Manifold Modeling of Pentagon Spaces Using Laplacian Eigenfunctions

Quincy Alston, Elise Alvarez-Salazar, Kiyanna Porter

July, 2023

Abstract

The algorithmic sampling of conformation spaces is a problem with applications across computational chemistry and biology. Often these spaces represent the configurations of a molecule at different energies. We present an algorithm for sampling the moduli space of pentagons with mathematical justification. Furthermore, we use persistent cohomology and Eilenberg-MacLane spaces to identify the underlying manifold of said moduli spaces. We use said identification as well as eigenfunctions of the Laplacian to create best fit manifolds. In the process we recover a 2D visualization of our space. The developed pipeline allows for the visualization of complicated high dimensional spaces in lower dimension.

Contents

1 Introduction

1.1 Background

Conformation spaces are known as the space encompassing all possible configurations of a molecule. These spaces are most commonly found in the scientific fields of computational chemistry and biology, especially in the study of covalent bonds between molecules and carbon atoms. The algorithmic sampling of these conformation spaces advances scientists' understanding of the higher dimensional data these molecules produce [4]. Our project presents an algorithm for the sampling of the moduli space of pentagons with mathematical justification and a mapping of their energies.

1.2 Contributions

Klaus and Kojima use cell decomposition to prove the moduli space of equilateral pentagons is a closed oriented 2-manifold of genus 4 using cell decomposition. We take inspiration from this decomposition to sample the moduli space in its entirety. Klaus and Kojima showed that moduli spaces with different edge lengths can produce different 2-manifolds (See Figure 1). Klaus and Kojima use the Surfer software to visualize these manifolds but their visualization lacks the identification of a pentagon to a point on the manifold. We work with the edge lengths which realize surfaces of genus 3, genus 4, and a sphere. Specifically, we use distances $(0.6, 0.6, 0.6, 1, 1)$ for the genus 3 surface, $(1, 1, 1, 1, 1)$ for the genus 4 surface, and $(0.4, 0.4, 0.4, 0.4, 1)$ for the sphere [6].

We work to identify the topological structure of these samplings from pentagon conformation spaces using persistent cohomology as well as the Topological Dimensionality Reduction (TDR) package DREiMac. DREiMac uses Eilenberg-MacLane spaces and persistent cohomology to find the underlying 3D representation of the moduli space. Additionally we are interested in the planar visualization of these spaces. It is known that genus g 2-manifolds can become a $4g$ sided or $8g$ sided polygon once mapped to a planar model. By taking persistent cohomology on the 3D visualization we find cuts to attain planar decompositions.

1.3 Goal

The goal of this paper is to present an algorithm which will allow for the visualizations of high dimensional spaces into lower dimension to then unfold the manifolds into the underlying representations of the moduli spaces. This will then hopefully allow for the unfolding of these manifolds into a 2D planar decomposition to help map out the specific coordinates of a point on the manifold while preserving inherent data.

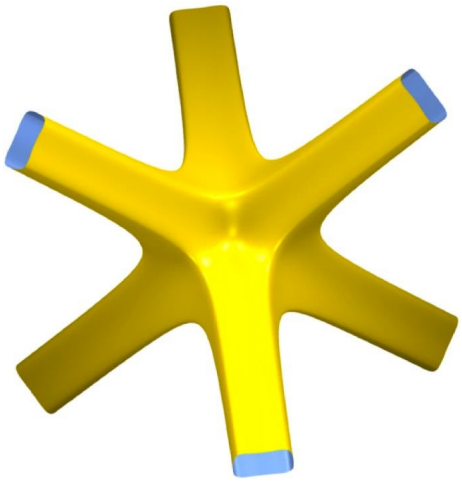


Figure 1: Pentagon conformation space produced by side lengths $(0.6, 0.6, 0.6, 1, 1)$. [6]

2 Preliminaries

2.1 Definitions and Background

2.1.1 Pentagons and Pentagon Moduli Spaces

We will begin by providing a formalization of pentagons and pentagon spaces as the solutions to a set of equations in affine complex space. Complex space allows us to define pentagons as solutions to polynomials and makes our notion of an equivalence relation cleaner.

Definition 2.1. A pentagon p is a vector $(z_1, z_2, z_3, z_4, z_5) \in \mathbb{C}^5$ such that $d_i = |z_{i+1} - z_i| > 0$ for all $1 \leq i \leq 5$ where $z_6 = z_1$.

Definition 2.2. Let $(d_1, d_2, d_3, d_4, d_5) \in \mathbb{R}^5_+$. The moduli space of pentagons with side lengths $(d_1, d_2, d_3, d_4, d_5)$ is the quotient

$$\mathcal{M}_5(d_1, d_2, d_3, d_4, d_5) := \{p = (z_1, z_2, z_3, z_4, z_5) \in \mathbb{C}^5 \mid d_i = |z_{i+1} - z_i| > 0, \forall 1 \leq i \leq 5, \}/\sim$$

where $z_6 = z_1$ and $p = (z_1, z_2, z_3, z_4, z_5)$, $p' = (z'_1, z'_2, z'_3, z'_4, z'_5)$ are equivalent if and only if there exists $a, b \in \mathbb{C}$ with $|a| = 1$ such that $z_i = az'_i + b$ for all $1 \leq i \leq 5$.

For our purposes, it can be assumed without loss of generality that $d_5 = 1$ because for any $(d_1, d_2, d_3, d_4, d_5) \in \mathbb{R}^5_+$ there is a bijection $\mathcal{M}_5(d_1, d_2, d_3, d_4, d_5) \leftrightarrow \mathcal{M}_5(d_1/d_5, d_2/d_5, d_3/d_5, d_4/d_5, d_5/d_5 = 1)$ sending each equivalence class representative p to the equivalence class of p with side lengths scaled by $\frac{1}{d_5}$. Intuitively, there is no difference in the topological features of a pentagon moduli space and its scaled version.

Definition 2.3. Let $a = \frac{1}{z_1 - z_5}$ and $b = \frac{-z_5}{z_1 - z_5}$. Calculate that $|a| = 1$ and $az_1 + b = 1$, $az_5 + b = 0$. Therefore every pentagon is equivalent to one of the form $(1, w_1, w_2, w_3, 0)$. A pentagon in this form is said to be in standard form.

Note that two pentagons are equivalent if and only if there is an affine linear automorphism on $\mathbb{C}$ which sends one pentagon to the other. Intuitively, this means two pentagons are equivalent if we can transform one onto the other via a translation and rotation. There is a unique affine linear automorphism which sends $z_1 \to 1$ and $z_5 \to 0$ because affine linear automorphisms are determined by their images on two distinct vectors. This shows that the standard form of a pentagon is unique.

2.1.2 Persistent Cohomology

Persistent cohomology is a computational method for discerning the topological features of data. There are many existing libraries which provide optimized persistent cohomology computation.

Definition 2.4. Let (X, d_X) be a metric space and $\alpha > 0$. Then the Rips Complex is the simplicial complex $R_\alpha(X, d_X) := \{\sigma \in X \mid 0 < |\sigma| < \infty, \text{diam}(\sigma) < \alpha\}$.

Definition 2.5. Let (X, d_X) be a metric space and let $R_\alpha(X, d_X)$ be the Rips complex on X for an $\alpha > 0$. The Rips filtration is the set $\mathcal{R}(X, d_X) := \{R_\alpha(X, d_X)\}_{\alpha \geq 0}$.

Define $K^{(i)} := \{\sigma \in R_\alpha(X, d_X) \mid |\sigma| = i + 1\}$. Let $C^i(R_\alpha(X, d_X); \mathbb{F})$ be the $\mathbb{F}$ vector space with basis $K^{(i)}$ and set $\alpha_1 < \alpha_2 < ... < \alpha_n$ to be the α at which the Rips complex gains new edges. Then the Rips filtration induces a chain complex

$$C^i(R_{\alpha_1}(X, d_X); \mathbb{F}) \leftarrow C^i(R_{\alpha_2}(X, d_X); \mathbb{F}) \leftarrow ... \leftarrow C^i(R_{\alpha_n}(X, d_X); \mathbb{F}) \leftarrow 0 \tag{1}$$

The ith cohomology functor $H^i(-; \mathbb{F})$ applied to (1) induces the diagram

$$H^i(R_{\alpha_1}(X, d_X); \mathbb{F}) \leftarrow H^i(R_{\alpha_2}(X, d_X); \mathbb{F}) \leftarrow ... \leftarrow H^i(R_{\alpha_n}(X, d_X); \mathbb{F}) \leftarrow 0 \tag{2}$$

[1]

Definition 2.6. An ith persistent cohomology class denoted $[\alpha_i, \alpha_j)$ with $i < j$ is a cohomology class $\eta \in H^i(R_\alpha(X, d_X); \mathbb{F})$ such that η is a cocycle for $\alpha < \alpha_j$ and a coboundary for $\alpha < \alpha_i$. [1]

Definition 2.7. The ith persistence diagram is the plot where the set of all ith persistent cohomology classes $\{[\alpha_i, \alpha_j)\}$ are plotted as points (α_i, α_j).

Persistence diagrams allow us to visually interpret which topological features are most persistent in the data. Using appropriate coefficients and nth cohomologies one can infer numerous topological invariants of the data including genus and Betti numbers. Further, because persistence diagrams exhibit the comparative persistence of different classes, one can infer the size of different topological features in the data.

In our case, the metric space X is going to be a finite compact set of points sampled from a pentagon moduli space. The metric will often be the Euclidean metric on $\mathbb{C}^5 \cong \mathbb{R}^{10}$ as an $\mathbb{R}$ vector space.

On occasion, the metric used will be the geodesic distance on the data set. Let (X, d_X) be a Euclidean metric space and fix $k \in \mathbb{N}$. Build a graph G_k by connecting the k nearest neighbors under the Euclidean metric for each $x \in X$. Give each edge $E(x, y)$ between $x, y \in X$ in the graph a weight which is the Euclidean distance between vertices.

Definition 2.8. Let $P(x, y)$ be the set of paths in G_k beginning at x and ending at y. Define for $p \in P(x, y)$ the value $|p|$ to be the sum of the weights of all edges in the path p. Then the geodesic distance is $d_{G_k}(x, y) := \inf_{p \in P(x,y)}\{|p|\}$.

Geodesic distance is often useful when calculating persistent cohomology because certain holes and features in data will show up as more persistent than with the Euclidean metric. This is because sometimes cycles which are narrow or distorted may not be as persistent under the Euclidean metric as with geodesic distance. The drawbacks however are that geodesic distance can be extremely expensive to compute.

2.1.3 The Laplacian and the Spectral Theorem

Once data is sampled, using persistence diagrams, we can detect what topological space our pentagons create. A question to ask is how can we do manifold modeling since this can allow us to have better control over our data. There is some premise to modeling data in this way. In the context of 2D periodic data and what are known as Fourier approximations, periodic data can be decomposed into linear combinations of trigonometric functions. We now do something similar with specific manifolds. Laplacian eigenfunctions allow us to decompose certain functions from a manifold to $\mathbb{C}$ as a convergent infinite series. In summary, these well understood and explicit functions can be used to model noisy data.

Definition 2.9. $\mathcal{M} \subseteq \mathbb{R}^\ell$ is a smooth k-manifold if for all $x \in \mathcal{M}$ there exists an open set $U \subset \mathbb{R}^k$ and diffeomorphism $\varphi : U \to \mathcal{M}$ such that $0 \mapsto x$.

For the remainder of the paper, all manifolds are smooth k-manifolds.

Definition 2.10. A Hilbert space H is a real or complex inner product space that is also a complete metric space with respect to the metric induced by the inner product.

Definition 2.11. For $\Omega \subseteq \mathbb{R}^k$ a bounded domain, a function $f : \Omega \to \mathbb{C}$ is said to be square integrable if $\int_\Omega \mid f(\mu) \mid^2 \, d\mu < \infty$ where μ is the Lebesgue measure.

Consider a smooth k-manifold $\mathcal{M}$ which is bounded in $\mathbb{R}^l$. Note that $\mathcal{M}$ can sit in $\mathbb{C}^n$ and still be viewed as sitting in a higher dimension $\mathbb{R}^{2n}$. Define the complex Hilbert space $L^2(\mathcal{M}) := \{f : \mathcal{M} \to \mathbb{C} \mid f \text{ is square integrable}\}$ with norm $\langle u, v \rangle := \int_\Omega uv \, d\mu$. $L^2(\mathcal{M})$ is infinite dimensional.

Definition 2.12. The Laplacian operator $\Delta : L^2(\mathcal{M}) \to C^0(\mathcal{M})$ is defined as the linear partial differential equation taking $f \mapsto \nabla^2 f$.

Theorem 2.13. The eigenfunctions of the Laplacian operator Δ form an orthonormal basis for $L^2(\mathcal{M})$ [7].

The above follows from the complex spectral theorem on compact self-adjoint operators. Since $L^2(\mathcal{M})$ is infinite dimensional, Δ will have infinitely many eigenfunctions.

Theorem 2.14. For any $f \in L^2(\mathcal{M})$ there exists $\hat{h}_n \in \mathbb{C}$ such that,

$$\lim_{N \to \infty} \sum_{n=1}^{N} \hat{h}_n u_n = f$$

where u_n are the eigenfunctions of Δ [7].

These results indicate that for square integrable functions from a manifold into $\mathbb{C}$ we can approximate using Laplacian eigenfunctions, and the only remaining question to answer is which choices of $\hat{h}_n$ values behave best. If we can solve for the $\hat{h}_n$ then we can create a best fit model for manifold valued data in high dimension.

2.2 Previous Work

In this section we discuss the previous work from which we draw inspiration. We begin with findings from Klaus and Kojima in their studies of the pentagon space. By fixing two points of the pentagon in the complex plane to 0 and 1, there are eight real variables and six real polynomial conditions which define the space $\mathcal{M}_5$. Therefore, $\mathcal{M}_5$ is also a real affine variety in $\mathbb{R}^8$. In the case of non-equilateral pentagons, there are specific choices of edge lengths such that the manifold is diffeomorphic to a sphere (See Fig 2). This manifold is given by a polynomial of degree 12 [5].

One can define the moduli space of equilateral pentagons by taking $\mathcal{M}_5(1,1,1,1,1)$. One can recover an equivalent definition of $\mathcal{M}_5(1,1,1,1,1)$ as an algebraic variety over $\mathbb{C}$ via $\mathcal{M}_5(1,1,1,1,1) := T^4 \cap H \in \mathbb{C}^4$ where $H(z_1, z_2, z_3, z_4) := z_1 + z_2 + z_3 + z_4 + 1 = 0$ is a hyperplane and T^4 is the 4-torus sitting in $\mathbb{C}^4$ [6].

Theorem 2.15 (Klaus–Kojima, 2019)**.** $\mathcal{M}_5(d_1, d_2, d_3, d_4, d_5)$ is a connected orientable closed 2-manifold.

Theorem 2.16 (Klaus–Kojima, 2019)**.** $\mathcal{M}_5(1,1,1,1,1)$ is a 2-manifold of genus 4.

2.3 DREiMac and Ripser

The Python function Ripser takes in data X and a coefficient ring R and computes $\mathcal{R}(L) = \{\mathcal{R}_\alpha(L)\}_{\alpha \geq 0}$, where $L \subseteq X$ is a sample of landmarks for X, and the set of persistent cohomology classes is $PH^i(\mathcal{R}(L); \mathbb{Z}/q\mathbb{Z})$, for q prime and $0 \leq i \leq 2$.

$K(A, n)$ is the Eilenberg-MacLane space for an abelian group A and integer n. $K(A, n)$ is also a cell complex.

Definition 2.17. For a CW-complex, the space $[X \to K(A, n)]$ is the set of homotopy equivalence classes of continuous maps between X and $K(A, n)$.

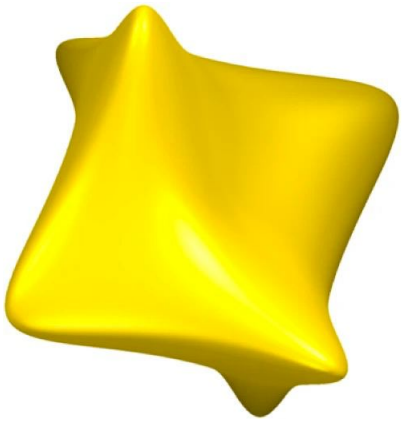


Figure 2: Visualization of of the moduli space of pentagons with side lengths $(0.4, 0.4, 0.4, 0.4, 1)$ [6].

For a definition of a CW-complex see Hatcher [3]. There is a general result regarding CW-complexes on which the algorithm is based:

Theorem 2.18. For CW-complexes X, abelian groups A, and integers $n \geq 0$, there are natural isomorphisms

$$H^n(X; A) \cong [X \to K(A, n)]$$

[3].

Particularly, for $R = \mathbb{Z}$ then $K(\mathbb{Z}, 1) = S^1$ and $K(\mathbb{Z}, 2) = \mathbb{C}P^\infty$. The Python package DREiMac contains algorithms that intake sampled data of a space and do the following:

1. Run the data through Ripser to compute persistent cohomology.
2. For each cohomology class in $PH^i(\mathcal{R}(L); \mathbb{Z}/q\mathbb{Z})$, DREiMac lifts the cohomology class to $PH^i(\mathcal{R}(L); \mathbb{Z})$.
3. DREiMac uses $PH^i(\mathcal{R}(L); \mathbb{Z})$ and Theorem 2.18 to generate:

$$f_\eta : \bigcup_{x \in L} B_\alpha(x) \to K(\mathbb{Z}, i)$$

   in the corresponding homotopy class such that the coordinates the function assigns to the data are topologically meaningful, where α is chosen depending on the persistent cohomology class and $B_\alpha(x)$ is the open ball of radius α centered at x.

In the $K(\mathbb{Z}, 1) = S^1$ case, the DREiMac function is called `CircularCoords` and allows high dimensional data to be plotted in $\mathbb{R}^3$ using 3 linear combinations of persistent classes. In the $K(\mathbb{Z}, 1) = \mathbb{C}P^\infty$ case, the DREiMac function is called `ComplexProjCoords` and it uses a series of projections to map a data set X onto S^2.

3 Sampling Pentagon Conformation Spaces

In this section we present our algorithm for sampling from different pentagon moduli spaces. In the applied setting, conformation spaces containing the set of all configurations of a geometric or physical object are extremely difficult to sample evenly [4]. Often the objects which are at a higher energy level or are degenerate are much less likely to be measured. We present a proof that our algorithm can recover every pentagon in the moduli space.

3.1 Algorithm

Fix a set of side lengths $(d_1, d_2, d_3, d_4) \in \mathbb{R}^5_+$ and input a sample size n. Without loss of generality let $d_5 = 1$. We first confirm that $1 \leq \sum_{i=1}^4 d_i$ because otherwise there are no valid pentagons in the space. Once confirmed, fix $z_1 = 1, z_5 = 0$ and set $r_1 = d_3 + d_4$ and $r_2 = d_1 + d_2$. Let $B(z_5, r_1)$ be the closed disk of radius r_1 around z_5 and define $B(z_1, r_2)$ analogously. Now, if $d_1 > d_2$ take $B_{d_1 - 2d_2}(z_1)$ as the open disk and if $d_2 > d_1$ take $B_{d_2 - d_1}(z_1)$. Similarly do the same for $\max(d_3, d_4)$ and z_5. Take $j_1 = \max(d_1, d_2) - \min(d_1, d_2)$ and $j_2 = \max(d_3, d_4) - \min(d_3, d_4)$. Call:

$$C = B(z_5, r_1) \cap B(z_1, r_2) \setminus B_{j_1}(z_1) \cup B_{j_2}(z_5) \tag{3}$$

Take a evenly spaced and finite set of points $P \subseteq C$ along a grid. Instantiate the list `sample=[]`. For each $z_3 \in P$ do the following.

1. Let $S(z_1, d_1)$ be the circle of radius d_1 around z_1 define $S(z_3, d_2)$ analogously.
   (a) If there are finitely many intersection points then append them all to the set A.
   (b) If there are infinitely many intersection points then take k intersection points evenly spaced around $S(z_1, d_1)$ and append them to the set A.
2. Repeat the above process using $S(z_5, d_4)$ and $S(z_3, d_3)$ and append points to the set B.
3. For each choice of $z_2 \in A$ and $z_4 \in B$ append to `sample` the pentagon $(z_1 = 1, z_2, z_3, z_4, z_5 = 0)$.

The algorithm then returns the list `sample` which contains approximately $4n$ pentagons in standard form.

3.2 Mathematical Justification of Sampling Algorithm

We claim that the above presented algorithm not only always generates valid pentagons but also is capable of sampling every possible configuration of the pentagon. The sampling algorithm relies on the following claims.

- We assume that a pentagon is valid if and only if z_3 lies in C
- We assume that $S(z_5, d_4)$ and $S(z_3, d_3)$ always intersect at least once and the same for $S(z_1, d_1)$ and $S(z_3, d_2)$
- For determined z_3 we assume that a pentagon is valid if and only if $z_2 \in A$ and $z_4 \in B$.

We justify these claims via the following proposition.

Proposition 3.1. Fix side lengths $(d_1, d_2, d_3, d_4) \in \mathbb{R}^5_+$. If a vector $(1, z_2, z_3, z_4, 0) \in \mathbb{C}^5$ is a standard form pentagon then $z_3 \in C$. Conversely, if $z_3 \in C$ then there exists z_2, z_4 which construct a valid standard form pentagon.

Proof. Suppose that $(z_1 = 1, z_2, z_3, z_4, z_5 = 0) \in \mathbb{C}^5$ is a valid pentagon in standard form with side lengths (d_1, d_2, d_3, d_4). We know $|z_3 - z_2| = d_2$ and $|z_2 - z_1| = d_1$ so by the triangle inequality $|z_3 - z_1| \leq d_2 + d_1 = r_2$. Thus $z_3 \in B(z_1, r_2)$. Symmetrically, $|z_4 - z_3| = d_3, |z_5 - z_4| = d_4$ so $|z_5 - z_3| \leq d_3 + d_4 = r_1$ and $z_3 \in B(z_5, r_1)$. So $z_3 \in B(z_5, r_1) \cap B(z_1, r_2) \subseteq C$.

Suppose that $z_3 \in C$. First, consider z_2. We separate the possible intersection cases of $S(z_1, d_1)$ and $S(z_3, d_2)$ into the following cases: disjoint, one circle inside the other, the same circle, tangent, and twice intersecting.

The disjoint case is ruled out because if $S(z_1, d_1), S(z_3, d_2)$ were disjoint then $r_2 + d_1 + d_2 < |z_3 - z_1|$. The one circle inside the other case is ruled out because if $S(z_1, d_1)$ were inside $S(z_3, d_2)$ or vice versa then $|z_3 - z_1| + \min(d_1, d_2) < \max(d_1, d_2)$. Then $|z_3 - z_1| < \max(d_1, d_2) - \min(d_1, d_2)$ which is a contradiction because $z_3 \notin B_O(z_1, \max(d_1, d_2) - \min(d_1, d_2))$. In the case that one circle is inside the other there exist infinitely many sufficient z_2. In the tangent and twice intersecting cases there are 1 and 2 sufficient z_2 respectively. The existence of z_4 follows from a symmetric argument on $S(z_5, d_4)$ and $S(z_3, d_3)$. □

To further address the last claim, the points z_2 which satisfy both $|z_2 - z_1| = d_1$ and $|z_2 - z_3| = d_2$ are exactly the intersection points of the circles $S(z_1, d_1)$ and $S(z_3, d_2)$. Therefore the circle intersection method does recover every possible pentagon conformation.

One enduring issue is the even sampling of the conformation space. We are able to prove that that C is exactly the image of the projection map $\rho : \mathcal{M}_5 \to \mathbb{C}$ defined by $\rho([(1, z_2, z_3, z_4, 0)]) \mapsto z_3$. For different $z_3 \in C$ there are a few possibilities for $\rho^{-1}(z_3)$. We know $|\rho^{-1}(z_3)| \in \{\infty, 4, 2, 1\}$. The case that $|\rho^{-1}(z_3)| = \infty$ comes from when $z_3 = z_1$ or $z_3 = z_5$ and $d_1 = d_2$. Then $\rho^{-1}(z_3)$ is either 1 or 2 loops in $\mathcal{M}_5$. The $|\rho^{-1}(z_3)| \in \{4, 2, 1\}$ cases are generic and come from when $S(z_1, d_1)$ and $S(z_3, d_2)$ have 1-2 intersections and the same for $S(z_5, d_4)$ and $S(z_3, d_3)$. These cases can present a problem for the algorithm because if C is randomly sampled then there is probability 0 that the algorithm will compute the infinitely many pentagons in $\rho^{-1}(z_1)$. We create visualizations of the pentagon space in 3D which allow us to see directly whether the algorithm completely samples the space. We then manually alter the algorithm to produce the pentagons which we miss.

4 Topological Dimensionality Reduction

Topological Dimensionality Reduction is the transformation of high dimensional data to low dimensional data. For our data, the moduli space of pentagons is 5 dimensional and forms various surfaces. We used the algorithm from Section 3 to sample enough evenly dispersed data points to fully form the surface.

For the surfaces of genus 2, 3 and 4, we used the the circular coordinates function in DREiMac to produce visualizations. We took the data X in 5 dimensions and computed the persistent cohomology classes $H^1(X; \mathbb{Z})$ by means of Ripser. By Theorem 2.18, these cohomology classes correspond bijectively to homotopy classes of continuous functions from $X \to S^1$. These classes can be considered 1 dimensional circles in the 5 dimensional data; these can be represented with cohomology classes. To obtain the homotopy classes which capture the most topological features in the data, we use the most persistent cohomology classes to assign circular coordinates and plot the surfaces on T^3. If this does not produce the desired surface, we then take linear combinations of the cohomology classes to get better homotopy classes and thus a better representation of our data.

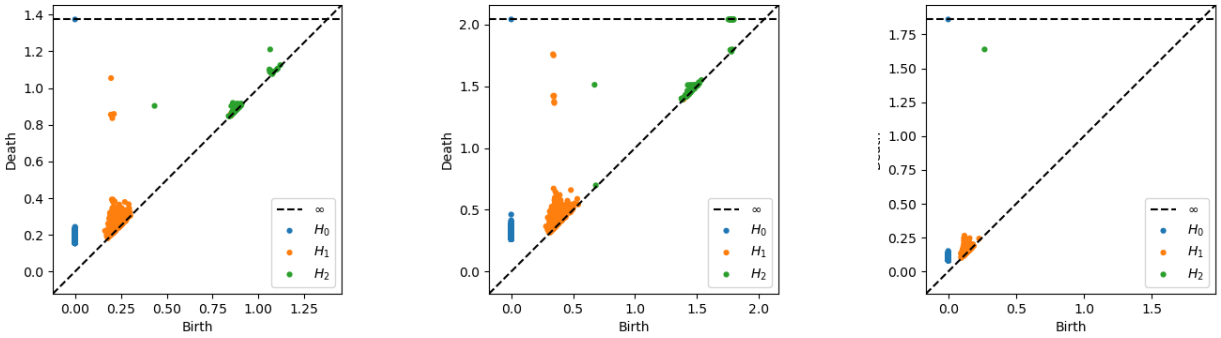
For the surface homeomorphic to a sphere, the above method does not work because the sphere data does not have any persistent 1 cohomology classes in principle. Therefore, we use the DREiMac complex projective coordinates function which applies Eilenberg-MacLane spaces to $H^2(X; \mathbb{Z})$. Once we compute the persistent 2nd cohomology, by Theorem 2.18, the Eilenberg-MacLane space for $H^2(X; \mathbb{Z})$ is $\mathbb{C}P^\infty$. DREiMac uses the most persistent 2nd cohomology class and returns a continuous map f from X to $\mathbb{C}P^\infty$. Because X is a compact set, the image of f is contained in $\mathbb{C}P^N$ for some finite N. There is then a canonical projection from $\mathbb{C}P^N$ to S^2. We then plot the projection of the data onto S^2.

4.1 Persistence Diagrams

Persistence diagrams show the birth and death of cohomology classes computed from data. Persistent cohomology classes are plotted on the upper diagonal where the x-axis is the α in the Rips filtration such that the class was created and the y-axis is the α at which the class vanishes. We present persistence diagrams computed from from the pentagon space samplings which generate genus 3, genus 4, and sphere manifolds.

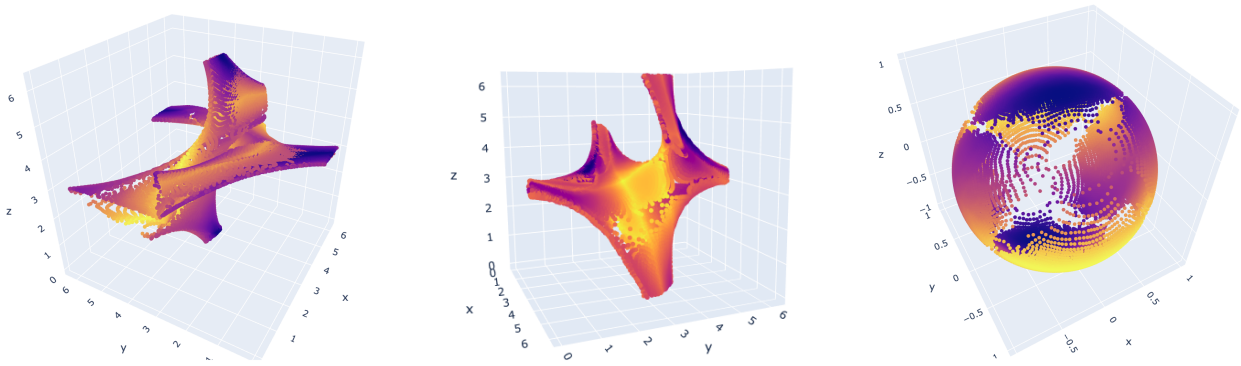
Total number of sampled pentagons for each surface:

- sphere: 38303
- genus 3: 37167
- genus 4: 40735



((a)) Genus 3 ((b)) Genus 4 ((c)) Sphere

Figure 3: Persistence Diagrams



((a)) Genus 3 ((b)) Genus 4 ((c)) Sphere

Figure 4: DREiMac Visualizations

4.2 DREiMac Usage

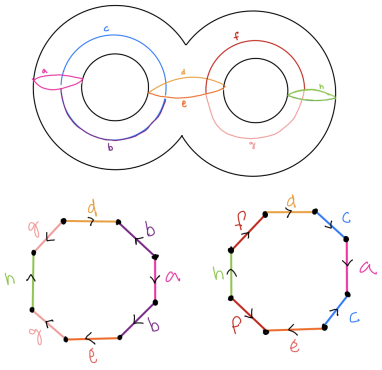
In this section we look at the visualizations the DREiMac package provides us. We show images of the genus 3, genus 4, and sphere data which we plotted using various DREiMac methods.

4.2.1 Genus 4 Data

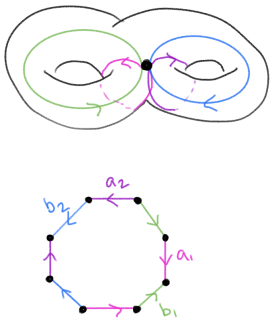
For the genus 4 data, we use the specific edge lengths of $d = (1, 1, 1, 1, 1)$. We take the 4 most persistent cohomology classes in $PH^1(\mathcal{R}(L); \mathbb{Z})$ which comes from a Ripser calculation with 900 landmarks. We then take linear combinations of the four most persistent cohomology classes in descending order from most persistent to least persistent, $[[1, 1, 0, 0], [0, 1, 0, 0], [0, 0, 1, 1]]$. We then apply circular coordinates and plot the data in T^3. This Figure 4(b) shows our TDR visualization of our data.

4.2.2 Genus 3 Data

Genus 3 data is formed from the distinct edge lengths $d = (0.6, 0.6, 0.6, 1, 1)$. Using circular coordinates we have the three cohomology classes chosen out of the six most persistent cohomology classes from a Ripser calculation with 900 landmarks. With the linear combinations we apply circular coordinates and plot in T^3. Figure 4(a) shows our TDR visualization of our data.



((a)) 16-gon model of a genus 2 manifold.



((b)) Octogon model of a genus 2 manifold.

Figure 5: Planar Models

4.2.3 Sphere Data

Sphere data uses complex projective coordinates and edge lengths $d = (0.4, 0.4, 0.4, 0.4, 1)$ Using a single $PH^2(\mathcal{R}(L); \mathbb{Z})$ class instead of 1st cohomology classes from before but still uses the Ripser calculation with 900 landmarks. This Figure 4(c) shows our TDR visualization of our data.

5 Planar Models for Parameterization

Planar models allow us to associate a sampled data point to the point on the 2-manifold on which its lies by creating $\mathbb{R}^2$ coordinates for the manifold. One application of these coordinates is to associate a path on the manifold to a continuous deformation of a pentagon.

5.1 $4g$-gon Model

The $4g$-gon model uses a bouquet of loops on a genus g manifold to create a flat model of the manifold. The model is created by cutting along the loops and unfolding onto the plane. Sides with the same color are identified as in Figure 5(b). While the bouquet of loops uses minimal loops to decompose a surface, computationally it is difficult to compute loops with common intersection because of how distorted the loops must be for complicated surfaces.

5.2 $8g$-gon Model

The $8g$-gon model uses a chain of loops which sprawl over the genus g surface. The $8g$-gon model requires separating the surface into 2 different sides in order to decompose. Unlike the $4g$-model, not all of the vertices in the $8g$-model are identified because they correspond to different intersection points of loops. Sides with the same color are identified as in Figure 5(a).

5.3 Stereographic Model

In the case of a sphere, one can split the sphere into hemispheres and use stereographic projection to lay each hemisphere onto the plane. The result is two disks with each circumference identified to the other disk's.

6 Unfolding Onto Planar Models

6.1 Circular Coordinates on Dimensionality Reduced Data

Using DREiMac, we obtain a visualization of our data plotted on the 3-torus. Our data lies in $\mathbb{C}^5$ and DREiMac calculates a persistent cohomology class which corresponds to a 1-cycle or a circle on our 5-dimensional manifold. Planar models use particular loops in a 2-manifold visualized in 3D to cut along and then unfold. DREiMac allows us to draw loops in manifold valued data by assigning points circular coordinates. After using DREiMac a first time to create our 3D visualization, we run persistent cohomology via DREiMac on the new 3D circular coordinates to get new persistent classes. This second round of circular coordinates gives 1-cycles in 3 dimensions. These circular coordinates represent loops in our 3D model which we manually alter by using different linear combinations of persistent classes.

This process is highly dependent on the quality of the 3D visualization obtained from circular coordinates. In particular, there could not be any topological features which are folded or pinched. Otherwise, the persistent cohomology calculation will not recover every 1st or 2nd cohomology class which is present in the original data. Further, the cohomology classes would not be able to create even cuts across the entire 2-manifold. For persistent cohomology under the Euclidean metric which does not recover a reasonable persistence diagram, we use geodesic distance instead.

6.2 Algorithm

Once we obtain satisfactory cuts on the manifold as circular coordinates, we use an array to associate each data point with its circular coordinates according to each cut. We then run the following algorithm for fixed $k \in \mathbb{N}$:

1. Compute a graph G_k by connecting each point to its k nearest neighbors under the Euclidean metric.
2. For each vertex x in G_k and for each of its k nearest neighbors y, if for any cut the circular coordinates of x and y have difference more than π, remove the edge between them.
3. Compute the geodesic distance matrix of G_k and use multi-dimensional scaling to compute a data set in $\mathbb{R}^2$ with a similar distance matrix.

7 Modeling with Laplacian Eigenfunctions

From the above sections, we have a mapping from X to the manifold which is possible via TDR. We will call $h : X \to \mathcal{M}$ where $X \subseteq \mathbb{C}^5$ and $\mathcal{M}$ is a smooth k-manifold. Now, we are hoping to better model the manifold our data creates. Once we have equations that better model the manifold associated with our data, we can fill in the gaps of our sampling algorithm as well as make predictions about how a pentagon transforms with respect to energy levels. In this section, we present a modeling strategy for the pentagon conformation space which is diffeomorphic to $\mathcal{M}$.

7.1 Setting

To begin our approximation, we start with our mapping h from X to the manifold $\mathcal{M}$. For our data set X, define D to be the cardinality of X. To define a function $H : \mathcal{M} \to \mathbb{C}^5$ such that $H \circ h(x) \approx x$, we first require $H \in L^2(\mathcal{M})$. Since our space lies in higher dimensions, we can decompose $H = (H_1, H_2, H_3, H_4, H_5)$ where $H_i : \mathcal{M} \to \mathbb{C}$. From Theorem 2.14, we can construct a representation of H using the eigenfunctions of the Laplacian notated $\varphi_i^D : \mathcal{M} \to \mathbb{C}$ for $1 \leq d \leq D$ such that:

$$H_i(\theta_1, ..., \theta_k) = \lim_{N \to \infty} \sum_{d=0}^{N} \hat{h}_i(d) \varphi_i^D(\theta_1, ..., \theta_k)$$

where $\hat{h}_d : \mathbb{N} \to \mathbb{C}$ for $i \in \mathbb{N}$ and the θ_i are a parameterization of $\mathcal{M}$.

7.2 Defining Loss Function

In these formulas for H_i, the coefficients are undetermined. In order to approach this problem, recall that our function H will need to be local identity of the function h. To measure how much H inverts h, we will check how closely $H \circ h(x)$ and x lie in $\mathbb{C}^5$ by checking the norm of the difference between the two vectors and taking the square: $\|x - H \circ (x)\|^2$. Take N to be a user defined input indicating the number of parameters taken. To solve for the values of $\hat{h}_d$ we must minimize the following loss function E and for a better perspective, we write our loss function in terms of matrices.

$$\begin{aligned} E(\hat{h}_j) &= \sum_{x_i \in X} \|x - H \circ h(x_i)\|^2 \\ &= \sum_{x_i \in X} \left\| \begin{bmatrix} x_i^1 \\ \cdot \\ \cdot \\ \cdot \\ x_i^5 \end{bmatrix} - \begin{bmatrix} H_1 \circ h(x_i) \\ \cdot \\ \cdot \\ \cdot \\ H_5 \circ h(x_i) \end{bmatrix} \right\|^2 \\ &= \sum_{x_i \in X} \sum_{d=1}^{5} | \, x_i^d - H_d(h(x_i)) \, |^2 \\ &= \sum_{x_i \in X} \sum_{d=1}^{5} | \, x_i^d - \sum_{|n| \leq N} \hat{h}_d(n) \varphi_n^D(x) \, |^2 \end{aligned} \tag{4}$$

The final line results from applying our function representation via Theorem 2.14. From here, we can begin viewing our error function as $\|\text{Mat}(X) - \Phi \cdot \hat{\mathcal{H}}\|$ where:

$$\text{Mat}(X) = \begin{bmatrix} x_i^1 \\ \cdot \\ \cdot \\ \cdot \\ x_D \end{bmatrix} \quad \Phi = \begin{bmatrix} \varphi_1^D(x_1) & \dots & \varphi_N^D(x_1) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \varphi_1^D(x_D) & \dots & \varphi_N^D(x_D) \end{bmatrix} \quad \hat{\mathcal{H}} = \begin{bmatrix} \hat{h}_1(1) & \dots & \hat{h}_D(1) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \hat{h}_1(N) & \dots & \hat{h}_D(N) \end{bmatrix}$$

We see that $\hat{h}_d(n)$ are solutions to $E(\hat{h}_j) \approx 0$ which is a linear least squares problem. A common method for solving such problems is usage of the Moore-Penrose inverse, which in python is notated `numpy.linalg.pinv()`. We apply this function to our defined matrix Φ which we call Φ'. From here, we compute $\hat{\mathcal{H}} = \Phi' \cdot \text{Mat}(X)$ where $\hat{\mathcal{H}}$ is the approximations for $\hat{h}_d(n)$.

7.3 Results and Error

Recall from Klaus and Kojima that the space of pentagons of edge lengths $d = (0.4, 0.4, 0.4, 0.4, 1)$ will form the surface of a sphere. For this section, we solely deal with approximating $h(X)$ which is $\mathbb{S}^2$. For our generated sphere data, $\mathbb{S}^2$, the eigenfunctions of the Laplacian operator are called spherical harmonics notated:

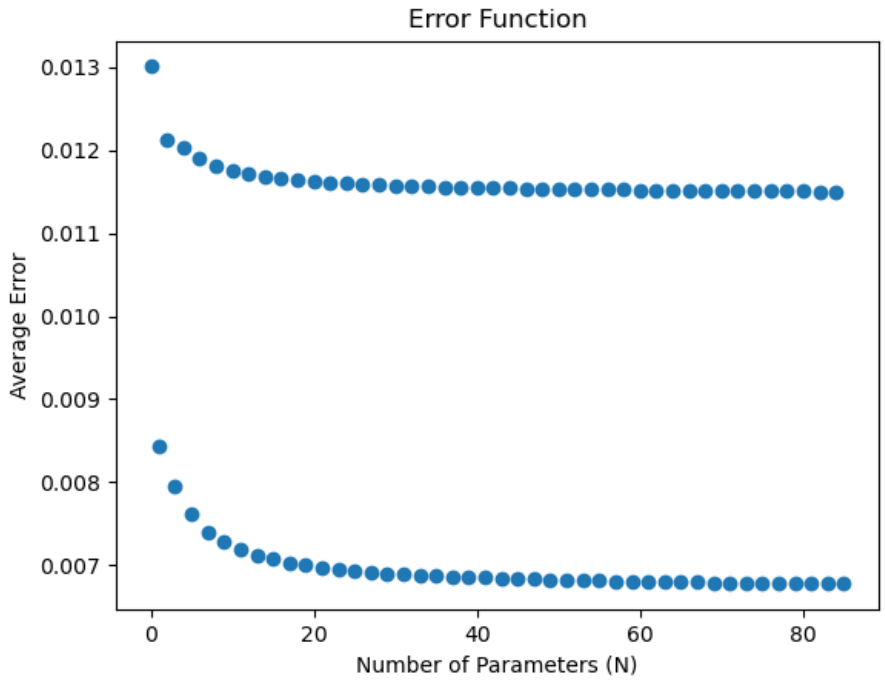
$$\varphi_\ell^m(\theta, \psi) = \sqrt{\frac{2\ell + 1(\ell - m)!}{4\pi(\ell + m)!}} P_\ell^m(\cos\theta) e^{im\psi} \tag{5}$$

where $P_\ell^m(x) = \frac{(-1)^m}{2^\ell \ell!}(1 - x^2)^{m/2} \frac{d^{\ell+m}}{dx^{\ell+m}}(x^2 - 1)^\ell$ [2].

Take $N = 85$ parameters. We follow the method described in the above section: generate our matrix Φ where each element is an evaluated spherical harmonic, apply `numpy.linalg.pinv()` to Φ to generate Φ' and generate $\hat{\mathcal{H}} = \Phi' \cdot \text{Mat}(X)$. Our sphere data X had 38,303 data points and its 3D representation was located along the unit sphere $h(X)$ which has diameter 2. The computed average error came out to be:

$\sqrt{E(\hat{h}_j)} \approx 0.0067\ldots$. We can use the image of H to fill the gaps of our sampled pentagon data.

Below, we present plotting of parameters against the average Euclidean distance between a data point an its corresponding point on the model curve:



8 Discussions and Further Work

8.1 Identifying Underlying Spaces

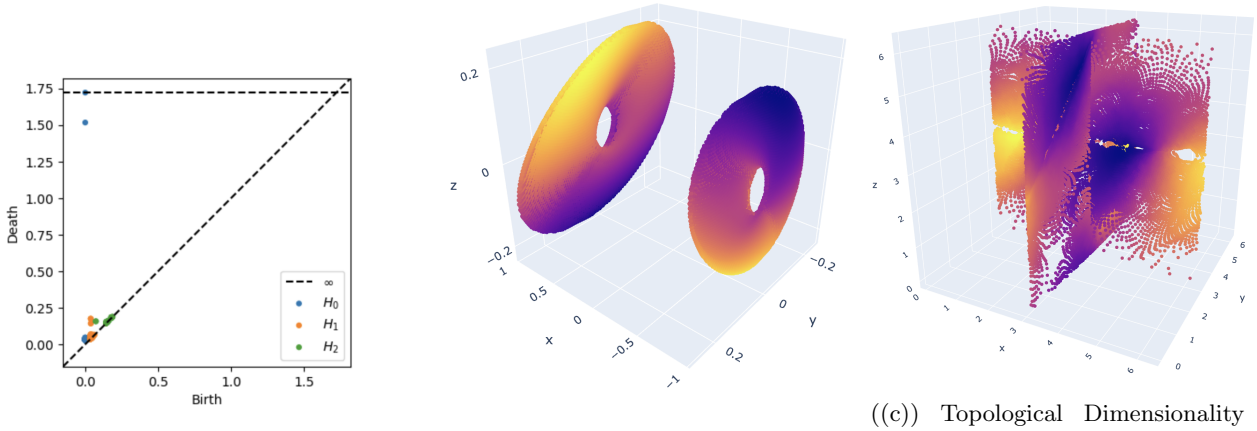
From our generated data, we are curious about what surfaces we can generate from different choices of edge lengths. A question to explore is whether there is a correlation between the complexity of the surface and its edge lengths. Additionally, we see that for edge lengths of smaller size, our sampling algorithm has trouble creating a uniform sampling.

8.2 n-gons and Pentagons in 3-D

Currently, the sampling algorithm we created will input edge lengths and generate a sampling of planar pentagons with these chosen edge lengths. These pentagon spaces may be n-manifolds instead of simple 2-manifolds. Another application we are curious about is creating an algorithm that will uniformly sample the space of pentagons living in 3-D. With the additional dimension, it may be possible to extend our above theorem and idea regarding a 3-D general region where the z_3 vertex can lie.
Being able to generate such pentagons in three dimensions is useful in computational chemistry and biology. Scientists wishing to generate better sampled data of different molecules in physical science have trouble tracking the high energy compounds that are not easy to isolate in a lab.

8.3 Disjoint Tori

An additional space that we sampled was the space of pentagons X with edge lengths $[0.1, 0.1, 1, 1, 1]$ which creates two disjoint tori. Running the data through Ripser, we get the persistence diagram in Figure 6(a).



((a)) Persistence Diagram

((b)) PCA

((c)) Topological Dimensionality Reduction

Figure 6: Disjoint Tori

A quick check reveals that four cohomology classes in $PH^1(X;\mathbb{Z}/q\mathbb{Z})$ and two cohomology classes in $PH^2(X;\mathbb{Z}/q\mathbb{Z})$ persist, which matches the fact that our space is two disjoint tori, where q is a large prime. Furthermore, when running principal component analysis on our data, we retrieve the projection of our data shown in Figure 6(b).

As promising as it is, there is still an issue with the space being flattened. Moving onto attempts at applying TDR, we achieve the visual in Figure 6(c).

Further attempts to separate the two clearly present tori were unsuccessful. The next hope is to separate the data by referencing the PCA done above. From there, the data should display two tori without intersecting with each other.

9 Acknowledgements

We sincerely thank the Simons Laufer Mathematical Science Institute which allow us the space and resources to work on this research. We also wish to thank the Alfred P. Sloan (Sloan grant: G-2020-12592) and National Science Foundation grant (REU Site Award No. 2149642) which sponsor our work and allow us to continue down this path of research. We also acknowledge and thank our professors Dr. Jose Perea and Dr. Mercedes Franco who push and aid us throughout our journey. We express our gratitude.

References

[1] Vin de Silva and Mikael Vejdemo-Johansson. Persistent Cohomology and Circular Coordinates. *Proceedings of the twenty-fifth annual symposium on Computational geometry*, 25:227–236, jun 2009.

[2] *NIST Digital Library of Mathematical Functions*. https://dlmf.nist.gov/, Release 1.1.10 of 2023-06-15. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[3] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 1 edition, 2002.

[4] Burak T Kaynak and et al. Sampling of Protein Conformational Space Using Hybrid Simulations: A Critical Assessment of Recent Methods. *Frontiers in Molecular Biosciences*, 9, 2022.

[5] Stephan Klaus. Möbius Strips, Knots, Pentagons, Polyhedra, and the SURFER Software. *Mathematisches Forschungsinstitut Oberwolfach*, pages 9–11, 2017.

[6] Stephan Klaus and Sadayoshi Kojima. On the Moduli Space of Equilateral Plane Pentagons. *Beiträge zur Algebra und Geometrie*, 60:487–497, 2019.

[7] Mihai Nica. Eigenvalues and Eigenfunctions of the Laplacian. *The Waterloo Mathematics Review*, 1:23–31, 2011.