Mapper Stability and Visualization Methods

Comprehensive Exam

Jacob Miller

Department of Mathematics

August 26th, 2025

Outline

- Background & Motivation
- Research Questions
- Simplexification
- Persistence and Stability
- Visualization Methods and Tools
- Preliminary Results
- Future Work
- Timeline

Topological Data Analysis

The use of topology to extract and study the shape of data.

Relies on the *Manifold Hypothesis*

- Naturally occuring datasets lie along manifolds inside the ambient space.

Three Themes

1. Simplicial Approximation
2. Persistence and Stability
3. Visualization Methods

Constructing Topological Spaces from Data

Problem: Raw data lacks inherent topological structure.

Simplicial Approximation: Methods for producing topological spaces from point-cloud data

General Framework:

Given a point-cloud $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_n\} \mathrel{\substack{\subset \\ \sim}} \bigcup M_{i}^{d_{i}} \subset \mathbb{R}^d$

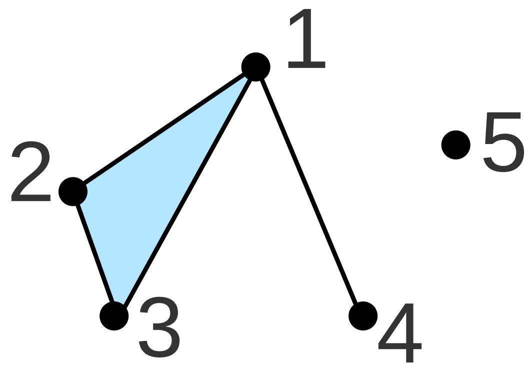
- Build simplicial complex which approximates the topology of M.

Simplicial Complex:

- A simplicial complex is a family of sets which is closed under taking subsets:

$\text{Vertices} = \Delta^{0} = \{ 1\}, \{ 2\}, \{ 3\}, \{ 4\}, \{ 5\}$ $\text{Edges} = \Delta^{1} = \{ 1, 2\}, \{ 2, 3\}, \{ 1, 3\}, \{ 1, 4\}$ $\text{Faces} = \Delta^{2} = \{ 1, 2, 3\}$

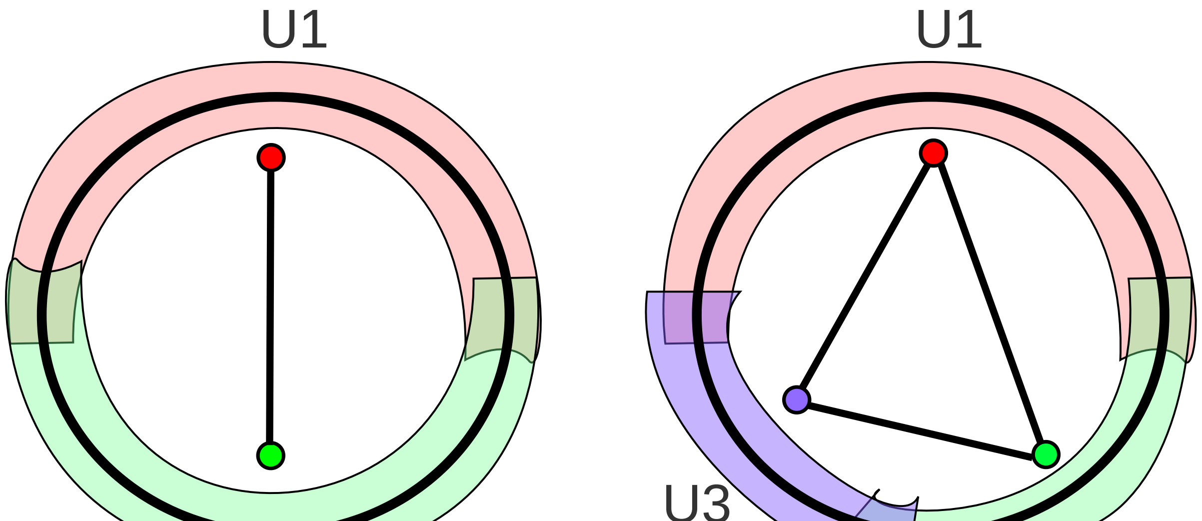
Geometric Realization



Nerve Complex:

The nerve of a family of sets $\mathcal{U}$ is a simplicial complex which records intersection information.

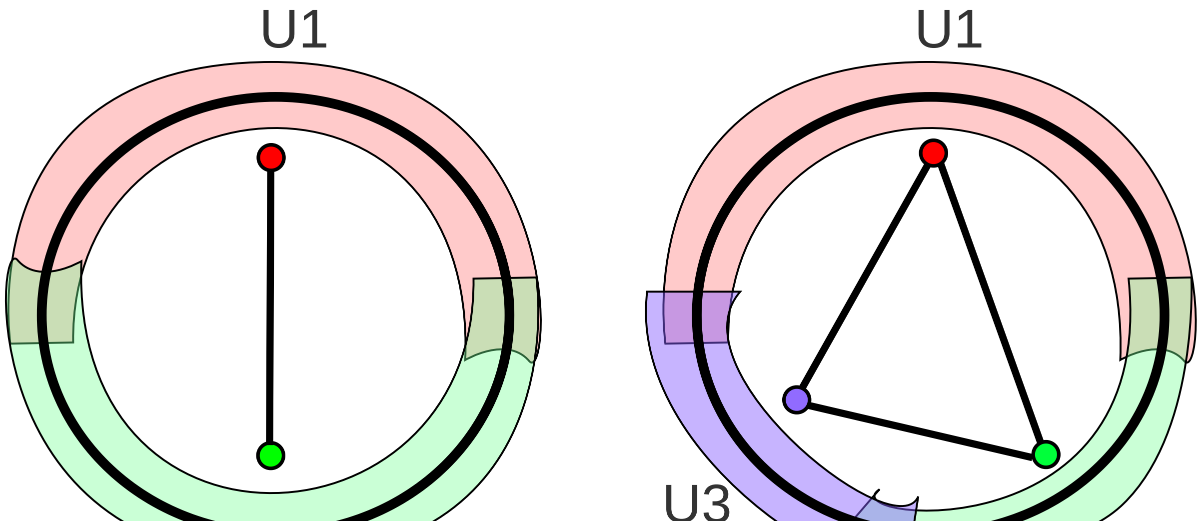
$ \sigma = \{ i_{0}, \dots, i_{k}\} \in N(\mathcal{U}) \iff \bigcap_{n=0}^{n=k} U_{i_{n}} \neq \emptyset$



Nerve Theorem:

Given an open cover $\mathcal{U}$ of a manifold M when is the topology of $N(\mathcal{U})$ equivalent to the topology of M?

The two are equivalent if each pairwise intersection $U_i \cap U_j$ is either empty or contractible.



Applying Nerve Theorem

We know how to construct the nerve from an open cover of a manifold.

Question: How do we use the nerve theorem when given a sampling from a manifold?

Čech Complex

Construction:

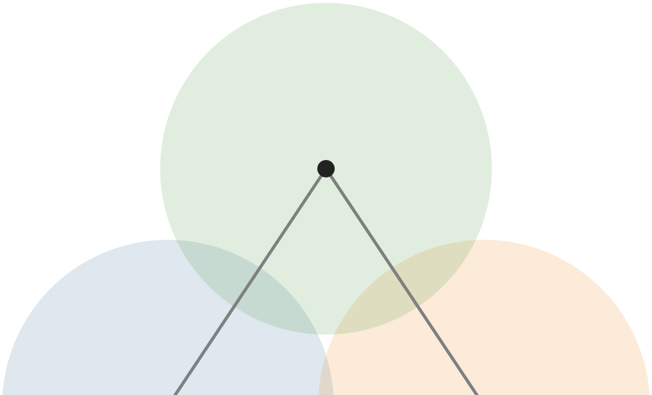
1. Create an open cover of our sampling X with balls of radius $\frac{\epsilon}{2}$ centered at each $\vec{x} \in X$.
2. $\check{C}_\epsilon (X)$ is then the nerve of this open cover.

$$ \sigma \in \check{C}_{\epsilon}(X) \iff \bigcap_{\vec{x}_{k} \in \sigma} B(\vec{x}_{k}, \frac{\epsilon}{2}) \neq \emptyset $$

Image of Cech Complex

Definition: For point cloud X and threshold $\epsilon \in \mathbb{R}$:

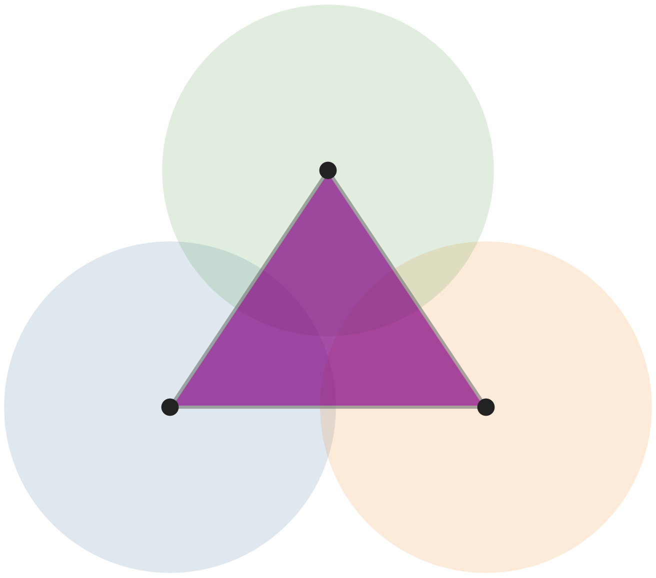
$\check{C}_\epsilon(X) = \left\{\sigma \subseteq X : \bigcap_{\vec{x} \in \sigma} B\left(\vec{x}, \frac{\epsilon}{2}\right) \neq \emptyset \right\}$



Vietoris-Rips Complex

$\sigma = {(x_{i_0}, x_{i_1}, \ldots, x_{i_k})} \in VR_\epsilon(X)$ if all pairwise distances satisfy:

$d(x_{i_j}, x_{i_k}) \leq \epsilon$



VR_{ϵ} Definition:

For point cloud X in a metric space and a given threshold $\epsilon \in \mathbb{R}^{\geq 0}$:

$$VR_\epsilon(X) = \{\sigma \subseteq X : \text{diam}(\sigma) \leq \epsilon\}$$

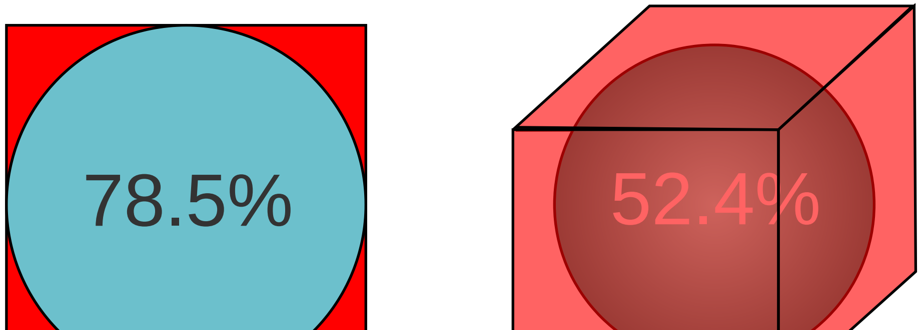
VR_{ϵ} Properties:

- $\check{C}_\epsilon (X) \subseteq VR_\epsilon(X)$ as a subcomplex
- 1-skeleton (vertices + edges) are equal for Rips and Cech complex.
- The Rips Complex may include additional "hollow" simplices:

Curse of Dimensionality

In high dimensions, geometric intuition breaks down and common computations are exponentially more expensive.

$$\frac{V_\text{hypersphere}}{V_\text{hypercube}} = \frac{\pi^{d/2}}{d2^{d-1}\Gamma(d/2)} \to 0 \text{ as } d$$



Complexity of VR_ϵ and $\check{C}_{\epsilon}$

Cech complex k-skeleton in $O(n^{k+1})$ time.

Rips Complex distance comparisons in $O(n^2)$ time.

- The number of cliques grows exponentially so the enumeration step will dominate.

Zomorodian, A. (2010). *Fast construction of the Vietoris-Rips complex*. Computers & Graphics, 34(3), 263–271.

The (Strict) Witness Complex

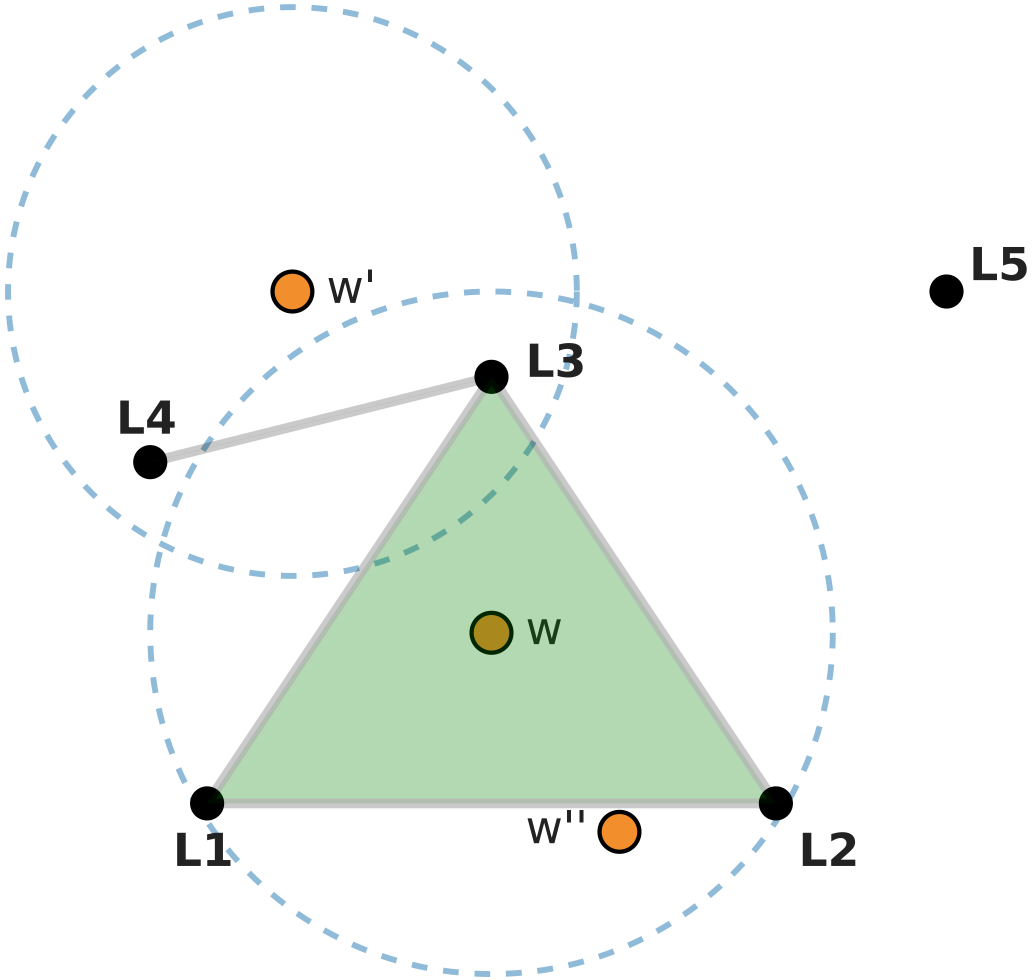
$$\text{Wit}_\infty(L, W)$$

Definition: Given landmarks and witnesses $$L, W \subset X$$

$\sigma \subseteq L$ is included as a simplex in $\text{Wit}_\infty(L, W)$ if it is 'witnessed' by some point $w \in W$.

Silva, V., & Carlsson, G. (2004). *Topological estimation using witness complexes*. Proc. Sympos. Point-Based Graphics.

A Witness Complex:



$\text{Wit}_\infty (L, W)$ Construction:

1. The vertex set is L.

2. Add an edge (l_i , l_j) if they are the two closest landmarks to some witness (ties allowed)

3. Add the k-simplex $(l_{i_0} , \dots, l_{i_{k}})$ if all of its faces exist and it is the ($k+1$)-nearest neighbors to some witness (ties allowed)

The (Lazy) Witness Complex:

$\text{Wit}(L, W)$

Lazy is to Strict as Rips is to Cech

- Same 1-skeleton as $\text{Wit}_\infty(L, W)$
- Include the simplex σ if all of its edges are included.

The (Laziest) Witness Complex:

$\text{Wit}_{\epsilon}(L, W; \nu)$ where $\epsilon \in \mathbb{R}^{+}$ and $\nu \in \mathbb{N}$

The edge (l_i, l_j) is included if there is a witness w_i whose $(d^{\nu}(w_i) + \epsilon)$-neighborhood contains l_i and l_j

where $d^\nu (w_i)$ is the distance from w_i to its

ν^{th}-nearest neighbor in L.

Then fill in all possible higher dimensional simplices.

Laziest Properties:

When $\nu = 0$ is closely related to VR_ϵ

When $\nu = 2$ and $\epsilon = 0$ we have:

$$\text{Wit}_{0}(L, W; 2) = \text{Wit}(L, W)$$

Silva, V., & Carlsson, G. (2004). *Topological estimation using witness complexes*. Proc. Sympos. Point-Based Graphics.

Persistence and Stability

Each method of obtaining a topological space from data typically has a set of parameters.

Different choices of parameters may yield different complexes.

Filtrations:

A nested sequence of simplicial complexes: $$\Delta_0 \subseteq \Delta_1 \subseteq \Delta_2 \subseteq \cdots \subseteq \Delta_n$$

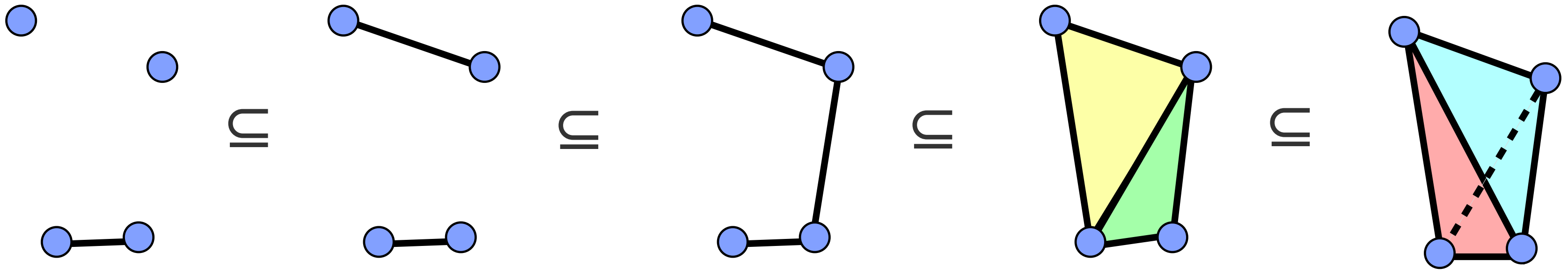
- Required in order to compare complexes in the parameter space.

Examples:

If $0 \leq \epsilon' \leq \epsilon < \infty$ then:

$VR_{\epsilon'}(X) \subseteq VR_\epsilon(X)$,

$\check{C}_{\epsilon'}(X) \subseteq \check{C}_\epsilon(X)$, $\text{Wit}_{\epsilon'}(L, W; \nu) \subseteq \text{Wit}_{\epsilon}(L, W; \nu)$



Homology

Homology measures topological features of a space X
by identifying a sequence of abelian groups:

- $H_0(X)$ ~ Connected components (0-dim holes)
- $H_1(X)$ ~ Loops/circles (1-dim holes)
- $H_2(X)$ ~ Voids/cavities of spheres (2-dim holes)
- $H_k(X)$ ~ k-dim holes

Persistent Homology

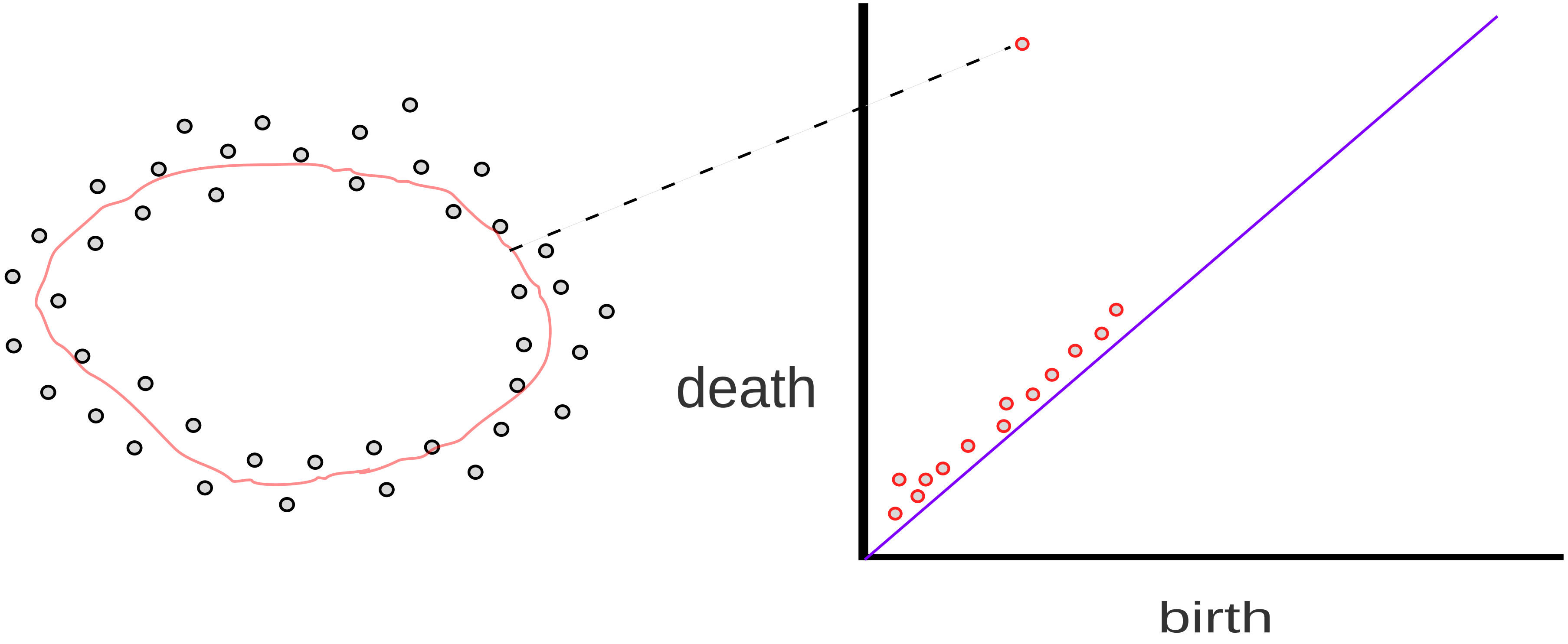
Tracks how homological features evolve across a filtration of simplicial complexes.

- Computes the homology at each stage.
- Identifies topological features that persist.

Persistence: $$\text{P}(\text{f}) = \epsilon_\text{f-death} - \epsilon_\text{f-birth}$$

Persistence Diagram

Birth-Death: Feature f_i appears at x_i (birth) and disappears at y_i (death).



Approximating Cech Complex Persistence

Given a point-cloud $X \subset \mathbb{R}^{d}$ then there is a chain of inclusions

$$ \check{C}_{\epsilon'} (X) \subseteq VR_\epsilon (X) \subseteq \check{C}_{\epsilon}(X) $$

Whenever $\frac{\epsilon}{\epsilon'} \geq \sqrt{\frac{2d}{d + 1}}$

Silva, V., & Ghrist, R. (2007). *Coverage in sensor networks via persistent homology*. Algebraic & Geometric Topology, 7.

Approximating via Witnesses

If L is an ϵ-net landmark set of $X \subset \mathbb{R}^{d}$ for $\epsilon > 0$, then

$$ VR_{\alpha / 3}(L) \subseteq \text{Wit}_{\alpha}(L, X \setminus L; \nu = 1) \subseteq VR_{3 \alpha}(L)$$ for $\alpha \geq 2\epsilon $

Arafat, N. A., Basu, D., & Bressan, S. (2019). *Topological Data Analysis with ε-net Induced Lazy Witness Complex*. arXiv preprint arXiv:1906.06122.

Stitching Together

So in theory with a representative landmark set satisfying the ϵ-net condition you can approximate the persistence in the Cech complex filtration via a laziest witness complex filtration.

Problem: May still require a large high-dimensional Landmark set for it to be representative.

Problem: Curse of dimensionality makes it so that nearest neighbors is not stable under noise.

Mapper

"You're still trying to replace $\check{C}_{\epsilon}(X)$. I told you we can't do it. Now, what we might be able to do is re-create it in the aggregate."

Welch, D. (n.d.). *We're going to recreate Toney in the aggregate.* Medium.
https://medium.com/@duncanwelch31/were-going-to-recreate-toney-in-the-aggregate-4a4dfe370b1c

Mapper Ingredients

Singh, G., Mémoli, F., & Carlsson, G. E. (2007). *Topological methods for the analysis of high dimensional data sets and 3d object recognition*. In SPBG (pp. 91–100).

1. A dataset X with a metric

2. A filter function $f:X \to \mathbb{R}^{m}$

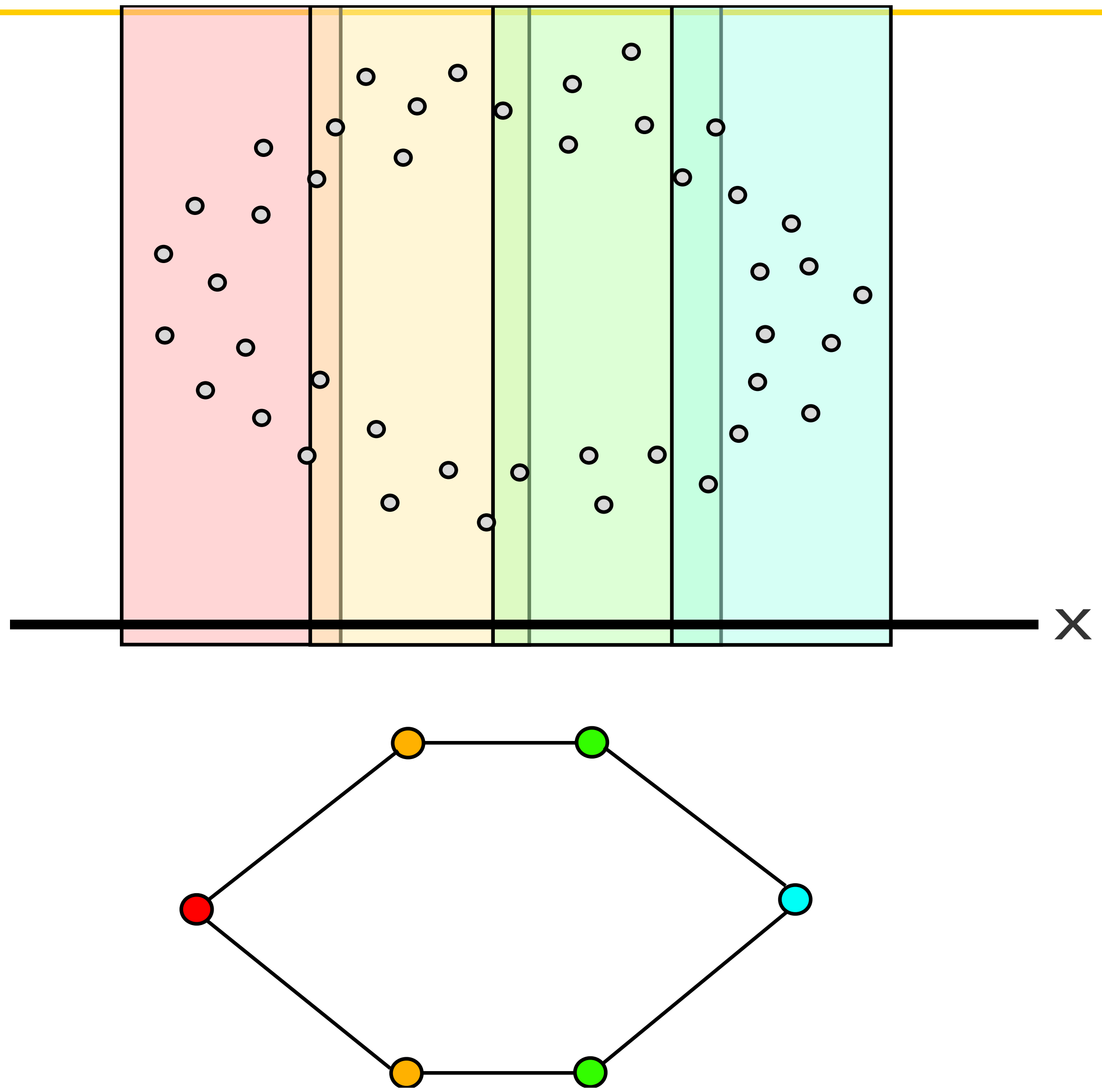
3. An open cover $\mathcal{U}$ of $f[X]$

Construction

Pullback each cover element $$\bar{U}_i = f^{-1}[U_i] \subseteq X$$ $$\downarrow$$

Cluster points in each $\bar{U}_i$ $$\downarrow$$ Build the nerve from the set of all resulting clusters.

Mapper



Mapper filtration:

Bungula, W., & Darcy, I. (2024). *Bi-Filtration and Stability of TDA Mapper for Point Cloud Data*. arXiv:2409.17360. https://arxiv.org/abs/2409.17360

Stability in Mapper Parameter Choices

(do last)

Maper Similarity Metric

Previous work on estimating parameters

Visualization/Analysis Methods

1. Multi-scale summaries of homology
    - Persistence Diagrams
2. Fixed-scale nerve representations
    - Simplicial representations (Cech complex, etc.)
    - Mapper

Current Tools:

- Ripser
- CeREEBrus
- Kepler Mapper
- Dionysus
- GHUDI
- Zen Mapper
- TDAView
- MappeR

Natural Image Space Setup

Given a grid of grayscale pixels, a major focus of computer vision research is studying the subset of natural images.

Van Hateren Image: imk00001.imc
(Correct Endian + Percentile Scaling)

van Hateren, J. H., & van der Schaaf, A. (1998). *Independent component filters of natural images*

Major Problem

The space of such images globally has as many dimensions as there are pixels which makes analysis intractible.

To reduce this dimensionality problem, most analysis is done locally.

Natural Images Analysis

Carlsson, G., Ishkhanov, T., Silva, V., & Zomorodian, A. (2008). *On the Local Behavior of Spaces of Natural Images*. International Journal of Computer Vision, 76, 1–12.

Considered the space of 3×3 *high-contrast* patches from natural images.

Previous Results

Have shown evidence for a high-density manifold whose 1-skeleton follows the three circle model:

Pre-processing Flowchart:

- Randomly sample 5000 patch vectors in $\mathbb{R}^9$
- Take logarithm
- Subtract average of all coordinates from each coordinate
- Compute contrast or "D-norm" of the vector
- Keep the patch if it is among the top 20% of all patches
- Normalize by D-norm to place on a 7-dim ellipsoid

Final Analysis

- Dataset X of 4×10^6 high-contrast patches in $\mathbb{R}^{8}$

Goal: Study high-density subset of these high-contrast patches.

Density Filter

Let $X(\nu, p)$ to be the $p\%$ of points in X with the smallest distance to their ν^th-nearest neighbor.

Five Circles

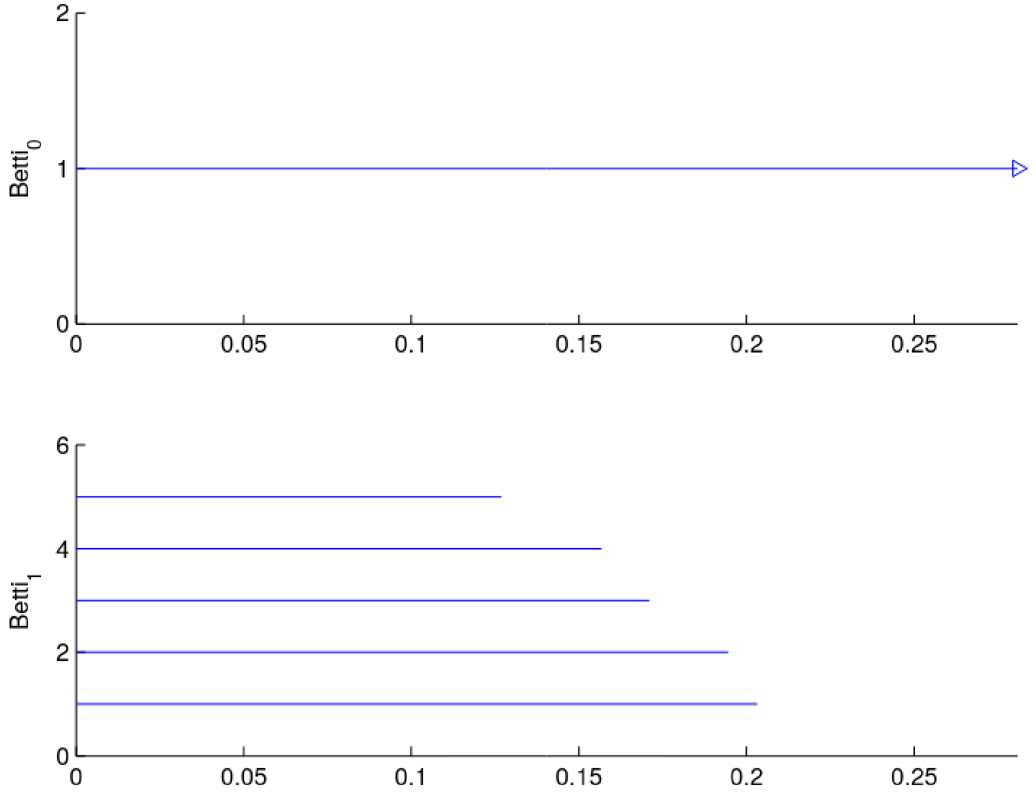
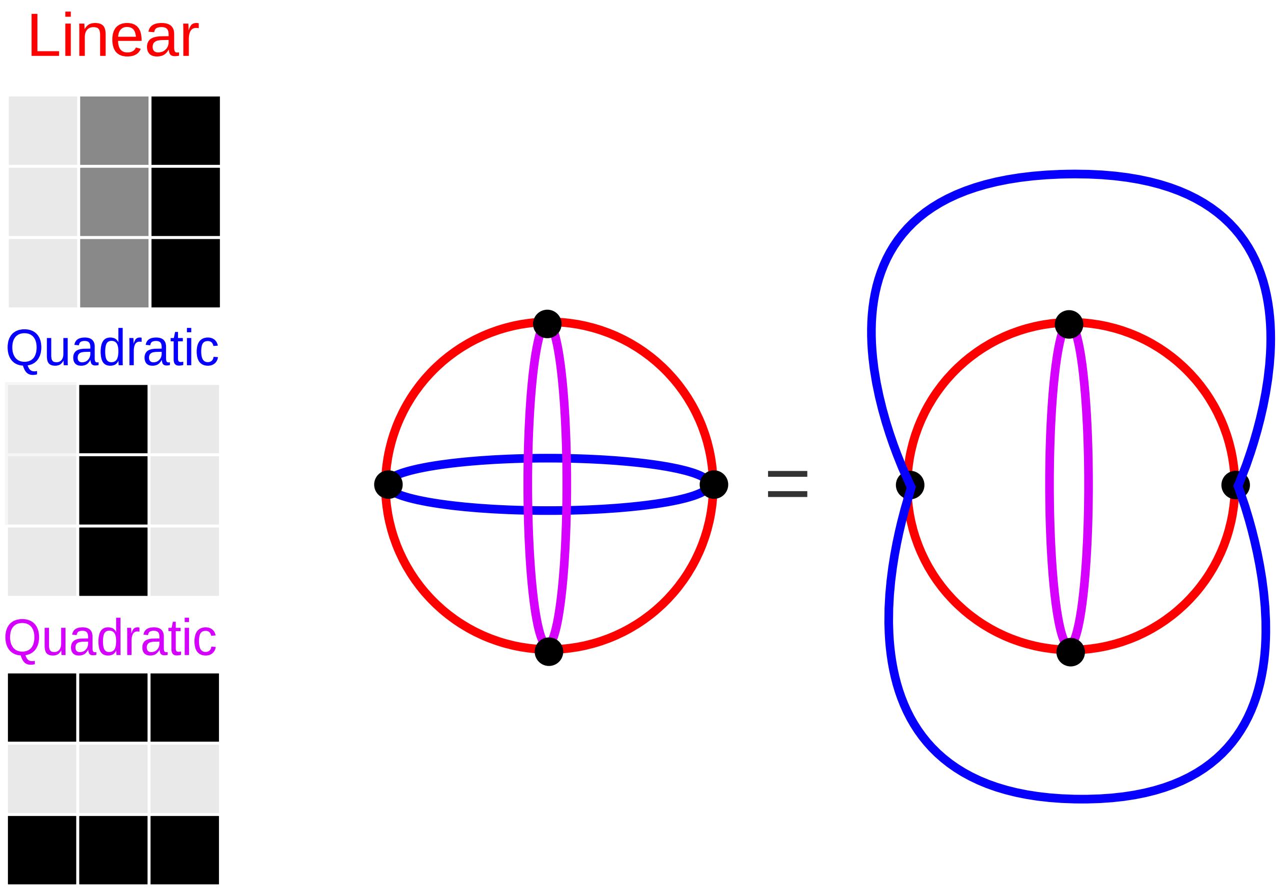
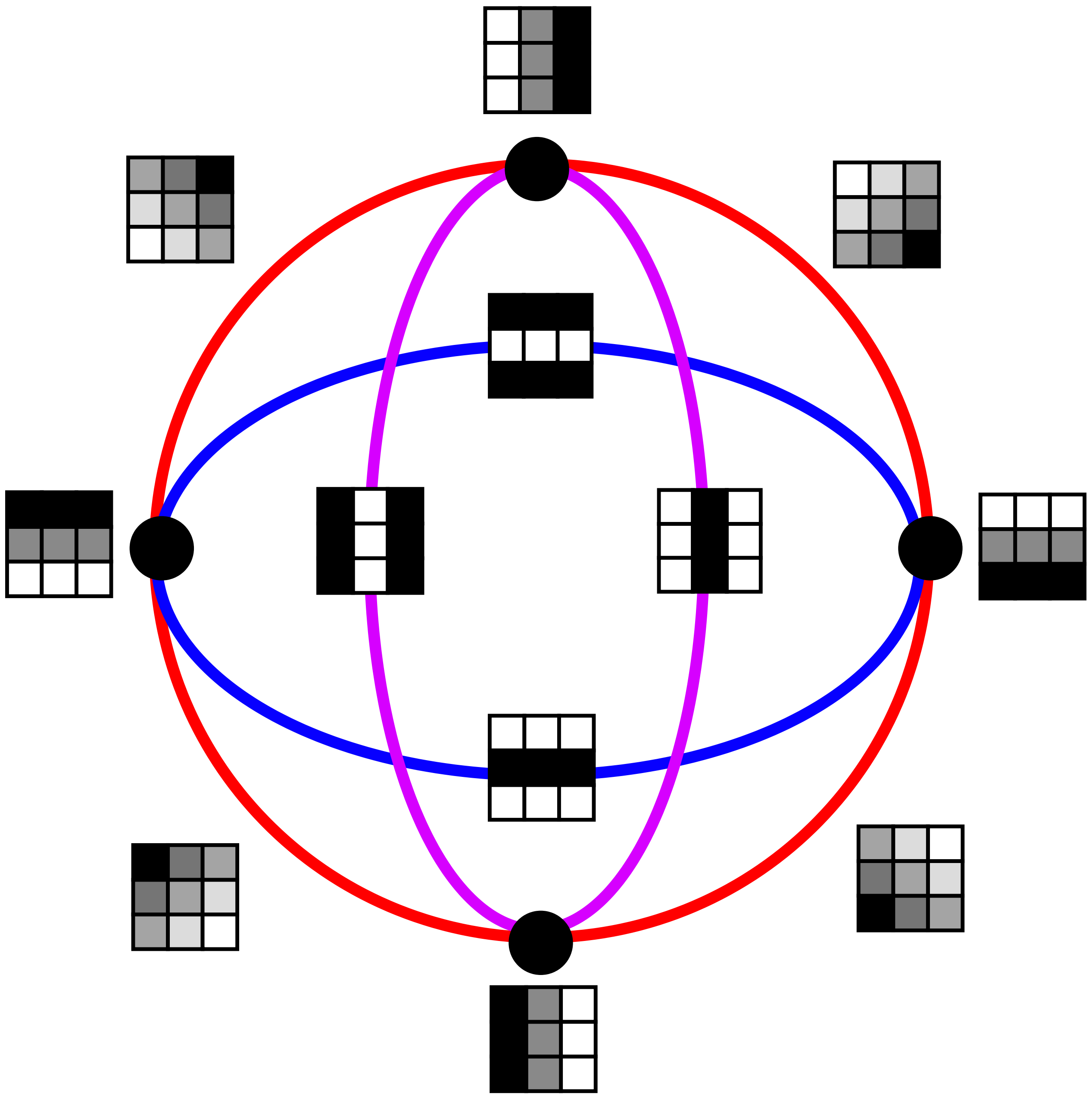


Figure 7: $PLEX$ results for $X(15, 30)$

Three Circle Model:

After analyzing patch data-points





2-Manifold

$\beta_{2} = 1$ is evidence that we have a surface.

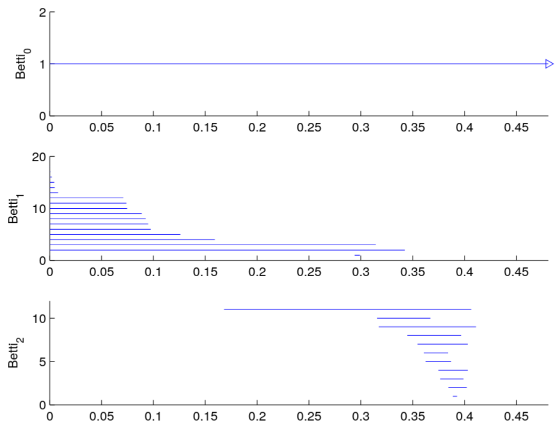


Figure 9: $PLEX$ results for $X(100, 10) \cup Q$

- So this space is either a Torus or a Klein Bottle.

The Klein Bottle:

They conclude you can parameterize this space with a Klein bottle:

Figure 3: Klein bottle immersion in $\mathbb{R}^3$

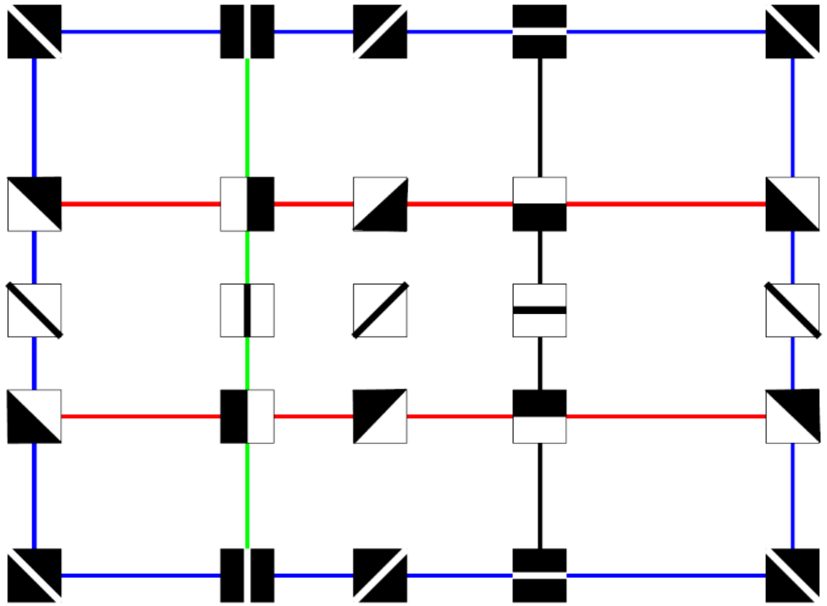


Figure 6: 3 by 3 patches parametrized by the Klein bottle

Goal:

Implement a similar qualitative analysis with mapper.

- This would require an interactive exploratory tool for mapper
    - Manipulate and decompose simplicial complex

Zen-Sight

Interactive Exploration

- **Filtrations**: Animate across parameter ranges
- **Mapper Graph**: Manipulate and decompose simplicial complex
- **Data Integration**: Include with original data for analysis

Future Work

Filtration Visualization

Fin