n, \mathbb{F}) \subseteq gl(n, \mathbb{F})$ denote the set of all $n \times n$ matrices such that the matrices have trace equal to zero. Then, we can define the Lie bracket

$$[x, y] := xy - yx, \ \forall x, y \in sl(n, \mathbb{F})$$

and $sl(n, \mathbb{F})$ is a Lie algebra, known as the special linear algebra. The two properties of Lie brackets are inherited from $gl(n, \mathbb{F})$.

Subalgebras

A **Lie subalgebra** of a Lie algebra L is defined to be a vector space $K \subseteq L$ such that

$$[x, y] \in K, \ \forall x, y \in K.$$

In our previous example of Lie algebras, particularly example 2, we saw that $sl(n, \mathbb{F})$ is a subalgebra of the Lie algebra $gl(n, \mathbb{F})$. Another way to think about the definition of a subalgebra is as follows: a subalgebra L' of a Lie algebra L is a subset of elements of L that themselves form a Lie algebra with the same commutator and field as that of L.

Ideals

Given a Lie algebra L a subalgebra I of L is defined to be an **ideal** if

$$[x, y] \in I \quad \text{for all} \quad x \in L, y \in I$$

If I and J are both ideals of a Lie algebra L, then $I \cap J$, $I + J$, and $[I, J]$ are also ideals, where

$$I + J := \{x + y \mid x \in I, y \in J\}$$
$$[I, J] := \text{Span}\{[x, y] \mid x \in I, y \in J\}$$

Examples of ideals:

(1) L is an ideal of itself.

(2) $\{0\}$ is the **trivial ideal**.

(3) A frequently non-trivial ideal is the **center of** L, which is defined to be

$$Z(L) := \{x \in L \mid [x, y] = 0 \ \text{for all} \ y \in L\}$$

(4) If $I = J = L$, then we write $L' = [L, L]$. This ideal is the **derived algebra of** L**.**

Given an ideal I we can define a new Lie algebra by considering the cosets $z + I = \{z + x \mid x \in I$ and then define the corresponding quotient Lie algebra

$$L/I := \{z + I \mid z \in L\}$$

where the bracket relation on L/I is given by

$$[w + I, z + I] := [w, z] + I \ \text{ for } \ w, z \in L$$

Theorem: There is a bijective correspondence between the ideals of L/I and the ideals of L that contain I.

Proof. If J is an ideal of L and $I \subseteq J$, then J/I is an ideal of L/I. If K is an ideal of L/I, then $J := \{z \in L \mid z + I \in K\}$ is an ideal of L that contains I.

Homomorphisms

If L_1 and L_2 are Lie algebras over a field F, then a map $\varphi : L_1 \to L_2$ is a **homomorphism** if φ is a linear map and

$$\varphi([x, y]) = [\varphi(x), \varphi(y)] \ \text{ for all } \ x, y \in L_1$$

If φ is also bijective, then we say that it is an *isomorphism.*

If $L_1 = L_2$, then any homomorphism between them can be referred to as an endomorphism. An isomorphism between a Lie algebra and itself is called an automorphism.

Adjoint

For a Lie Algebra L we define its **adjoint homomorphism** $\text{ad} : L \to gl(L)$ by

$$(\text{ad}(x)(y) := [x, y] \ \text{ for all } \ x, y \in L$$

Example: Consider the case when $L = sl(2, \mathbb{C})$. A basis for this Lie algebra is

$$h = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad e = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad f = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

with structure equations

$$[h, e] = 2e, \ [h, f] = -2f, \ [e, f] = h$$

Then for any $x \in L$ we know $x = a_1 h + a_2 e + a_3 f$ with $a_i \in \mathbb{C}$. Then to understand $\text{ad}(x) \in gl(L)$ we see that

$$\text{ad}(x)(h) = [x, h] = [a_1 h + a_2 e + a_3 f, h] = -2a_2 e + 2a_3 f$$
$$\text{ad}(x)(e) == [x, e] = [a_1 h + a_2 e + a_3 f, e] = 2a_1 e - a_3 h$$
$$\text{ad}(h)(f) = [x, f] = [a_1 h + a_2 e + a_3 f, f] = -2a_1 f + a_2 h.$$

Lie Algebra Isomorphism Theorems

(1) Let $\varphi : L_1 \to L_2$ be a homomorphism of Lie algebras. Then $\ker \varphi$ is an ideal of L_1, im φ is a subalgebra of L_2 and

$$L_1 / \ker \varphi \cong \text{ im } \varphi.$$

(2) If I and J are ideals of a Lie algebra J, then

$$(I + J)/J \cong I/(I \cap J).$$

(3) Suppose that I and J are ideals of a Lie algebra L such that $I \subseteq J$. Then J/I is an ideal of L/I and

$$(L/I)/(J/I) \cong L/J.$$

Example: Fix a field F and consider the linear map $\text{tr} : gl(n, F) \to F$ which is defined by taking the trace of a matrix. It can be shown that tr is a surjective Lie algebra homomorphism and that $\ker$ tr $= sl(n, F)$. Thus by the first isomorphism theorem we can conclude that

$$gl(n, F)/sl(n, F) \cong F.$$

References

[1] Karin Erdmann and Mark J. Wildon. *Introduction to Lie Algebras.* Springer, 2006.

A Brief Introduction to Machine Learning

Tyler Ramil and Terry Wang
University of California Santa Barbara

Intro to The Statistical Learning Framework

The basic statistical learning setting involves a learner who is given a set of objects (the "domain set"), a set of labels, and a training set consisting of labeled examples. As an example, consider the domain to be the set of all papayas where each papaya can be represented by its color and softness. So these papayas may be labeled with $\mathcal{Y} = \{0, 1\}$, where 0 means not tasty and 1 means tasty. The learner's goal is to output a **prediction rule**, $h : \mathcal{X} \to \mathcal{Y}$ that can be used to predict the label of new objects. In our papaya example, this would predict whether a papaya is tasty or not. The training data is generated by sampling objects from a probability distribution and then labeling them according to an unknown labeling function. The error of a predictor is defined as the probability that it predicts the wrong label for a randomly chosen object (papaya) from the distribution. The learner is blind to the underlying distribution and labeling function and can only interact with the environment through the training set.

ERM and Overfitting

The concept of Empirical Risk Minimization (**ERM**) can be used to analyze how a learning algorithm performs. The algorithm receives a training set S, which is sampled from an unknown distribution $\mathcal{D}$ and labeled by some target function f. It outputs a predictor $h_S : \mathcal{X} \to \mathcal{Y}$, where the subscript S emphasizes that the predictor depends on S. The goal is to find a predictor that **minimizes** the error with respect to the unknown $\mathcal{D}$ and f. Since the true error is unknown to the learner, a useful notion of error is the training error, which is the error the predictor incurs over the training sample:

$$L_S(h) \overset{\text{def}}{=} \frac{|\{i \in [m] : h(x_i) \neq y_i\}|}{m},$$

where $[m] = \{1, \ldots, m\}$.
The idea behind ERM is to find a predictor that minimizes the training error, which is a measure of how well the predictor fits the training data. However, this approach can lead to overfitting, where the predictor fits the training data too well and does not generalize well to new data. This means that the predictor may have a low training error but a high error on new data.

ERM with Inductive Bias

Even though the ERM rule may lead to **overfitting**, there are ways to address this problem. One solution is to apply the ERM learning rule over a hypothesis class, denoted by $\mathcal{H}$. Each $h \in \mathcal{H}$ maps from input space $\mathcal{X}$ to output space $\mathcal{Y}$. The ERM$_\mathcal{H}$ learner chooses a predictor $h \in \mathcal{H}$ with the lowest possible error over the training sample S,

$$\text{ERM}_\mathcal{H}(S) \in \underset{h \in \mathcal{H}}{\text{argmin}} L_S(h)$$

where argmin stands for the set of hypotheses in $\mathcal{H}$ that achieve the minimum value of $L_S(h)$ over $\mathcal{H}$.[1] This is an **inductive bias**, which biases the learner toward a specific set of predictors based on prior knowledge about the problem. For example, relating back to the papaya example given earlier, we might assume that there is a softness threshold after which a papaya is too hard to be tasty. However, choosing a more restricted hypothesis class may introduce a stronger inductive bias.
A fundamental question in learning theory is which hypothesis classes lead to ERM learning that does not overfit. To answer this is dependent on the problem at hand and the choice of our hypothesis class. The tradeoff between restricting the hypothesis class and avoiding overfitting is something that is highly considered in machine learning.
Overall, ERM is a very useful tool that is used in machine learning, but given its limitations, such as overfitting, we must pay attention to how we work to resolve such issues.

A Formal Learning Model

PAC Learning Probably Approximately Correct (PAC) learning is a framework for supervised learning that allows us to quantify the trade-off between the accuracy of a learned model and the number of training examples required to achieve that accuracy. In PAC learning, we seek to find a hypothesis that is "probably approximately correct" with respect to an unknown target function, given a finite set of training examples drawn from some distribution over the input space. Formally, we have the following definition:

Definition (PAC Learnability):
A hypothesis class $\mathcal{H}$ is *PAC learnable* if there exist a function $m_\mathcal{H} : (0,1)^2 \to \mathbb{N}$ and a learning algorithm with the following property: For every $\epsilon, \delta \in (0,1)$, for every distribution $\mathcal{D}$ over $\mathcal{X}$, and for every labeling function $f : \mathcal{X} \to \{0,1\}$, if the realizable assumption holds with respect to $\mathcal{H}, \mathcal{D}, f$, then when running the learning algorithm on $m \geq m_\mathcal{H}(\epsilon, \delta)$ i.i.d. examples generated by $\mathcal{D}$ and labeled by f, the algorithm returns a hypothesis h such that, with probability of at least $1-\delta$ (over the choice of the examples), $L_{(\mathcal{D},f)}(h) \leq \epsilon$ [1]

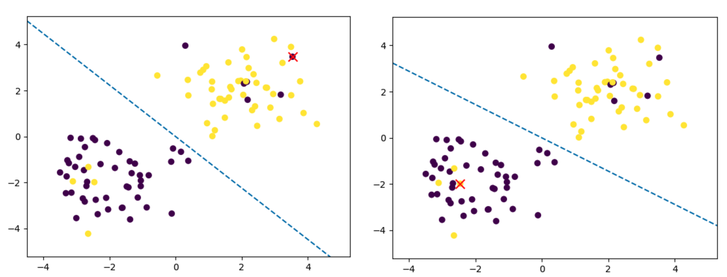
A limitation of PAC learning is that it assumes that the target function belongs to the hypothesis class being considered, which may not always be the case. Also, the PAC framework assumes that the training examples are drawn independently and identically from some fixed distribution, which may not be true in most real-world scenarios. Finally, the sample complexity bounds for PAC learning can be quite loose in some cases, leading to a large number of training examples being required to achieve a desired level of accuracy.

Perceptron Algorithm

Perceptron is an implementation of ERM rule. It's an iterative algorithm that output a sequence of vectors $w^{(1)}, w^{(2)}, \ldots$. At iteration i, the algorithm would update $w^{(i)}$ based on one misclassified data point by adding to it the instance x_i times its label y_i. The following is an example of Perceptron algorithm implemented in Python and screenshots of two iterations. The misclassified point that is causing the update is marked with a red x.

```python
def perceptron_learning_algorithm(X, y, max_iter=1000):
    w = np.zeros(X.shape[1])
    w_hist = [w.copy()]
    iter_count = 0
    errors = []
    while iter_count < max_iter:
        for i in range(X.shape[0]):
            if y[i] * np.dot(w, X[i]) <= 0:
                w += y[i] * X[i]
                w_hist.append(w.copy())
                errors.append(i)
        iter_count += 1
        if len(errors) == 0:
            break
    return w_hist, errors

def plot_hyperplane(w, ax):
    x_vals = np.array(ax.get_xlim())
    y_vals = (-w[0] * x_vals) / w[1]
    ax.plot(x_vals, y_vals, '--')

def update_plot(frame, X, y, w_hist, errors, scat, line, scat_error, ax):
    w = w_hist[frame]
    plot_hyperplane(w, ax)
    line.set_data(ax.lines[-1].get_data())
    ax.lines[-1].remove()
    scat.set_offsets(X)
    scat.set_array(y)
    scat_error.set_offsets(X[errors[frame], :])
    return scat, line, scat_error
```

```python
def animate_perceptron(X, y, max_iter=1000):
    w_hist, errors = perceptron_learning_algorithm(X, y, max_iter)

    fig, ax = plt.subplots()
    scat = ax.scatter(X[:, 0], X[:, 1], c=y, cmap='viridis')
    line, = ax.plot([], [], '--')
    scat_error = ax.scatter([], [], c='red', marker='x', s=100)

    # Set axis limits
    ax.set_xlim(X[:, 0].min() - 1, X[:, 0].max() + 1)
    ax.set_ylim(X[:, 1].min() - 1, X[:, 1].max() + 1)

    ani = animation.FuncAnimation(fig, update_plot, frames=len(w_hist), fargs=(
        X, y, w_hist, errors, scat, line, scat_error, ax), repeat=False, interval=2000)
    plt.show()

# Create a non-realizable dataset
np.random.seed(0)
X_positive = np.random.randn(50, 2) + np.array([2, 2])
X_negative = np.random.randn(50, 2) + np.array([-2, -2])
X = np.vstack((X_positive, X_negative))

y_positive = np.ones(50)
y_negative = -np.ones(50)
y = np.hstack((y_positive, y_negative))

# Add noise to the labels to make the dataset non-realizable
noise_indices = np.random.choice(X.shape[0], size=10, replace=False)
y[noise_indices] = -y[noise_indices]

animate_perceptron(X, y)
```



Boosting

Weak Learnability

The limitations of PAC learning, which ignores the computational aspect of learning, motivates a new type of learnability: γ-Weak-Learnability.

Definition (γ-Weak-Learnability):

- A learning algorithm, A, is a γ-weak-learner for a class $\mathcal{H}$ if there exists a function $m_\mathcal{H} : (0,1) \to \mathbb{N}$ such that for every $\delta \in (0,1)$, for every distribution $\mathcal{D}$ over $\mathcal{X}$, and for every labeling function $f : \mathcal{X} \to \{\pm 1\}$, if the realizable assumption holds with respect to $\mathcal{H}, \mathcal{D}, f$, then when running the learning algorithm on $m \geq m_\mathcal{H}(\delta)$ i.i.d. examples generated by $\mathcal{D}$ and labeled by f, the algorithm returns a hypothesis h such that, with probability of at least $1-\delta, L_{(\mathcal{D},f)}(h) \leq 1/2 - \gamma$
- A hypothesis class $\mathcal{H}$ is γ-weak-learnable if there exists a γ-weak-learner for that class. [1]

This is a weaker definition than the PAC learning definition. While PAC learning implies the ability to find a classifier with error rate at most an arbitrary small number $\epsilon > 0$, our weak learning only need to find a hypothesis with error rate less than $\frac{1}{2} - \gamma$. With the Weak Learnability, we can now introduce the AdaBoost algorithm.

AdaBoost

AdaBoost, short for Adaptive Boosting, is an algorithm using a weak learner to find a hypothesis with relative low empirical error. The boost proceeds in several consecutive rounds. Eventually, the result is a relative strong hypothesis that's the weighted combination of the several weak hypotheses. At each round i, the algorithm defines a distribution D_i over the input training set. Then the weak learner passes this distribution and the training set and supposedly it should return a hypothesis h_i with error ϵ_i:

$$\epsilon_i \overset{\text{def}}{=} L_{D_i}(h_i) \overset{\text{def}}{=} \sum_{j=1}^{m} D_{ij} 1_{h_i(x_j) \neq y_j} \leq \frac{1}{2} - \gamma$$

After each round, the algorithm would assign a weight, $w_i = \frac{1}{2} log(\frac{1}{\epsilon_i} - 1)$ to h_i and then it will update D_i so that in the next round, the probability of appearance of the samples on which h_i errs is higher than it is in the current round. This ensures that additional rounds create a stronger learner overall.

Acknowledgements

We thank Lucas Fagan for his guidance as well as the UCSB Directed Reading Program for the opportunity to work on this project.

References

[1] S. Ben-David. *Understanding Machine Learning*. New York, NY: Cambridge University Press, 2014.

Algebraic Geometry

Sebastian Nunez and Scott Ricardo Figueroa Weston - Mentor: Gregory McGrath

University of California, Santa Barbara

Structure from motion

Structure from Motion is an interdisciplinary field that merges the principles of algebraic geometry and computer vision. Its primary objective is to reconstruct three-dimensional structure from a sequence of two-dimensional images. Through techniques such as camera calibration and feature extraction, we can estimate the camera parameters and identify common points across multiple images. By triangulating these correspondences, we obtain a set of 3D points that represent the object's structure. Algebraic geometry provides a mathematical framework to handle perspective distortions and depth ambiguities inherent in 2D image captures.

Algebraic Geometry

Let $R = k[x_1, \ldots, x_n]$ where k is algebraically closed.

Affine Variety: Given polynomials $f_1, \ldots, f_s \in R$, an affine variety is the set

$$V = \{(a_1, \ldots, a_n) \in k^n \mid f_i(a_1, \ldots, a_n) = 0 \text{ for all } 1 \leq i \leq s\}$$

Finitely Generated Ideal: Given polynomials $f_1, \ldots, f_s \in R$ we call an ideal generated by $f_1, \ldots, f_s$,

$$\langle f_1, \ldots, f_s \rangle = \left\{ \sum_{i=1}^{s} h_i f_i \;\middle|\; h_1, \ldots, h_s \in R \right\}.$$

Hilbert's basis theorem assures that every ideal $I \leq R$ is finitely generated. An ideal I can be classified as a radical ideal $\sqrt{I}$ if

$$I = \sqrt{I} = \{f \in R \mid f^n \in I \text{ for some } n\}$$

Additionally, we would like to define for an affine variety $V \subseteq k^n$ the ideal associated with the variety is

$$\mathbf{I}(V) = \{f \in k^n \mid f(a_1, \ldots, a_n) = 0 \text{ for all } (a_1, \ldots, a_n) \in V\}$$

Ideal–Variety Correspondence:

1. The correspondence

   $$\textit{affine varieties} \xrightarrow{\mathbf{I}} \textit{radical ideals} \quad \text{and} \quad \textit{radical ideals} \xrightarrow{\mathbf{V}} \textit{affine varieties}$$

   gives an inclusion reversing bijection.
2. For any variety V it holds that $\mathbf{V}(\mathbf{I}(V)) = V$. Similarly for any ideal I, it satisfies $\mathbf{I}(\mathbf{V}(I)) = \sqrt{I}$

n-Dimensional Projective Space: The n-dimensional projective space over k is denoted $\mathbb{P}^n(k)$. This is,

$$\mathbb{P}^n(k) = (k^{n+1} \backslash \{0\}) / \sim .$$

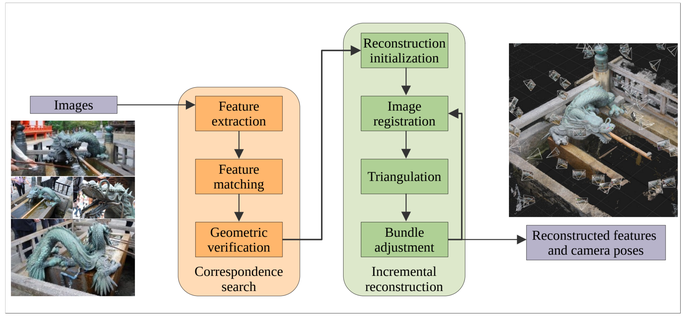
Where the equivalence relationship $\sim$ is denoted by

$$(x'_0 : \cdots : x'_n) \sim (x_0 : \cdots : x_n) \Longleftrightarrow (x'_0, \ldots, x'_n) = \lambda(x_0, \ldots, x_n) \quad \text{for some } \lambda \neq 0.$$

We will say that $(x_0 : \cdots : x_n)$ are homogeneous coordinates.

Reconstruction

Overview - Reconstruction is achieved by taking a series of 2D input images and matching features to reconstruct camera positions and 3D points.

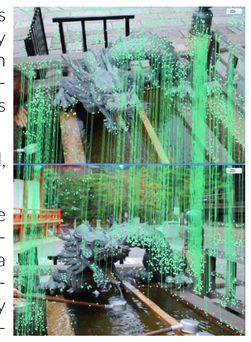


Feature Extraction/ Feature Matching - Detects prominent points or lines in each image using the SIFT algorithm (Scale-Invariant Feature Transform). Matches the extracted features across different images to identify common parts of the scene.

Geometric Verification - Not all feature matches from the previous step are correct, many are outliers. For verification, it is necessary to compute a geometric transformation to map features between images. If such a transformation exists, the features are geometrically verified. This is an algebraic geometry problem that involves solving systems of polynomial equations.

To account for outliers, the RANSAC sampling method is used, which increases robustness.

Reconstruction - Using a pair of geometrically verified images, the features of these images are matched to their corresponding 3D coordinates. Next, a new image is registered by estimating its camera pose using 2D-3D correspondences and solving polynomial equations. The 3D coordinates of the new image are reconstructed by triangulation and considering the Euclidean distance degree. To improve robustness, bundle adjustment is used to refine the camera poses and 3D coordinates, enhancing the overall reconstruction process.



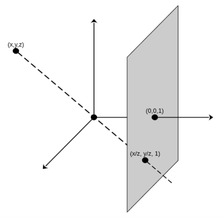
Camera Model

In order to produce a mapping between 3D and 2D we must specify a camera model used. We have selected a **pinhole camera** which uses homogeneous coordinates to classify points on the same ray. This produces a map, $\mathbb{P}^3 \to \mathbb{P}^2$.

Algebraically each camera C is given as a 3×4 matrix A of rank 3. We will use this notion of cameras to produce a fiber of a **joint camera map**

$$\Phi = \mathcal{X} \times \mathcal{C}_m \to \mathcal{Y}.$$

This maps a collection of $X \in \mathcal{X}$ 3D points and an n-tuple of cameras $(C_1, \ldots, C_n) \in \mathcal{C}_m$ to the n 2D images.



Multiview Geometry for Verification

Using the previously defined camera model, the **joint image** of the cameras is given by $\Phi_C(\mathbb{P}^3_{\mathbb{R}})$. The Zaraski closure of the joint imaage gives the **joint image variety**

$$M_C := \Phi_C(\mathbb{P}^3_{\mathbb{R}})$$

Muliview Constraints

Let $A_1, \cdots, A_m$ be the 3×4 matrices of rank three that define the cameras. $x_i = (x_{i1} : x_{i2} : x_{i3})$ represent the homogeneous coordinates on the i-th image of for the homogeneous. This produces the following $3m \times (m + 4)$ matrix:

$$M_A := \begin{bmatrix} A_1 & x_1 & 0 & \cdots & 0 \\ A_2 & 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_m & 0 & 0 & \cdots & x_m \end{bmatrix}$$

For 2 cameras ($m = 2$), M_A is a 6×6 matrix. Its bilinear determinant defines the multiview hypersurface M_C in $\mathbb{P}^2_{\mathbb{R}} \times \mathbb{P}^2_{\mathbb{R}}$. The matrix representation $F \in R^{3\times 3}$ of this bilinear form $x_2^\top F x_1$ is known as the **fundamental matrix**.

Symmetry of Joint Camera Map

The joint camera map Φ carries a symmetry, where a group G acts on the fibers of Φ.

Suppose that Cm consists of tuples of calibrated cameras $[R_1|t_1], \ldots, [R_m|t_m]$ where $R_i \in \mathrm{SO}(3)$ and $t_i \in \mathbb{R}^3$. Take $X = (\mathbb{P}^3)^n$ again to consist of tuples of 3D projective points. Now the relevant group is

$$G = \left\{ g \in \mathrm{GL}(4, \mathbb{R}) : g = \begin{bmatrix} R & t \\ 0 & \lambda \end{bmatrix} \text{ for some } R \in \mathrm{SO}(3), t \in \mathbb{R}^3, \lambda \in \mathbb{R} \setminus \{0\} \right\}$$

Two projective cameras example

Fix two projective pinhole cameras $C_1, C_2 : \mathbb{P}^3 \to \mathbb{P}^2$ represented by matrices A_1, A_2. Consider the set of corresponding point pairs defined as the multiview variety:

$$M_c = \{(x_1, x_2) \in \mathbb{P}^2 \times \mathbb{P}^2 : \exists X \in \mathbb{P}^3 \text{ s.t. } C1(X) = x_1, C2(X) = x_2\}.$$

This equation for M_c may be written as $x^T F x_1 = 0$ where for $1 \leq i, j \leq 3$,

$$F_{ij} = (-1)^{i+j} \det\left(\text{submatrix}\left([A_1|A_2], \{1,2,3,4\}, \{1,2,3,4,5,6\} \setminus \{i, j+3\}\right)\right).$$

References and Acknowledgments

[1] David A. Cox, John B. Little, and Donal O'Shea. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer, 2015.

[2] Joe Kileel and Kathlén Kohn. Snapshot of algebraic vision. *arXiv.org*, 2022.

[3] Miles Reid. *Undergraduate Algebraic Geometry*. Cambridge University Press, 2001.

A special thanks to our fantastic mentor, Greg McGrath, who made this project possible.

An Application of the Sparse Identification of Nonlinear Dynamics (SINDy) Algorithm

Han Zhang

University of California Santa Barbara

Intro to SINDy

Accurately modeling the nonlinear dynamics of a system from measurement data is a challenging yet vital topic. **Sparse Identification of Nonlinear Dynamics (SINDy)** is a data-driven method used for uncovering the underlying dynamics of a nonlinear system from observational data. The key idea behind SINDy is to express the dynamics of the system as a **sparse linear combination of a library of candidate functions or basis functions**. These basis functions can be chosen based on prior knowledge or intuition about the system, such as polynomials, trigonometric functions, or exponential functions.
SINDy has found applications in various fields, including **physics, biology, engineering, and finance**. It enables the discovery of simplified mathematical models that capture the essential dynamics of complex systems, even in the absence of complete knowledge about the system's governing equations.

The SINDy Method

Consider the following initial value problem

$$\frac{dx}{dt} = f(x,t), \qquad x(t_0) = x_0 \in \mathbb{R}^n$$

where f is a Lipschitz continuous function in x. SINDy addresses the problem of inferring the function f from data and takes advantage of the fact that many of these systems have dynamics with only **a few active terms** in the space of potential functions f. This avoids the intractable combinatorial search across all possible model structures. SINDy approximates f by a generalized linear model

$$f(x) \approx \Sigma_{k=1}^{p} \theta_k(x)\xi_k = \Theta(x)\xi$$

with the fewest nonzero terms in ξ as possible. In the formula above, $\theta_k(x)$ represents the candidate functions we fit to the data and ξ_k is the corresponding coefficients of these functions that demonstrate the weight of these functions on the overall dynamics. $\Theta(x)$ is a library of candidate nonlinear functions and may be constructed from **X**, e.g.,

$$\Theta(X) = [1\ X\ X^2\ \cdots\ X^d \cdots\ sin(X)\ e^X \cdots]$$

Then, it is possible to use **sparse regression** to solve for the relevant terms that are active in the dynamics.The dynamical system can then be represented as

$$\dot{X} = \Theta(X)\Xi$$

where Ξ contains a column vector ξ_k that represents the coefficients determining the active terms in the k^{th} row. A parsimonious model will provide an accurate model in fit with as few terms as possible in Ξ. Such a model may be identified using a **convex ℓ_1-regularized sparse regression**:

$$\xi_k = argmin_{\xi_k'} ||\dot{X}_k - \Theta(X)\xi_k'||_2 + \lambda||\xi_k'||_1$$

Here $\dot{X}_k$ is the k^{th} column of $\dot{X}$, and λ is a sparsity-promoting knob.
The sparse vector ξ_k may be synthesized into a dynamical system:

$$\dot{x}_k = \Theta(x)\xi_k$$

Note that x_k is the k^{th} element of **x** and Θ(x) is a row vector of symbolic functions of **x**.

Applying SINDy to the Duffing Equation

The Duffing equation is a mathematical model that describes the motion of a **damped, driven oscillator**. It is named after the German engineer **Georg Duffing**, who first introduced it in the early 20th century. The simplified version of this equation we explore takes the form:

$$\ddot{x} + \gamma\dot{x} + \beta x + \epsilon x^3 = 0.$$

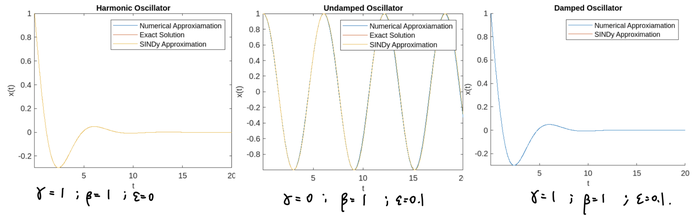
In this equation, x represents the displacement of the oscillator, t is time, overhead dot represents differentiation. γ controls the amount of **damping**, β controls the linear **stiffness** and ϵ controls the amount of **non-linearity** in the restoring force. We consider three dynamical systems generated by the equation. Our goal is to test **how well SINDy performs** in three different situations.
Case I: Damped Linear Oscillator(ϵ=0)
A harmonic oscillator is a fundamental concept that refers to a system exhibiting simple harmonic motion. The mathematical description of this system is:

$$\ddot{x} + \gamma\dot{x} + \beta x = 0$$

This ordinary differential equation can be solved algebraically. By assuming γ>0, β>0, $\gamma^2 - 4\beta$<0, and setting the initial conditions to be $x(0)$=1 and $\dot{x}$=0, the exact solution for this ODE is:

$$x(t) = e^{-\frac{\gamma t}{2}}\left(\cos\omega t + \frac{\gamma}{4\omega}\sin\omega t\right)$$

where $\omega^2 = \gamma^2 - 4\beta$. By comparing the $x-t$ graph of the **exact solution** to the $x-t$ graph generated by **SINDy algorithm** and the $x-t$ graph approximated by the **Runge-Kutta Method(ode45)** with the same initial conditions, we conclude that SINDy **is efficient** in identifying the dynamics of harmonic oscillators.



Case II: Undamped Nonlinear Oscillator(γ=0)
An undamped oscillator is a system that exhibits oscillatory motion without any dissipation or damping effects. The mathematical description of this system is:

$$\ddot{x} + \beta x + \epsilon x^3 = 0$$

This equation cannot be solved algebraically due to the involvement of the nonlinearity ϵx^3. Considering ϵ to be a small parameter, we apply **multiple-scale analysis** to construct a uniformly valid approximation to the solution of the undamped nonlinear oscillator:

$$x(t) = \cos\left(1 + \frac{3\epsilon}{8}\right)t + O(\epsilon) \qquad (1)$$

By comparing the $x-t$ graph of the **leading-order approximation** to the undamped nonlinear oscillator with the $x-t$ graph obtained by **SINDy** and numerics, we conclude that SINDy performs well in the undamped case.
Case III: Duffing oscillator($\gamma \neq 0,\ \epsilon \neq 0$)
The last case is more general with both γ and ϵ not equal to 0. Having built confidence in our methodology in the previous two cases, we rely solely on numerical simulations of the differential equation to test the SINDy method. We then compare the graph generated by **SINDy** with the **numerical approximation**. The two graphs suggest that SINDy yet again performs well.

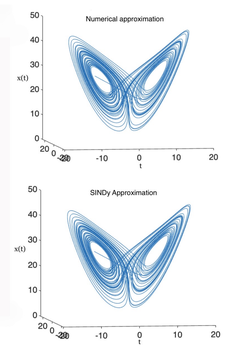
Applying SINDy to the Lorenz system

Lorenz system
The Lorenz system refers to a set of three differential equations that were discovered by the mathematician and meteorologist Edward Lorenz in 1963. These equations are used to describe a simplified model of atmospheric convection, which is the process by which heat is transferred through the motion of a fluid. It is a well-known model in chaos theory.
The system is defined by the following equations:

$$\begin{aligned} x' &= \sigma(y - x), \\ y' &= x(\rho - z) - y, \\ z' &= xy - \beta z. \end{aligned}$$

In these equations, x, y, and z represent the variables that describe the state of the system over time, and t represents time itself. σ, ρ, and β are parameters that determine the behavior of the system. We simulate the system with **ode45**, and show the resulting chaotic trajectory of the system upright corner of the figure below.
We then apply SINDy to the data generated by the numerics. We choose all polynomials up to 3^{rd} order as our library of functions. The output of the SINDy algorithm is a sparse matrix of coefficients Ξ. We show the coefficient matrix Ξ found by the SINDy algorithm below:

	'xdot'	'ydot'	'zdot'
'1'	[0]	[0]	[0]
'x'	[-10.0000]	[28.0000]	[0]
'y'	[10.0000]	[-1.0000]	[0]
'z'	[0]	[0]	[-2.6667]
'xx'	[0]	[0]	[0]
'xy'	[0]	[0]	[1.0000]
'xz'	[0]	[-1.0000]	[0]
'yy'	[0]	[0]	[0]
'yz'	[0]	[0]	[0]
'zz'	[0]	[0]	[0]
'xxx'	[0]	[0]	[0]
'xxy'	[0]	[0]	[0]
'xxz'	[0]	[0]	[0]
'xyy'	[0]	[0]	[0]
'xyz'	[0]	[0]	[0]
'xzz'	[0]	[0]	[0]
'yyy'	[0]	[0]	[0]
'yyz'	[0]	[0]	[0]
'yzz'	[0]	[0]	[0]
'zzz'	[0]	[0]	[0]



Comparing the two graphs, we conclude that SINDy works efficiently to identify the dominant terms that account for the observed behaviors of the Lorenz system.

Acknowledgements

I express my gratitude to my mentor Jimmie Adriazola for his invaluable guidance and unwavering support throughout the project as well as the UCSB Directed Reading Program for the opportunity to work on this project.

References

1. S.L. Brunton, J.L. Proctor, J.N. Kutz(2016), *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, National Acad Sciences.
2. S.L. Brunton J.N. Kutz (2019), *Data-driven Science and engineering: Machine Learning, Dynamical Systems, and Control*, Control-Cambridge University Press, NY.

An Introduction to Clifford Algebra

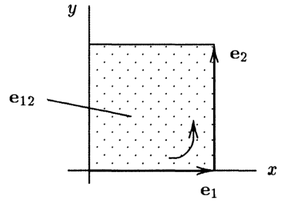
Jiayu Bian, Teo Zeng, Yoonsoo Cha
University of California Santa Barbara

Definitions and Notations

- $\mathbf{e_1}, \mathbf{e_2} \in \mathbb{R}^2$ are said to be **orthonormal unit vectors** if
$$|\mathbf{e_1}| = |\mathbf{e_2}| = 1, \quad \langle \mathbf{e_1}, \mathbf{e_2} \rangle = 0,$$
where $|\cdot|$ is the norm and $\langle \cdot, \cdot \rangle$ is the inner product. They form a basis of $\mathbb{R}^2$, and geometrically this means that $\mathbf{e_1}$ and $\mathbf{e_2}$ have length 1 and are perpendicular to each other.
- For $\mathbf{a} = a_1\mathbf{e_1} + a_2\mathbf{e_2}$ and $\mathbf{b} = b_1\mathbf{e_1} + b_2\mathbf{e_2}$, the **dot product** is defined by
$$\mathbf{a} \cdot \mathbf{b} = a_1b_1 + a_2b_2.$$
- For $\mathbf{a} = a_1\mathbf{e_1} + a_2\mathbf{e_2}$ and $\mathbf{b} = b_1\mathbf{e_1} + b_2\mathbf{e_2}$, the **bivector** is defined by
$$\mathbf{a} \wedge \mathbf{b} = (a_1b_2 + a_2b_1)e_{12}$$

Bivectors on $\mathbb{R}^2$

Given a vector space $\mathbb{R}^2$, the addition and scalar multiplication of the vectors are well-defined. Then a natural question to ask is: can we define a vector multiplication on $\mathbb{R}^2$? Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, we want the multiplication preserves the norm, i.e.,
$$|\mathbf{ab}| = |\mathbf{a}||\mathbf{b}|,$$
and it is distributive and associative.
Let $\mathbf{e_1}, \mathbf{e_2}$ be the orthonormal unit vectors on $\mathbb{R}^2$. The norm of a vector
$$\mathbf{r} = r_1\mathbf{e_1} + r_2\mathbf{e_2}$$
is defined by
$$|\mathbf{r}| = \sqrt{r_1^2 + r_2^2}.$$
If $\mathbf{r}$ is multiplied by itself, we require that $\mathbf{r}^2 = |\mathbf{r}|^2$, that is, we want
$$\begin{aligned}(r_1\mathbf{e_1} + r_2\mathbf{e_2})^2 &= (r_1\mathbf{e_1} + r_2\mathbf{e_2})(r_1\mathbf{e_1} + r_2\mathbf{e_2}) \\ &= r_1^2\mathbf{e_1}^2 + r_1r_2(\mathbf{e_1e_2} + \mathbf{e_2e_1}) + r_2^2\mathbf{e_2^2} \\ &= r_1^2 + r_2^2.\end{aligned}$$
This equality is satisfied if the orthonormal unit vectors $\mathbf{e_1}$ and $\mathbf{e_2}$ satisfy the following properties:
$$\mathbf{e_1}^2 = \mathbf{e_2}^2 = 1, \quad \mathbf{e_1e_2} = -\mathbf{e_2e_1}.$$
Consequently, we have
$$(\mathbf{e_1e_2})^2 = \mathbf{e_1e_2e_1e_2} = \mathbf{e_1}(-\mathbf{e_1e_2})\mathbf{e_2} = -1,$$
which follows that $\mathbf{e_1e_2}$ is neither a scalar nor a vector of $\mathbb{R}^2$. This product is called a **bivector**, representing the oriented plane area of the square with sides $\mathbf{e_1}$ and $\mathbf{e_2}$. Denote $\mathbf{e_1e_2} = \mathbf{e_{12}}$, it is illustrated by the following figure from [1]



We defined the **Clifford product** of two vectors $\mathbf{a} = a_1\mathbf{e_1} + a_2\mathbf{e_2}$ and $\mathbf{b} = b_1\mathbf{e_1} + b_2\mathbf{e_2}$ to be
$$\mathbf{ab} = a_1b_2 + a_2b_2 + (a_1b_2 - a_2b_1)\mathbf{e_{12}} = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \wedge \mathbf{b}.$$

Reflections

Given two vectors $\mathbf{a}$ and $\mathbf{r}$ in $\mathbb{R}^2$, $\mathbf{r}$ has a parallel component to $\mathbf{a}$ denoted as $\mathbf{r}_\parallel$ that is given by the dot product of $\mathbf{r}$ and $\mathbf{a}$ multiplied by the vector $\mathbf{a}^{-1} = \frac{\mathbf{a}}{|\mathbf{a}|^2}$, that is
$$\mathbf{r}_\parallel = (\mathbf{r} \cdot \mathbf{a})\frac{\mathbf{a}}{|\mathbf{a}|^2} = (\mathbf{r} \cdot \mathbf{a})\mathbf{a}^{-1}$$
Also, $\mathbf{r}$ has a perpendicular component to $\mathbf{a}$ denoted as $\mathbf{r}_\perp$ that is given by
$$\mathbf{r}_\perp = \mathbf{r} - \mathbf{r}_\parallel = \mathbf{r} - (\mathbf{r} \cdot \mathbf{a})\mathbf{a}^{-1} = (\mathbf{ra} - \mathbf{r} \cdot \mathbf{a})\mathbf{a}^{-1} = (\mathbf{r} \wedge \mathbf{a})\mathbf{a}^{-1}$$
Thus the reflection of $\mathbf{r}$ denoted as $\mathbf{r}'$ can be obtained by decomposing $\mathbf{r} = \mathbf{r}_\parallel + \mathbf{r}_\perp$ and sending it to $\mathbf{r}' = \mathbf{r}_\parallel - \mathbf{r}_\perp$. Note that since $\mathbf{r}_\perp$ is a bivector,
$$\mathbf{r}_\perp = (\mathbf{r} \wedge \mathbf{a})\mathbf{a}^{-1} = -\mathbf{a}^{-1}(\mathbf{r} \wedge \mathbf{a}) = \mathbf{a}^{-1}(\mathbf{a} \wedge \mathbf{r}) = -(\mathbf{a} \wedge \mathbf{r})\mathbf{a}^{-1}$$
Then we can find two direct formulas for $\mathbf{r}'$ as
$$\begin{aligned}\mathbf{r}' &= \mathbf{r}_\parallel + \mathbf{r}_\perp \\ &= (\mathbf{r} \cdot \mathbf{a})\mathbf{a}^{-1} - (\mathbf{r} \wedge \mathbf{a})\mathbf{a}^{-1} \\ &= (\mathbf{r} \cdot \mathbf{a} - \mathbf{r} \wedge \mathbf{a})\mathbf{a}^{-1} \\ &= (\mathbf{a} \cdot \mathbf{r} + \mathbf{a} \wedge \mathbf{r})\mathbf{a}^{-1} \\ &= \mathbf{ara}^{-1}\end{aligned}$$
$$\begin{aligned}\mathbf{r}' &= (\mathbf{r} \cdot \mathbf{a} - \mathbf{r} \wedge \mathbf{a})\mathbf{a}^{-1} \\ &= (2\mathbf{r} \cdot \mathbf{a} - \mathbf{ra})\mathbf{a}^{-1} \\ &= 2\frac{\mathbf{a} \cdot \mathbf{r}}{\mathbf{a}^2}\mathbf{a} - \mathbf{r}\end{aligned}$$
With the commutative properties of Clifford products,
$$\mathbf{ar}_\parallel\mathbf{a}^{-1} = \mathbf{r}_\parallel\mathbf{aa}^{-1} = \mathbf{r}_\parallel$$
and
$$\mathbf{ar}_\perp\mathbf{a}^{-1} = -\mathbf{r}_\perp\mathbf{aa}^{-1} = -\mathbf{r}_\perp$$
which yields to formula $\mathbf{r}' = \mathbf{ara}^{-1}$.

Reflections and Rotation in 3D

In the Euclidean space $\mathbb{R}^3$ the vectors $\mathbf{r}$ and $\mathrm{ar}^{-1} = 2(\mathbf{a} \cdot \mathbf{r})\mathbf{a}^{-1} - \mathbf{r}$ are symmetric with respect to the axis $\mathbf{a}$. The opposite of $\mathbf{ar}^{-1}$, the vector
$$-\mathbf{ara}^{-1} = \mathbf{r} - 2\frac{\mathbf{a} \cdot \mathbf{r}}{\mathbf{a}^2}\mathbf{a}$$
is obtained by reflecting $\mathbf{r}$ across the mirror perpendicular to $\mathbf{a}$ [reflection across the plane $\mathbf{ar}_{123}$]
Consider a vector $\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$ and the bivector $\mathbf{ae}_{123} = a_1\mathbf{e}_{23} + a_2\mathbf{e}_{31} + a_3\mathbf{e}_{12}$ dual to $\mathbf{a}$. The vector a has positive square
$$\mathbf{a}^2 = |\mathbf{a}|^2, \quad \text{where} \quad |\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2},$$
but the bivector $\mathbf{ae}_{123}$ has negative square
$$(\mathbf{ae}_{123})^2 = -|\mathbf{a}|^2$$
It follows that
$$\exp(\mathbf{ae}_{123}) = \cos\alpha + \mathbf{e}_{123}\frac{\mathbf{a}}{\alpha}\sin\alpha$$
where $\alpha = |\mathbf{a}|$. A spatial rotation of the vector $\mathbf{r} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$ around the axis a by the angle α is given by
$$\mathbf{r} \to a\mathbf{r}a^{-1}, \quad a = \exp\left(\frac{1}{2}\mathbf{ae}_{123}\right).$$
The sense of the rotation is clockwise when regarded from the arrow-head of a. The axis of two consecutive rotations around the axes $\mathbf{a}$ and $\mathbf{b}$ is given by the Rodrigues formula
$$\mathbf{c}' = \frac{\mathbf{a}' + \mathbf{b}' + \mathbf{a}' \times \mathbf{b}'}{1 - \mathbf{a}' \cdot \mathbf{b}'} \quad \text{where} \quad \mathbf{a}' = \frac{\mathbf{a}}{\alpha}\tan\frac{\alpha}{2}$$
This result is obtained by dividing both sides of the formula
$$\exp\left(\frac{1}{2}\mathbf{ce}_{123}\right) = \exp\left(\frac{1}{2}\mathbf{be}_{123}\right)\exp\left(\frac{1}{2}\mathbf{ae}_{123}\right)$$
by their scalar parts and then by inspecting the bivector parts.

Linear Space of Bivectors In $\mathbb{R}^3$

Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in \mathbb{R}^3$ be orthonormal unit vectors, so they form a basis of $\mathbb{R}^3$. The bivectors
$$\mathbf{e}_1 \wedge \mathbf{e}_2, \mathbf{e}_1 \wedge \mathbf{e}_3, \mathbf{e}_2 \wedge \mathbf{e}_3$$
form a basis for the linear space of bivectors, denoted by $\bigwedge^2 \mathbb{R}^3$. The inner product on Euclidean space $\mathbb{R}^3$ can be extended to a symmetric bilinear product (an inner product) on $\bigwedge^2 \mathbb{R}^3$, given by
$$\langle \mathbf{x}_1 \wedge \mathbf{x}_2, \mathbf{y}_1 \wedge \mathbf{y}_2 \rangle = \begin{vmatrix} \mathbf{x}_1 \cdot \mathbf{y}_1 & \mathbf{x}_1 \cdot \mathbf{y}_2 \\ \mathbf{x}_2 \cdot \mathbf{y}_1 & \mathbf{x}_2 \cdot \mathbf{y}_2 \end{vmatrix}.$$
In particular, we have
$$\langle \mathbf{a} \cdot \mathbf{b} \rangle = |\mathbf{a}|^2|\mathbf{b}|^2 - (\mathbf{a} \cdot \mathbf{b})^2.$$

The Hodge Dual

The Hodge dual sending a vector $\mathbf{a} \in \mathbb{R}^3$ to a bivector $\star\mathbf{a} \in \bigwedge^2 \mathbb{R}^3$, defined by
$$\mathbf{b} \wedge \star\mathbf{a} = (\mathbf{b} \cdot \mathbf{a})\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3 \text{ for all } \mathbf{b} \in \mathbb{R}^3$$
The Hodge dual depends not only on the metric but also on the choice of orientation - it is customary to use a right-handed and orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ Thus, we have assigned to each vector
$$\mathbf{a} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3 \in \mathbb{R}^3$$
a bivector
$$\mathbf{A} = \star\mathbf{a} = a_1\mathbf{e}_2 \wedge \mathbf{e}_3 + a_2\mathbf{e}_3 \wedge \mathbf{e}_1 + a_3\mathbf{e}_1 \wedge \mathbf{e}_2 \in \bigwedge^2 \mathbb{R}^3.$$
Using the induced metric on the bivector space $\Lambda^2\mathbb{R}^3$ we can extend the Hodge dual to a mapping sending a bivector $\mathbf{A} \in \Lambda^2\mathbb{R}^3$ to a vector $\star\mathbf{A} \in \mathbb{R}^3$, defined by
$$\mathbf{B} \wedge \star\mathbf{A} = \langle \mathbf{B}, \mathbf{A} \rangle \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3 \text{ for all } \mathbf{B} \in \bigwedge^2 \mathbb{R}^3.$$
Using duality, the relation between the cross product and the exterior product can be written as
$$\mathbf{a} \wedge \mathbf{b} = \star(\mathbf{a} \times \mathbf{b}),$$
$$\mathbf{a} \times \mathbf{b} = \star(\mathbf{a} \wedge \mathbf{b}).$$

Acknowledgements

We would like to express my sincere gratitude to my mentor Alex for his guidance and support throughout this project. He has been very generous with his time and expertise, and has provided us with valuable feedback and insights. We are very fortunate to have such a dedicated and knowledgeable mentor.

References

[1] Pertti Lounesto. *Clifford Algebras and Spinors, Second edition*. The Press Syndicate of the University of Cambridge, 2001.

An Introduction to Hamiltonian Monte Carlo Methods

Siddharth Sabata
University of California Santa Barbara

Markov Chain Monte Carlo Methods

Markov Chain Monte Carlo methods are a type of algorithm that allow us to sample from a probability distribution without knowing what the distribution looks like. We'll begin by defining the individual parts that make up MCMC: Markov Chains and Monte Carlo methods.

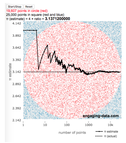
Markov Chains A Markov Chain is a random process that undergoes state changes. The chain also has a property, called the Markov property, where the probability of moving to the next state is only dependent on the current state. Another way to say this is "in order to know the future, the knowledge of the past does not add anything to the knowledge of the present" ([1, p. 45]).

More formally, the discrete, or finite, case for a Markov Chain is defined as a discrete-time process $\{X_n\}_{n\geq 0}$, i.e. a collection of random variables with the index n usually representing time, with values in a countable space E is a Markov chain if for all $n \geq 0$ and all states $i_0, i_1, \ldots, i_{n-1}, i, j \in E$,

$$\mathbb{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i),$$

where $\mathbb{P}(Y = j | X = i)$ means the conditional probability of the event $\{Y = j\}$ given $\{X = i\}$ ([1, p. 46 (Definition 2)]).

Monte Carlo Methods Monte Carlo methods are algorithms that estimate quantities that are too difficult to obtain analytically or by discretization. This is done through repeated random sampling and averaging the results. Some common examples are estimating π or integrals.



Markov Chain Monte Carlo Algorithm MCMC methods were proposed by Metropolis and his colleagues in 1953. They wanted to develop a method to estimate an unknown target distribution using Markov Chains.

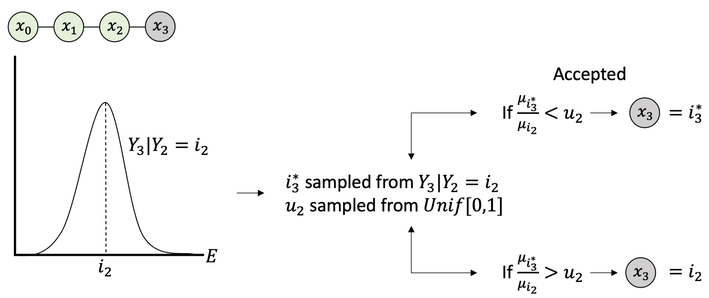
The algorithm for the discrete, or finite, case is defined as the following:

- Choose a value for i_0 for X_0 (randomly or e.g $i_0 = 0$).
- Once values $i_0, \ldots, i_n$ of $X_0, \ldots, X_n$, respectively, have been found:
  - Generate a proposed value $i^*_{n+1} \in E$ from an auxiliary distribution $Y_{n+1} | Y_n = i_n$.
  - If $\mu_{i^*_{n+1}}/\mu_{i_n} > u_n$ set $X_{n+1} = i^*_{n+1}$; in this case we say that the proposal is accepted. Else set $X_{n+1} = i_n$ and we say the proposal is rejected.

Here, E is a discrete state space, i.e., the set of all possible events, $\mu_{i^*_{n+1}}$ and μ_{i_n} are the value of the target distribution at i^*_{n+1} and i_n respectively, and u_n is drawn from the uniform distribution on $[0, 1]$ to carry out acceptance/rejection step with probability $\min\{1,\ \mu_{i^*_{n+1}}/\mu_{i_n}\}$. ([1, p. 55]).

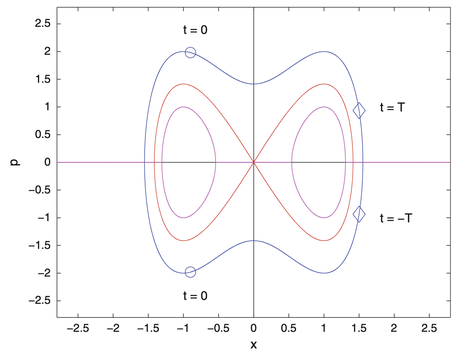
Typically, first few hundreds or thousands of samples will not be close to those drawn from the target distribution. Once the Markov Chain converges to what is called its *stationary distribution*, a point at which the probability distribution representing the movement from one state to the next does not change, the Markov Chain transition is just like "independent sampling" from our target distribution μ.

Visual Representation of MCMC Algorithm



Hamiltonian Dynamics

Visual Representation of the Phase Space [1, p. 74 (Figure 13)]



Hamiltonian Dynamics Hamiltonian dynamics is another way to look at classical mechanics in physics. This framework describes how a system changes over time based on the Hamiltonian function H which represents the total mechanical energy of a system, which equals the sum of the kinetic and potential energy (represented by V) of the system of ν particles:

$$H = \sum_i \frac{1}{2m_i} \mathbf{p}_i^T \mathbf{p}_i + V(\mathbf{r}_1, \ldots, \mathbf{r}_\nu),$$

where $\mathbf{r_i}$, $\mathbf{p_i}$ and m_i correspond to the position, momentum, and mass of i-th particle ([1, p. 70]).

Hamiltonian Equations From now on, we discuss the case with only one particle for simplicity of presentation. In the phase space space $\mathbb{R}^D$, $D = 2d$, $(\mathbf{p}, \mathbf{x}) \in \mathbb{R}^D$, to each smooth real-valued function $H = H(\mathbf{p}, \mathbf{x})$ (Hamiltonian), the corresponding system of first order differential equations, called the *canonical* or *Hamilton's* equations describes the time evolution of the system ([1, p. 71]):

$$\frac{d}{dt} p_j = -\frac{\partial H}{\partial x_j}, \quad \frac{d}{dt} x_j = +\frac{\partial H}{\partial p_j}, \quad j = 1, \ldots, d.$$

Flow The flow of a Hamiltonian system is denoted with $\{\Phi_t\}_{t\in\mathbb{R}}$. Φ_t is a map in the phase space, $\Phi_t : \mathbb{R}^D \to \mathbb{R}^D$, that is defined as follows: $\Phi_t(\mathbf{p}_0, \mathbf{x}_0)$ is the solution $(\mathbf{p}(t), \mathbf{x}(t))$ at time t of the canonical equation with the initial value $(\mathbf{p}_0, \mathbf{x}_0)$ at $t = 0$. Basically, Φ_t tells us how the system evolves over time ([1, p. 71-72]).

Property 1: Conservation of Energy The function H is a conserved quantity of the Hamilton's equations. Along solutions, we have

$$\frac{d}{dt} H(\mathbf{p}(t), \mathbf{x}(t)) = \sum_j \left(\frac{\partial H}{\partial p_j} \frac{d}{dt} p_j + \frac{\partial H}{\partial x_j} \frac{d}{dt} x_j \right) = \sum_j \left(-\frac{\partial H}{\partial p_j} \frac{\partial H}{\partial x_j} + \frac{\partial H}{\partial x_j} \frac{\partial H}{\partial p_j} \right) = 0.$$

Therefore,

$$H(\mathbf{p}(t), \mathbf{x}(t)) = H(\mathbf{p}(0), \mathbf{x}(0)).$$

In other words, when $d = 1$, the dynamics follows a contour curve of H on the phase space [1, p. 72] (see the figure above).

Property 2: Conservation of Volume If we take the area of a specific region in the phase space and evolve it through time, the area remains the same even though the shape of the area might change ([1, p. 72-73]).

Property 3: Reversibility This property states that if we go backwards in the time evolution of our system, the system's motion will also go backwards. This means that if we knew the state of the system at any point in time, we could run it backwards and find the state of the system at a previous point in time ([1, p. 73]).

These properties lead to many desirable aspects, when combined with the MCMC, including the existence of a stationary distribution and high acceptance rate of proposed states.

Hamiltonian Monte Carlo

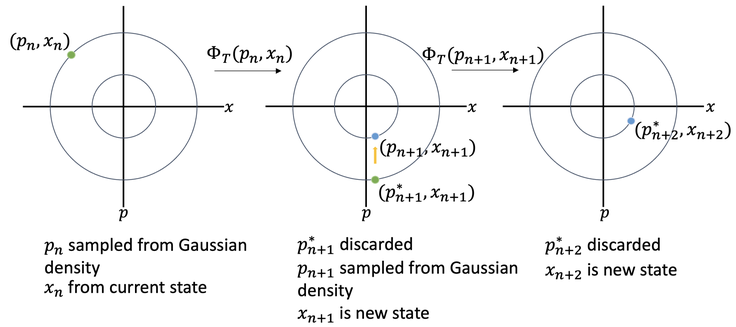
We can combine MCMC and Hamiltonian dynamics together to form Hamiltonian Monte Carlo methods. These methods use Hamiltonian dynamics for the Markov chain transition. We write the target density $\pi(\mathbf{x})$ in the state space $\mathbb{R}^D$ as $\exp(-V(\mathbf{x}))$. We can think of $\mathbf{x} \in \mathbb{R}^d$ as the position of our mechanical system, $V(\mathbf{x})$ as corresponding potential energy, and $\mathbf{p} \in \mathbb{R}^d$ as momentum. $\mathbf{p}$ is usually endowed a Gaussian density with the identity covariance matrix, which corresponds to mass being 1 in the canonical equation. With this set up, we can construct a Markov Chain in $\mathbb{R}^d$ ([1, p. 78]). For the simplicity of presentation, we further assume $d = 1$ below.

HMC Algorithm (analytic flow version) Define the transitions $x_n \mapsto x_{n+1}$ in the state space $\mathbb{R}^d$ by the following procedure:

- Draw p_n from a Gaussian density.
- Find $(p^*_{n+1}, x_{n+1}) = \Phi_T(p_n, x_n)$, where Φ_T is the T-flow of the canonical system (the Hamiltonian equations) with Hamiltonian function H.

Then $x_n \mapsto x_{n+1}$ defines a Markov chain in $\mathbb{R}^d$ that has the target $\pi(x) \propto \exp(-V(x))$ as an invariant probability distribution ([1, p. 78-79 (Theorem 8)]).

Visual Representation of Sampling from the Hamiltonian Phase Space



Why is this helpful? MCMC algorithms are typically inefficient due to low acceptance rates, resulting in the Markov Chain converging to its stationary distribution taking a long time. On top of this, MCMC algorithms usually suffer from high autocorrelation which comes from each sample being too close to the previous one. This also leads to the Markov Chain taking a longer time to converge to its stationary distribution.

Notice that the momentum p^*_{n+1} is refreshed every iteration. This makes it possible to explore different energy levels, which likely leads to wider range of state space. Also note that HMC does not even feature acceptance/rejection step because the acceptance rate is always 1: the ratio of the current and proposed energy levels. This dramatically reduces autocorrelation, which is high when the states are repeated as is the case when the proposed states are often rejected.

References

[1] J. M. Sanz-Serna. "Markov chain Monte Carlo and numerical differential equations". In: *Current challenges in stability issues for numerical differential equations*. Vol. 2082. Lecture Notes in Math. Springer, Cham, 2014, pp. 39–88.

Acknowledgements

Thank you to Jea-Hyun Park for his mentorship and guidance and the UC Santa Barbara Directed Reading Program for this opportunity.

Optimal Transport

Chunting Zheng[1] Mentor: Đorđe Nikolić[1]

[1]University of California - Santa Barbara

Introduction

Optimal transport (OT) originates from Gaspard Monge who considered the problem of redistributing (transporting and reshaping) a pile of sand to form a fortification with minimal effort.

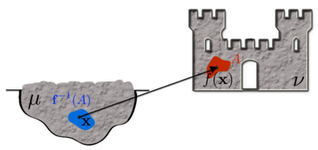


Fig. 1: Optimal transport problem [1].

Assignment and Monge Problem

Given an $n \times n$ cost matrix $C = [c_{ij}]$, the optimal assignment problem is to find a permutation σ of $\{1, \dots, n\}$ that solve

$$\min_{\sigma \in \mathrm{Perm}(n)} \frac{1}{n} \sum_{i=1}^{n} \mathbf{C}_{i,\sigma(i)}$$

Note that this assignment is not unique.

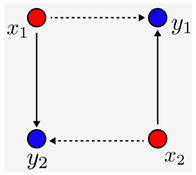


Fig. 2: Non-unique assignment [2].

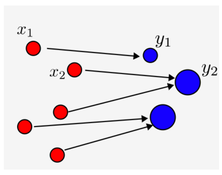


Fig. 3: Monge problem [2].

Extending the above definition to a slightly larger family of histograms, we obtain the Monger problem. It seeks a surjective map $T : \{x_1, \dots, x_n\} \to \{y_1, \dots, y_m\}$, T must verify:

$$\mathbf{b}_j = \sum_{i:T(x_i)=y_i} \mathbf{a}_i, \quad \forall j \in \{1, \dots, m\}$$

where $\mathbf{a}$ and $\mathbf{b}$ are weights associated with $\mathbf{x}$ and $\mathbf{y}$, respectively.

The Monge formulation does not allow for the splitting of mass.

Kantorovich Relaxation

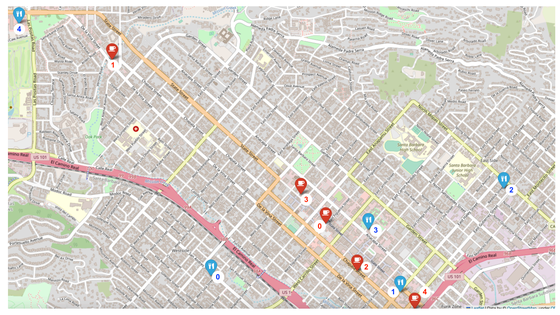
The key idea of Kantorovich formulation is to **relax the deterministic nature of transportation!** Replacing the permutation σ by a coupling matrix $\mathbf{P} = \mathbf{P}_{ij} \in \mathbb{R}_+^{n \times m}$, where $\mathbf{P}_{ij}$ is the mass found at x_i toward y_j. Kantorovich's optimal transport problem is

$$\mathbf{L}_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b})} \sum_{i,j} \mathbf{C}_{ij}\mathbf{P}_{ij} = \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{C}, \mathbf{P} \rangle \quad (\star)$$

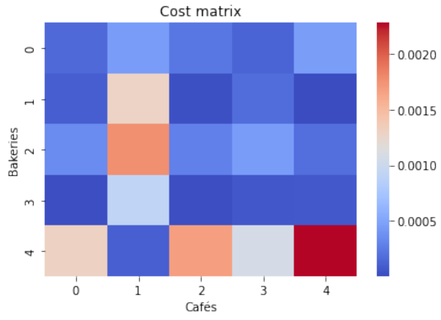
where

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

Bakeries and Cafés in Santa Barbara

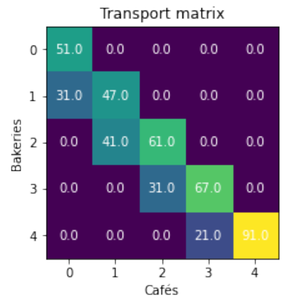


- Bakery production: [51, 78, 102, 98, 112]
- Cafe sale: [82, 88, 92, 88, 91]
- Total croissants: 441



North-West Corner Rule

The rule starts by setting $\mathbf{P}_{1,1}$ to $\min(\mathbf{a}_1, \mathbf{b}_1)$. At each step, the entry $\mathbf{P}_{i,j}$ is chosen to saturate either the row constraint at i, the column constraint at j, or both if possible. The rule proceeds until $\mathbf{P}_{n,m}$ has received a value.



Auction Algorithm

The auction algorithm maintains a partial matching (σ, S) and prices function p that together satisfy **ε-complementary slackness** (ε-CS),

$$c\big(x, \sigma(x)\big) + p\big(\sigma(x)\big) \leq \min_{y \in Y} [c(x, y) + p(y)] + \varepsilon$$

for all $x \in S \subset X$ and $\sigma : S \to Y$ is an injective map.
At the end of the execution, σ is a bijection, and $(\sigma, \mathbf{p})$ satisfy the ε-CS condition. The following algorithm is taken from [3].

Algorithm 1 Auction algorithm

Require: $c, \varepsilon, p = 0$
  $S \leftarrow \emptyset$
  while $\exists x \in X \setminus S$ **do**
    $y_0 \leftarrow \arg\min_{y \in Y} c(x, y) + p(y)$
    $y_1 \leftarrow \arg\min_{y \in Y \setminus \{y_0\}} c(x, y) + p(y)$
    $p(y_0) \leftarrow p(y_0) + \big(c(x, y_1) + p(y_1)\big) - \big(c(x, y_0) + p(y_0)\big) + \varepsilon$
    if $\exists x' \in X$ s.t. $\sigma(x') = y_0$ **then**
      $S \leftarrow S \setminus \{x'\}$
    end if
    $S \leftarrow S \cup \{x\}, \quad \sigma(x) \leftarrow y_0$
  end while
  return σ, p

Entropy Regularization

The **entropy of a coupling matrix $\mathbf{P}$** is defined as

$$\mathbf{H}(\mathbf{P}) = -\sum \mathbf{P}_{i,j} \left(\log(\mathbf{P}_{i,j}) - 1 \right)$$

Use $-\mathbf{H}$ as a regularizing function to approximate solutions to the original transport problem ($\star$):

$$\mathbf{L}^{\varepsilon}_{\mathbf{C}(\mathbf{a},\mathbf{b})} = \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P}, \mathbf{C} \rangle - \varepsilon \mathbf{H}(\mathbf{P})$$

The solution is unique and has the form

$$\mathbf{P}_{i,j} = \mathbf{u}_i \mathbf{K}_{i,j} \mathbf{v}_j$$

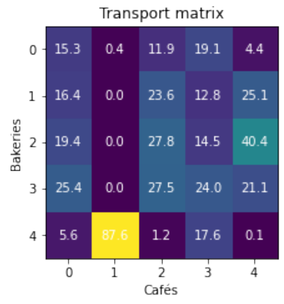
$\forall (i, j) \in \{1, \dots, n\} \times \{1, \dots, m\}$, where $\mathbf{K}_{i,j} = e^{-\frac{\mathbf{C}_{i,j}}{\varepsilon}}$, $\mathbf{u}$ and $\mathbf{v}$ are unknown scaling variables.

Sinkhorn's Algorithm

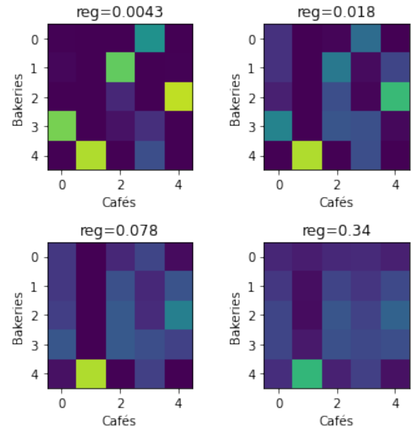
These two updates define Sinkhorn's algorithm:

$$\mathbf{u}^{(\ell+1)} = \frac{\mathbf{a}}{\mathbf{K}\mathbf{v}^{\ell}} \quad \text{and} \quad \mathbf{v}^{(\ell+1)} = \frac{\mathbf{b}}{\mathbf{K}^T\mathbf{u}^{(\ell+1)}},$$

initialized with an arbitrary positive vector $\mathbf{v}^{(0)} = \mathbb{1}_m$.



Varying the regularization parameter in Sinkhorn,



References

[1] Chad Atalla, Sai Aditya, and Xiao Sai. Lecture 2/28: Optimal Transport. 2019.

[2] Gabriel Peyré and Marco Cuturi. Computational optimal transport. arXiv:1803.00567, 2018.

[3] Remi Flamary, Nicolas Courty, and Aurelie Boisbunon. Introduction to Optimal Transport with Python.

[4] Quentin Merigot and Boris Thibert. Optimal transport: discretization and algorithms. arXiv:2003.00855, 2020.

[5] Filippo Santambrogio. Optimal transport for applied mathematicians. Springer, 2015.

Palindromic Parking Functions Polytope

Yanru Liu, Mentored by **Sam Sehayek**

2023 Mathematics Direct Reading Program. University of California, Santa Barbara

Palindromic Parking Functions

A **parking function** of length n is a function $f : \{1, \cdots, n\} \to \{1, \cdots, n\}$ which has the property that the list $(f(1), \cdots, f(n))$ can be rearranged into ascending order $(a_1, \cdots, a_n)$ such that $a_i \leq i$, where $f(j) = a_i$ for some $j \in \{1, \cdots, n\}$. Let $\mathcal{PF}_n$ denote the set of parking functions of length n.

Parking functions have applications in various areas of combinatorics, such as graph theory and algebraic combinatorics. They are used to study problems related to permutations, lattice paths, and Catalan numbers, among others.

The symmetric group Sym_n acts on $\mathcal{PF}_n$ by permutating coordinates. The set fixed by the permutation

$$(1, n)(2, n-1)\cdots\left(\left\lfloor\frac{n}{2}\right\rfloor \left\lceil\frac{n}{2}\right\rceil\right)$$

is called the **palindromic parking function** (or **symmetric parking function**). That is, the palindromic parking functions are parking functions that read the same from both of the forward and backward directions. Let $\mathcal{PPF}_n$ denote the set of palindromic parking functions of length n.

Based on the above definition, a conclusion follows immediately:

Theorem. *The order of palindromic parking functions of length n is*

$$|\mathcal{PPF}_n| = (n+1)^{\lfloor\frac{n+1}{2}\rfloor}.$$

Consider a $\sigma \in \mathrm{Sym}_n$ fixing $f \in \mathcal{PF}_n$. Note that there are $\lfloor(n+1)/2\rfloor$ cycles in σ. Therefore, the number of f fixed by σ is

$$\frac{(n+1)^{\lfloor\frac{n+1}{2}\rfloor}}{n+1} = (n+1)^{\lfloor\frac{n+1}{2}\rfloor - 1} = (n+1)^{\lfloor\frac{n-1}{2}\rfloor}.$$

Parking Function Polytopes

In 2020, in the journal *American Math Monthly*, Richard Stanley proposed questions about **parking function polytope** $\mathcal{P}_n$, an n-dimensional polytope defined as the convex hull in $\mathbb{R}^n$ of all parking functions of length n. The questions are:

1. the number of vertices of $\mathcal{P}_n$;
2. the number of $(n-1)$ dimensional faces (facets) of $\mathcal{P}_n$;
3. the number of integer points in $\mathcal{P}_n$;
4. the n-dimensional volume of $\mathcal{P}_n$.

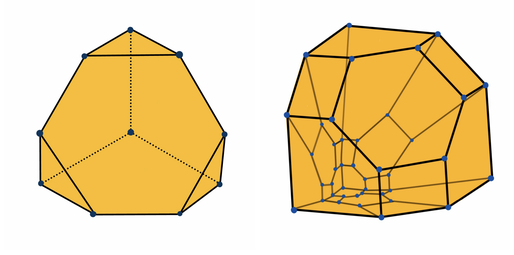


Fig. 1: Parking function polytope $\mathcal{P}_3$ (left), where the hexagonal facet is the regular permutahedron and the three triangular facets are copies of $\mathcal{P}_2$; and the Schlegel diagram of $\mathcal{P}_4$ (right).

In Aruzhan Amanbayeva and Danielle Wang's paper *The Convex Hull of Parking Functions of Length n*, they proved that $\mathcal{P}_n$ is a simple polytope and found the f-vector of $\mathcal{P}_n$.

Palindromic Parking Function Polytope

Motivated by Richard Stanley, we are interested in the f-vectors of **palindromic parking function polytopes** $\mathcal{PP}_n$. The ***f*-vector** of $\mathcal{PP}_n$ is the sequence

$$(1, f_0, f_1, \cdots, f_k),$$

where $k = \lfloor(n+1)/2\rfloor!$, and f_i denotes the number of i-dimensional faces. For example, the f-vector of $\mathcal{PP}_5$ and $\mathcal{PP}_6$ is $(1, 10, 15, 7, 1)$.

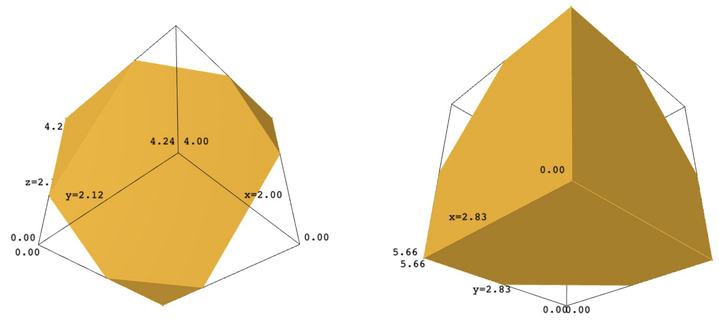


Fig. 2: Visualizations of $\mathcal{PP}_5$ (left) and $\mathcal{PP}_6$ (right).
$\mathcal{PP}_5$ and $\mathcal{PP}_6$ are combinatorically equivalent whose dimensions are both 3 and with 10 vertices.

Compare figure 1 and 2, we noted that $\mathcal{P}_3$ and $\mathcal{PP}_5$ & $\mathcal{PP}_6$ share the same number of vertices ($\# = 10$). This pattern still holds if we increase the length of parking function. We conclude that

$$\mathcal{P}_{\lceil n\rceil} \text{ and } \mathcal{PP}_n$$

share the same number of vertices. Moreover, the number of vertices follows the recursion

$$a(0) = 0,$$
$$a(n) = n \cdot a(n-1) + 1.$$

i.e., the sequence $(0, 1, 3, 10, 41, 206, \cdots)$.

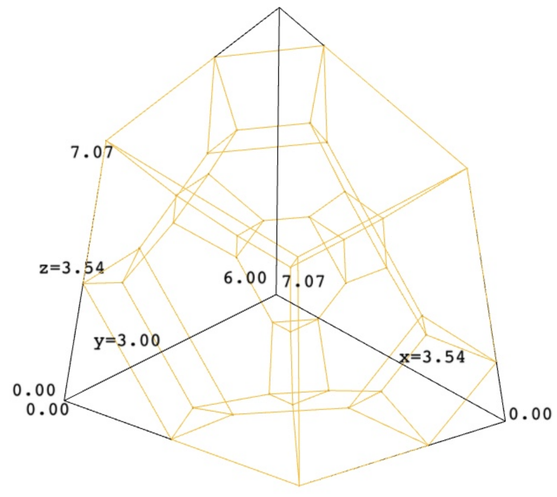


Fig. 3: The schlegel diagram of $\mathcal{PP}_7$ and $\mathcal{PP}_8$ of dimension 4 and 41 vertices.

A New Theorem

From the previous example, we know that $\mathcal{PP}_5$ has 10 vertices, while $|\mathcal{PPF}_5| = 36$, indicating that not all the palindromic parking functions are the vertices of $\mathcal{PP}_n$. The theorem we discovered provides which vertices are desired ones.

Theorem. *The vertices coming from palindromic parking functions are:*

- *if n is even, the vertices of $\mathcal{PPF}_n$ are $(\vec{a}, \overleftarrow{a})$, where $\vec{a}$ is of length $n/2$ and is a permutation of*

$$(\underbrace{1, \cdots, 1}_{k \text{ times}}, 2k+1, 2k+3, \cdots)$$

for $1 \leq k \leq n/2$.

- *if n is odd, the vertices are of the form $(\vec{a}, m, \overleftarrow{a})$, where $\vec{a}$ is of length $(n-1)/2$ and is a permutation of*

$$(\underbrace{1, \cdots, 1}_{k \text{ times}}, 2k+2, 2k+4, \cdots)$$

for $m = 1$; and

$$(\underbrace{1, \cdots, 1}_{k \text{ times}}, 2k+1, 2k+3, \cdots, \hat{m}, m+1, \cdots)$$

for $m \neq 1$. Where m is the middle term, an odd number smaller than n and $1 \leq k \leq (m-1)/2$.

To see how this theorem works, let us consider an example of length 5. $\mathcal{PP}_5$ is an odd case. For $m = 1$, we have vertices

$$(1,1,1,1,1), (4,2,1,2,4), (4,1,1,1,4), (2,4,1,4,2), (1,4,1,4,1);$$

and for $m \neq 1$, we have vertices

$$(4,1,3,1,4), (1,1,5,1,1), (3,1,5,1,3), (1,4,3,4,1), (1,3,5,3,1).$$

Based on this new theorem we discovered, the number of vertices of $\mathcal{PP}_n$ can be immediately counted:

Corollary. *The number of vertices of $\mathcal{PP}_n$ is given by*

$$|\mathcal{PP}_n| = \left\lfloor\frac{n+1}{2}\right\rfloor! \left(\frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{\left\lfloor\frac{n+1}{2}\right\rfloor!}\right).$$

Future Direction

We understand how to find the vertices and compute the number f_0. Using this, we are confident that we can compute the entire f-vector for $\mathcal{PP}_n$. Another future direction is to investigate the other polytopes that arise by considering different permutations.

Acknowledgements

I am immensely grateful to my mentor Sam Sehayek, for his profound knowledge, tremendous passion, and guidance in understanding the essence of mathematical research, which greatly influenced and shaped my life choices in this field.

Reference

[1] Aruzhan Amanbayeva and Danielle Wang. *The Convex Hull of Parking Functions of Length n*, 2021. DOI: 2212.06885.
[2] Mitsuki Hanada, John Lentfer, and Andres R. Vindas-Melendez. *Generalized Parking Function Polytope*, 2022. DOI: 104.08454.

Parameters Estimation of Cox-Ingersoll-Ross Model

Jiayue Chen, advised by Jimmie Adriazola

University of California, Santa Barbara

Cox–Ingersoll–Ross Model

The Cox–Ingersoll–Ross model (CIR), introduced by John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross in 1985, is a mathematical equation employed to simulate fluctuations in interest rates. CIR is a single-factor model as it describes interest movements as driven by a sole source of market risk. This model finds application in predicting future interest rate levels, allowing for the calculation of bond prices and the valuation of interest-rate financial derivatives.

The CIR model specifies that the instantaneous interest rate r_t evolves according to the stochastic differential equation:

$$dr_t = a(b - r_t)d_t + \sigma\sqrt{r_t}dW_t$$

- r_t = Instantaneous interest rate at time t
- W_t = Wiener Process (a Brownian motion random variable which models the random market risk factor)
- a = Rate of mean reversion
- b = Mean of the interest rate
- σ = Standard deviation of the interest rate(measure of volatility)

The square root element does not allow for negative rates and the drift factor, $a(b - r_t)$ ensures mean reversion of the interest rate towards the long-run value b, with the speed of adjustment governed by the strictly positive parameter a. Mean reversion in the CIR model suggests that when r is high, mean reversion tends to cause a negative drift; when r is low, mean reversion tends to cause it to have a positive drift, corresponding to real-life phenomenon.

Numerical Method to Solve Stochastic Differential Equations

The Euler–Maruyama method provides an approximate numerical solution for a stochastic differential equation (SDE). This method is the analog of the Euler method for ordinary differential equations. To develop an approximate solution on the interval $[c, d]$, assign a uniform grid of points:

$$c = t_0 < t_1 < t_2 < ... < t_n = d$$

Approximate x values

$$w_0 < w_1 < w_2 < ... < w_n$$

will be determined at the respective t points. Given the SDE initial value problem

$$\begin{cases} dX(t) = a(t, X)dt + b(t, X)dW_t \\ X(c) = X_c \end{cases}$$

Then we can compute the approximate solution as follows:

Euler-Maruyama Method:

$$w_0 = X_0$$

$$w_{i+1} = w_i + a(t_i, w_i)\Delta t_{i+1} + b(t_i, w_i)\Delta W_{i+1}$$

$$\Delta t_{i+1} = t_{i+1} - t_i$$

$$\Delta W_{i+1} = W(t_i + 1) - W(t_i)$$

Each random number ΔW_i is computed as

$$\Delta W_i = z_i\sqrt{\Delta t_i}$$

where z_i is chosen from $N(0, 1)$, normal distribution with mean 0 and standard deviation 1.

Numerical Solution to CIR model

Deterministic CIR: Without the Brownian motion term dW_t, we define the deterministic CIR model by the ODE:

$$\frac{dr}{dt} = a(b - r) + \sigma\sqrt{r}$$

Given an initial value $r(0)$, we are able to numerically solve it using the forward Euler Method

$$r_{n+1} = r_n + hf(t_n, r_n)$$
$$\text{where } f(t_n, r_n) = a(b - r_n) + \sigma\sqrt{r_n}$$

Stochastic CIR Using the Euler-Maruyama Method, we update the solution for the stochastic CIR model:

$$r_{i+1} = r_i + a(b - r_i)\Delta t_{i+1} + \sigma\sqrt{r_i}\Delta W_i$$

since $\Delta W_i = z_i\sqrt{\Delta t}$ and we have discretized the time interval with equal length of step and denoted as h, we can rewrite the formula in the form:

$$r_{i+1} = r_i + a(b - r_i)h + \sigma\sqrt{r_i h}z_i$$

Visualization of Numerical Solution to CIR model By implementing these two methods in Python, we simulate the deterministic and stochastic CIR models by setting parameters $a = 0.45, b = 1$, initial value $r(0) = 0.5$, step size $\delta t = 0.01$, ending time $T = 2$.

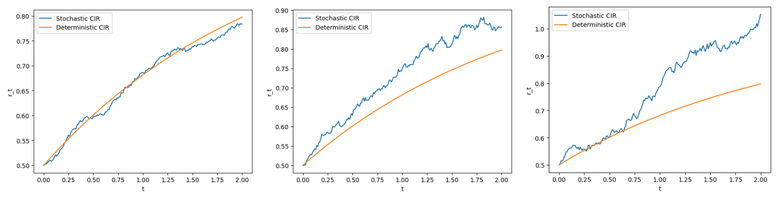


Figure 1. Simulation of stochastic CIR model with $\sigma = 0.03, 0.06, 0.09$

Parameters Estimation of CIR Model

We shift our focus to estimating the coefficients of the CIR model using historical data on interest rates. By addressing this inverse problem within the CIR model, it brings forth several implications, including:

1. **Forecasting:** Estimating the coefficients enables us to generate forecasts of future interest rate movements. By understanding the behavior of interest rates, we can make informed predictions about their future trajectory, aiding in decision-making and risk management.
2. **Calibration and Model Comparison:** Estimating the coefficients enables us to find optimal parameter values that minimize the disparity between the model's predictions and the actual data and allows for meaningful comparisons with alternative interest rate models, aiding in model selection and evaluation.

Figure 2. A real-life interest rate that can be modeled: FRED data of 30-Year Fixed Rate Mortgage Average in the United States

Method to Solve the Inverse Problem

Suppose we are given $T, \sigma,\ r(0)$ and N realizations of the stochastic CIR model, and we are interested in estimating the parameters a, b. That is, how do we estimate the respective paramters a and b? We propose to estimate a, b by using the following regression problem:

$$\min_{\xi} J(\xi) = \frac{1}{2N}\min_{\xi}\sum_{j=1}^{N}\int_0^T (r_j^{data} - r_\xi)^2 dt$$

- N - number of realizations of the stochastic CIR model that we have given
- r_j^{data} - the given realization of the stochastic CIR model
- r_ξ - solves the deterministic CIR model for parameter $\xi = (a, b)$
- $\int_0^T$ - sum the difference between the given data and the deterministic CIR model at each infinitesimal steps
- min - find ξ that minimizes the difference between r_j^{data} and r_ξ

Procedure

1. Write the functions to solve the stochastic and deterministic CIR model by using Euler-Maruyama and forward Euler method
2. Apply the trapezoidal rule to numerically integrate the cost function $J(\xi)$.
3. Create an anonymous cost function with respect to a, b and utilize the "scipy" package in Python to optimize and determine the values of a, b that minimize the cost function. These values of a, b will serve as our desired coefficient.

Result

Following the aforementioned procedure above, we successfully implemented the optimization on the cost function $J(\xi)$. As a result, we present the following graph, which illustrates our achieved outcome:

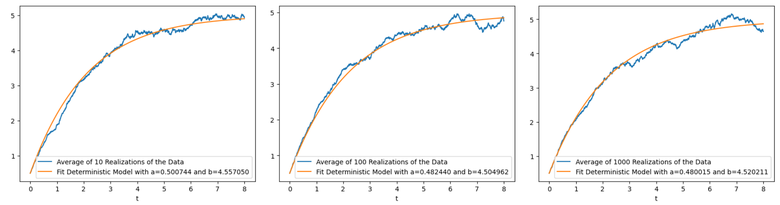


Figure 3. Performance of the Fitted Deterministic Models with Varying Number of Simulations

By setting our parameter σ to a fixed value of 0.1, we proceeded to simulate $10, 100,$ and 1000 times respectively data from the CIR model. The resulting images below demonstrate the close fit between the deterministic CIR model and the average simulated data. The estimated $\hat{a}, \hat{b}$ are fairly close to the true a, b despite the number of each simulations in each case.

Considering that the model imposes a condition of $2ab \geq \sigma^2$ to prevent negative values of r_t, we can establish an upper bound for the corresponding σ. Yet, It is important to note that a rigorous estimation for σ should be further developed, which holds the potential for enhancing our estimation of CIR model parameters.

Reference:

[1] 30-year fixed rate mortgage average in the united states [MORTGAGE30US]. Retrieved from FRED, Federal Reserve Bank of St. Louis, May 18 2023. Available at `https://fred.stlouisfed.org/series/MORTGAGE30US`.

[2] Timothy Sauer. Numerical solution of stochastic differential equations in finance. In *Handbook of computational finance*, pages 529–550. Springer, 2011.

Code for the project can be found: https://colab.research.google.com/drive/1veuUhQFOIJuJrNWOArr87EQ8PNPLW8iy?usp=sharing

jiayuechen@ucsb.edu

P-Colorability: Not Your Typical Knot Invariant

Miao Lei, Kriteen Shrestha, Cindy Zhao Mentor: Marcos Reyes

Department of Mathematics, University of California, Santa Barbara

;

Abstract

This poster explores the idea of distinguishing knots and studying their projections using Reidemeister moves. We will also discuss three interconnected knot invariants: Tricolorability, P-Colorability, and Knot Polynomials. More information can be found on our website, available after scanning the QR code on Figure 2.

Introduction

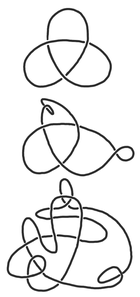


Figure 1: Three projections of the same knot

Loosely speaking, a **knot** is a tangled loop of string with connected ends. Mathematically, it is a closed curve embedded in 3-D space. The most fundamental knot is the **unknot**, or a circle. One knot can be drawn in different ways with different **projections**.

But how can we be sure that all knots are not just different projections of the unknot? Strategies called **invariants** can help us distinguish two knots. The ongoing study of **knot theory** is a point of curiosity for other sciences like cryptography and DNA sequencing.

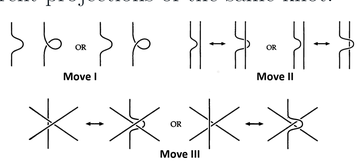
Let's now look at how we can transport one knot projection to another using *Reidemeister moves.*

Figure 2: Scan me!

Reidemeister Moves

Ambient isotopy is the idea of rearranging a knot in space without tearing it or having it intersect itself. If we can get from one projection to another by ambient isotopy, then the projections are of the same knot.

Reidemeister moves are valid changes on a knot that give different projections of the same knot.



Thus, two knots are equivalent if and only if one can be transformed into the other through a finite sequence of Reidemeister moves.

Invariants: An Overview

Figure 3: A wacky projection of the unknot.

Invariants help us answer the following question, fundamental in knot theory:

COULD ALL KNOTS JUST BE EQUAL TO THE UNKNOT?

We want to find an efficient way to tell whether two knots are distinct. Checking to see if no finite sequence of Reidemeister moves exists between two knot projections is a difficult task. Thus, we search for invariants—properties that hold true across all three Reidemeister movies—to distinguish knots.

Tricolorability

A knot is **tricolorable** if each strand of the knot diagram can be colored one of three colors, subject to the following rules:

1.At least two colors must be used.
2.At each crossing, the three incident strands are either exactly the same or completely different colors.

Note: In a knot diagram, a strand indicates a continuous piece that goes from one under-crossing to the next.

The rules of tricolorability hold true under each Reidemeister move, making it a valid knot invariant (see a proof on our website). As pictured on the right, the **trefoil knot** is tricolorable since it can be properly colored with three different colors.

Figure 4: Trefoil knot is tricolorable

Figure 5: The unknot is not tricolorable.

Because the unknot can only be colored by one color, it is *not equivalent* to the trefoil knot and any tricolorable knot. However, if a knot is not tricolorable, we cannot conclude that it is a projection of the unknot.

P-Colorability

Notice that with tricolorability, we were subject to just three "colors". **P-Colorability** is another knot invariant that generalizes the rules of tricolorability numerically so we can use more colors.

A knot, and any of its projections, will be **p-colorable** (where p is a prime > 2) if all strands of the knot can be labeled with variables such that c is the over-strand of a crossing, and a and b are the under-strands, following the condition $a+b-2c \equiv 0 \pmod{p}$.

We can use the determinant of the **coloring invariant matrix** of a knot to deduce which primes make a valid coloring system.

Theorem: A knot is p-colorable if and only if p divides the determinant.

Based on our p-colorable equation, each element of the matrix represents whether the strand is an overstrand (i.e., coefficient is -2) or an under-strand (coefficient is 1). If the coefficient is 0, that just means that the strand is not part of the three participating strands of that crossing.

Below is a coloring invariant matrix for the **figure-8 knot**.

$$\text{A} = \begin{bmatrix} -2 & 0 & 1 & 1 \\ 1 & 1 & 0 & -2 \\ 0 & 1 & -2 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \Rightarrow A_{\text{Cofactor}} = \begin{bmatrix} -2 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & -2 \end{bmatrix}$$

Figure 6: Figure 8 knot is 5-colorable.

Then $|det(A_{\text{Cofactor}})| = 5$, and thus, we can conclude that *the figure-8 knot is 5-colorable.* As shown on the left, we can now follow the coloring rules as we defined earlier on the figure-8 knot with at most five colors.

Also, because the determinant of our matrix is invariant on the labeling of the knot and the projection of the knot, we can conclude that the determinant of a knot's coloring matrix is a knot invariant.

Knot Polynomials

Knot Polynomials are numerical representations of knots and can be used as a knot invariant. The first polynomial in this discussion is the bracket polynomial. Calculation of the **bracket polynomial**, denoted <L>, follows three rules:

1. $\langle O \rangle = 1$
2. $\langle X \rangle = A\langle)(\rangle + A^{-1}\langle \asymp \rangle$
   $\langle X \rangle = A\langle \asymp \rangle + A^{-1}\langle)(\rangle$
3. $\langle L \cup O \rangle = (-A^2 - A^{-2})\langle L \rangle$

It can be shown that the Type II or III move on the knot's projection will result in the same bracket polynomial. However, making a Type I move will change the final bracket polynomial, so the bracket polynomial cannot be a knot invariant. Hence, we define a new polynomial to work around this issue.

The **X-Polynomial** is defined as follows:

$$X(L) = (-A^3)^{-w(L)} < L >$$

where w(L) is the writhe (pronounced like "faith") of the knot projection. Each strand in the projection has to be given a direction (we call this "oriented") to calculate the writhe. The trefoil knot has $X(L) = A^{-4} + A^{-12} - A^{-16}$.

Another polynomial, the **Alexander Polynomial** $\Delta(x)$, is found by labeling crossings. Interestingly enough, plugging $t = -1$ into $\Delta(x)$ gives the same value as the determinant of our coloring matrix!

While there are other knot invariants that give us more information, we focus on the most introductory ones to present knot theory in a palpable way.

Acknowledgements and References

We would like to thank our wonderful mentor Marcos for guiding us and inspiring our project, as well as the UCSB Math Department for this fantastic opportunity.

[1] Adam, C.C, .1994. The Knot Book. *Amer. Math. Soc*
[2] Oesper, L., 2005. p-Colorings of Weaving Knots. Pomona College. Claremont, CA
[3] Price, C., (2003). Coloring Invariant and Determinants. California State University, Chico
[4] Dixon, S., (2010). Composite Knot Determinants.

Persistent Homology Over $\mathbb{Q}$

Daniel Badilla Jr – Mentored by Troy Kling

University of California, Santa Barbara

Importance of Persistent Homology

When working with multidimensional data, it can be challenging to understand its underlying geometric structure. By constructing a sequence of simplicial complexes and computing the homology groups at each stage, one may gain insights into the connectivity and presence of holes in a given dataset. Furthermore, the structural knowledge gained is relatively robust to noise and sampling error.

What is Homology?

Simplicial homology is a fundamental concept from algebraic topology which aims to quantify the number of n-dimensional holes in a simplicial complex. Given a simplicial complex K, one first establishes bases for the cycles (linear combinations of simplices which "close up") and boundaries (linear combinations of simplices which bound something higher-dimensional) of K. The homology groups are then obtained by taking a quotient – cycles mod boundaries. This effectively removes all "trivial" cycles, i.e. those that bound a filled-in region, leaving behind the true holes in the space.

Definitions

Let K be a simplicial complex and G an abelian group (for us, $G = \mathbb{Q}$).

- An n-**simplex** is the convex hull of $n+1$ affinely independent points.

Figure 1. Examples of n-simplices, for $n = 0, 1, 2, 3$.

- A **simplicial complex** is a mathematical structure formed from a collection of simplices of various dimensions whose faces satisfy certain conditions.
- An n-**chain** of K is a formal G-linear combination of n-simplices of K.
- The n^{th} **chain group** of K over G, denoted $C_n(K;G)$, is the set of all n-chains of K, under the binary operation of (formal) addition.
- The n^{th} **boundary map** is defined as

$$\partial_n : C_n(K;G) \to C_{n-1}(K;G)$$
$$[v_0, \dots, v_n] \mapsto \sum_{i=0}^{n} (-1)^i [v_0, \dots, \widehat{v_i}, \dots, v_n],$$

  where $[v_0, \dots, \widehat{v_i}, \dots, v_n]$ is the $(n-1)$-simplex obtained by deleting the i^{th} vertex from $[v_0, \dots, v_n]$.
- n-**cycles** are elements of $\ker(\partial_n)$; n-**boundaries** are elements of $\operatorname{im}(\partial_{n+1})$.
- $\partial_n \circ \partial_{n+1} = 0$, so the boundary maps form a **chain complex**:

$$\cdots \to C_{n+1}(K;G) \xrightarrow{\partial_{n+1}} C_n(K;G) \xrightarrow{\partial_n} \cdots \xrightarrow{\partial_2} C_1(K;G) \xrightarrow{\partial_1} C_0(K;G) \xrightarrow{\partial_0} 0.$$

- The n^{th} **homology group** of K over G is defined to be

$$H_n(K;G) = \ker(\partial_n)/\operatorname{im}(\partial_{n+1}).$$

- Working over $\mathbb{Q}$ (or any field of characteristic zero) eliminates torsion, so the rational homology groups are actually vector spaces over $\mathbb{Q}$, since

$$H_n(K;\mathbb{Q}) \cong H_n(K;\mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{Q}.$$

- The n^{th} **Betti number** of K, denoted $\beta_n(K)$, is the rank of the n^{th} homology group of K. It can be computed as follows, using the rank-nullity theorem:

$$\begin{aligned}\beta_n(K) &= \dim(H_n(K;\mathbb{Q})) \\ &= \dim(\ker(\partial_n)/\operatorname{im}(\partial_{n+1})) \\ &= \dim(C_n(K;\mathbb{Q})) - \operatorname{rank}(\partial_n) - \operatorname{rank}(\partial_{n+1}).\end{aligned}$$

An Example

The boundary of the 2-simplex $[v_1, v_2, v_3]$ is

$$\partial_2([v_1, v_2, v_3]) = [v_2, v_3] - [v_1, v_3] + [v_1, v_2] \in C_1(K;G).$$

This can be seen geometrically in the simplicial complex, K, shown below.

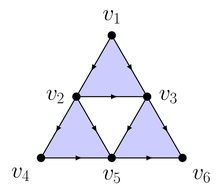


Figure 2. A simplicial complex, K, composed of 6 vertices, 9 edges, and 3 faces.

Ordered bases for the chain groups of the simplicial complex above are given by

$$\begin{aligned}C_0(K;\mathbb{Q}) &= \langle [v_1], [v_2], [v_3], [v_4], [v_5], [v_6] \rangle \cong \mathbb{Q}^6, \\ C_1(K;\mathbb{Q}) &= \langle [v_1, v_2], [v_1, v_3], [v_2, v_3], [v_2, v_4], \\ &\qquad [v_2, v_5], [v_4, v_5], [v_3, v_5], [v_3, v_6], [v_5, v_6] \rangle \cong \mathbb{Q}^9, \\ C_2(K;\mathbb{Q}) &= \langle [v_1, v_2, v_3], [v_2, v_4, v_5], [v_3, v_5, v_6] \rangle \cong \mathbb{Q}^3.\end{aligned}$$

Computing the boundaries of each basis element yields $\partial_0 = \partial_3 = 0$, and

$$\partial_1 = \begin{bmatrix} -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \qquad \partial_2 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

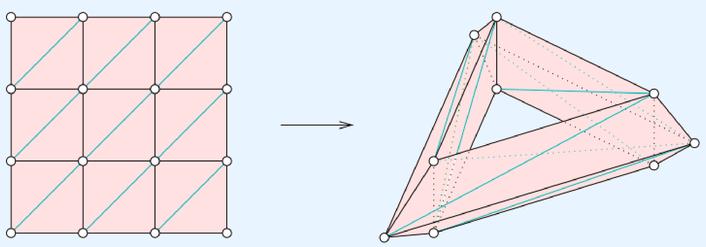
Bases for the null space of ∂_1 and the column space of ∂_2 can be computed as

$$\operatorname{null}(\partial_1) = \left\langle \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \right\rangle, \qquad \operatorname{col}(\partial_2) = \left\langle \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix} \right\rangle.$$

Continuing in this fashion, one can compute:

$$\begin{aligned} H_0(K;\mathbb{Q}) &= \mathbb{Q}^6/\mathbb{Q}^5 \cong \mathbb{Q}, & \beta_0(K) &= 1, \\ H_1(K;\mathbb{Q}) &= \mathbb{Q}^4/\mathbb{Q}^3 \cong \mathbb{Q}, & \beta_1(K) &= 1, \\ H_2(K;\mathbb{Q}) &= \mathbb{Q}/\mathbb{Q} \cong 0. & \beta_2(K) &= 0. \end{aligned}$$

Homology of the Torus



$$\begin{aligned} H_0(T^2;\mathbb{Q}) &= \mathbb{Q}, & \beta_0(T^2) &= 1, \\ H_1(T^2;\mathbb{Q}) &= \mathbb{Q} \oplus \mathbb{Q}, & \beta_1(T^2) &= 2, \\ H_2(T^2;\mathbb{Q}) &= \mathbb{Q}. & \beta_2(T^2) &= 1. \end{aligned}$$

Figure 3. Simplicial complex for the 2-dimensional torus, and its rational homology groups.

Alpha Complexes

Let K be a simplicial complex with vertex set S. If $\overline{B_r(v)}$ denotes the closed ball of radius r centered on a point $v \in S$, and $V(v)$ denotes the Voronoi region associated to v, then the **alpha complex** of radius r is the simplicial complex defined by

$$\alpha_r(K) = \left\{ \sigma \subseteq S : \bigcap_{v \in \sigma} (B_r(v) \cap V(v)) \neq \varnothing \right\}.$$

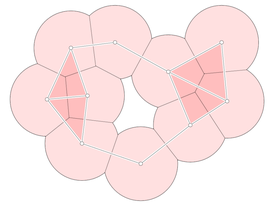


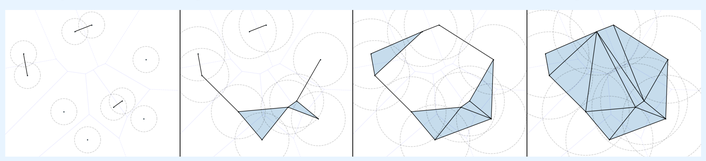
Figure 4. Example of an alpha complex, along with the associated union of closed balls.

The alpha complex has the following nice properties:

- The **Nerve Theorem** states that $\alpha_r(K)$ is homotopy equivalent to $\bigcup B_r(v)$. This implies that the two spaces have isomorphic homology groups.
- The dimension of the alpha complex can be no larger than the dimension of the ambient space to which the given data points belong.
- Computing the associated homology groups is computationally inexpensive.

Persistence & Topological Data Analysis

Given a point cloud of data, one may construct a **filtration** of alpha complexes by increasing the radius, r, of the closed balls surrounding the vertices. In doing so, **persistent** holes may be detected in the alpha complexes, thereby providing valuable insights into the structure of the underlying data.



$$\begin{aligned} &\beta_0(\alpha_{0.75}(K)) = 7, \quad \beta_0(\alpha_{1.5}(K)) = 2, \quad \beta_0(\alpha_2(K)) = 1, \quad \beta_0(\alpha_{2.5}(K)) = 1, \\ &\beta_1(\alpha_{0.75}(K)) = 0. \quad \beta_1(\alpha_{1.5}(K)) = 0. \quad \beta_1(\alpha_2(K)) = 1. \quad \beta_1(\alpha_{2.5}(K)) = 0. \end{aligned}$$

Figure 5. A filtration created by increasing r, along with the associated Betti numbers

Acknowledgements

I would like to acknowledge Troy, my mentor, for his invaluable guidance, expertise, and unwavering support throughout this research project. Without his assistance, I would not have been able to accomplish this project successfully. Thank you, Troy, for your dedication and for being an exceptional mentor.

References

Zomorodian, Afra J. *Topology for Computing*. Cambridge University Press, 2005.
Hatcher, Allen. *Algebraic Topology*. Cambridge University Press, 2001.
Otter, Nina, et al. "A roadmap for the computation of persistent homology." *EPJ Data Science* 6 (2017).

Popular Concepts in Functional Analysis

Neha Banerjee, Elizabeth Thomson, Xuanyi Zhou

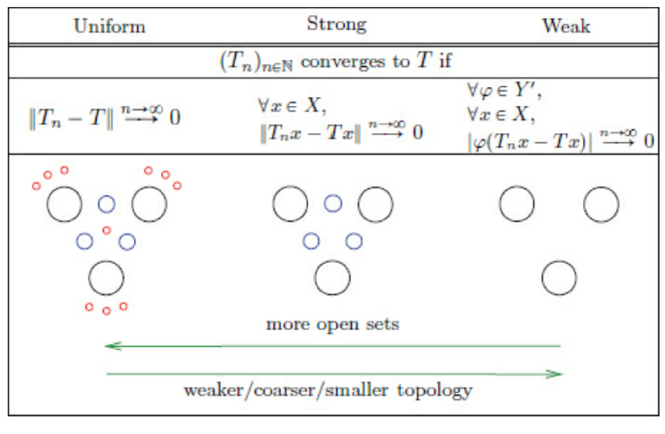
University of California - Santa Barbara

The Normed Space CL(X,Y)

$CL(X,Y)$ denotes the set of all continuous linear transformations from a normed space X to a normed space Y.
A continuous linear transformation is characterized by
Theorem 1.
Let X and Y be normed spaces, and T: X $\to$ Y be a linear transformation. Then the following properties of T are equivalent:
(1) *T is continuous.*
(2) *T is continuous at* ***0****.*
(3) *There exists an M>0 such that for all x* $\in X, ||Tx||_Y \leq M||x||_X$.
CL(X,Y) is a normed space, with pointwise operations and the "operator norm" ||.|| : CL(X,Y) $\to \mathbb{R}$ *given by:*
||T|| := sup{||Tx|| : x $\in X, ||x|| \leq 1\}, T \in CL(X,Y)$
A subspace Y of a normed space X is said to be an ***invariant subspace*** *with respect to a linear transformation T: X* $\to$ *X if TY* $\subset Y$.
Invariant subspaces are useful since they are helpful in studying complicated operators by breaking them down into smaller operators acting on invariant subspaces. This is similar to the diagonalization process in linear algebra where one decomposes the vector space into eigenspaces, where the linear transformation acts trivially. This also contributes to one of the open problems in functional analysis known as the ***invariant subspace problem****:*
Does every $T \in CL(H)$ on a separable complex Hilbert space H have a non-trivial invariant subspace?

Topology in Normed Spaces

There are three key topologies in the normed space, categorized into **uniform operator topology**, **strong operator topology**, and **weak operator topology** on CL(X,Y).



Banach Spaces

Definition
A normed space in which the set of *Cauchy sequences* is equal to the set of convergent sequences is called a **Banach space**. Sometimes, we also call it a **complete normed space**.
Some common examples of Banach spaces include: $(\mathbb{R}, |.|)$, $(\mathbb{C}, |.|)$, $(l^p, ||.||_p)$, $(C[a,b], ||.||_\infty)$, $(L^2[a,b], ||.||_2)$.

A Property of Banach Space

Theorem 2.
In a Banach space, absolutely convergent series converge, that is:
If $(x_n)_{n\in\mathbb{N}}$ is a sequence in a Banach space $(X, ||.||)$ such that $\sum_{n=1}^{\infty} ||x_n|| < \infty$, then $\sum_{n=1}^{\infty} x_n$ converges in X. Moreover, $||\sum_{n=1}^{\infty} x_n|| \leq \sum_{n=1}^{\infty} ||x_n||$.

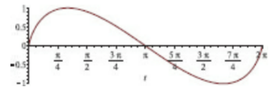
Example: $\sum_{n=1}^{\infty} \frac{sin(n.)}{n^2}$ converges in $C[0,2\pi], ||.||_\infty)$. (Here sin(n·) means the function t $\to sin(nt) : ([0,2\pi] \to \mathbb{R})$
Indeed, we have $\left|\left|\frac{\sin(n.)}{n^2}\right|\right|_\infty = \frac{1}{n^2}$, and $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$.
So $\boldsymbol{x} := \sum_{n=1}^{\infty} \frac{sin(n.)}{n^2}$ defines a continuous function on $[0, 2\pi]$.
We can approximate the limit by computing the first N terms and plotting the resulting function. The error can then be bounded as follows:

$$\left|\left|\sum_{n=N+1}^{\infty} \frac{\sin(n.)}{n^2}\right|\right|_\infty \leq \sum_{n=N+1}^{\infty} \left|\left|\frac{\sin(nt)}{n^2}\right|\right|_\infty \leq \sum_{n=N+1}^{\infty} \frac{1}{n^2}.$$

For example, if $N = 100$, then the error is bounded above by

$$\sum_{n=101}^{\infty} \frac{1}{n^2} \leq \int_{100}^{\infty} \frac{1}{x^2} dx = \frac{1}{100} = 0.01.$$

Using Maple, one can plot the partial sum of x with $N = 100$ and get:



Thus the sum converges to a continuous function that lies in the strip of width 0.01 around the graph shown in the figure. We can also use Theorem 2 to show that e^A converges (A belongs to $CL(X)$). $CL(X)$ denotes a certain Banach space, namely the space of all "continuous linear transformations" from X to itself, equipped with the "operator norm". For example, when $X = \mathbb{R}^d$, $CL(X)$ turns out to be the space of all $d \times d$ real matrices. Why is there a focus on e^A? The answer is that it plays a crucial role in differential equations. The initial value problem,

$$\frac{dx}{dt} = A(t)x(t), t \in \mathbb{R}, \quad x(0) = x_0 \in X,$$

has a unique solution given by $x(t) = e^{tA}x_0$, where $t \in \mathbb{R}$.

Spectral Theory

Spectrum and Resolvent

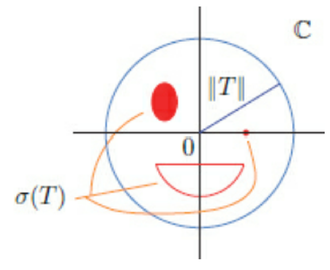
- For a linear transformation $T \in L(X)$ on a *finite* dimensional vector space X over $\mathbb{C}$, the set of eigenvalues is known as its *spectrum* $\sigma(T)$ with cardinality at most $dimX$. In infinite dimensional complex vector spaces, linear transformations may have no eigenvalues, finitely many eigenvalues, or infinite eigenvalues.
- Let X be a normed space and $T \in CL(X)$. We say that $\lambda \in \mathbb{C}$ belongs to the spectrum $\sigma(T)$ of T if $\lambda I - T$ is not invertible in $CL(X)$. Thus

$$\rho(T) := \mathbb{C}\backslash\sigma(T) = \{\lambda \in \mathbb{C} : \lambda \text{ is invertible in } CL(X)\}$$

- The set $\rho(T)$ is the *resolvent set*
- The set $\sigma_p(T)$ of all eigenvalues of T is called the *point spectrum of T*.
- We have that $\sigma_p(T) \subset \sigma(T)$, since if $\lambda \in \sigma_p(T)$ then there exists a nonzero vector x such that $Tx = \lambda x$, that is $(\lambda I - T)x = 0$, showing that $\lambda I - T$ is not injective, and hence cannot be invertible either.

Theorem 3.

- Let X be a Banach space and $T \in CL(X)$. Then
  (1) $\sigma(T) \subset \{\in \mathbb{C} : |\lambda| \leq ||T||\}$
  (2) $\rho(T)$ is an open subset of $\mathbb{C}$
  (3) $\sigma(T)$ is a compact subset of $\mathbb{C}$
  (4) $\sigma(T)$ is nonempty



Acknowledgements

Reference Material: "A Friendly Approach to Functional Analysis" by Amol Sasane
Thank you to the UCSB Directed Reading Program and to our mentor Melody Molander for making this project possible.

80 Knots...100 Knots...V1...Rotate! Prandtl's Method of Computing Lift

Daric Zhou, Lishan Shi, and Rodrick Zhu

Department of Mathematics, University of California, Santa Barbara

Introduction

Our project aims to calculate the lift and drag of an airfoil in two and three dimensions. To achieve this goal, we need to obtain the equation of motion of the fluid passing over the airfoil. Two approaches can be taken to address this problem: the microscopic approach utilizing Boltzmann's equation or the continuum approach that treats the fluid as a group of parcels occupying each point of space with well-defined thermodynamics properties. We will be utilizing the continuum approach, which relies on calculus to solve problems. This approach utilizes several fundamental equations to comprehend fluid motion, namely Euler's equation and Navier-Stokes equations. Specifically, Euler's equation applies to ideal fluids, while Navier-Stokes equations are applied to viscous fluids. These equations are derived from Newton's second law, which states that an object's acceleration is proportional to the net force applied and inversely proportional to its mass.

Preliminaries

Euler's and Navier-Stokes equations are fundamental for understanding fluid motion. In our case, we need to apply these equations to fluid flow around an obstacle, such as an airfoil. A stationary, ideal, and homogeneous fluid setting is a suitable context for studying an airfoil in two dimensions, as it allows us to simplify Euler's equation to Cauchy-Riemann equations in complex analysis. More specifically, consider the following setting:

Our velocity field $U = (u(x, y), v(x, y))$

Incompressible fluid $\Rightarrow \nabla \cdot U = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$, that is $\frac{\partial u}{\partial x} = -\frac{\partial v}{\partial y}$

Irrotational fluid $\Rightarrow \nabla \times U = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = 0$, that is $\frac{\partial v}{\partial x} = \frac{\partial u}{\partial y}$

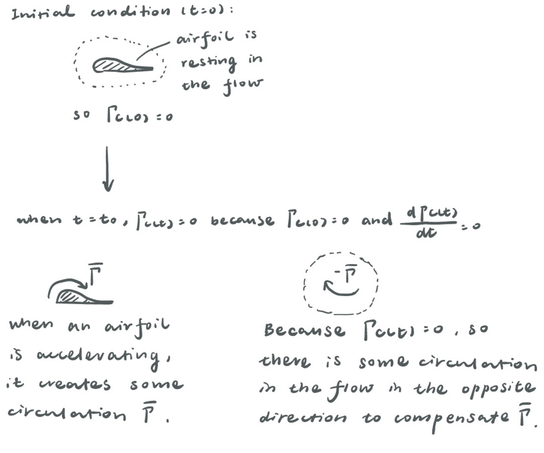
These two conditions imply that our force $F = u - iv$ is complex differentiable.

Now assume that $F = \frac{\partial W}{\partial z}$ where $W = \phi + i\psi$, then by Cauchy-Riemann equations, the force can be expressed as $F = \frac{\partial \phi}{\partial x} + i\frac{\partial \psi}{\partial x} = \frac{\partial \phi}{\partial x} - i\frac{\partial \psi}{\partial y} = u - iv$ where $u = \frac{\partial \phi}{\partial x}$ and $v = \frac{\partial \psi}{\partial y}$. In other words, $U = (u, v) = \nabla\phi$

We would also like to provide the following comments.

1. In a stationary, ideal, and homogeneous fluid, the original Euler's equation $\rho\frac{du}{dt} = -\nabla P + \rho b$ can be simply expressed as $P = -\rho\frac{\|u\|^2}{2}$. Furthermore, the Cauchy-Riemann equations further simplify the incompressible condition as $\nabla \cdot U = \nabla \cdot \nabla\phi = \triangle\phi = 0$. Lastly, we consider the boundary condition. Theoretically, the ideal flow would be tangential to the boundary and not penetrate the boundary of the domain, which we may express as $U \cdot \hat{n} = \nabla\phi \cdot \hat{n} = V \cdot \hat{n}$ on ∂D, where D refers to the domain.

2. Our assumption of ideal and incompressible flow implies that the density field ρ is constant by the continuity equation. This result allows us to apply the Kelvin's circulation theorem, and understand how circulation changes throughout time. Specifically, the theorem informs us that $\frac{d\Gamma_{c(t)}}{dt}$=0 where $\Gamma_{c(t)}$ denotes the circulation around contour $c(t)$. If the original circulation is zero, then the theorem tells us that the circulation around $c(t)$ will remain to be zero throughout the flow. See the following graph.



Lift of 2-dim Airfoil

To determine the lift and drag of a 2-dimensional airfoil, we follow a systematic approach that involves analyzing the force and potential of the flow. These steps outline the process:

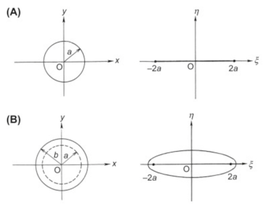
1. We begin by examining the force and potential of the flow in the absence of an airfoil. We express the initial velocity field as $U = (u, v) = (Q\cos\alpha, Q\sin\alpha)$. Because we proved in the previous section that the force is $F = u - iv$, we rewrite it as $F = Q\cos\alpha - iQ\sin\alpha = Qe^{-i\alpha}$ by Euler's formula. Then the complex potential is $W = Qze^{-i\alpha}$, since $F := \frac{\partial W}{\partial z}$.

2. Next, we investigate the force and potential of the flow after introducing a 2-dimensional disk-shaped airfoil. To calculate these values, we can utilize the Milne-Thomson circle theorem, which allows us to insert a circle into a 2-dimensional flow and understand how the complex potential changes. Applying this theorem, the new complex potential is given by

$$\hat{W} = Qze^{-i\alpha} + \overline{Qze^{-i\alpha}} = Qze^{-i\alpha} + Qe^{i\alpha}a^2/z^2 - i\Gamma\ln z/2\pi$$

where Γ refers to the circulation and a denotes the radius of a disk.

3. We then apply conformal mapping techniques to transform the airfoil shape. This transformation allows for easier analysis and computation of the flow properties around the airfoil. See the below for two examples of airfoils provided in [5].



Let us discuss the computation for the exterior of the disk of radius a. Mathematically, we denote the exterior of the disk as z, and z will map to the exterior of the plate ($|x| \le 2a$ in $\mathcal{R}^2$), $\hat{z} = z + a^2/z$. Then we analyze the following equation:

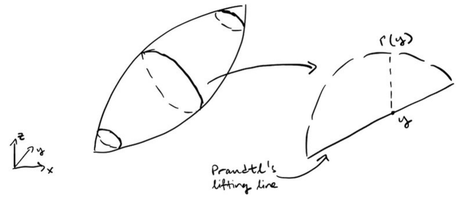
$$\frac{d\hat{W}}{d\hat{z}} = \frac{d\hat{W}}{dz}\frac{dz}{d\hat{z}} = \left(Qe^{-i\alpha} + \frac{Qe^{i\alpha}a^2}{\hat{z}^2} - \frac{i\Gamma}{2\pi\hat{z}}\right)\frac{\sqrt{\hat{z}^2 - 4a^2} + \hat{z}}{2\sqrt{\hat{z}^2 - 4a^2}}$$

We can obtain $\frac{dz}{d\hat{z}}$ by rewrite $\hat{z} = z + a^2/z$ in terms of z, which is $z = \frac{\hat{z}\pm\sqrt{\hat{z}^2-4a}}{2}$. However, only the solution with plus sign will be taken as we assume $z \approx \hat{z}$ at infinity.

Note that $\frac{d\hat{W}}{d\hat{z}}$ is not defined precisely at $\hat{z} = \pm 2a$. The Kutta-Joukowski condition allows us to eliminate one of the singularities at $\hat{z} = 2a$. We pick Γ such that $\left(Qe^{-i\alpha} + \frac{Qe^{i\alpha}a^2}{\hat{z}^2} - \frac{i\Gamma}{2\pi\hat{z}}\right) = 0$, and it turns out that this condition is satisfied when $\Gamma = -4\pi a\sin\alpha$. Applying Kutta-Joukowski theorem, we can calculate the force exerted on D. We obtain $\hat{F} = \rho\Gamma(V - iU) = -4\rho\pi Qa\sin\alpha + 4i\rho\pi Qa\sin\alpha\cos\alpha = \rho\Gamma(Q\sin\alpha - iQ\cos\alpha) = x + iy$, where $\hat{F} = (x, y)$ denotes the force field, and $(U, V) = (Q\cos\alpha, Q\sin\alpha)$ is obtained from $(u, v) \to (U, V)$ at infinity by our assumption.

4. The lift, therefore, is the magnitude of the complex number $\hat{F}$, which is $4\pi\rho\alpha Q^2\sin\alpha$.

Whereas the flat plate had a singularity at the left endpoint, a general K-J airfoil can have a smoother leading edge and a sharp trailing edge to which we apply the K-J condition. For our analysis of the three-dimensional wing, we consider each cross-section to be a general K-J airfoil in the coordinate system shown below:



Lift of 3-dim Wing

When studying 3-dimensional wings, we can slice the wing by a plane $y = \text{const.}$ to obtain a 2-dimensional airfoil section. We consider Joukowski airfoils - characterized by a smooth leading edge and a sharp trailing edge - for the cross-sections of our wing. We use Prandtl's model of a 3-dimensional wing to calculate the lift. In this model, the wing is regarded as long and thin; the aspect ratio of a wing is defined as wingspan^2/wing area, or $4b^2/A$ where b is the half-span. This model is an asymptotic approximation as the aspect ratio of the wing tends to ∞.

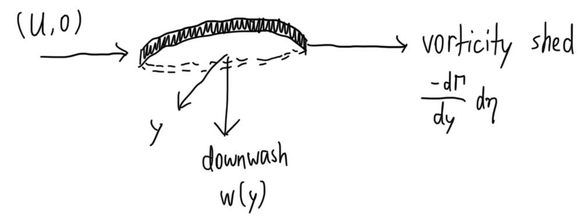
We start by determining $\Gamma(y)$, assuming each cross section is a Joukowski airfoil and that angles formed by the flow and the airfoil, α, β are small:

$$l(y) = 4\pi\rho c(y)Q^2\sin(\alpha + \beta)$$
$$\Longrightarrow\ \Gamma(y) = 4\pi c(y)U(\alpha(y) + \beta(y))$$

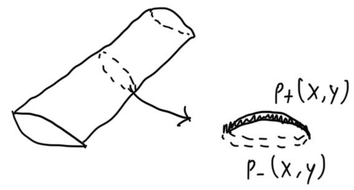
We denote the effective angle of attack as

$$\alpha_{\text{eff}} = \alpha + \frac{w(y)}{U}$$

where α is the fixed angle of attack and $w(y)$ is the downwash due to shed vorticity. A visualization of this downwash is below.



One computes the shed vorticity by an application of the Kutta-Joukowski theorem to an infinitesimal section of the wing.



We now have

$$\Gamma(y) = 4\pi c(y)U(\alpha + \frac{w(y)}{U} + \beta(y))$$

Note that a doubly infinite vortex creates a velocity given by the 2-dimensional point vortex flow: namely, a line that carries unit circulation and is parallel to the x-axis at a position $y = \eta$ in the $z = 0$ plane introduces the velocity

$$\frac{1}{2\pi(y - \eta)}$$

Since the shed vortex is semi-infinite, this vorticity is reduced by a factor of $\frac{1}{2}$. We can now determine $w(y)$ from $\Gamma(y)$:

$$w(y) = -\frac{1}{4\pi}\int_{-b}^{b}\frac{\frac{d\Gamma}{dy}(\eta)}{y - \eta}d\eta$$

We can also deduce an integral equation for $\Gamma(y)$:

$$\Gamma(y) = 4\pi c(y)U(\alpha + \beta(y) - \frac{1}{4\pi U}\int_{-b}^{b}\frac{\frac{d\Gamma}{dy}(\eta)}{y - \eta}d\eta$$

Solving this equation results in the lift on the wing:

$$L = \rho U\int_{-b}^{b}\Gamma(y)dy$$

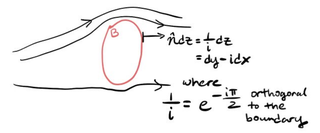
It is important to note that this method is quite basic and assumes an ideal fluid flow, which is unrealistic. In the real world, one major concept is that of the boundary layer: a thin layer adjacent to the surface of the object where the effect of viscosity is the most prominent. The boundary layer acts as another source of viscosity, and in the case of airfoils, it causes the fluid flow to "separate" or "detach" from the trailing edge into a wake.

Blasius' Formula and the Kutta-Joukowski Theorem

(Blasius' Formula) *Consider an ideal, incompressible, potential flow (with flow velocity $U = (u, v)$) around a rigid obstacle B, and suppose $F = u - iv$ is the complex velocity. Then the force $\mathcal{F}$ exerted on B is given by*

$$\mathcal{F} = -\int_{\partial B} p\mathbf{n}ds = \frac{-i\rho}{2}\overline{\left[\int_{\partial B} F^2dz\right]}$$

In short, Blasius' formula relates the force exerted by the pressure of the flow to the force exerted by the density of the flow, and the diagram below illustrates the infinitesimal normal displacement of along the boundary of B, which can be used to derive Blasius' formula from the expression of the force involving pressure.



(Kutta-Joukowski Theorem) *Consider an incompressible, potential flow exterior to an obstacle B, and suppose the velocity field (u, v) goes to some constant value (U, V) at infinity. Then the force $\mathcal{F}$ exerted on D is given by*

$$\mathcal{F} = -\rho\Gamma_{\partial B}||(U - iV)||\mathbf{n}$$

where $\Gamma_{\partial B}$ is the circulation about the obstacle and $\mathbf{n}$ is a unit vector orthogonal to (U, V).

Proof. Recall that in the statement of Blasius' formula, we defined F to be the complex velocity outside of B, so it is by definition analytic, and can therefore be written in a Laurent series like so:

$$F = \cdots + \frac{a_{-2}}{z^2} + \frac{a_{-1}}{z} + a_0 + a_1z + a_2z^2 + \dots$$

But because of the limiting value at $z \to \infty$, there are no positive powers (otherwise F diverges) and so

$$F = a_0 + \frac{a_{-1}}{z} + \frac{a_{-2}}{z^2} + \dots$$

Recall that $F = u - iv$ as in Blasius' formula, so we have

$$\begin{aligned}\int_{\partial B} Fdz &= \int_{\partial B}(u - iv)(dx + idy)\\ &= \int_{\partial B} udx + vdy = \int_{\partial B}(u, v)ds = \Gamma_{\partial B}\end{aligned} \tag{1}$$

and by the Cauchy Residue theorem, we have that $a_{-1} = \frac{\Gamma_{\partial B}}{2\pi i}$. Finally, using our derivation from (1) and Blasius' formula, we get

$$\begin{aligned}\mathcal{F} &= \frac{-i\rho}{2}\overline{\left[\int_{\partial B} F^2dz\right]} = \frac{-i\rho}{2}\overline{\left[\int_{\partial B}\left(a_0^2 + \frac{2a_0a_1}{z} + \dots\right)dz\right]}\\ &= \frac{-i\rho}{2}\overline{2a_0a_{-1}2\pi i} = \frac{-i\rho}{2}\cdot 2a_0\cdot\frac{\Gamma_{\partial B}}{2\pi i}\cdot 2\pi i\\ &= \frac{-i\rho}{2}\overline{2\Gamma_{\partial B}(U - iV)} = -\rho\Gamma_{\partial B}(V - iU)\\ &= -\rho\Gamma_{\partial B}||(U - iV)||\mathbf{n}\end{aligned}$$

as desired.

Acknowledgements

We would like to thank the DRP team for organizing the program this year, and we would like to thank our mentor Pranav for his unyielding patience and mentorship.

References

[1] Bak, Joseph, and Donald J. Newman. *Complex Analysis*. Undergraduate Texts in Mathematics. New York, NY: Springer New York, 2010.

[2] Childress, Stephen. *An Introduction to Theoretical Fluid Mechanics*. Vol. 19. Courant Lecture Notes. American Mathematical Society, 2009.

[3] Chorin, Alexandre J., and Jerrold E. Marsden. *A Mathematical Introduction to Fluid Mechanics*. Vol. 4. Texts in Applied Mathematics. New York, NY: Springer New York, 1993.

[4] Schoeffel, Laurent. *Lecture Notes in Fluid Mechanics*. arXiv, July 24, 2016.

[5] Nakayama, Yasuki. *Introduction to Fluid Mechanics*. 2018. Accessed though <https://www.sciencedirect.com/texttopics/engineering/conformal-mapping>.

Projective Geometry and Camera Geometry

Minxi Lin, Joy Luo, Mentor: Danning Lu

University of California - Santa Barbara

Geometry of Projective Transformations

A projective transformation is a geometric transformation that preserves certain properties such as collinearity (a straight line is imaged as a straight line). It models the distortion caused by perspective camera and is commonly used in computer vision and computer graphics.
Homogeneous representation: it is a vector with an additional coordinate which allows the representation of points at infinity and further transformations.
Points:
- 2D: An arbitrary homogeneous representation of a point is $\mathbf{x} = (x_1, x_2, x_3)^\mathbf{T}$, which means $\mathbf{x} = (x_1/x_3, x_2/x_3)$ in $\mathbb{R}^2$.
- 3D: Similarly, for a point $\mathbf{x}$ in 3-space, $\mathbf{x} = (x_1, x_2, x_3, x_4)^\mathbf{T}$ with $x_4 \neq 0$ represents $(x_1/x_4, x_2/x_4, x_3/x_4)^\mathbf{T}$ in $\mathbb{R}^3$.
Lines: In a plane, a line is represented by the equation $ax+by+c=0$, and therefore can be represented by the vector $(a, b, c)^\mathbf{T}$.
Planes: A plane in 3-space is shown as $\pi_1\mathbf{x}+\pi_2\mathbf{y}+\pi_3\mathbf{z}+\pi_4 = 0$, so its homogeneous representation can be written as $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)^\mathbf{T}$. Since only the three independent ratio $\{\pi_1 : \pi_2 : \pi_3 : \pi_4\}$ are significant, the dof of a plane in 3-space is 3.
Ideal points: For two parallel lines $\mathbf{l} = (a, b, c)^\mathbf{T}$ and $\mathbf{l}' = (a, b, c')^\mathbf{T}$, the intersection is $\mathbf{l}\times\mathbf{l}' = (c-c')(b, -a, 0)^\mathbf{T}$. Ignoring $(c-c')$, the scale factor, the inhomogeneous representation of the point $(b/0, -a/0)^\mathbf{T}$ does not make sense, except to indicate that the corresponding intersection point has coordinates that approach infinity. Hence, points with the homogeneous notation $(x_1, x_2, 0)^\mathbf{T}$ are called ideal points, also known as points at infinity.
Line at infinity: It is a set of all ideal points which the ratio of $x_1 : x_2$ is constant. It is denoted as the vector $\mathbf{l}_\infty = (0, 0, 1)^\mathbf{T}$ such that $(0, 0, 1)(x_1, x_2, 0)^\mathbf{T} = 0$
2D Projective transformation: A projective transformation is a general non-singular linear transformation of homogeneous coordinates. It generalizes an affine transformation, which encompasses a non-singular linear transformation of inhomogeneous coordinates combined with a translation. It can be decomposed as

$$\mathbf{H} = \mathbf{H_S}\mathbf{H_A}\mathbf{H_P} = \begin{bmatrix} s\mathbf{R} & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{K} & 0 \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ \mathbf{v}^T & v \end{bmatrix} = \begin{bmatrix} \mathbf{A} & t \\ \mathbf{v}^T & v \end{bmatrix}$$

where $\mathbf{A}$ is a non-singular matrix given by $\mathbf{A} = s\mathbf{R}\mathbf{K}+\mathbf{t}\mathbf{v}^T$, and $\mathbf{K}$ is a upper-triangular matrix normalized as $\det \mathbf{K} = 1$. The decomposition is valid if $v \neq 0$, and is unique if $s > 0$.
$\mathbf{H_P}$ (2 dof) is an elation that moves line at infinity. $\mathbf{H_A}$ (2 dof) affects the affine properties but not move the line at infinity. $\mathbf{H_S}$ (4 dof) is the general similarity transformation that does not affect the affine or projective properties.

The Direct Linear Transformation (DLT) Algorithm

The Direct Linear Transformation Algorithm is used for estimating the camera projection matrix from corresponding 2D and 3D points. The transformation for the 2D to 2D point correspondence $x_i \leftrightarrow x_i'$ is given by the equation $\mathbf{x}_i' = \mathbf{H}\mathbf{x}_i'$. The equation is then expressed using the vector cross product $\mathbf{x}_i' \times \mathbf{H}\mathbf{x}_i = 0$. Let the j-th row of the matrix $\mathbf{H}$ be denoted as $\mathbf{h}^{j\mathbf{T}}$ and $\mathbf{x}_i' = (x_i', y_i', w_i')^\mathbf{T}$, then:

$$\mathbf{H}x_i = \begin{pmatrix} \mathbf{h}^{1\mathbf{T}}\mathbf{x}_i \\ \mathbf{h}^{2\mathbf{T}}\mathbf{x}_i \\ \mathbf{h}^{3\mathbf{T}}\mathbf{x}_i \end{pmatrix} \text{ and } \mathbf{x}_i' \times \mathbf{H}\mathbf{x}_i = \begin{pmatrix} y_i'\mathbf{h}^{3\mathbf{T}}\mathbf{x}_i - w_i'\mathbf{h}^{2\mathbf{T}}\mathbf{x}_i \\ w_i'\mathbf{h}^{1\mathbf{T}}\mathbf{x}_i - x_i'\mathbf{h}^{3\mathbf{T}}\mathbf{x}_i \\ x_i'\mathbf{h}^{2\mathbf{T}}\mathbf{x}_i - y_i'\mathbf{h}^{1\mathbf{T}}\mathbf{x}_i \end{pmatrix}$$

Since $\mathbf{h}^{j\mathbf{T}}x_i = x_i^\mathbf{T}\mathbf{h}^j$ for $j = 1, 2, 3, ...$, then the equations can be written as: $\begin{bmatrix} 0^\mathbf{T} & -w_i'\mathbf{x}_i^\mathbf{T} & y_i'\mathbf{x}_i^\mathbf{T} \\ w_i'\mathbf{x}_i^\mathbf{T} & 0^\mathbf{T} & -x_i'\mathbf{x}_i^\mathbf{T} \\ -y_i'\mathbf{x}_i^\mathbf{T} & x_i'\mathbf{x}_i^\mathbf{T} & 0^\mathbf{T} \end{bmatrix} \begin{pmatrix} \mathbf{h}^1 \\ \mathbf{h}^2 \\ \mathbf{h}^3 \end{pmatrix} = 0$
These equations have the form of $A_i\mathbf{h} = 0$ where A_i is a 3×9 matrix and h is a 9-vector made up of the entries in matrix H.

The Gold Standard Algorithm

Estimation of the projection matrix P from 3D to 2D when 3D points are accurately known:
Objective:
Given $n \geq 6$ world to image point correspondence $\mathbf{X_i} \leftrightarrow \mathbf{x_i}$, determine the Maximum Likelihood estimate of the camera projection matrix P, i.e. the P which minimized the geometric error $\sum_i d(x_i, PX_i)^2$.
Algorithm:
(i) Compute an initial estimate of P using linear methods such as:
(a) **Normalization**: Using a similarity transformation T to normalize the image points ($\hat{x} = Tx_i$), and a second similarity transformation to normalize the space points ($\hat{X} = UX_i$).
(b) **DLT**: Form the $2n \times 12$ matrix A by stacking equations generated by each correspondence $\mathbf{X_i} \leftrightarrow \mathbf{x_i}$ (similar to 2D but with vector $\mathbf{P}$ instead of vector $\mathbf{h}$).
(ii) Use the linear estimate as a starting point to minimize the geometric error over $\tilde{P}$: $\min_P \sum_i d(\mathbf{x}_i, P\mathbf{X}_i)^2$
(iii) Obtain the camera matrix for the original coordinates: $P = T^{-1}\tilde{P}U$.

A quantity such as $\mathbf{x}$ represents a measured point
Estimated quantities are represented with a hat
True values are represented by a bar
Quantities of the transformed version are represented with the tidle symbol.

Camera Model and Computation

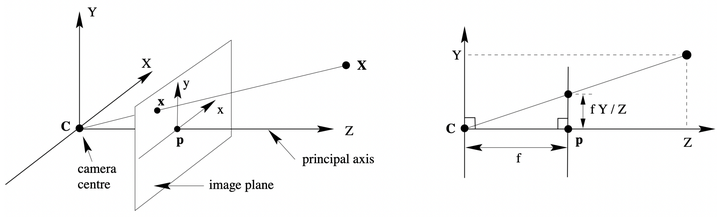
Camera centre: The camera centre $\mathbf{C}$ is the point for which $\mathbf{PC} = 0$. The algebraic expression is $\mathbf{C} = (\mathbf{X}, \mathbf{Y}, \mathbf{Z}, \mathbf{T})$, where

$$\mathbf{X} = det([p_2, p_3, p_4]), \mathbf{Y} = -det([p_1, , p_3, p_4])$$
$$\mathbf{Z} = det([p_1, p_2, p_4]), \mathbf{T} = -det([p_1, , p_2, p_3])$$

Camera orientation and internal parameters: The camera matrix can be decomposed into $\mathbf{P} = [\mathbf{M}| - \mathbf{M}\tilde{\mathbf{C}}] = \mathbf{K}[\mathbf{R}| - \mathbf{R}\tilde{\mathbf{C}}]$
The matrix $\mathbf{R}$ represents the orientation of the camera, and the matrix $\mathbf{K}$ is the calibration matrix. As the diagonal entries in $\mathbf{K}$ are positive, the ambiguity in the decomposition is removed. The matrix representation of $\mathbf{K}$ is

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$$

α_x and α_y are the scale factors in the x- and y-coordinate directions respectively. s is the skew factor, and $(x_0, y_0)^T$ are the coordinates of the principal point, which is the point of intersection between the camera's image plane and the axis.



Acknowledgements

Reference Material: "Multiple View Geometry in Computer Vision" by Richard Hartley and Andrew Zisserman
Thank you to the UCSB Directed Reading Program and to my mentor Danning Lu for making this project possible.

Quantum Advantages Over Classical Cryptography

Ponokela DeMarzo and Natalya Rodriguez

University of California, Santa Barbara

Classical Cryptography

Cryptosystems are cryptographic protocols that allow for secure communication between two parties across the internet (or any medium where eavesdroppers are a concern). AES is one such protocol and is widely considered to be the most secure classical cryptosystem available today. The premise behind AES is as follows: both parties are in possession of a secret private key of either 128, 192, or 256 bits in length which is used for encryption and decryption. Its usefulness, however, is dependent on both parties knowing the secret key, which may be difficult to share in a secure manner. So, what if you wanted to communicate with someone you had never met?

Public Key Cryptosystems

Public key cryptosystems allow parties to generate and distribute a public key, which others may use to encrypt and send them messages, as well as a private key which they keep secret and use to decrypt messages received. The most popular of these protocols is RSA which utilizes the difficulty of finding prime factors of large numbers.

RSA Key Creation

1. Choose two large prime numbers p and q.
2. Find $n = pq$ and $\varphi(n) = (p-1)(q-1)$.
3. Choose a relatively small number e such that e and $\varphi(n)$ are coprime.
4. Find $d \equiv e^{-1} \mod \varphi(n)$.
5. The RSA public key is $P = (e, n)$ and RSA private key is $S = (d, n)$.

Security Considerations

Although RSA is a secure cryptosystem, it is not unbreakable. In cryptography, a break refers to method of determining the secret key faster than using brute force. The protocol relies on the inability of classical computers to factor large products of primes fast enough to be considered a security concern.

However, when quantum computers become large and powerful enough, they will be able to solve problems which are difficult for classical computers. Shor's algorithm, for example, is a quantum factoring algorithm which runs in polylogarithmic time. This is much faster than the fastest known classical algorithms. If a large enough quantum computer were to be built, RSA would no longer be secure.

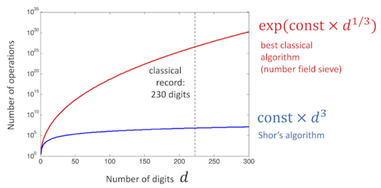


Figure 1: Classical vs Quantum factoring algorithm runtimes [2]

References

[1] Federico Grasseli. *Quantum Cryptography, From Key Distribution to Conference Key Agreement*. Springer Nature Switzerland AG, 2021.

[2] IBM. *Shor's Algorithm - IBM Quantum Computing Composer Documentation*. IBM (International Business Machines Corporation), 2021.

[3] Michael Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, New York, 2010.

Guided by

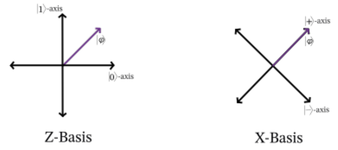
Sanjay Kumar
UCSB

Quantum Bits

A classical bit can be prepared in one of two states: 0 or 1. A quantum bit, or **qubit** for short, can be prepared in any superposition of those two states. If we define the 0 state as the vector $|\mathbf{0}\rangle = [\mathbf{1}, \mathbf{0}]$ and the 1 state as $|\mathbf{1}\rangle = [\mathbf{0}, \mathbf{1}]$, a superposition of these two states is a linear combination of the two vectors. We call $|0\rangle$ and $|1\rangle$ basis vectors, and these particular vectors form the computational basis, also referred to as the **Z-basis**.

Another common basis is the **X-basis**, with basis vectors $|+\rangle = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]$ for the 0 state, and $|-\rangle = [\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]$ for the 1 state. Consider the qubit represented by the state $|\varphi\rangle = \frac{1}{\sqrt{2}}[1, 1]$, and see what happens when it is measured in either basis.

$$|\varphi\rangle = \frac{1}{\sqrt{2}}[1,1] =$$



In the Z-basis, $|\varphi\rangle$ can be expressed as $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$, so it collapses to either state with equal probability. However, in the X-basis, $|\varphi\rangle = |+\rangle + 0\,|-\rangle$, so it collapses to the 0 state every time.

Non-Orthogonal Advantages

With classical bits, the two states they can be prepared in are orthogonal to each other, meaning there is no interaction between the 0 state and 1 state. If a classical bit is measured as a 0, there is no possibility it was prepared as a 1.

Two vectors, $\vec{U}$ and $\vec{V}$, are orthogonal if the following condition is true.

$$\vec{U} \cdot \vec{V} = 0$$

Non-orthogonality allows for the creation of qubits that can only be read in a specific way. In the example above, $|\varphi\rangle$ was prepared in a state not orthogonal to the Z-basis.

$$|\varphi\rangle \cdot |0\rangle = \frac{1}{\sqrt{2}}$$
$$|\varphi\rangle \cdot |1\rangle = \frac{1}{\sqrt{2}}$$

When $|\varphi\rangle$ gets measured in the Z-basis, it collapses to 0 or 1 with equal probability. This does not tell you if $|\varphi\rangle$ was intended to be a 0 or 1 to begin with, illustrating the main advantage of non-orthogonal states. Measuring a qubit as a 0 (or 1) does not mean it was prepared as a 0 (or 1) unless you know in which basis it was prepared in and measure it accordingly.

The No-Cloning Theorem

The no-cloning theorem states that it is impossible to create a perfect copy of an unknown quantum state without altering the original state in some way.

Suppose a perfect cloning machine exists.

- Apply it on two quantum states $|\psi\rangle \neq |\phi\rangle$ which are non-orthogonal.
- The action of the cloning machine is represented by a unitary operation U.
  U copies the input state on some auxiliary system initially in a normalized state $|s\rangle$:
  $$U(|\psi\rangle \otimes |s\rangle) = |\psi\rangle \otimes |\psi\rangle$$
  $$U(|\phi\rangle \otimes |s\rangle) = |\phi\rangle \otimes |\phi\rangle$$
- Taking the inner product of the equations, we obtain:
  $$\langle\psi|\phi\rangle = (\langle\psi|\phi\rangle)^2$$
- This is only true when $|\psi\rangle$ and $|\phi\rangle$ are either equivalent or orthogonal.

Thus, by contradiction, such a cloning machine does not exist.

Quantum Key Distribution

Quantum key distribution **(QKD)** takes advantage of the no-cloning theorem and non-orthogonality principle to distribute cryptographic keys securely. By starting with a sequence of 0s and 1s, a qubit string can be prepared which encodes the classical states allowing for secure transmission.

The use of non-orthogonal states in QKD protocols ensure even if someone were to measure a qubit, without knowing which basis to use, the measurement will be useless. The no-cloning theorem prevents a malicious party from copying the key without leaving fingerprints.

A common QKD protocol is BB84, which, in contrast to classical protocols, is not simply difficult to crack, but impossible.

BB84 Protocol

1. Alice *(the sender)* chooses $4n$ random data bits.
2. Alice randomly encodes each bit in the Z or X basis, keeping track of which one she uses.
3. Alice sends the resulting product of qubits to Bob *(the receiver)*.
4. Bob receives the $4n$ qubits, announces this fact, and measures each qubit in the X or Z basis at random.
5. Alice and Bob compare the bases they used for each qubit, and discard any bits for which their bases don't match. With high probability, there are at least $2n$ bits left, which they keep. *If there are less than $2n$ bits, they abort the protocol.*
6. Alice selects n of these bits to check for interference and tells Bob which bits she selected.
7. Alice and Bob compare the values of the check bits. *If more than an acceptable number disagree, they abort the protocol.* Otherwise, the n non-check bits are kept and are the private key.

Even if there is an eavesdropper, Eve, she will not be able to successfully intercept the private key. If she attempts to measure an intercepted qubit, since she won't know which basis it was prepared in, the non-orthogonality principle ensures that Eve will not be able to determine if it is a 0 or 1.

Additionally, the no-cloning theorem prevents Eve from cloning an intercepted qubit and waiting until Alice announces the basis she used to prepare it. If she attempts to, she will change the state the qubit is in when Bob receives it and Alice and Bob might not have matching bits even if their bases match. In step 7, Alice and Bob check for this type of interference, and if they find enough, they will abandon this attempt.

BB84 Protocol Example

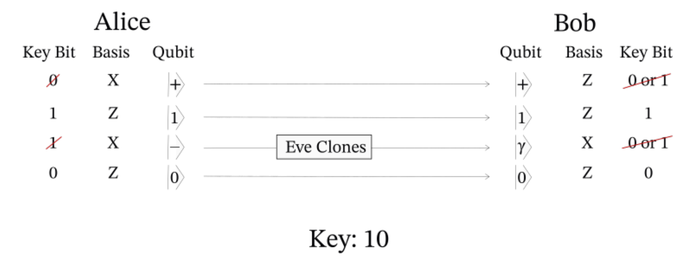


Figure 2: Here, Alice and Bob will discard the first bit since they used different bases. Alice has chosen the third bit as the interference test bit, which happens to be the a qubit Eve tried to clone. Alice and Bob will compare their third bits, possibly realizing they don't match. If this exceeds their tolerance for interference, they will abandon the protocol. Otherwise, the key will be 10.

Quivers

Sogol Cyrusian and Amy Somers

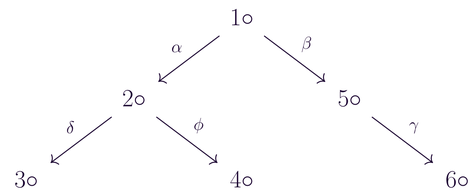
University of California, Santa Barbara - Directed Reading Program 2023

Quivers

- A **quiver** $Q = (Q_0, Q_1, s, t)$ consists of the data associated with a directed graph where loops, as well as multiple edges between vertices, are allowed. Q_0 is the set of vertices of Q, and Q_1 is the set of arrows, $s, t : Q_1 \to Q_0$ are maps associating to each arrow $\alpha \in Q_1$ its **source**, $s(\alpha) \in Q_0$, and its **target**, $t(\alpha) \in Q_0$.
- A **path** with source a and target b is a sequence
$$(a|\alpha_1, \alpha_2, ..., \alpha_\ell|b)$$
where $\alpha_i \in Q_1$ for all $1 \leq i \leq \ell$ and ℓ is the **length** of the path. The path $(a|\alpha_1, \alpha_2, ..., \alpha_\ell|b)$ can be visualized as follows:
$$a = a_0 \xrightarrow{\alpha_1} a_1 \xrightarrow{\alpha_2} a_2 \longrightarrow \cdots \xrightarrow{\alpha_\ell} a_\ell = b$$
Each vertex $a \in Q_0$ has a **trivial path** $\varepsilon_a = (a||a)$ of length $\ell = 0$ associated with it.
- The **path algebra** KQ of a quiver Q is a vector space over the field K with a basis consisting of the set of all paths $(a|\alpha_1, \alpha_2, ..., \alpha_\ell|b)$ in Q where the product of two basis vectors
$\gamma = (a|\alpha_1, \alpha_2, ..., \alpha_\ell|b), \varphi = (c|\beta_1, \beta_2, ..., \beta_k|d)$ is
$\varphi\gamma = (a|\alpha_1, \alpha_2, ..., \alpha_\ell, \beta_1, \beta_2, ..., \beta_k|d)$ if $t(\alpha_\ell) = s(\beta_1)$ and 0 otherwise.

We give the following examples of quivers and path algebras:

(a) Let Q be the quiver



Then the basis of the path algebra KQ is
$\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \alpha, \delta, \phi, \beta, \gamma, \delta\alpha, \phi\alpha, \gamma\beta\}$.

(b) Let Q be the quiver

Then the basis of the path algebra KQ is $\{\varepsilon_1, \alpha, \alpha^2, \alpha^3, ..., \}$.
Note: if we consider the mapping defined by $\varepsilon_1 \mapsto 1, \alpha \mapsto t$, we get that $KQ \cong K[t]$!

Quiver Representations

A **representation** M of a quiver Q is defined by the following data:

(i) To each vertex $a \in Q_0$ is associated a K-vector space V_a,

(ii) To each arrow $\alpha \in Q_1$ is associated a linear transformation $V_{s(\alpha)} \to V_{t(\alpha)}$ between the vector spaces associated to the source and target of α.

Examples of quiver representations:

- One representation of any quiver Q is the zero representation given by associating each vertex to the zero vector space and each arrow to the zero map.
- The Kronecker Quiver is $1\circ \;\substack{\longleftarrow\\ \longleftarrow}\; \circ 2$
An example of a representation of this quiver is:
$$K^2 \;\underset{(0,1)^T}{\overset{(1,0)^T}{\longleftarrow\!\longleftarrow}}\; K$$
Another representation is:
$$K^2 \;\underset{B}{\overset{A}{\longleftarrow\!\longleftarrow}}\; K^2$$
Where $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$

If $M = (V_a, f_\alpha), \tilde{M} = (\tilde{V}_a, \tilde{f}_\alpha)$ are representations of Q, then $\phi : M \to \tilde{M}$ is a **morphism** if it consists of the data of a collection of linear maps $\{\phi_a : V_a \to \tilde{V}_a\}$ such that the following diagram commutes:

$$\begin{array}{ccc} V_a & \xrightarrow{f_\alpha} & V_b \\ {\scriptstyle \phi_a}\downarrow & & \downarrow{\scriptstyle \phi_b} \\ \tilde{V}_a & \xrightarrow{\tilde{f}_\alpha} & \tilde{V}_b \end{array}$$

For example there is a morphism ϕ between the two representations of the Kronecker Quiver given above given by $\phi_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\phi_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$$\begin{array}{ccc} K^2 & \underset{(0,1)^T}{\overset{(1,0)^T}{\longleftarrow\!\longleftarrow}} & K \\ {\scriptstyle \phi_1}\downarrow & & \downarrow{\scriptstyle \phi_2} \\ K^2 & \underset{B}{\overset{A}{\longleftarrow\!\longleftarrow}} & K^2 \end{array}$$

Modules Correspond to Representations

- For any representation of a quiver Q we have a collection of vector spaces $\{V_a\}$ and a collection of maps $\{V_a \to V_b\}$. Consider the vector space
$$V = \bigoplus_{a \in Q_0} V_a$$
with the action of a path $a_0 \to \cdots \to a_\ell$ given by the composition of the corresponding linear maps
$$V \twoheadrightarrow V_{a_0} \to \cdots \to V_{a_\ell} \hookrightarrow V$$
Then V is a module over KQ, or KQ-module, corresponding to the given representation.
- If V is any KQ-module, then we may define a representation of Q by, $V_a = \varepsilon_a V$ and for all basis elements α of KQ corresponding to the arrow $a \to b$, α is a map $V_a \to V_b$ because $\alpha\varepsilon_a = \alpha$ and $\varepsilon_b\alpha = \alpha$.

Arrow Ideals & Quotients of the Path Algebra

Let Q be a quiver. The **arrow ideal** R_Q of the path path algebra KQ is the subspace of KQ generated by paths of length $\ell \geq 1$. And so, R_Q is closed under multiplication by any elements in KQ.
Notice that there is a direct sum decomposition
$$R_Q = KQ_1 \oplus KQ_2 \oplus \cdots \oplus KQ_\ell \oplus \cdots$$
where Q_ℓ is the set of all paths in Q of length ℓ.
For all $\ell \geq 1$,
$$R_Q^\ell = \bigoplus_{m \geq \ell} KQ_m$$
is the ideal where all paths are of length greater than or equal to ℓ, as it is a subspace of KQ and is closed under multiplication from elements of KQ.
Note: it follows that $R_Q^\ell / R_Q^{\ell+1}$ is a vector space over K with basis Q_ℓ, and so is isomorphic to KQ_ℓ.
KQ/I is said to be a **bound quiver algebra** if I is an ideal such that $R_Q^m \subseteq I \subseteq R_Q^2$.

Proposition

Let Q be a finite quiver, and KQ/I be a bound quiver algebra. Then, KQ/I is finite dimensional.

Proof:
Since KQ/I is a bound quiver algebra, there exists $m \geq 2$ such that $R^m \subseteq I$, where R is the arrow ideal R_Q of KQ. But then there exists a natural surjective homomorphism $KQ/R^m \to KQ/I$. And so, it suffices to show that KQ/R^m is finite-dimensional. Now the equivalence classes of paths of length less than m form a basis of KQ/R^m as a K-vector space. Since there are finitely many such paths, we are done. ■

References & Acknowledgments

Thanks to our DRP mentor Mychelle Parker and the DRP Organizing Committee.

[1] Ibrahim Assem and Daniel Simson and Andrzej Skowroński. *Elements of the Representation Theory of Associative Algebras Volume I: Techniques of Representation Theory*. Cambridge University Press, 2006.

[2] D. J. Benson. *Representations and Cohomology I*. Cambridge University Press, 1995.

[3] Karin Erdmann and Mark J. Wildon. *Introduction to Lie Algebras*. Springer Science & Business Media, 2006.

Solvable and Semisimple Lie Algebras

Elise Alvarez-Salazar and Ava Brückner-Kockel
University of California Santa Barbara

Abstract

An *algebra* A is a vector space over a field together with a bilinear map from $A \times A \to A$. A Lie algebra is a non-associative algebra and can be visualized as the tangent plane to a Lie group. Two key properties of a Lie algebra are solvability and semisimplicity. Understanding when a Lie algebra is solvable or semisimple is helpful in determining its structure and characteristics. It also helps classify finite-dimensional Lie algebras. This poster lays out the different ways in which solvability and semisimplicity can be classified.

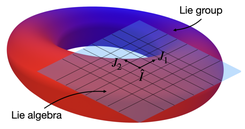
Preliminary Information

DEFINITION 1. A **Lie algebra,** L, over a field F is an F-vector space together with a bilinear map called the **Lie bracket**:

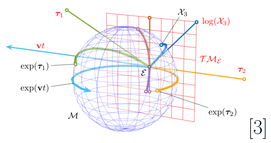
$$L \times L \to L$$
$$(x, y) \longmapsto [x, y]$$

such that: i) $\forall x \in L,\ [x, x] = 0$
ii) $\forall x, y, z \in L,\ [x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$

We can visualize our Lie algebra as a tangent plane of a Lie group:



[2]



[3]

DEFINITION 2. Let L_1 and L_2 be Lie algebras over a field F. Then, a map $\phi : L_1 \to L_2$ is a **Lie algebra homomorphism** if and only if ϕ is a linear map where

$$\phi([x, y]) = [\phi(x), \phi(y)] \quad \forall\, x, y \in L_1$$

Example 1. Let V be a finite dimensional vector space over F. We call the vector space gl(V) the set of all all linear maps from $V \to V$, and together with the Lie bracket operation

$$[x, y] = x \circ y - y \circ x \quad \text{for } x, y \in \text{gl}(V), \text{ makes gl}(V) \text{ into a Lie algebra.}$$

A particularly important Lie algebra homomorphism is the adjoint homomorphism.

DEFINITION 3. For a Lie algebra, L, define the **adjoint homomorphism** as follows:

$$\text{ad: } L \to \text{gl}(V) \quad \text{where for } x, y \in L$$
$$(\text{ad } x)(y) := [x, y]$$

Analogs to Groups and Rings

An interesting tangent is that Lie algebras are structurally similar to two well-known algebraic objects: groups and rings.

DEFINITION 4. An **ideal** of a Lie algebra L is a subspace, $I \subseteq L$, such that for all $x \in L$, and for all $y \in I$, $[x, y] \in I$.

DEFINITION 5. A Lie algebra is **abelian** if for all $x, y \in L,\ [x, y] = 0$

DEFINITION 6. A non-abelian Lie algebra L is **simple** if it has no ideals other than 0 and itself.

DEFINITION 7. A Lie algebra is said to be **solvable** if for some $m \in \mathbb{N}$, m ≥ 1, we have $L^{(m)} = 0$ where $L^{(m)} = [L^{(m-1)}, L^{(m-1)}]$ for $m \geq 2$ and $L^{(1)} = L' = [L, L]$.

Solvable Lie Algebras

Another condition which gives solvability of a Lie algebra, in terms of ideals, is: If L is a Lie algebra with ideals I_j such that for some $m \in \mathbb{Z}_{\geq 0}$

$$L = I_0 \supseteq I_1 \supseteq \cdots \supseteq I_{m-1} \supseteq I_m = 0$$

where I_{k-1}/I_k is abelian for all $1 \leq k \leq m$, then L is solvable. An interesting characteristic of solvable ideals is that there is a *unique* solvable ideal of L which contains every other solvable ideal of L.

DEFINITION 8. The largest such solvable ideal is called the **radical of** L, rad L.

A third way to test for solvability, which becomes more useful when working with complex problems, is defined below:

DEFINITION 9. The **Killing form** on a complex Lie algebra L is a symmetric, bilinear form $\kappa : L \times L \to \mathbb{C}$ defined as:

$$\kappa(x, y) = \text{tr}(\text{ad}x \circ \text{ad}y), \text{ for } x, y \in L.$$

THEOREM 1 (Cartan's First Criteria)**.** *The complex Lie algebra L is solvable if and only if $\kappa(x, y) = 0$ for all $x \in L$ and $y \in L'$.*

Example 2. Any 2-dimensional non-abelian Lie algebra, L, has basis $\{x, y\}$ where $[x, y] = x$. Take L to be a 2-dimensional complex non-abelian Lie algebra. Consider: (ad x)(x) = $[x, x] = 0$, (ad x)(y) = $[x, y] = x$, (ad y)(x) = $[y, x] = -[x, y] = -x$, and (ad y)(y) = $[y, y] = 0$. Notice that this also gives us a basis of L', namely $\{x\}$. With respect to our basis for L, ad x and ad y are:

$$\text{ad } x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad \text{ad } y = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$$

Since $\{x\}$ is a basis for L', $\{x\}$ spans L', that is, all elements of L' are scalar multiplies of x. By Cartan's First Criteria, to prove solvability, it is sufficient to prove that $\kappa(y, x) = 0 = \kappa(x, x)$.

$$\kappa(y, x) = tr(\begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix} \circ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}) = tr(\begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}) = 0$$

$$\kappa(x, x) = tr(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \circ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}) = tr(\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}) = 0$$

Thus, any two dimensional non-abelian complex Lie algebra is solvable.

Semisimple Lie Algebras

Besides its usefulness in terms of solvability, the ideals of a Lie algebra help determine whether a Lie algebra is semisimple.

DEFINITION 10. Let L be a nonzero Lie algebra. L is **semisimple** if it has no nonzero solvable ideals. Alternatively, L is semisimple if rad L = 0.

DEFINITION 11. For $S \subseteq V$ and a bilinear, symmetric form $\beta : V \times V \to V$, we define:

$$S^{\perp} = \{x \in V \mid \beta(x, y) = 0,\ \forall y \in V, \}$$

If $V^{\perp} = 0$, β is said to be **non-degenerate**.

DEFINITION 12. The Lie algebra L is **nilpotent** if for some $m \geq 1$, we have $L^m = 0$, where $L^m = [L, L^{m-1}]$.

THEOREM 2 (Cartan's Second Criteria)**.** *The complex Lie algebra, L, is **semisimple** if and only if the Killing form κ of L is non-degenerate.*

Proof Sketch:
$\Rightarrow$: $L^{\perp} \subseteq L$ is an ideal of L. By definition of an ideal, $(L^{\perp})' \subseteq L^{\perp}$. By Cartan's First Criterion $L^{\perp}$ is solvable since $\kappa(x, y) = 0$ for all $x \in L^{\perp}$ and for all $y \in (L^{\perp})'$. Since L is semisimple, $L^{\perp} = 0$ and thus κ is non-degenerate.

Semisimple Lie Algebras: Continued

$\Leftarrow$: Suppose L is not semisimple, then rad $L \neq 0$. L has a nonzero abelian ideal, A. For nonzero $a \in A$ and $x \in L$, im(ad$x \circ$ ad$a$) $\subseteq A$, thus $(\text{ad}a \circ \text{ad}x)^2 = 0$, which means ad$a \circ$ ad$x$ is a nilpotent map. Nilpotent maps have trace 0, thus tr(ad$a \circ$ ad$x$) $= \kappa(a, x) = 0$. This holds true for all $x \in L$ and so $a \in L^{\perp}$, which implies κ is degenerate. Thus, L is semisimple.

Further Ideas

Another perspective to Lie algebras is in the form of representations. Informally, a representation is a function which compresses the information of a Lie algebra L into a matrix representation.

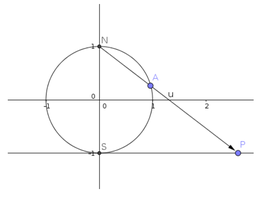
DEFINITION 13. A **representation** of L is a *Lie algebra homomorphism*

$$\phi : L \to gl(V)$$

where V is a finite-dimensional vector space over F.

We note that the image of ϕ is a Lie subalgebra of gl(V) and $\ker(\phi) \subseteq L$ is an ideal. This leads to the question: Can we discover conditions for solvability based solely off a representation, ϕ, for the Lie algebra?

Lie groups are groups that are also smooth manifolds. To begin understanding the Lie algebra of a Lie group we can start with studying the tangent spaces of the manifold and a defined Lie bracket. The unit circle is an example of a smooth manifold.



The relationship between Lie groups and Lie algebras is something worth exploring and allows for overlapping subject matter.

More on Lie Theory:

Acknowledgements

We thank Melody Molander for all her hard work, kindness, patience, and expert guidance while working on this project. We would also like to thank the Directed Reading Program and all its administrative personnel for the opportunity to participate in this program.

References

[1] Karin Erdmann and Mark J. Wildon. *Introduction to Lie Algebras*. London, UK: Springer-Verlag, 2006.

[2] HashPi. "Lie Groups: Intuition and Geometrical Interpretation". In: *HashPi* (2023).

[3] Joan Solà, Jeremie Deray, and Dinesh Atchuthan. *A Micro Lie Theory for State Estimation in Robotics*. Dec. 2018.

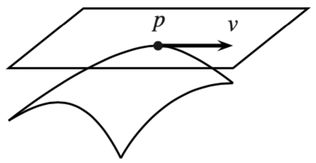
Tangent Vectors as Derivations

Christian Reynaldo Mentor: Daniel Halmrast

University of California, Santa Barbara

Introduction

In elementary calculus, the tangent space is typically introduced as the vector space orthogonal to the gradient of a function at a point. Intuitively, one can visualize any tangent vector as an arrow emanating from the given point, typically satisfying an equation of tangency. This approach works well when considering surfaces given by a single smooth function because we can imagine how a surface might sit in $\mathbb{R}^3$. However, on the more abstract subject of manifolds we aim for a more intrinsic definition of the tangent space.



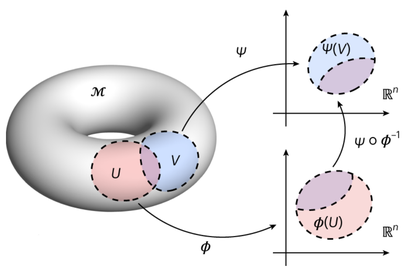
Definitions

Germ: A germ at $p \in \mathbb{R}^n$ is an equivalence class of C^∞ real-valued functions wherein two functions are equivalent if they agree on some neighborhood of p. In this way any directional derivative can be thought to operate on the set of germs at p. This set is an algebra over $\mathbb{R}$ denoted by C_p^∞.

Derivation: A derivation is a linear map $D : C_p^\infty \to \mathbb{R}$ satisfying the Leibniz rule $D(fg) = (Df)g(p) + f(p)Dg$. The set of all derivations of this kind is the real vector space $\mathcal{D}_p\mathbb{R}^n$.

Coordinate Chart: A topological space M is locally Euclidean of dimension n if every point $p \in M$ has a neighborhood U such that there is a homeomorphism ϕ from U into an open subset of $\mathbb{R}^n$. The pair (U, ϕ) is called a coordinate chart.

Smooth Manifold: A topological space M is said to be a smooth manifold if it is Hausdorff, second countable, and has a C^∞ atlas. An atlas is a collection of coordinate charts that cover M, and we call it C^∞ if the transition functions are smooth. Some classical examples include the n-sphere, the torus, and perhaps the most elementary is $\mathbb{R}^n$ itself.



Smooth Map: A map of manifolds $F : M \to N$ is said to be smooth at $p \in M$ if, for coordinate charts (U, ϕ) containing p and (V, ψ) containing $F(p)$, we have $\psi \circ F \circ \phi^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ being smooth at $\phi(p) \in \mathbb{R}^m$.

Tangent Vectors as Derivations in $\mathbb{R}^n$

In an intuitive sense tangent vectors might best be thought of as directions of travel. For this reason a tangent vector to $p \in \mathbb{R}^n$ is any n-dimensional vector $v = \langle v_1, ..., v_n \rangle$. The set of tangent vectors forms a vector space $T_p\mathbb{R}^n$. For any tangent vector v at p, the directional derivative $D_v : C_p^\infty \to \mathbb{R}$ is linear and satisfies the Leibniz rule. Hence, it is in fact a derivation.

Theorem

The map $\varphi : T_p\mathbb{R}^n \to \mathcal{D}_p\mathbb{R}^n$ given by $v \mapsto D_v$ is an isomorphism of vector spaces.

Proof

Injectivity: Suppose $D_v = 0$ for some $v \in T_p\mathbb{R}^n$. If we apply D_v to the coordinate function r^j then we have

$$0 = D_v(r^j) = \sum_i v^i \frac{\partial}{\partial r^i}\bigg|_p r^j = v^j$$

Since this is true for $1 \leq j \leq n$ we have $v = 0$, and so φ is injective.

Surjectivity: Let D be an arbitrary derivation at p, and let $f : \mathbb{R}^n \to \mathbb{R}$ be the representation of some germ in C_p^∞. By Taylor's theorem with remainder there exists smooth functions $g_i(x)$ in a neighborhood of p such that

$$f(x) = f(p) + \sum (r^i(x) - p^i) g_i(x), \quad g_i(p) = \frac{\partial f}{\partial r^i}(p)$$

Now, applying D to both sides we get by the Leibniz rule,

$$Df = \sum \left(Dr^i g_i(p) + (p^i - p^i) Dg_i \right) = \sum (Dr^i) \frac{\partial f}{\partial r^i}(p)$$

$$D = \sum (Dr^i) \frac{\partial}{\partial r^i}\bigg|_p$$

Thus $D = D_v$ where $v = \langle Dr^1, ..., Dr^n \rangle$. This shows that every derivation is the directional derivative with respect to some vector, and so φ is a bijection.

With this in mind, we will redefine a tangent vector at p in $\mathbb{R}^n$ to be a derivation at p, and the tangent space $T_p\mathbb{R}^n$ is the vector space of derivations with basis $\{\partial/\partial r^i|_p\}_{i=1}^n$.

Generalizing to Manifolds

It is rather straightforward now to extend our idea of a tangent space to manifolds. We simply tweak our derivation definition to be a map $D : C_p^\infty(M) \to \mathbb{R}$, where $C_p^\infty(M)$ denotes the set of germs at any $p \in M$. These derivations form the tangent space T_pM, and the above becomes the particular case $M = \mathbb{R}^n$.

Our goal has thus been reached, as the tangent space has been defined in a way that does not depend on any coordinate chart. However, each coordinate chart (U, ϕ) containing p can yield a basis for T_pM as follows. We define the derivation $\partial/\partial x^i|_p$ such that for any $f \in C_p^\infty(M)$ we have

$$\frac{\partial}{\partial x^i}\bigg|_p f = \frac{\partial}{\partial r^i}\bigg|_{\phi(p)} (f \circ \phi^{-1})$$

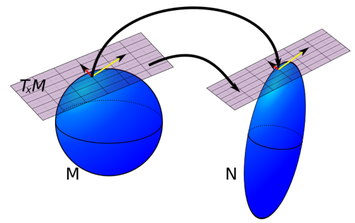
The collection of these derivations are linearly independent, and hence form a basis.

References

[1] Loring W. Tu. Manifolds. *An Introduction to Manifolds*. Springer, 2nd edition, 2011.

The Pushforward

Given a smooth map $F : M \to N$, the pushforward of F at $p \in M$ is a linear map $F_* : T_pM \to T_{F(p)}N$ such that for any $v \in T_pM$ and $f \in C_{F(p)}^\infty(N)$ we have $F_*(v)f = v(f \circ F)$.



Any coordinate chart inverse ϕ^{-1} provides a smooth map of manifolds between an open subset of $\mathbb{R}^n$ and a manifold M. For a point $p \in M$ we can then consider the pushforward $(\phi^{-1})_* : T_{\phi(p)}\mathbb{R}^n \to T_pM$. If we attempt to apply this map to our basis vectors $\partial/\partial r^i|_p$ we get the following result.

$$(\phi^{-1})_* \left(\frac{\partial}{\partial r^i}\bigg|_{\phi(p)} \right) f = \frac{\partial}{\partial r^i}\bigg|_{\phi(p)} (f \circ \phi^{-1}) = \frac{\partial}{\partial x^i}\bigg|_p f$$

Properties

Matrix Representation: Being linear, the pushforward can be represented by a matrix. This matrix is the Jacobian $[\partial F^i/\partial x^j(p)]$.

The Chain Rule: One final property of the pushforward that will be used in the next section is its chain rule. Some elementary linear algebra gives us the following powerful result: If $F : M \to N$ and $G : N \to P$ are both smooth maps of manifolds, then we have $(G \circ F)_* = G_* \circ F_*$.

Applications to Calculus

The usual chain rule taught in calculus can be proven as a particular case for when we consider smooth maps from $\mathbb{R}^m$ to $\mathbb{R}^n$.

As an example, let $F : \mathbb{R} \to \mathbb{R}^3$ and $G : \mathbb{R}^3 \to \mathbb{R}$ be smooth functions and let w be such that

$$w = (G \circ F)(t) = G\big(F^1(t), F^2(t), F^3(t)\big)$$

The pushforwards F_*, G_*, and $(G \circ F)_*$ are given by the following matrices.

$$F_* = \begin{bmatrix} dF^1/dt \\ dF^2/dt \\ dF^3/dt \end{bmatrix} \quad G_* = \begin{bmatrix} \frac{\partial w}{\partial F^1} & \frac{\partial w}{\partial F^2} & \frac{\partial w}{\partial F^3} \end{bmatrix} \quad (G \circ F)_* = \frac{dw}{dt}$$

The chain rule for the pushforward gives us $(G \circ F)_* = G_* \circ F_*$, or equivalently through multiplication of the matrices above,

$$\frac{dw}{dt} = \frac{\partial w}{\partial F^1}\frac{dF^1}{dt} + \frac{\partial w}{\partial F^2}\frac{dF^2}{dt} + \frac{\partial w}{\partial F^3}\frac{dF^3}{dt}$$

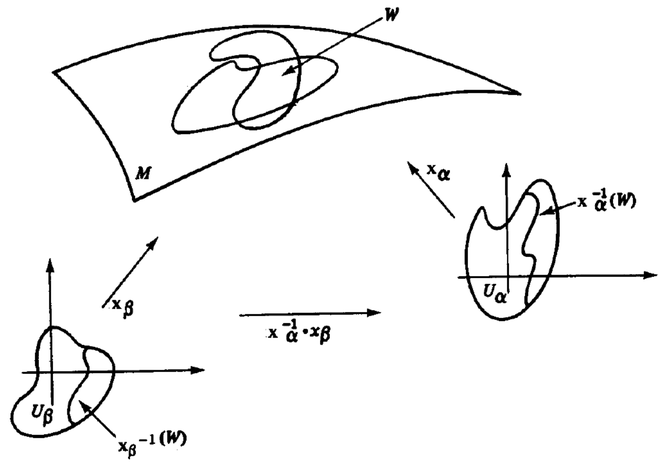
The Earth is (Locally) Flat

An Introduction to Riemannian Geometry

Jeremy Lauro and Fabio Ricci

What is a Manifold?

The backbone of differential geometry is the idea of a manifold. Intuitively, a manifold can be thought of as a arbitrary 3D shape. However, in reality, manifolds can be any dimension.



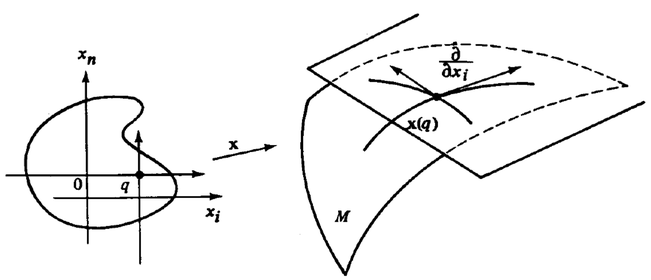
More rigorously, A *differential manifold* of dimension n is a set $\mathbf{M}$ and collection of smooth mappings $\mathbf{x}_\alpha : U_\alpha \subset \mathbf{R}^n \to M$ of open set such that:

1. $\bigcup_\alpha \mathbf{x}_\alpha(U_\alpha) = \mathbf{M}$.
2. if $\mathbf{x}_\alpha(U_\alpha) \cap \mathbf{x}_\beta(U_\beta) = W \neq \emptyset$, then $\mathbf{x}^{-1}(W)$ are open in $\mathbb{R}^n$ and $\mathbf{x}_\beta^{-1} \otimes \mathbf{x}_\alpha^{-1}$ and differentiable.

Tangent Vectors & Tangent Spaces

Now, we would like to extend the idea of **tangent vectorc** in the manifold settting. Intuitively a tangent vector is an infinitesimal displacement at a specific point on a manifold.

We think of a **tangent vector** at a point p $\in \mathbf{M}$ as the *directional derivative* of some curve $\alpha : (-\epsilon, \epsilon) \to \mathbf{M}$ with $\alpha(0) = p$. The set of all tangent vectors to $\mathbf{M}$ at p is know as the **tangent space** to $\mathbf{M}$ at p ($T_p\mathbf{M}$)



It can be shown that the set ($T_p\mathbf{M}$) with the usual operations of functions, forms a vector space of equal dimension to the manifold. This vector space is isomorphic to $\mathbb{R}^n$. Finally, the choice of parameterization $\mathbf{x} : U \to \mathbf{M}$ determines an **associated basis**

$$\left\{ \left(\frac{\partial}{\partial x^1}\right)_0, ..., \left(\frac{\partial}{\partial x^n}\right)_0 \right\}.$$

Riemannian Manifolds

After constructing a differential manifold, one can introduce riemannian structure by defining a metric. A Riemannian metric on $\mathbf{M}$ is a correspondence that associates each $p \in \mathbf{M}$ to an inner product $\langle\ ,\ \rangle_p$ (i.e., a symmetric, bilinear, positive-define 2-from) on $\boldsymbol{T_pM}$ that varies smoothly in any coordinate neighborhood.

Example: The Euclidean Metric

If our manifold $\mathbf{M} = \mathbb{R}^n$, the the elements of the tangent space of $\mathbf{M}$ are given by $\{\frac{\partial}{\partial x^i}\} \equiv \{\mathbf{e}_i\}$ and our metric is

$$\mathbf{g}_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}.$$

The geometry induced by this metric on $\mathbb{R}^n$ is called euclidean geometry.

Connections & Covariant Derivative

To generalize the notion of "directional derivatives" to vector fields, and by extension, tensor fields, we define a new notation for directional derivatives (∇) and consider the properties we wish the operator to have.

1. $\nabla_X(f) = Xf$.
2. $\nabla_X(T + S) = \nabla_X(T) + \nabla_X(S)$.
3. $\nabla_X(T \otimes S) = \nabla_X T \otimes S + T \otimes \nabla_X S$.
4. $\nabla_{fX+Z}T = f\nabla_X T + \nabla_Z T$.

With these definitions, it turns out the connection of tensor fields depend on connection coefficients Γ^k_{ij} in the following way: Let $X, Y \in \mathcal{X}(M)$.

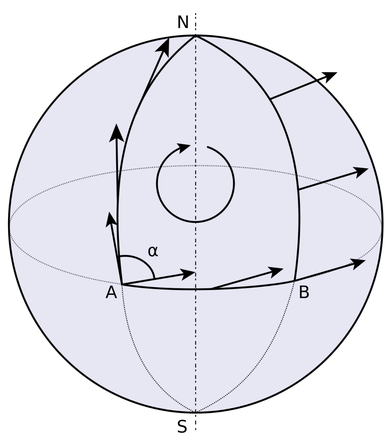
$$\nabla_X Y = \left(X^i \frac{\partial Y^k}{\partial x^i} + X^i Y^j \Gamma^k_{ij} \right) \frac{\partial}{\partial x^k} \qquad \Gamma^k_{ij} = \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}.$$

As it turns out, a Riemannian metric **g** uniquely specifies the connection coefficients such that

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl} \left(\partial_i g_{ji} + \partial_j g_{il} - \partial_l g_{ij} \right)$$

In this case, we refer to the connection coefficients as the Christoffel symbols and the connection as a covarient derivative.

Geodesics & Distance



With a Riemannian metric, it is possible to calculate the length of a curve. Let $c : I \to \mathbf{M}$ be restricted to a closed interval $[a, b] \subset I$. The length of this curve from $t = a$ to $t = b$ is defined as

$$\ell^a_b(c) = \int_a^b \left\langle \frac{dc}{dt}, \frac{dc}{dt} \right\rangle^{1/2} dt.$$

With this, we define a **geodesic** to be a curve between two points $a, b \in \mathbf{M}$ that minimizes ℓ^a_b. However, this definition of a geodesic is impractical to use. Instead, we favor an alternative definition: a geodesic is a curve of *zero acceleration* with respect the covariant derivative.

$$0 = \nabla_{v_\gamma} v_\gamma \iff 0 = \ddot{\gamma}^m + \dot{\gamma}^n \dot{\gamma}^m \Gamma^q_{mn}$$

As it turns out, these two definitions are equivalent.

The Many Types of Curvature

The Riemannian Curvature Tensor

The Riemannian curvature tensor R is a multi-linear mapping that associates every $X, Y \in \mathcal{X}(M)$ a mapping $R(X, Y) : \mathcal{X}(M) \to \mathcal{X}(M)$ given by

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z + \nabla_{[X,Y]} Z$$

Algebraically, the curvature tensor measures the failure of the covariant derivative to commute.

Example: The Curvature of Flat Space

In flat space, two vector fields will always commute as the basis vector fields commute. As such, for all $X, Y, Z \in \mathcal{X}(M)$

$$R(X, Y)Z = 0$$

Sectional Curvature

Given any two $X, Y \in T_p\mathbf{M}$ that form a subspace $\sigma \subset T_p\mathbf{M}$, the section curvature is defined as

$$K(X, Y) = \frac{\langle R(X, Y)X, Y \rangle}{\sqrt{|X|^2|Y|^2 - \langle X, Y \rangle}}$$

Ricci Curvature

Let $\{z_1, ..., z_j, ..., z_n\}$ be an orthonormal basis for $T_p\mathbf{M}$, then the Ricci curvature is

$$Ric_p(z_j) = \frac{1}{n-1} \sum_{i \neq j} \langle R(z_j, z_i)z_j, z_i \rangle$$

Scalar Curvature

Taking the average of the Ricci curvature for each Z_j in the orthonormal basis of $T_p\mathbf{M}$ gives the scalar curvature

$$Sc(p) = \frac{1}{n} \sum_j Ric_p(z_j) = Ric_p(z_j) = \frac{1}{n(n-1)} \sum_j \sum_i \langle R(z_j, z_i)z_j, z_i \rangle$$

The Bonnet-Meyers Theorem

The **Bonnet-Meyers theorem** is a classic theorem in Riemannian geometry connecting the Ricci curvature of a manifold to its diameter. Suppose $\mathbf{M}$ is a complete Riemannian manifold whose Ricci curvature satisfies

$$Ric_p(v) \geq \frac{1}{r^2} > 0$$

for any $p \in \mathbf{M}$ and all $v \in T_p\mathbf{M}$. Then M is compact and $diam(\mathbf{M}) \leq \pi r$, where

$$diam(\mathbf{M}) = \sup_{a,b} \left(\ell a^b(c) \right)$$

and the curve $C(t)$ is a geodesic with $C(0) = a$ and $C(1) = b$. In essence, the Bonnet-Meyers theorem states the larger the curvature, smaller the size.

References & Acknowledgements

- Do Carmo, Manfredo Perdigaõ. *Riemannian Geometry*. Birkhäuser, 1993.
- Lee, John M. *Introduction to Riemannian Manifolds*. 2nd ed., Springer, 2018.
- Lee, John M. *Introduction to Smooth Manifolds*. Springer, 2013.
- Lauro Jeremy, Ricci Fabio, our own Overaleaf notes.

Special thanks to the Directed Reading Program at UCSB and for my mentor Fabio Ricci for his dedication and patience (beyond all reasonable limits).

The Gauss-Bonnet Theorem and its Applications

Emma Opper, advised by Malik Tuerkoen

University of California Santa Barbara

Introduction

The Gauss-Bonnet Theorem says that for a compact, orientable surface S, its total curvature depends only on the surface's Euler characteristic, a topological concept which relates a surface's faces, vertices, and edges. The theorem itself takes form as

$$\iint_S K d\sigma = 2\pi\chi(S).$$

In all, Gauss-Bonnet presents the surprising and extremely beautiful connection between the worlds of differential geometry and topology.

Curvature

Normal and Principal Curvatures

The **shape operator** of a surface at a point p with tangent vector v

$$S_p(v) = -DN_p(v)$$

measures how fast the normal vectors on a surface change. The second fundamental form is defined as

$$II_p(v) = \langle S_p(v), v\rangle.$$

If C is a curve on a surface S, the normal curvature relates the normal vector n of C to the normal vector N of S

$$k_n = k\cos(\theta)$$

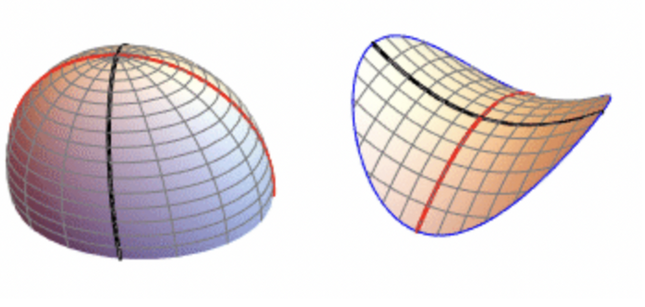
such that $\cos\theta = \langle n, N\rangle$. The maximum and minimum normal curvatures are the principal curvatures with principal directions e_1, e_2.

Gaussian Curvature

The product of the principal curvatures is the Gaussian curvature:

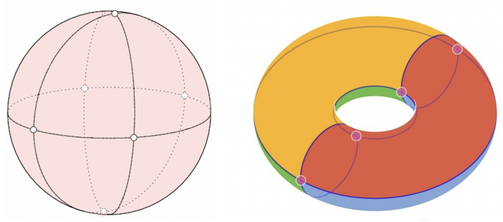
$$K = k_1 k_2$$

and is also the determinant of the shape operator with respect to the orthonormal basis. The sign of the Gaussian curvature can be used to chracterize the surface. At an elliptic point, shown on the left below, the Gaussian curvature is positive. At a hyperbolic point, on the right, the Gaussian curvature is negative.



Euler Characteristic

A **regular region** R is compact with boundary made up of a finite union of simple, closed, piecewise, regular curves that don't cross. Every regular region of a regular surface admits a **triangulation**, which is a finite family of triangles whose union covers R with empty intersection, i.e. is simply a way of splitting up a surface:



The Euler Characteristic of a surface, given by the formula

$$F - E + V = \chi(S),$$

is a number that describes a topological space's shape or structure regardless of the way it is bent or deformed, where F represents its faces, E its edges, and V its vertices.

For the sphere above, notice that $F = 8, E = 12$, and $V = 6$, giving $\chi(S) = 2$. For the torus, $\chi(S) = 0$.

There is actually a direct relation between the Euler characterstic of a surface and its holes, or genus g:

$$\chi(S) = 2 - 2g$$

The Global Theorem

Theorem and Outline of Proof

If R is a regular region of S, then

$$\sum_{i=1}^{n}\int_{c_i} k_g(s)ds + \iint_R K d\sigma + \sum_{l=1}^{P}\theta_l = 2\pi\chi(R).$$

where the geodesic curvature is the covariant derivative of a parametrization $\alpha(s)$ of C on S

$$k_g = [D\alpha'(s)/ds]$$

and θ_i are the external angles of the curves that make up the boundary of R.

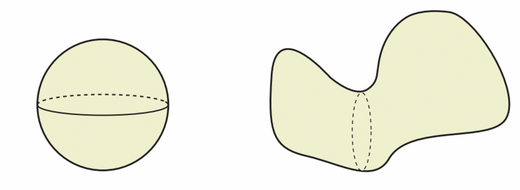
Proof. Since every regular region has a triangulation [1], we can use the local Gauss-Bonnet Theorem on each triangle:

$$\sum_{i=0}^{k}\int_{s_i}^{s_{i+1}} k_g(S)ds + \iint_R K d\sigma + \sum_{i=0}^{k}\theta_i = 2\pi$$

When summing over every triangle, the interior angles of these triangles can be related to the external and internal edges and vertices of the triangulation, producing $\chi(S)$. □

Compact and Orientable Surfaces

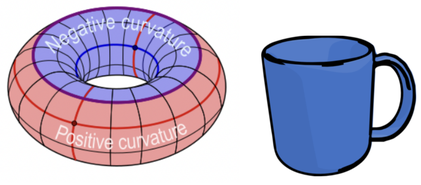
When a surface is compact, it has empty boundary, sending the terms of geodesic curvature and external angles in the global theorem to zero, meaning a compact surface's total curvature *only* depends on its Euler characteristic.



For example, though extreme positive and negative curvatures are evident after deforming the sphere above [2], the theorem tells us that the total curvature of these two surfaces are the same:

$$\iint_S K d\sigma = 2\pi\chi(S) = 2\pi(2) = 4\pi$$

Since a torus and a coffee mug have only one genus and thus are the same from a topological perspective, and because the Euler characteristic of the torus is zero, the total curvature of both objects must also be zero.



Acknowledgements

I would like to thank the UCSB Directed Reading Program for the opportunity to work on this project. None of this would have been possible without the substantial guidance of my mentor Malik Tuerkoen, whose continued help I am extremely grateful for.

References and More Information

[1] Manfredo P. Do Carmo. *Differential Geometry of Curves Surfaces*. Mineola, New York: Dover Publications, Inc., 2016.

[2] Theodore Shifrin. *A First Course in Curves and Surfaces*. Theodore Shifrin, Summer 2016.

Visit this website for more information. ↗

The Mathematics of the Standard Model

Logan Joseph, Keryn Jung, and Billy Wang
University of California, Santa Barbara

Vector and Principle Bundles

A principle bundle, denoted

$$G \longrightarrow P \xrightarrow{\pi} M$$

is a fibre bundle with a Lie group G as its fibre and a smooth right action $P \times G \to P$. The action is transitive and preserves the fibres of P. For every principle bundle, there exists a principle bundle atlas with bundle charts, $\phi_i : P_{U_i} \to U_i \times G$. These charts satisfy

$$\phi_i(p \cdot g) = \phi_i(p) \cdot g \quad \forall p \in P_{U_i}, g \in G$$

where G acts on $(x, a) \in U_i \times G$ by

$$(x, a) \cdot g = (x, ag)$$

Physically, this is "attaching" a Lie group to every point in spacetime. Given a representation (ρ, V) of the structure group G, we can build a vector bundle out of a principle bundle P which takes the form

$$E = P \times_\rho V = (P \times V)/G$$

A connection on a principle bundle is a $\mathfrak{g}$-valued 1-form while curvature is a $\mathfrak{g}$-valued 2 form given by

$$A \in \Omega^1(P, \mathfrak{g}) \qquad F = dA + \frac{1}{2}[A, A] \in \Omega^2(P, \mathfrak{g})$$

respectively. Both are required to be ad-invariant. A must act as the identity on fundamental vector fields associated to the action of G on P. Connection 1-forms correspond to gauge fields in the SM. Curvature 2-forms physically represent the field strength, an example being the EM field strength tensor: $F_{\mu\nu}$. We can realize these sections as physical fields by specifying a local section (choice of gauge) and pulling back the forms under the section. We will also need a specific vector bundle called the spinor bundle. These are defined using spin structures, denoted $\text{Spin}^+(M)$, which are a special class of principle bundle. The spinor bundle is then defined by

$$S = \text{Spin}^+(M) \times_\kappa \Delta$$

Here (κ, Δ) is the appropriate spinor representation. By definition, sections of these spinor bundles are spinor fields, which describe fermions. In 4 dimensions, the spinor bundle also decomposes into chiral spinor bundles. The Dirac operator D acts on spinors via

$$D = \Sigma \text{Cl}_{e_a} \circ \nabla_{e_a}$$

Mass Generation and Charged Matter

The SM Lagrangian is expressed as

$$\mathscr{L}_{SM} = \mathscr{L}_{YM} + \mathscr{L}_D + \mathscr{L}_H + \mathscr{L}_Y$$

where $\mathscr{L}_{YM}$ are Yang-Mills terms, $\mathscr{L}_D$ are Dirac terms, $\mathscr{L}_H$ are Higgs terms, and $\mathscr{L}_Y$ are Yukawa terms.
The Yang-Mills Lagrangian for a principal bundle $P \to M$ is given by

$$\mathscr{L}_{YM} = \frac{1}{2}\langle F_M^A, F_M^A \rangle_{Ad(P)}$$

Here, the twisted two-form $F_M^A \in \Omega^1(M, Ad(P))$ is induced from the curvature of the principle bundle connection. The Higgs action for a scalar field Φ coupled to a gauge field A is defined by

$$\mathscr{L}_H[\Phi] = \langle d_A\Phi, d_A\Phi \rangle_E - V(\langle \Phi, \Phi \rangle_E)$$

Here, the scalar fields are sections of an associated vector bundle of the Yang-Mills principle bundle defined using an appropriate representation. Likewise, the Dirac Lagrangian for a twisted spinor field Ψ of mass m coupled to gauge field A on principal bundle P is defined by

$$\mathscr{L}_D[\Psi] = Re\langle \Psi, D_A\Psi \rangle_{S\otimes E} - m\langle \Psi, \Psi \rangle_{S\otimes E}$$

Here, the spinors are sections of $S \otimes E$, where E is also an associated vector bundle of the Yang-Mills principle bundle.
Spontaneous symmetry breaking occurs when the vacuum (lowest energy field configuration) has a proper subgroup of symmetries of the Lagrangian. In such a context, expanding the Yang-Mills-Higgs action about the vacuum and keeping low-order terms implies that gauge Bosons may "acquire" mass. This is called the Higgs Mechanism. This mechanism provides a way to give mass to gauge Bosons by introducing massive scalar fields, which are called Higgs fields. These are the scalars which appear in the SM.
We also need to be able to give masses to twisted chiral fermions. This is done using scalar fields and quantities called Yukawa forms, which are maps from unitary representations V_L, W, and V_R of G.

$$\tau : V_L \times W \times V_R \longrightarrow \mathbb{C}$$

This form is invariant under the action of G, complex antilinear in V_L, real linear in W, and complex linear in V_R. Given a Yukawa form τ, the Yukawa coupling g_Y defines a gauge invariant Lagrangian

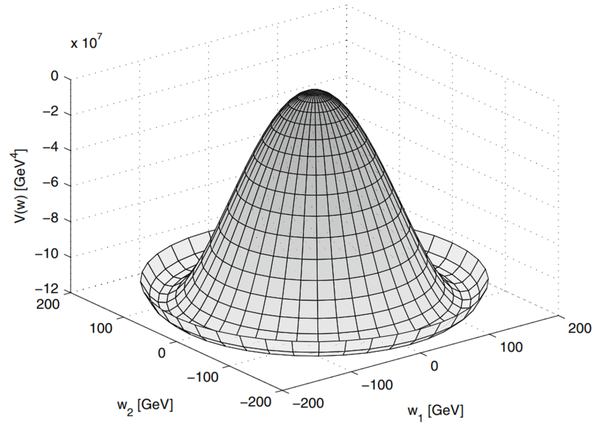
$$\mathscr{L}_Y[\Psi_L, \Phi, \Psi_R] = -g_Y(\bar{\Psi}_L\Phi\Psi_R) - g_Y(\bar{\Psi}_L\Phi\Psi_R)^*$$

Standard Model Particle Content

The gauge group for the SM is $SU(3) \times SU(2) \times U(1)$. The gauge Boson content follows from the adjoint representation of the Lie algebra $\mathfrak{g}$ of this group, which splits into three subrepresentations:

$$\mathfrak{g} = \mathfrak{su}(3)_C \oplus \mathfrak{su}(2)_L \oplus \mathfrak{u}(1)_Y$$

These are the gluon sector, the weak sector, and the hypercharge sector, respectively. The 8 generators of $SU(3)$ corresponds to 8 gluons, the 3+1 generators of $SU(2)$ and $U(1)$ corresponds to the $W^\pm/Z$ Bosons and the photon γ.
However, without the Higgs field, the gauge Bosons would all be massless. The SM then features multiplets of charged scalar fields called Higgs fields to address this issue. The specific choice of Higgs potential used in the action is depicted below!
The shifted Higgs Potential (due to spontaneous symmetry breaking) leads to Higgs Mechanism. Some of the gauge Bosons, including Z and $W^\pm$, acquire mass through this mechanism.
Building the twisted chiral spinor bundle to model spinors requires a choice of representation of the SM gauge group, which is ultimately determined by experiment. The correct choice of representation results in 3 generations of so-called quarks and leptons. Of these, there are 6 charged quarks, 3 charged leptons, and 3 neutral leptons (neutrinos). The Yukawa couplings present in the SM then feature terms for both these quarks and leptons.



Acknowledgements

Reference Material: Mathematical Gauge Theory by Mark J.D. Hamilton
We would like to thank our mentor Aaron Kennon and Professor David Morrison for their assistance with this reading course.

The Topology of the Chair Tiling Space

Miles Gould, Niels Wang

University of California Santa Barbara

Abstract

Tilings are characterized by covering Euclidean space with a set of tiles, traditionally polygons, in which two adjacent tiles meet full edge to edge. Given a tiling T, we can generate a complete metric space, and we can study such a space using the tools of cohomology. We will discuss a famous example, the chair tiling.

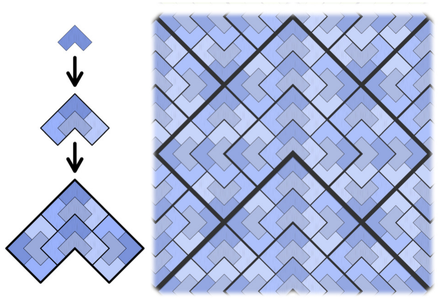


Fig. 1

This tiling is an example of a substitution tiling, as one can cover the entire plane by iterating the substitution shown on the left.

The Tiling Metric

Consider two tilings T, T' of $\mathbb{R}^m$. We can assess their "closeness" using the tiling metric. Heuristically, two tilings are close if, after a small translation, their patterns overlap on a large ball around the origin. As seen in figure 2, because the blue patch centered at x is the same as the blue patch centered at y, the tilings centered at these points are $\frac{1}{2}$-close but not $\frac{1}{3}$-close because the corresponding red patches differ on their lower left corners. Equipped with this metric, we can construct a complete metric space Ω_T around T, called the hull of T, by taking the orbit of T under translation and completing this space by taking limits of Cauchy sequences with respect to the tiling metric.

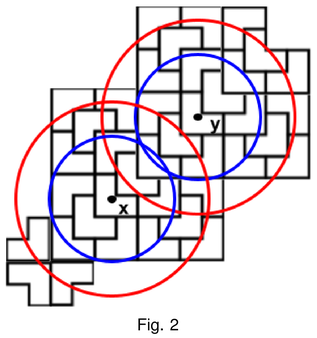


Fig. 2

$$\Omega_T \coloneqq \text{closure}\left\{T - x \middle| x \in \mathbb{R}^m\right\}$$

Acknowledgments and References

We thank Kyle Hansen for his time and guidance as well as the UCSB Directed Reading Program for the opportunity to study this subject.

[1] *Chair tiling - Wikipedia — en.wikipedia.org.* https://en.wikipedia.org/wiki/Chair_tiling. [Accessed 20-May-2023].

[2] Jeong-Yup Lee and Robert Moody. “Lattice Substitution Systems and Model Sets”. In: *Discrete & Computational Geometry* 25 (Mar. 2000). DOI: 10.1007/s004540010083.

[3] Lorenzo Adlai Sadun. *Topology of tiling spaces.* Vol. 46. American Mathematical Soc., 2008.

A Tale of Two Cohomologies

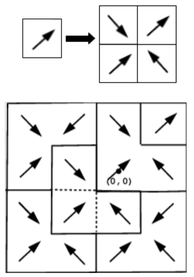


Fig. 3

To simplify our study, we will consider another substitution called the arrow substitution. As it turns out, we can group these arrow tiles in threes to achieve the chair tiling, so these tilings are equivalent. We will use C to denote the arrow tiling and Ω_C to denote its hull. In contrast to the complete construction of the space Ω_C, we can construct the hull as an inverse limit, as follows: Let Γ_0 be the set of instructions for laying down a tile over the origin. Inductively, let Γ_{n+1} be the set of instructions for laying down a border of tiles around those patches in Γ_n. Linking these are the morphisms, $f_n : \Gamma_{n+1} \to \Gamma_n$, a sequence of forgetful functions which forgets the outermost layer of tiles. In the case of C we can think of these functions as forgetting the last iteration of the substitution, and is in some sense dual to the substitution. For example, in figure 3, the 4×4 square is sent to the 2×2 square. To summarize, we can represent the hull as the inverse limit:

$$\Omega_C \cong \varprojlim(\Gamma_n, f_n) = \left\{(x_0, x_1, ...) \in \prod_{n\in\mathbb{N}} \Gamma_n \middle| x_n = f_n(x_{n+1})\right\},$$

which can be thought of as the set of sequences in which each component has "forgotten" the structure of the component after it. To study the structure of this space, we can use two types of cohomologies: the Cech cohomology and the PE variant of the De Rham cohomology. A special form of De Rham's theorem allows us to relate these two cohomologies:

$$H^*_{\text{PE}}(T) \cong \frac{\text{Closed PE forms on T}}{\text{d(PE forms on T)}} \cong \frac{\varinjlim \text{Closed forms on } \Gamma_n}{\varinjlim \text{Exact forms on } \Gamma_n} \cong \varinjlim H^*_{\text{deRham}}(\Gamma_n) \cong \varinjlim \check{H}^*(\Gamma_n, \mathbb{R}) \cong \check{H}^*(\Omega_T, \mathbb{R})$$

Čech Cohomology

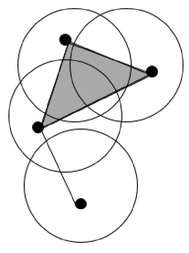


Fig. 4

The nerve $\mathcal{N}(\mathcal{U})$ of a cover $\mathcal{U} = \{U_\alpha\}$ of a general space X is a simplicial complex consisting of an n-simplex for each nonempty intersection of n elements of $\mathcal{U}$. In general, $\check{H}^*(X) \coloneqq \varinjlim H^*_\Delta(\mathcal{N}(\mathcal{U}))$ is the direct limit of the simplicial cohomology of $\mathcal{N}(\mathcal{U})$ over all open covers $\mathcal{U}$ of X. As we can view $\Omega_C \cong \varprojlim(\Gamma_n, f_n)$, and each Γ_n is a CW complex, there is a nice property:

$$\check{H}^*(\Omega_C) \cong \varinjlim \check{H}^*(\Gamma_n)$$

This works because for any open cover of a CW complex, we can create a "good" refinement such that finite intersections of open sets are contractible. Sketch of proof: Inductively take a "good" cover U_n of each Γ_n as a refinement of $f_{n-1}^{-1}(U_{n-1})$. Let $\pi_n : \Omega_C \to \Gamma_n$ be the projection. Then each U_n induces a cover $V_n = \pi_n^{-1}(U_n)$ of Ω_C with $\mathcal{N}(V_n) = \mathcal{N}(U_n)$. We can pick U_n to make the sequence V_n cofinal on covers of Ω_C. So $\check{H}^*(\Omega_C) \cong \varinjlim \check{H}^*(V_n) \cong \varinjlim \check{H}^*(U_n) \cong \varinjlim \check{H}^*(\Gamma_n)$.

Calculating $\check{H}^1(\Omega_C)$

Down to Earth, view Ω_C as $\varprojlim(\Gamma_n, f_n)$ where each f_n is the map induced by substitution. As each Γ_n is a finite Δ-complex, $\check{H}^*(\Omega_C) \cong \varinjlim H^*_\Delta(\Gamma_n)$. We will compute $\check{H}^1(\Omega_C)$ using the chain complex of Γ_n and the homomorphism from $H^1_\Delta(\Gamma_n)$ to $H^1_\Delta(\Gamma_{n+1})$ induced by the substitution. Up to rotation r by $\frac{\pi}{2}$ and the ways we identify tiles, there is one tile A, one edge α, and two vertices a, b in each Γ_n (see Fig.5), with the relations: $A = r^4 A$, $\alpha = r^4\alpha$, and $a = ra, b = rb$. The boundary map $\partial_i : C_i \to C_{i-1}$ is $\partial_2 A = (1 - r + r^2 - r^3)\alpha$, and $\partial_1 \alpha = b - a$. Now we compute $\varinjlim H^1_\Delta(\Gamma_n)$ by decomposing it into 3 irreducible representations of $\mathbb{Z}_4$, as the group of rotations: The cases $r = 1$ and $r = -1$ contribute no cohomology in degree 1. The case $r^2 = -1$ is 2-dimensional, so we have $C^0 = 0$ and $C^1 = C^2 = \mathbb{Z} \oplus \mathbb{Z}$. We have $\text{Im}(\partial_1^T) = 0$, and $\ker(\partial_2^T) = C^1$ because $\partial_2(A) = (1 - r - 1 + r)A = 0$. Hence for each Γ_n we have $H^1_\Delta(\Gamma_n) \cong \frac{ker(\partial_2^T)}{Im(\partial_1^T)} \cong \mathbb{Z} \oplus \mathbb{Z}$. Moreover, substitution induces multiplication by 2 in the direct limit, since $\sigma(\alpha) = (1 - r^2)\alpha = 2\alpha$. Therefore, as direct limits commute with direct sums, we have

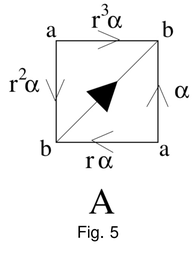


Fig. 5

$$\check{H}^1(\Omega_C) \cong \varinjlim H^1_\Delta(\Gamma_n) \cong \mathbb{Z}[1/2] \oplus \mathbb{Z}[1/2].$$

Pattern Equivariant De Rham Cohomology

To understand the PE variant of the De Rham cohomology, we must understand its namesake, pattern equivariant functions. Let T be a tiling on $\mathbb{R}^m$. A smooth function $f : \mathbb{R}^m \to \mathbb{R}$ is called pattern equivariant on T if, for some radius R, whenever two balls $B_R(x), B_R(y)$ of radius R share the same pattern in T, $f(x) = f(y)$. In figure 2, we can see that a PE function f on a chair tiling with blue radius must have $f(x) = f(y)$, however a function PE on the red radius need not have $f(x) = f(y)$, because the red balls differ on the bottom left corner. In the arrow tiling case, these will be functions of the form γ (0-forms), $\alpha_1 dx_1 + \alpha_2 dx_2$ (1-forms), and $\zeta dx_1 dx_2$ (2-forms), where $\alpha_1, \alpha_2, \gamma, \zeta$ are pattern equivariant on the arrow tiling. We will denote by $\Lambda_k(T)$ the set of k-forms pattern equivariant on a tiling T. The pattern equivariant cohomology is constructed as the quotient

$$H^k_{\text{PE}}(T) \cong \frac{\{\alpha \in \Lambda_k(T) | d\alpha = 0\}}{\{\alpha \in \Lambda_k(T) | \alpha = d\gamma, \gamma \in \Lambda_{k-1}(T)\}}$$

We will make sense of this in the section below.

Understanding $\check{H}^1(\Omega_C)$ Using $H^1_{\text{PE}}(C)$

PE cohomology helps us interpret the result calculated for $\check{H}^1(\Omega_C)$. To do this, we will find pattern equivariant 1-forms on the chair tiling C which represent the cohomology classes we computed previously. These will be functions on the horizontal and vertical edges of figure 3. By symmetry, we can focus on the horizontal components. The quotient from the cohomology induces the equivalence relation $\sim$ on the set of PE functions on the horizontal edges of an order n supertile such that $\alpha_n \sim \beta_n$ iff there exists some 0-form γ_n (a PE function on vertices) such that $d\gamma_n = \alpha_n - \beta_n$.

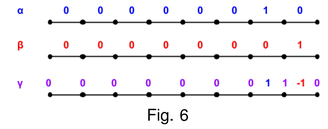


Fig. 6

The representatives ν_n of each of these classes will form the basis for the horizontal component of $H^1_{\text{PE}}(C)$, and they will satisfy $\nu_n \sim 2\nu_{n+1}$. The order 0 supertile is simply one tile, so ν_0 is a natural representative for 1 in the horizontal component of $H^1_{PE}(C)$ and thus of $\check{H}^1(\Omega_C)$. Inductively, $2^n \nu_n \sim \nu_0$, and so ν_n is a natural representative for $\frac{1}{2^n}$ in the horizontal component of $H^1_{PE}(C)$. These representatives of $(1, 1/2, ..., 1/2^n, ..)$ form a basis for the horizontal component of $\check{H}^1(\Omega_C)$ and by the above symmetry, the vertical component will be the same. So we now have a greater understanding of the original result

$$\check{H}^1(\Omega_C) \cong \mathbb{Z}[1/2] \oplus \mathbb{Z}[1/2].$$

The Zeta Function and Prime Number Theorem

Jianing Zhang | Directed by Christian Hong
University of California - Santa Barbara

Abstract

After Euler found that $\Sigma_p \frac{1}{p}$ diverges, Legendre and Gauss found that approximately,

$$\pi(x) \sim \frac{x}{\log x},$$

where $\pi(x)$ = number of primes less than or equal to x.

Background

The zeta function is related to some of the facts of numbers, like $\Sigma_p 1/p$ diverges, **Dirichlet Theorem** (for any a, q that are coprime, there exists infinitely many primes with the form $p = an + q$), and so on. [1]

Some theorems and definitions that are related to the proof of prime number theorem are:

The initial definition for zeta function when $s > 1$ is real is

$$\zeta(s) = \Sigma_{n=1}^{\infty} \frac{1}{n^s},$$

by using analytic continuation, we could define $\zeta(s)$ on $\mathbb{C}$.

An alternative form for zeta function is

$$\zeta(s) = \Pi_p (1 - p^{-s})^{-1}$$

Theorem. The function ζ is holomorphic for $Re(s) > 1$ and has an analytic continuation to all of $\mathbb{C}$ as a meromorphic function with simple poles at $s = 0$ and $s = 1$. Moreover,

$$\zeta(s) = \zeta(1 - s)$$

Theorem. The zeta function has no zeros on the line $Re(s) = 1$.
After Riemann introduced the analytic continuation of ζ, he begun to believe: The zeros of $\zeta(s)$ in the critical strip lie on the line $Re(s) = 1/2$, which is called **Riemann Hypothesis**.

During the studies, Tchebychev found another approximation for $\pi(x)$ that is easier to compute.
Definition. Tchebychev's ψ-function is defined by

$$\psi(x) = \Sigma_{p \le x} \log p$$

Definition. Define function ψ_1 by

$$\psi_1(x) = \int_1^x \psi(x) du.$$

Proposition. For all $c > 1$,

$$\psi_1(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{x^{s+1}}{s(s+1)} (-\frac{\zeta'(s)}{\zeta(s)}) ds.$$

Relation between ψ,ψ_1, and π

Theorem. If $\psi(x) \sim x$ as $x \to \infty$, then $\pi(x) \sim \frac{x}{\log x}$ as $x \to \infty$.
Proof. It suffices to show:

$$1 \le \liminf_{x\to\infty} \pi(x)\frac{\log x}{x} \quad \text{and} \quad \limsup_{x\to\infty} \pi(x)\frac{\log x}{x} \le 1. \tag{1}$$

Using the trick that:

$$\psi(x) = \Sigma_{p\le x} [\frac{\log x}{\log p}] \log p \le \Sigma_{p \le x} \frac{\log x}{\log p} \log p = \pi(x) \log x.$$

the first inequality in (1) holds.
Fix $0 < \epsilon < 1$, note that

$$\psi(x) \ge \Sigma_{p \le x} \ge \Sigma_{x^\epsilon < p \le x} \ge (\pi(x) - \pi(x^\epsilon)) \log x^\epsilon,$$

thus

$$\psi(x) + \epsilon \pi(x^\epsilon) \log x \ge \epsilon \pi(x) \log x.$$

Dividing both side by x, and noting that $\pi(x^\epsilon) \le x^\epsilon, \epsilon < 1, \psi(x) \sim x$, we have

$$\frac{\epsilon \pi(x^\epsilon) \log x}{x} \le \frac{\epsilon \log x}{x^{1-\epsilon}} \to 0 \quad \text{as x} \to \infty,$$

thus

$$1 \ge \epsilon \limsup_{x\to\infty} \frac{\pi(x) \log x}{x}.$$

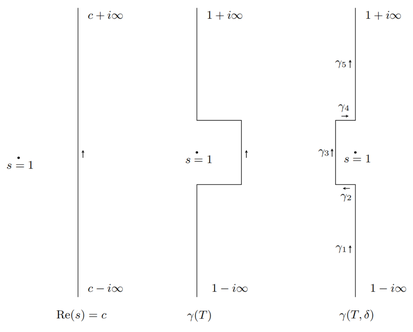
Since ϵ is arbitrary, the proof is complete.
Proposition. If $\psi_1(x) \sim \frac{x^2}{2}$, then $\psi(x) \sim x$ as $x \to \infty$.
Sketch. Let $\alpha < 1 < \beta$, the proof follows from the inequality:

$$\frac{1}{(1-\alpha)x} \int_{\alpha x}^{x} \psi(u) du \le \psi(x) \le \frac{1}{(\beta - 1)x} \int_x^{\beta x} \psi(u) du,$$

since ψ is increasing. we get $\limsup_{x\to\infty} \frac{\psi(x)}{x} \le 1$. By a similar argument, $\liminf_{x\to\infty} \frac{\psi(x)}{x} \ge 1$.

The Asymptotics for ψ_1

Proposition. $\psi_1(x) \sim \frac{x^2}{2}$ as $x \to \infty$
Proof.



[2]: Page 195

Proof Continued

Let $F(x) = \frac{x^{s+1}}{s(s+1)}(-\frac{\zeta'(s)}{\zeta(s)})$. Let $\gamma(T) = \gamma(T, \delta)$. We have

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} F(s) ds = \frac{1}{2\pi i} \int_{\gamma(T)} F(s) ds.$$

Next, we pass from $\gamma(T)$ to $\gamma(T, \delta)$. For fixed T, we choose $\delta > 0$ small enough such that ζ has no zeros in the box

$$\{s = \sigma + it, 1 - \delta \le \sigma \le 1, |t| \le T\}.$$

Thus ζ does not vanish on the line $\sigma = 1$. Note that $F(s)$ has a simple pole at $s = 1$ with residue is $\frac{x^2}{2}$. Thus

$$\frac{1}{2\pi i} \int_{\gamma(T)} F(s) ds = \frac{x^2}{2} + \frac{1}{2\pi i} \int_{\gamma(T,\delta)} \frac{x^{s+1}}{s(s+1)} F(s) ds.$$

Decompose the contour $\gamma(T, \delta)$ as $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5$. Note that for T large,

$$|\int_{\gamma_1} F(s) ds| \le \frac{\epsilon}{2} x^2 \quad \text{and} \quad |\int_{\gamma_5} F(s) ds| \le \frac{\epsilon}{2} x^2.$$

Fix T, let δ be small. On γ_3, there exists a constant C_T such that

$$|\int_{\gamma_3} F(s) ds| \le C_T x^{2-\delta}.$$

Finally, on γ_2(similarly on γ_4), estimates the integrals as:

$$|\int_{\gamma_2} F(s) ds| \le C'_T \int_{1-\delta}^{1} x^{1+\delta} d\sigma \le C'_T \frac{x^2}{\log x}.$$

Thus there exists C_T, C'_T (possibly different from the ones above) such that

$$|\psi_1(x) - \frac{x^2}{2}| \le \epsilon x^2 + C_T x^{2-\delta} + C'_T \frac{x^2}{\log x}.$$

Thus, dividing $\frac{x^2}{2}$, and for x large,

$$|\frac{2\psi_1(x)}{x^2} - 1| \le 2\epsilon + 2C_T x^{-\delta} + 2C'_T \frac{1}{\log x} \le 4\epsilon,$$

which leads that

$$\psi_1(x) \sim x^2/2 \quad \text{as } x \to \infty.$$

Thus, we finished sketching the proof of the prime number theorem.

References

[1]: "Pricenton Lecture Notes, Volume 1, Fourier Analysis An Introduction" by Elias M. Stein and Rami Shakarchi.

[2]: "Pricenton Lecture Notes, Volume 2, Complex Analysis" by Elias M. Stein and Rami Shakarchi.

Unraveling the Enigma of Topological Structures

Genevieve Yao and Taylor Mayton, Mentored by Katherine Merkl

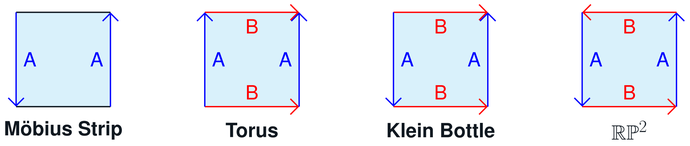
University of California Santa Barbara

Introduction

Topology provides the framework needed to describe continuous deformations and to characterize the differences between spaces that appear the same locally, but have a fundamentally different global structure. A topology on a space is a structure that defines how elements of a set are spatially related. In this poster, we explore the role of quotient topologies and how to understand them.

Quotient Topology

It turns out that we can construct new spaces from old ones via some sort of a "gluing" operation. For example, consider taking a rectangle and "gluing" two opposing edges together to form a Cylinder, or adding a half-twist to get a Möbius Strip. This act of "gluing" (identification) is formally defined using the language of quotient topologies.



Möbius Strip **Torus** **Klein Bottle** $\mathbb{RP}^2$

Definition 1. Suppose that $(X, \mathcal{T})$ is a topological space and that $\sim$ is an equivalence relation on X. Denote the set of equivalence classes by $X/\sim$, and let $p : X \to X/\sim$ be the function that assigns to each point of X the equivalence class it is in. Let $\tilde{\mathcal{T}}$ be the family of all subsets $\tilde{U} \subseteq X/\sim$ such that $p^{-1}(\tilde{U}) \in \mathcal{T}$. Then $\tilde{\mathcal{T}}$ is a topology for $X/\sim$, called the **quotient topology**, $(X/\sim, \tilde{\mathcal{T}})$ is called a **quotient space** (of X), and $p : X \to X/\sim$ is called the **quotient map**. [2]

Above are four examples of quotient spaces, each given by defining different equivalence relations on the square. Geometrically, the quotient map p sends points on the square to their equivalence class by aligning arrows of the same color. For the Torus and Klein Bottle, the image of p is shown below.

Figure 1. Torus(*left*) and Klein Bottle(*right*)

The Real Projective Plane, $\mathbb{RP}^2$, can be difficult to visualize, so we outline equivalent forms of the space to help build intuition.

Proposition 2. Let S^2 denote the unit 2-Sphere in $\mathbb{R}^3$ and let D be the unit disc in $\mathbb{R}^2$. The following quotient spaces are all homeomorphic:

(a) $(\mathbb{R}^3 \backslash \{0\})/\sim$ where $x \sim y$ iff $y = \lambda x$ for some non-zero $\lambda \in \mathbb{R}$;

(b) $S^2/\sim$ where $\sim$ identifies each pair of antipodal points of S^2;

(c) $D/\sim$ where $\sim$ identifies each pair of antipodal points on the boundary of D;

(d) The square with given identifications seen above.

Proof. We provide the intuition for each homeomorphism. (a) describes the space of lines in $\mathbb{R}^3$ passing through the origin. Each such line intersects S^2 exactly twice, at antipodal points. For (b), each point in the upper hemisphere of S^2 has its antipodal point in the lower hemisphere, so we can discard the lower hemisphere and then project our points down to the unit disk in $\mathbb{R}^2$. For (c) and (d), center D and the square at the same point and project the square radially onto D. (Detailed proof, p.162 [2]) □

Path Homotopy and Fundamental Groups

Definition 3. Let $f, g : X \to Y$ be maps. Then f is **homotopic** to g if there exists a map $F : X \times I \to Y$ such that $F(x, 0) = f(x)$ and $F(x, 1) = g(x)$ for all points $x \in X$, where I is the unit interval. The map F is called a **homotopy** from f to g. [1]

Intuitively, the homotopy F is a continuous deformation of f into g.

Lemma 4. The relation of 'homotopy relative to a subset A of X' is an equivalence relation on the set of all maps from X to Y which agree with some given map on A. (For proof see p.90 [1])

Since it is an equivalence relation, naturally, we want to discuss its equivalence class:

Theorem 5. *Let $\langle\alpha\rangle$ denote the homotopy class of a loop α (Where homotopy classes are equivalence classes from the relation of homotopy relative to I). The set of homotopy classes of loops in X based at p forms a group under the multiplication*

$$\langle\alpha\rangle \cdot \langle\beta\rangle = \langle\alpha \cdot \beta\rangle$$

(For proof see p.92 [1])

Intuitively, if there's a hole inside a loop, then it is not in the same equivalence class with the point, if two loops are around different holes, they're not in the same class.

Definition 6. The group constructed in Theorem 5 is called the **fundamental group** of X (based at point p) and denoted $\pi_1(X, p)$. [1]

It turns out that for any path-connected space X, the fundamental group is independent of base point, and so we simply write the fundamental group of X as $\pi_1(X)$ (instead of $\pi_1(X, p)$.) Furthermore, it happens to be a topological invariant which makes it very important. Intuitively, the fundamental group of a space describes all the different kinds of loops you can have on a space which can't be continuously deformed into each other. In figure 2 we see an example of the fundamental group of S^1. We can see how the fundamental group here comes from winding number with orientation.

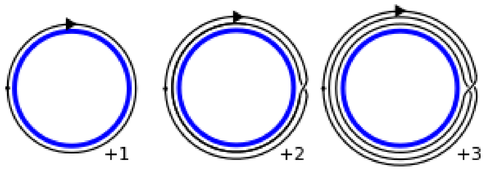


Figure 2. $\pi_1(S^1) = \mathbb{Z}$

It is not difficult to see that $\pi_1(S^2)$ is trivial. We also know $\pi_1(\mathbb{RP}^2) = \mathbb{Z}/2\mathbb{Z}$. This makes sense if you think about $\mathbb{RP}^2$ as in (b) from Proposition 2. We can go around the sphere and end up at either our base point p or $-p$. These correspond to the 2 different homotopy classes.

Manifolds and Embedding in Euclidean Space

Since in topology the shape and distance of objects (usually) don't preserve, so the thing called manifolds are more usual. But since manifolds vary a lot, it's hard for us to directly study on then. Therefore, by embedding the manifolds to $\mathbb{R}^n$, we can use tools in $\mathbb{R}^n$ on manifolds, which is very helpful (and so important) in our study to manifolds.

Definition 7. A **Manifold** of dimension n (n-manifold for short) is a second-countable Hausdorff space, each point of which has a neighbourhood homeomorphic to $\mathbb{R}^n$ (n dimensional Euclidean space). [1]

One can accept it roughly as a locally Euclidean smooth space.

Definition 8. If $f : X \to Y$ is a one-one map, and if $f : X \to f(X)$ is a homeomorphism when we given $f(X)$ the induced topology from Y, we call f an **embedding** of X in Y. [1]

Embedding the $\mathbb{RP}^2$ and Klein Bottle in $\mathbb{R}^4$ [2]

Embedding a manifold M to $\mathbb{R}^n$ means the same as "a subset of $\mathbb{R}^n$ is homeomorphic to M". So one way to prove embedding is to explicitly construct a smooth function in $\mathbb{R}^n$ and prove that it has the same identification as M, and we'll do this to two examples in the next section.

Embedding Real Projective Plane in $\mathbb{R}^4$

Proposition 9. There's a homeomorphism from real projective plane $\mathbb{RP}^2$ to a subspace of $\mathbb{R}^4$.

Proof. Let $S^2 = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 + z^2 = 1\}$, and construct the function $f : S^2 \to \mathbb{R}^4$ as

$$f(x, y, z) = (x^2 - y^2, xy, yz, zx)$$

detailed proof in omitted here, but one can check that f is an identification map results in the same identification space to $\mathbb{RP}^2$ we discussed on the left! (refer to the detail proof in page 164 of [2]) □

Embedding Klein Bottle in $\mathbb{R}^4$

We've seen the picture of Klein Bottle in the part of quotient topology, and there's some fancy models of Klein bottles. However, all of these are just Klein Bottle "immersed" in $\mathbb{R}^3$, which means it is roughly embedded but has self-intersections.[2] In fact, it cannot be embedded in $\mathbb{R}^3$:

Proposition 10. Klein Bottle can't be embedded in $\mathbb{R}^3$.

Proof. We prove this not rigorously here by argue about the orientability of Klein bottle. There is a theorem states that all smooth hypersurface(it means a $n-1$-manifold in a $\mathbb{R}^n$, one can perceive it as a similar thing to a surface as in $\mathbb{R}^3$, but in a higher dimension perspective) with no boundary in $\mathbb{R}^n$ is orientable, and Klein bottle, however, as a non-orientable 2-manifolds, which means it cannot be embedded in $\mathbb{R}^3$. □

And on the other hand, it can be embedded in $\mathbb{R}^4$:

Proposition 11. The Klein Bottle can be embedded in $\mathbb{R}^4$.

Proof. We write $X = [0, 2\pi] \times [0, \pi]$, then let $f : X \to \mathbb{R}^4$ defined as

$$(x, y) \mapsto ((2 + \cos x)\cos y, (1 + \cos x)\sin y, \sin x \cos y, \sin x \sin y)$$

Detailed proof is also omitted here,but one can check that f is an identification map results in the same identification space to Klein Bottle as we discussed on the left. (Refer to the detail proof in page 165 of [2]). □

Now, one might want to ask: wait, but more generally, what would happen for other 2-manifolds? Can they also be embedded in $\mathbb{R}^4$? The answers is: yes!

Theorem 12. *All 2-manifolds can be embedded in $\mathbb{R}^4$.*

And if keep asking, this road will leads us to the final theorem:

Theorem 13. ***(Strong Whitney Embedding Theorem)*** *Any smooth real m-manifolds can be smoothly embedded in $\mathbb{R}^{2m}$.*

Acknowledgements

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

References

[1] Mark Anthony Armstrong. *Basic Topology*. Springer Science & Business Media, 2013.

[2] Wilson A Sutherland. *Introduction to Metric and Topological Spaces*. Oxford University Press, 2009.

Visualizations in Number Theory

Jingyun Zhang and Ziyue Gao
Mentored by Sarah Mantell
University of California, Santa Barbara

Gaussian and Eisenstein Integers

The **Gaussian integers**, denoted $\mathbb{Z}[i]$, is the collection of complex numbers $x+yi$ where x and y are integers. For example, $1+2i$ and $5i$ are Gaussian integers, but $2+\pi i$ is not. The **Gaussian units** are ± 1 and $\pm i$.

The **Eisenstein integers** is the collection of complex numbers $x+y\omega$ where x and y are integers, and $\omega = e^{2\pi i/3}$. For example, $-1+\omega$ and $\omega^2 = -1-\omega$ are Eisenstein integers, but $1/2+\omega$ is not. The **Eisenstein units** are $\pm 1, \pm\omega$, and $\pm\omega^2$.

We abbreviate the collection of Gaussian/Eisenstein integers as **G/E integers**. A non-unit G/E integer q is **prime** if the only divisors of q in the G/E integers are units.

For example, 2 is an Eisenstein prime, but not a Gaussian prime since $2 = (1+i)(1-i)$.

Categorizing G/E Primes

The **primagon** of a Gaussian integer z is the 4-gon with vertices $\pm z, \pm iz$. Likewise, the primagon of an Eisenstein integer z is the 6-gon with vertices $\pm z, \pm\omega z, \pm\omega^2 z$. All vertices of a primagon have the same absolute value.

Using the primagon, one can classify the G/E integers into three types:

- **Type I:** The primagon has integer vertices and coincides with its complex conjugate.
- **Type R:** The primagon coincides with its complex conjugate, but does not have integer vertices.
- **Type S:** The primagon does not coincide with its complex conjugate.

Automatically Classifying G/E integers

Given any G/E integer, we automatically classify the integer and plot the primagon as follows:

1. **Input G/E integer** Specify whether the input is a Gaussian or Eisenstein integer by typing 'G' or 'E'. The user is then prompted to input integers a and b to form the Gaussian integer $a+bi$ or the Eisenstein integer $a+b\omega$.

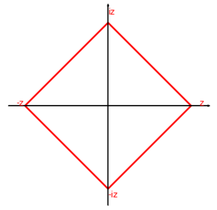
2. **Find the primagon** Let $z = a+bi$ or $z = a+b\omega$ be the inputted integer. Find three (five) new vertices by rotating z by $\frac{\pi}{2}$ ($\frac{\pi}{3}$) about the origin. Connect the four (six) vertices to obtain the primagon of z. Compute the primagon's conjugate (blue) by conjugating each vertex individually and connecting the new vertices.

3. **Classify** If the primagon (plotted in red) doesn't coincide with its conjugate (plotted in blue), then z is type S. Otherwise, determine whether the primagon has integer vertices by computing $|z|^2 = z\cdot\overline{z}$. If $z\cdot\overline{z}$ is a perfect square, then z is type I. If not, then z is type R.

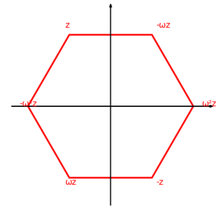
Visualizing Gaussian and Eisenstein integers

Some examples of G/E primes of type I/R/S:

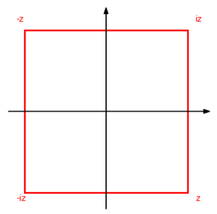
z=3 is an Gaussian prime integer of type I



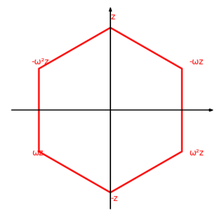
z=0+2ω is an Eisenstein prime integer of type I



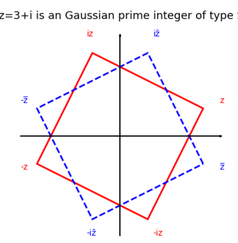
z=1-i is an Gaussian prime integer of type R



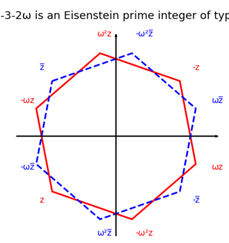
z=1+2ω is an Eisenstein prime integer of type R



z=3+i is an Gaussian prime integer of type S



z=-3-2ω is an Eisenstein prime integer of type S



Kissing Fractions

Let $\mathbb{Q}$ be the collection of rational numbers. An element $\frac{a}{b} \in \mathbb{Q}$ is a **fraction** if $\gcd(a,b) = 1$. Two fractions $\frac{a}{b}$ and $\frac{c}{d}$ are **kissing** if

$$ad - bc = \pm 1.$$
$$\frac{a}{b} \heartsuit \frac{c}{d}$$

The **mediant** of $\frac{a}{b}$ and $\frac{c}{d}$ is

$$\frac{a}{b} \vee \frac{c}{d} = \frac{a+c}{b+d}.$$

Note that $\frac{a}{b} \vee \frac{c}{d}$ kisses both $\frac{a}{b}$ and $\frac{c}{d}$.

If $\frac{a}{b}$ is a fraction, its **Ford circle** is the circle centered at $\left(\frac{a}{b}, 0\right)$ of diameter $\frac{1}{b^2}$. Then, the fractions $\frac{a}{b}$ and $\frac{c}{d}$ are kissing if the Ford circle of $\frac{a}{b}$ is tangent to the Ford circle of $\frac{c}{d}$.

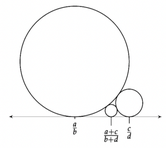


Figure 1. Ford circles of fractions and their mediants

Computing Kissing Fractions

Suppose $\frac{a}{b}$ be a fraction and consider the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then, $\frac{c}{d}$ is kissing $\frac{a}{b}$ if $da-bc = \pm 1$. That is, by setting a value for c, we obtain two values of d that guarantee $\frac{c}{d}$ is kissing $\frac{a}{b}$. Now consider the fraction $\frac{1}{2}$. For each $c \in \{\pm 1, \pm 2, \pm 3, \pm 4, \pm 5\}$ we obtain two possible values for d. After eliminating any $\frac{c}{d}$ that are not fractions, we obtain a list of fractions that are kissing $\frac{1}{2}$:

$$\left\{ \begin{array}{c} (-5,11),(-5,9),(-4,9),(-4,7),(-3,7),(-3,5), \\ (-2,5)(-2,3),(-1,3),(-1,1),(1,-1),(1,-3), \\ (2,-3),(2,-5),(3,-5),(3,-7),(4,-7),(4,-9) \end{array} \right\}$$

Kissing Fraction Results

Proposition: Let $\frac{a}{b}$ be a fraction. Then $\frac{a}{b}$ kisses infinitely many fractions.

Proposition: If $b > 1$, then there are exactly two fractions kissing $\frac{a}{b}$ with denominator smaller than b. These two fractions kiss each other and have mediant $\frac{a}{b}$.

Visualizing Kissing Fractions

```
def kiss(a,b,n):
    for c in range(-n,n):
        d = (1-b*c)/a
        e = (-1-b*c)/a
        if d is integer and greatest common divisor of c and d = 1 and c ≠0:
            add (c,d) to the kissing fractions list
        if e is integer and greatest common divisor of c and e = 1 and e ≠0:
            add (e,d) to the kissing fractions list
    return the kissing fractions list
```

References

Weissman, Martin H. An Illustrated Theory of Numbers. American Mathematical Society, 2017.

WALLPAPER GROUPS

Jack Storrs
University of California - Santa Barbara

Background

There are countless different unique patterns one can use to make a wallpaper. Despite this if we categorize them by their symmetries there are in fact only 17 different kinds. To properly discuss these kinds we will be considering wallpaper groups. Each of the 17 different wallpaper groups has different group actions which account for its unique symmetries.

The specific type of group action that we will need is called an isometry. An isometry is a distance preserving mapping from a metric space onto itself. Isometries of euclidean space are the euclidian group E_2. These isometries are translations, rotations, reflections and glide reflections. Any of these isometries can be denoted with an ordered pair (v, M) with $v \in \mathbb{R}^2$ and $M \in O_2$ Where O_2 is the group of orthogonal isometries. These are rotations about the origin or reflections over lines through the origin.

With this understanding of isometry we can now identify for any subgroup of isometries both the translation subgroup, which are the translations of the origin, and the point group, which is the group of all M.

Definition of Wallpaper Group

A Wallpaper group is defined as a subgroup G of E^2 with a translation subgroup H generated by two independent translations and a finite point group J.

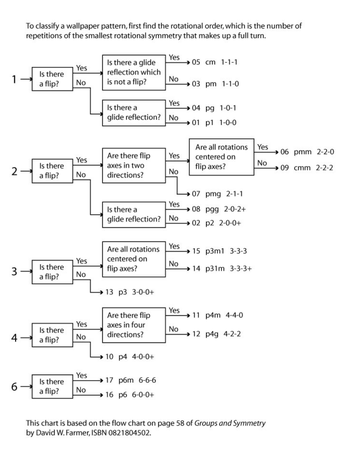
The Lattice

The Lattice of the wallpaper group is L the orbit of the origin under the wallpaper group. Let $a \in G$ be of minimum length and $b \in G$ be linearly independent with a and of minimum length. a and b span L. Different relationships between $\mid a \mid$, $\mid b \mid$, $\mid a - b \mid$ and $\mid a + b \mid$ establish five distinct types of lattice. These are oblique, rectangular, centered rectangular, square and hexagonal.

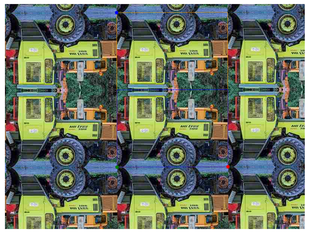
Notation for Wallpaper Groups

A common notation used to describe wallpaper groups is called crystallographic notation. It is named this because the same notation is used in crystallography to describe the 230 different space groups which are three dimensional analogues to wallpaper groups. The notation for each wallpaper group will have first either a p or a c. This denotes the type of cell either primitive or centered. Next there will be a number denoting the highest order or rotational symmetry. If the order is 1 it is often omitted. Finally there will be some number of m's and/or g's if the group contains reflections or glide reflections respectively.
As an example the wallpaper group p4m has primitive cells, rotational symmetry of highest degree 4 and mirror symmetries.

Flowchart for Classifying Wallpaper Groups

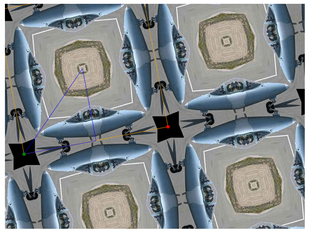


Examples of Wallpaper Patterns Generated With Wallpaper Groups



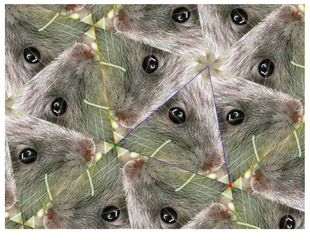
Wallpaper Group pm

The point group of this group consists of order 1 rotations and 1 axis of mirror reflections. The transnational subgroups is determined by the lattice and this example has a rectangular lattice.



Wallpaper Group p4m

The point group of this group consists of order 1,2 and 4 rotations and 2 axes of mirror reflections. The lattice in this example is square.



Wallpaper Group p6

The point group of this group consists of order 1,2 and 6 rotations. The lattice for the wallpaper group p6 must be hexagonal.

Acknowledgements

Reference Material: "Groups and Symmetry" by M. A. Armstrong
Thank you to the UCSB Directed Reading Program and to my mentor Jaime Vandeveer for making this project possible.