LAWSON, AUSTIN, M.A. Multi-Scale Persistent Homology. (2016)
Directed by Dr. Greg Bell. 62 pp.

Data has shape and that shape is important. This is the anthem of Topological Data Analysis (TDA) as often stated by Gunnar Carlsson. In this paper, we take a common method of persistence involving the growing of balls of the same size and generalize this situation to one where the balls have different sizes. This helps us to better understand the outlier and coverage problems. We begin with a summary of classical persistence theory and stability. We then move on to generalizing the Rips and Čech complexes as well as generalizing the Rips lemma. We transition into 3 notions of stability in terms of bottleneck distance. For the outlier problem, we show that it is possible to interpolate between persistence on a set with no noise and a set with noise. For the coverage problem, we present an algorithm which provides a cheap way of covering a compact domain.

MULTI-SCALE PERSISTENT HOMOLOGY

by

Austin Lawson

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
2016

Approved by

____
Committee Chair

APPROVAL PAGE

This thesis written by Austin Lawson has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair ______________________
Greg Bell

Committee Members ______________________
Cliff Smyth

____
Maya Chhetri

____
Date of Acceptance by Committee

____
Date of Final Oral Examination

ACKNOWLEDGMENTS

I would like to take this time to express my sincere gratitude to my advisor, Dr. Greg Bell, for his immeasurable patience, suggestions, and hours of discussion. I would also like to thank Dr. Cliff Smyth for his guidance and helpful critiques. Finally, I thank Joshua Martin for the countless hours dedicated to discourse and comprehension of the material.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

In this chapter we will give a historical overview of persistent homology. Persistent homology is a tool used in Topological Data Analysis (TDA) to extract topological information from a data set. Conclusions are then drawn from the information extracted to describe patterns that occur in the data set. The original results in this thesis were obtained through collaboration with Dr. Greg Bell, Dr. Cliff Smyth, Joshua Martin, and James Rudzinski all from UNC Greensboro.

1.1 History

The concept of persistent homology was developed independently through the work of Frosini, Robins, and Edelsbrunner. Frosini's size functions and the theory introduced in 1990 [Fro90], are equivalent to 0-dimensional persistent homology. In 1999, Robins [Rob99] studied homology of sampled spaces and described images of homomorphisms induced by inclusion. This was developed in terms of "persistent holes." In 2000, Edelsbrunner et al [ELZ02] formally introduced persistent homology with an algorithm and a persistence diagram. See Section 2.6. In the years that followed persistent homology blew up as hot topic in mathematics. Research in medical imaging [CBK09, LKC+12], sensor networks [DSG07], sports analysis [Gol14], and many other fields make use of this data analysis tool.

1.2 Motivation and Preliminary Results

We are motivated by two classical problems in network and data analysis: the outlier problem and the coverage problem. The outlier problem deals with detecting noise in data sets and then dealing with it appropriately. Unlike classical persistence (which does not necessarily remove noise) and modern data analysis (which typically removes all things considered to be noise), our method attempts to provide a medium where we do not remove noise, just reduce the importance of it.

Our result yields an interpolation between the two methods. In the coverage problem, we assume that we are given a compact region of interest to cover with predetermined sensor locations. We base the cost of covering a region on the range of the sensors. We then seek to produce ranges that yield a low coverage cost. Again, the classical notion is that all sensors have the same power, but we seek to reduce the power of dense points. In the next section we will cover a few preliminaries and introduce the theory of persistent homology. Our first result, Theorem 3.3, generalizes the classical Rips lemma to the situation with multiple radii. The next result, Theorem 3.10, strengthens a classical idea of stability, which says structure is preserved if points are perturbed slightly. Theorem 3.8 says that the structure is also preserved if the radii are changed slightly. Finally, Theorem 3.12 combines the previous two theorems and says we can preserve structure by moving points and changing radii. The last two results, Theorem 3.14 provides the interpolation and Theorem 3.21 shows that we can produce the ranges that reduce coverage cost. Finally, at the end we present experimental results of our methods.

CHAPTER II

INTRODUCTION TO PERSISTENCE

In this chapter we will develop enough persistence theory to be able to understand the main subject matter. Often we will be working inside the real vector space $\mathbb{R}^n$, though some notions can be defined over more general spaces. A basic understanding of algebra (especially linear algebra) as well as basic topology will be assumed. We will begin with categories.

2.1 Categories

The main problem in topology is to determine when two spaces are homeomorphic. Constructing such equivalences or proving that they do not exist is difficult in general. So we'd like to translate the problem to a simpler algebraic one. The tool that allows this translation can be found within category theory. This section is developed by following [Rot98].

Definition 2.1. A category $\mathcal{C}$ consists of a class of objects denoted $\mathrm{Obj}(\mathcal{C})$ and for any two objects $A, B \in \mathrm{Obj}(\mathcal{C})$ there corresponds a set of morphisms $\mathrm{Hom}(A, B)$, whose elements are denoted as $f : A \to B$,with the following properties.

- The family of morphism sets, $\mathrm{Hom}(\mathcal{C})$, is pairwise disjoint.
- There is a notion of **composition** $\circ : \mathrm{Hom}(A, B) \times \mathrm{Hom}(B, C) \to \mathrm{Hom}(A, C)$ such that,
  - if $f : A \to B$ and $g : B \to C$ are morphisms then $g \circ f : A \to C$ is a morphism and;

– composition is associative when defined, i.e. if $f : A \to B$, $g : B \to C$, and $h : C \to D$ are morphisms then $(h \circ g) \circ f = h \circ (g \circ f)$.

- To every object $A \in \mathrm{Obj}(\mathcal{C})$ there corresponds an **identity** morphism $id_A : A \to A$ so that if $f : A \to B$ and $g : C \to A$ are morphisms we have $f \circ id_A = f$ and $id_A \circ g = g$.

Example 2.2. The following is a list of categories.

- **Top** is the category whose objects are topological spaces and whose morphisms are continuous maps.

- $\mathbf{Top}_*$ has pointed topological spaces (spaces (X, x_0) where $x_0 \in X$ is fixed) as objects and continuous maps $f : (X, x_0) \to (Y, y_0)$ where $f(x_0) = y_0$. as morphisms.

- **Groups** is the category consisting of all groups as objects and whose morphisms are homomorphisms.

- **Sets** is the category whose objects are sets and whose morphisms are functions.

- **Ab** is the category whose objects are abelian groups and whose morphisms are homomorphisms.

- Let F be a field. Then $\mathbf{Vect}_F$ is a category whose objects are finite dimensional vector spaces over F and whose morphisms are linear maps.

- A preordered set $(P, \leq)$ is a set P along with a relation $\leq$ so that for every $x, y, z \in P$ we have $x \leq x$ and $x \leq y, y \leq z \Rightarrow x \leq z$. If P is a preordered set, then $\mathcal{P}$ forms a category whose objects are elements of P. The set of morphisms consists of $x \to y$ whenever $x \leq y$. We call such a category a **preordered category**.

Definition 2.3. If $\mathcal{A}$ and $\mathcal{C}$ are categories then a **functor** $T : \mathcal{A} \to \mathcal{C}$ satisfies

1) $T : \text{Obj}(\mathcal{A}) \to \text{Obj}(\mathcal{C})$ is a function; i.e. $A \in \mathcal{A}$ implies $T(A) \in \mathcal{C}$.

2) If $f : A \to A'$ is a morphism of $\mathcal{A}$ then $T(f) : T(A) \to T(A')$ is a morphism of $\mathcal{C}$ satisfying:

   – whenever $g \circ f$ is defined for two morphisms g and f in $\mathcal{A}$ we have $T(g \circ f) = T(g) \circ T(f)$; and

   – $T(1_A) = 1_{T(A)}$ for every $A \in \mathcal{A}$.

Example 2.4. 1) For a category $\mathcal{C}$ the **identity functor** $J : \mathcal{C} \to \mathcal{C}$ is defined by $J(A) = A$ for $A \in \mathcal{C}$ and $J(f) = f$ for $f \in \text{Hom}(\mathcal{C})$.

2) The **forgetful functor** F : **Top** $\to$ **Sets** assigns to each topological space its underlying set, and assigns each continuous function to itself as a function of sets ("forgetting" continuity). One can define a forgetful functor from any category to **Sets**.

3) Fix an object $A \in \mathcal{C}$. Then $\text{Hom}(A, \cdot) : \mathcal{C} \to$ **Sets** is a functor assigning to each object B the set $\text{Hom}(A, B)$ and to each morphism $f : B \to B'$ it assigns the **induced map** $\text{Hom}(A, f) : \text{Hom}(A, B) \to \text{Hom}(A, B')$ which is defined by $g \mapsto f \circ g$. We denote the induced map by f_*. This functor is called the **covariant Hom functor**.

4) The **fundamental group** is a functor π_1 : **Top**$_*$ $\to$ **Group**.

Definition 2.5. An **equivalence** in a category $\mathcal{C}$ is a morphism $f : A \to B$ for which there is a morphism $g : B \to A$ with $f \circ g = 1_B$ and $g \circ f = 1_A$.

For example, a homeomorphism is an equivalence in **Top** and a group isomorphism is an equivalence in **Group**. The following theorem formalizes the process of turning a topological problem into an algebraic one.

Theorem 2.6. *If $\mathcal{A}$ and $\mathcal{C}$ are categories and $T : \mathcal{A} \to \mathcal{C}$ is a functor then if f is f an equivalence in $\mathcal{A}$ implies that $T(f)$ is an equivalence in $\mathcal{C}$.*

Proof. For a functor T we see $1 = T(1) = T(f \circ g) = T(f) \circ T(g)$ and $1 = T(1) = T(g \circ f) = T(g) \circ T(f)$ Hence $T(f)$ is an equivalence in $\mathcal{C}$. □

Hence, if two topological spaces X and Y are homeomorphic and T : **Top** $\to$ **Group** is any functor, then $T(X)$ and $T(Y)$ are isomorphic.

2.2 Simplicial Complexes

Simplicial complexes provide a computable approximation of many topological spaces. Since many spaces of interest can be approximated in this way, we use simplicial homology (which is more easily computable) to extract information from data sets. We use [Rot98] to develop the simplicial homology theory.

Definition 2.7. A subset A of $\mathbb{R}^n$ is **convex** if for every pair of points $x, x' \in A$ the line segment determined by x and x' is contained in A. In symbols we have $\{tx + (1-t)x' \mid 0 \leq t \leq 1\} \subset A$.

Example 2.8. The following two figures involve an example and non-example of convex sets.

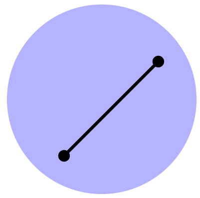


Figure 1. A Convex Set

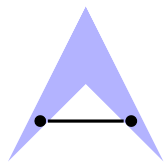


Figure 2. A Non-Convex Set

We'd like to be able to talk about the smallest convex set containing a set of points. But to do this we need to know whether intersections retain convexity.

Theorem 2.9. *Let J be an indexing set. If $\{X_j : j \in J\}$ is a family of convex subsets of $\mathbb{R}^n$ then $\cap X_j$ is also convex.*

Proof. Let $x, x' \in \cap X_j$. Then $x, x' \in X_j$ for every $j \in J$. Since each X_j is convex the line segment determined by x and x' is containted in X_j for every $j \in J$. Therefore the line segment is contained in the intersection of the X_j. Therefore the intersection is convex. □

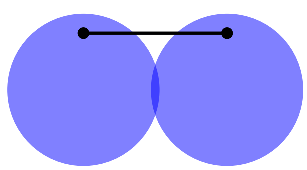


Figure 3. The Union of Convex Sets is Not Convex

Definition 2.10. Suppose $X \subset \mathbb{R}^n$. Then the **convex hull** of X in $\mathbb{R}^n$ is the intersection of all convex sets containing X. This is also called the **convex set** in $\mathbb{R}^n$ **spanned** by X.

We denote the convex hull of the points $p_0, p_1, ..., p_m \in \mathbb{R}^n$ by $[p_0, p_1, ..., p_m]$. We seek to prove that the convex hull is the union of all possible line segments in a set X. To show this we need a precise definition of line segment in Euclidean space.

Definition 2.11. Let $p_0, p_1 ..., p_m \in \mathbb{R}^n$. A **convex combination** these points is a point x with $x = \sum_{i=0}^{m} t_i p_i$ where $\sum t_i = 1$ and $t_i \geq 0$ for all $0 \leq i \leq m$.

Theorem 2.12. *If $p_0, p_1, ..., p_m \in \mathbb{R}^n$ then the convex hull, $[p_0, p_1, ..., p_m]$ is the set of all convex combinations of $p_0, p_1, ..., p_m$.*

Proof. Let S be the set of all convex combinations of $p_0, p_1, ..., p_m$. We will proceed by using the double containment argument.

First we show $[p_0, p_1, ..., p_m] \subset S$. It will suffice to show that S is a convex set containing $\{p_0, p_1, ..., p_m\}$. It is easy to see $p_j \in S$ by setting $t_j = 1$ and $t_i = 0$ where $i \neq j$. Now we show S is convex. Let $\alpha = \sum_{i=1}^{m} a_i p_i$ and $\beta = \sum_{i=1}^{m} b_i p_i$ be convex combinations of $p_0, p_1, ..., p_m$. Then $a_i, b_i \geq 0$ for all i and $\sum_{i=1}^{m} a_i = 1 = \sum_{i=1}^{m} b_i$. Now suppose $0 \leq t \leq 1$. Then $t\alpha + (1-t)\beta = \sum_{i=1}^{m} [ta_i + (1-t)b_i]p_i$. Since $\sum_{i=1}^{m} [ta_i + (1-t)b_i] = 1$ we have a convex combination and hence we have $[p_0, p_1, ..., p_m] \subset S$.

Next we show $S \subset [p_0, p_1, ..., p_m]$. If X is any convex set containing $\{p_0, p_1, ..., p_m\}$, we show by induction on m. If $m = 0$ then $S = \{p_0\}$ and we are done. Suppose $m > 0$. Suppose then $p = \sum_{i=0}^{m} t_i p_i$ is a convex combination. We may assume $t_0 \neq 0$ (otherwise we just relabel) and $t_0 \neq 1$ (else $S = \{p_0\}$). By induction

$$q = \frac{t_1}{1-t_0}p_1 + ... + \frac{t_m}{1-t_0}p_m \in X$$

(since this is a convex combination), and so $p = t_0p_0 + (1-t_0)q \in X$. $\square$

Definition 2.13. A set $\{p_0, ..., p_m\} \subset \mathbb{R}^n$ is **affine independent** if the set $\{p_j - p_0 \mid j = 1, ..., m\}$ is a linearly independent subset of $\mathbb{R}^n$.

Consider $\{p_0, p_1\}$. If $p_1 \neq p_2$ we see this set is affine independent. The set $\{p_0, p_1, p_2\}$ is affine independent if the three points are not collinear. In the same way a four point set is affine independent if they are not coplanar.

Definition 2.14. Let $\{p_0, ..., p_m\}$ be an affine independent subset of $\mathbb{R}^n$. Then $[p_0, ..., p_m]$ is called the $m-$**simplex** with **vertices** $\{p_0, ..., p_m\}$.

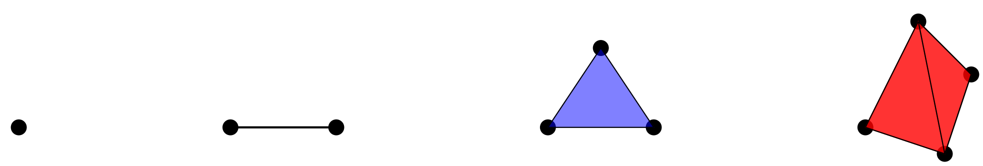


Figure 4. A 0-simplex (Point), 1-simplex (Line Segment), 2-simplex (Triangle), and 3-simplex (Tetrahedron).

Definition 2.15. If $\{p_0, ..., p_m\}$ is an affine independent subset of $\mathbb{R}^n$, then we define the **barycenter** of $\sigma = [p_0, ..., p_m]$ to be the average of its vertices. That is, $b^\sigma = \frac{1}{m+1}(p_0 + p_1 + ... + p_m)$.

The barycenter of a point is itself. The barycenter of a line segment is its midpoint. If we have a triangle or tetrahedron, the barycenter is what we call the center of gravity.

Definition 2.16. Let $[p_0, ..., p_m]$ be an m-simplex. The **face opposite** p_i is

$$[p_0, ..., \hat{p}_i, ..., p_m] = \{\sum t_j p_j \mid t_j \geq 0, \sum t_j = 1 \ and \ t_i = 0\}.$$

(The $\hat{p}_i$ means delete p_i). The **boundary** of $[p_0, ..., p_m]$ is the union of the faces opposite each p_i. A k-**face** is a k-simplex spanned by $k+1$ of the vertices $\{p_0, ..., p_m\}$ We refer to a k-face as a **face** when k is understood.

Definition 2.17. A finite **simplicial complex** K is a finite collection of simplices in some Euclidean space, so that

1) if $\sigma \in K$ and τ is a face of σ then $\tau \in K$; and

2) if $\sigma, \tau \in K$ then $\sigma \cap \tau$ is either empty or a face of σ and τ.

We denote the **vertex set** of K by Vert(K). It is the set of all 0-simplices of K.

Property 1) is often called the downward closure property. Property 2) could be seen as a minimal incidence property.

Example 2.18. The following figure is an example of a simplicial complex.

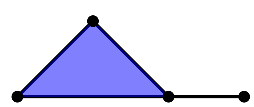


Figure 5. A Simplicial Complex

The following figure shows a failure of property (2).

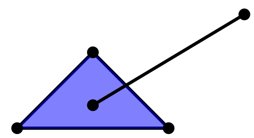


Figure 6. Property 2) Fails

In this figure we have a 0-simplex that intersects a 2-simplex which violates property (2).

Definition 2.19. The **underlying space** of a simplicial complex K is the union of all simplices in K and is denoted by $|K| = \cup_{\sigma \in K} \sigma$.

Definition 2.20. A topological space X is a **triangulable** if there exists a simplicial complex K and a homeomorphism $h : |K| \to X$. The ordered pair (K, h) is called a **triangulation**.

Definition 2.21. Let K and L be a simplicial complexes. A **simplicial map** $\phi : K \to L$ is a function $\phi : \text{Vert}(K) \to \text{Vert}(L)$ such that whenever $\{p_0, p_1, ..., p_m\}$ spans a simplex of K, then $\{\phi(p_0), \phi(p_1), ..., \phi(p_m)\}$ spans a simplex of L. Note that ϕ does not ncessarily have to be one-to-one.

The collection of all finite simplicial complexes with the set of all simplicial maps forms a category $\mathcal{K}$.

Definition 2.22. Let $\{p_0, p_1, ..., p_m\} \subset \mathbb{R}^n$ be affine independent. Let A denote the convex hull of the points. Then a map $T : A \to \mathbb{R}^k$ (for some $k \geq 1$) is said to be **affine** if $T(\sum t_j p_j) = \sum t_j T(p_j)$ where $\sum t_j = 1$ and $t_i \geq 0$.

Theorem 2.23. *If $[p_0, ..., p_m]$ is an m-simplex and $[q_0, ..., q_n]$ is an n-simplex and $f : \{p_0, ..., p_m\} \to \{q_0, ..., q_n\}$ is any function then there exists a unique affine map $T : [p_0, ..., p_m] \to [q_0, ..., q_n]$ so that $T(p_i) = f(p_i)$ for every $0 \leq i \leq m$*

Proof. Define $T(\sum t_j p_j) = \sum t_j f(p_j)$ where $t_j \geq 0$ and $\sum t_j = 1$. T is clearly an affine map. To show uniqueness suppose T' also satisfies this. Then with $\sum t_j p_j$ being a convex combination, linearity gives $T'(\sum t_j p_j) = \sum T'(t_j p_j) = \sum t_j T'(p_j) = \sum t_j f(p_j) = T(\sum t_j p_j)$. □

Theorem 2.24. *A simplicial map $\phi : K \to L$ induces a continuous map $|\phi| : |K| \to |L|$.*

Proof. We will define $|\phi|$ as follows. For each $\sigma \in K$, define $f_\sigma : \sigma \to |L|$ to be the affine map determined by ϕ restricted to $\mathrm{Vert}(\sigma)$ (by previous theorem). By condition 2) in the definition of a simplicial complex we have that the f_σ must agree on overlaps. Then we apply the gluing lemma [Rot98] to conclude that $|\phi|$ is continuous map from $|K|$ to $|L|$. □

This theorem implies that $|\ | : \mathcal{K} \to \mathrm{Top}$ by taking a simplicial complex $K \mapsto |K|$ and a simplicial map $\phi \mapsto |\phi|$ is a functor.

We can define a partial order on a simplicial complex K by saying $\sigma \leq \tau$ if σ is a face of τ. So, $\sigma < \tau$ if $\mathrm{Vert}(\sigma) \subsetneq \mathrm{Vert}(\tau)$.

Definition 2.25. If σ is a simplex let b^σ denote the barycenter of σ. If K is a simplicial complex define the **barycentric subdivision**, $\mathrm{Sd}\, K$ of K, to be the simplicial complex with

$$\mathrm{Vert}(\mathrm{Sd}\, K) = \{b^\sigma \mid \sigma \in K\}$$

and with simplices $[b^{\sigma_0}, ..., b^{\sigma_q}]$ where $\sigma_0 < \sigma_1 < ... < \sigma_q \in K$.

Figure 7. A Simplicial Complex K (Left) and $\mathrm{Sd}\,K$ (Right)

Definition 2.26. If σ is an m-simplex then we say the **dimension** of σ is m. Furthermore the **dimension** of a simplicial complex K is $\max\{\dim(\sigma) \mid \sigma \in K\}$

Definition 2.27. For any $q \geq -1$ the q**-skeleton**, $K^{(q)}$ of a simplicial complex K is the simplicial complex consisting of all simplices with dimension no greater than q. That is, $K^{(q)} = \{\sigma \in K \mid \dim(\sigma) \leq q\}$ As a convention we set $\dim \emptyset = -1$.

Now that we have a concrete notion of a simplicial complex, we will abstract the idea. In this manner we will be able to more easily define homology groups. The idea is to work with these objects abstractly, then worry about fitting them into a nice space later.

2.3 Abstract Simplicial Complexes

Definition 2.28. Let V be a finite set. An **abstract simplicial complex** K is a family of nonempty subsets of V, called **simplices** so that

1) $v \in V$ implies $\{v\} \in K$; and

2) $\sigma \in K$ and $\sigma' \subset \sigma$ implies $\sigma' \in K$.

Definition 2.29. Let K and L be a abstract simplicial complexes. A **simplicial map** $\phi : K \to L$ is a function $\phi : \mathrm{Vert}(K) \to \mathrm{Vert}(L)$ such that whenever $\{v_0, v_1, ..., v_m\}$ spans a simplex of K, then $\{\phi(v_0), \phi(v_1), ..., \phi(v_m)\}$ spans a simplex of L. Note that ϕ is not necessarily one-to-one.

Note that the definition of face is the same and hence the partial order is the same.

Definition 2.30. The **barycentric subdivision** $\operatorname{Sd} K$ of an abstract simplicial complex K is defined as follows: $\operatorname{Vert}(\operatorname{Sd} K) = \{\sigma \in K\}$; we define a simplex of $\operatorname{Sd} K$ to be a set $\{\sigma_0, ..., \sigma_q\}$ with $\sigma_0 < ... < \sigma_q \in K$.

Observe that all abstract simplicial complexes and simplicial maps form a category $\mathcal{K}^a$ and every simplicial complex K gives rise to an abstract simplicial complex K^a with the same vertex set.

Theorem 2.31. *There is a functor $u : \mathcal{K} \to \mathcal{K}^a$ such that $K \cong u(K^a)$ for all $K \in \mathcal{K}$ and $L \cong (u(L))^a$ for all $L \in \mathcal{K}^a$*

Proof. Let $L \in \mathcal{K}^a$ and let $V = \operatorname{Vert}(L) = \{v_0, ..., v_n\}$. The standard n-simplex Δ^n is a simplex with vertices $\{e_0, ..., e_n\} \subset \mathbb{R}^n$ where the e_i are the standard basis vectors. If $s = \{v_{i_0}, ..., v_{i_q}\}$ is a q-simplex in L, define $|s| = [e_{i_0}, ..., e_{i_q}]$ to be the q-simplex spanned by the mentioned vertices. Finally, we define $u(L)$ to be the family of all $|s|$ for $s \in L$.

Now suppose $\phi : L \to L'$. Then $u(\phi) : u(L) \to u(L')$ corresponds to the obvious simplicial map. From here it is easy to see that u is a functor and that the isomorphisms exist. □

Definition 2.32. A **geometric realization** of an abstract simplicial complex L is a space homeomorphic to $|u(L)|$.

For $K \in \mathcal{K}$, $|K|$ is a geometric realization. Theorem 2.31 is an important one as it will allow us to not worry about making the distinction between simplicial complexes and abstract simplicial complexes. From here on we shall drop the adjective "abstract." We will also not distinguish between the categories $\mathcal{K}$ and $\mathcal{K}^a$ and we will just write $\mathcal{K}$. We are heading towards defining homology but before we move on we shall cover two important simplicial complexes and a relationship between them. This will be the focus of our main matter later. In $\mathbb{R}^n$, given $\epsilon > 0$ the closed ball around x of radius ϵ is $B_\epsilon(x) = \{y \in \mathbb{R}^n \mid d(x, y) \leq \epsilon\}$. The material for the rest of this section can be found in [EH10a].

Definition 2.33. Suppose U is a collection of sets. Then the **nerve** of U is the abstract simplicial complex $\mathcal{N}(U) = \{\sigma \subset U \mid \bigcap \sigma \neq \emptyset\}$.

We will consider our set as a set of open balls. Then by taking the nerve we obtain a special type of simplicial complex known as the Čech complex.

Definition 2.34. Suppose $X \subset \mathbb{R}^n$ is finite and $\epsilon > 0$. Let $B = \{B_\epsilon(x) \mid x \in X\}$ The **Čech complex** at scale ϵ is the abstract simplicial complex $\check{C}_\epsilon(X) = \mathcal{N}(B)$.

From the definition we see that $\mathcal{N}(B) = \{\tau \subset B \mid \bigcap_{B_\epsilon(x) \in \sigma} B_\epsilon(x) \neq \emptyset\}$. Thus we can identify the simplex τ with the simplex σ whose vertices are the centers of the balls in τ in this way we can write $\check{C}_\epsilon(X) = \{\sigma \subset X \mid \bigcap_{x \in \sigma} B_\epsilon(X) \neq \emptyset\}$.

The Čech complex is hard to compute because the condition requires us to test that a collection of balls has a common intersection. The difficulty increases greatly as dimension increases. Thus, we seek something that is quick to compute and that approximates the Čech complex well enough. The next complex we look at satisfies this and is known as the Vietoris-Rips complex or just Rips complex. We could easily define the Rips complex as the flag of the Čech complex, however, this would not be useful computationally. So we define it in the following way.

Definition 2.35. Let $X \subset \mathbb{R}^n$ be finite and $\epsilon > 0$. The **Rips complex** at scale ϵ is the set

$$R_\epsilon(X) = \{\sigma \subset X \mid d(x, y) \leq 2\epsilon \ \forall x, y \in \sigma\}.$$

Example 2.36. We will illustrate the difference between the Rips and Čech complexes here.

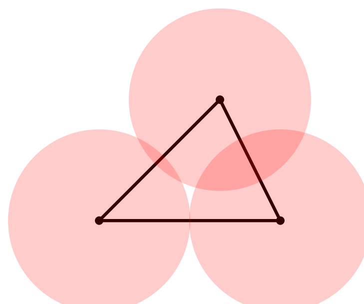


Figure 8. The Union of the Segments is the Čech Complex

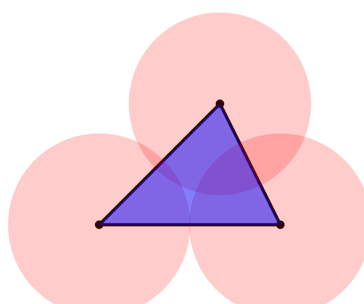


Figure 9. The Rips Complex Includes the 2-simplex

If we were to enlarge the radius of balls in the Čech complex just a little we would obtain another Čech complex that contains the Rips complex at the original scale. This actually happens for every Čech complex and hence we get a result that tells us that the Rips complex approximates the Čech complex. The following theorem can be found in [DSG07].

Theorem 2.37 (Rips Lemma)**.** *Suppose $X \subset \mathbb{R}^d$ is finite and $\epsilon, \epsilon' > 0$ so that $\epsilon \geq \epsilon' \cdot \sqrt{\frac{2d}{d+1}}$. Then $R_{\epsilon'} \subset \check{C}_\epsilon \subset R_\epsilon$*

We will omit this proof as we will give a proof of our more general result later in Theorem 3.3.

2.4 Homology and Betti Numbers

In this section we seek to understand simplicial homology. As the name suggests this will be a crucial tool in persistent homology. In particular, we will use simplicial homology to make an inference as to the shape of a data set. We will develop homology first as groups and then mention a generalization to modules.

Definition 2.38. An **oriented** simplicial complex K is a simplicial complex with a partial order on $\text{Vert}(K)$ whose restriction to the vertices of any simplex is a linear order.

Definition 2.39. Let $k \geq 0$ A k-**chain** $C_k(K)$ on an oriented simplicial complex K is the abelian group with the following presentation:

- Generators: all $(k+1)$-tuples $(p_0, \cdots, p_k)$ where $p_i \in \text{Vert}(K)$ such that $\{p_0, ..., p_k\}$ spans a simplex in K;
- Relations:
  1) $(p_0, ..., p_k) = 0$ if a vertex is repeated;
  2) $(p_0, ..., p_k) = sgn(\pi)(p_{\pi(0)}, \cdots, p_{\pi(k)})$ where π is a permutation of $\{0, ..., q\}$ and $sgn(\pi) = \pm 1$ depending on the parity of π.

Lemma 2.40. *If K is an oriented simplicial complex of dimension m then*

1) $C_k(K)$ is a free abelian group with basis all symbols $\langle p_0, ..., p_k \rangle$, where $\{p_0, ..., p_k\}$ spans a k-simplex in K and $p_0 < ... < p_k$. Moreover, $\langle p_{\pi(0)}, ..., p_{\pi(k)} \rangle = sgn(\pi)\langle p_0, ..., p_k \rangle$.

2) $C_k(K) = 0$ for all $k > m$

Thus, we have separated the simplicial complex K into its k-simplex pieces. We'd like to know how each k-chain relates to say the $k-1$ and $k+1$-chains. For this we have the boundary maps.

Definition 2.41. Suppose $C_k(K)$ and $C_{k-1}(K)$ are chain spaces. Then the *kth* **boundary map** is the homomorphism $\partial_k : C_k(K) \to C_{k-1}(K)$ given by

$$\partial_k(\langle p_0, p_1, ..., p_k \rangle) = \sum_{i=0}^{k} (-1)^i \langle p_0, p_1, ..., \hat{p}_i, ..., p_k \rangle.$$

Theorem 2.42. *Given the chains $C_{k+1}(K)$, $C_k(K)$ and $C_{k-1}(K)$. We have $\partial_k \partial_{k+1} =$ 0.*

Proof. The proof is just a routine computation which we show now.

$$\partial_k(\partial_{k+1}(\langle p_0, ..., p_{k+1}\rangle) = \partial_k\left(\sum_{i=0}^{k+1}(-1)^i\langle p_0, ..., \hat{p}_i, ..., p_{k+1}\rangle\right)$$
$$= \sum_{i=0}^{k+1}\partial_k((-1)^i\langle p_0, ..., \hat{p}_i, ..., p_{k+1}\rangle)$$
$$= \sum_{i=0}^{k+1}\left(\sum_{j<i}(-1)^{i+j}\langle p_0, ...\hat{p}_j, ..., \hat{p}_i, ..., p_{k+1}\rangle\right.$$
$$\left.+\sum_{j>i}^{k+1}(-1)^{i+j-1}\langle p_0, ..., \hat{p}_i, ...\hat{p}_j, ..., p_{k+1}\rangle\right).$$

Notice that for $j > i$ we have $j-1$ in the exponent. This is due to the fact that we are looking at a k simplex hence $j \leq k$. Further notice that for each simplex obtained while $j < i$ there is one of opposite orientation arising while $j > i$ hence everything cancels and we are left with $\partial_k(\partial_{k+1}(\langle p_0, ..., p_{k+1}\rangle) = 0$. □

Definition 2.43. The collection of all chains over K together with the boundary maps is called a **chain complex** and is denoted $C_\bullet(K)$.

Let K and L be oriented simplicial complexes. If $\phi : K \to L$ is a simplicial map then we will define the map $\phi_\# : C_q(K) \to C_q(L)$ by $\phi_\#(\langle p_0, p_1, ..., p_q\rangle) = \langle\phi(p_0), \phi(p_1), ..., \phi(p_q)\rangle$.

Lemma 2.44. *$\phi_\#\partial_k = \partial_k\phi_\#$ for each $k \geq 0$*

Proof. Suppose σ is a k-simplex. Then

$$\begin{aligned}\phi_{\#}\partial_k(\sigma) &= \phi_{\#}\left(\sum_{i=0}^{k}(-1)^i\langle p_0, p_1, ..., \hat{p}_i, ..., p_k\rangle\right)\\ &= \sum_{i=0}^{k}(-1)^i\phi_{\#}(\langle p_0, p_1, ..., \hat{p}_i, ..., p_k\rangle)\\ &= \sum_{i=0}^{k}(-1)^i\langle\phi(p_0), \phi(p_1), ..., \phi(\hat{p_i}), ..., \phi(p_q)\rangle.\end{aligned}$$

On the other hand, $\partial_k\phi_{\#}(\sigma) = \partial_k(\langle\phi(p_0), \phi(p_1), ..., \phi(p_q)\rangle)$
$= \sum_{i=0}^{k}(-1)^i\langle\phi(p_0), \phi(p_1), ..., \phi(\hat{p_i}), ..., \phi(p_q)\rangle$. □

We will now look at cycles and boundaries, which will allow us to define homology.

Definition 2.45. Given a chain complex $C_\bullet(K)$ we define the k-cycles $Z_k(K) = \ker\partial_k$. We define the k-boundaries $B_k(K) = \mathrm{im}\partial_{k+1}$.

Since the boundary maps are homomorphisms, $Z_k(K)$ and $B_k(K)$ are both subgroups of the abelian $C_k(K)$. Furthermore since $\partial_k\partial_{k+1} = 0$ we see $B_k(K)$ is a normal subgroup of $Z_k(K)$. Hence we get the following definition.

Definition 2.46. The *kth* **homology group** is $H_k(K) = Z_k(K)/B_k(K)$.

Theorem 2.47. *For each $k \geq 0$, $H_k : \mathcal{K} \to \mathrm{Ab}$ is a functor.*

Proof. We already know how H_k deals with objects of $\mathcal{K}$. So, for a simplicial map $\phi : K \to L$ we define $H_k(\phi) = \phi_* : H_k(K) \to H_k(L)$ by $\phi_*(z + B_k(K)) = \phi_{\#}(z) + B_q(L)$. From here it is a routine check to see that H_k is a functor. □

Although we developed homology in terms of groups, one can actually speak more generally about homology over an R-module (left or right) where R is a ring with unit [Rot08]. By writing $[p_0, ..., p_m]$ as the m-simplex spanned by the presented vertices we obtain the following definitions.

Definition 2.48. Let R be a ring with unit and K be an oriented simplicial complex. Then a k-**chain module** $C_k(K)$ is a collection of linear combinations (which are finite sums) of the form $\sum_i t_i\sigma_i$ where $t_i \in R$ and $\sigma_i \in K$. Any oriented k-simplex is equal to -1 times the simplex of opposite orientation. That is $[p_0, ..., p_m] = -[p_m, ..., p_0]$.

Definition 2.49. Suppose $C_k(K)$ and $C_{k-1}(K)$ are chain spaces. Then the *kth* **boundary map** is a linear map $\partial_k : C_k(K) \to C_{k-1}(K)$ given by

$$\partial_k([p_0, p_1, ..., p_k]) = \sum_{i=0}^{k}(-1)^i[p_0, p_1, ..., \hat{p}_i, ..., p_k].$$

Theorem 2.50. *Given the chain spaces $C_{k+1}(K)$, $C_k(K)$ and $C_{k-1}(K)$. We have $\partial_k\partial_{k+1} = 0$.*

Definition 2.51. The collection of all chains over K together with the boundary maps is called a **chain complex** and is denoted $C_\bullet(K)$.

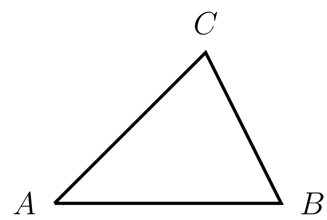
Definition 2.52. Given a chain complex $C_\bullet(K)$ we define the k-cycles $Z_k(K) = \ker \partial_k$. We define the k-boundaries $B_k(K) = \mathrm{im}\partial_{k+1}$.

Linearity of the boundary maps give us $B_k(K) \subseteq Z_k(K) \subseteq C_k(K)$ as submodules. Hence we finally have the following definition.

Definition 2.53. Let R be a ring and K be an oriented simplicial complex. Then the *kth* **homology module** is $H_k(K) = Z_k(K)/B_k(K) = \ker \partial_k/\mathrm{im}\partial_{k+1}$.

In this way, by setting $R = \mathbb{Z}$, we immediately obtain the definition of homology groups. Often we will take R to be a field which creates homology vector spaces. One field commonly used for its computability is $\mathbb{Z}/2\mathbb{Z}$. That is the field of two elements. By using this field we forgo orientation which allows us to easily compute homology.

Example 2.54. Let us compute the homology groups of a triangle that is not filled in.



For ease we will assume our coefficients are coming from $\mathbb{Z}_2$. To begin we compute the chain spaces $C_2 = 0,\ C_1 = span\{[AB], [AC], [BC]\}$, and $C_0 = span\{[A], [B], [C]\}$. Next we compute the rank and nullity of the boundary maps. $rk(\partial_2) = 0 = Nul(\partial_2)$ this one is easy since we have a 0 space. In a similar manner we have $rk(\partial_0) = 0$ and $Nul(\partial_0) = 3$ Now let us consider the matrices arising from the remaining boundary map.

$$\partial_1 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

This matrix is obtained by labeling the columns as the edges and the rows as the vertices then placing a 1 wherever a vertex meets an edge. After a quick reduction we see $rk(\partial_1) = 2, Nul(\partial_1) = 1$ Thus, we see $H_1 = Z_1/B_1 \cong \mathbb{Z}_2$, and $H_0 = Z_0/B_0 \cong \mathbb{Z}_2$

The following theorem can be found in [EH10b]

Theorem 2.55 (Nerve Theorem)**.** *If U is a union of convex sets then $H_k(\mathcal{N}(U)) \cong H_k(U)$.*

We note that the usual Nerve theorem is stated in terms of homotopy, but since we do not use homotopy, and homotopy type implies isomorphic homology, we just use the version that fits our case. The Nerve Theorem tells us that given a nice collection sets (such as a finite collection of closed or open balls) the topological information encoded in the union of the collection is also encoded in the nerve of the collection. This tells us that the Čech complex accurately represents the space formed by the union of balls.

Definition 2.56. The *kth* **Betti number**, β_k is the rank of the kth homology module.

Since simplicial complexes are finite, each k-chain has a basis which consists of exactly all of the k-simplices in K. Now consider the quotient map $\phi : Z_k \to Z_k/B_k$. By the Rank-Nullity Theorem we know $\dim Z_k = \dim(\mathrm{im}\phi) + \dim(\ker \phi)$. Note that $\ker \phi = B_k$ and $\mathrm{im}\phi = H_k$. Hence we obtain $\dim(H_k) = \dim(Z_k) - \dim(B_k)$. Informally, the Betti numbers count the number of k-dimensional holes a space has.

2.5 Persistence Modules

Imagine a set of points in some Euclidean space. Place balls around these points and allow them to grow. At each step compute the Čech complex. Notice that as the balls grow, the simplices that appear at scale ϵ are present at each scale $\epsilon' > \epsilon$. This notion of an increasing chain of inclusions is called a filtration, which we now define.

Definition 2.57. Let K be a simplicial complex. A **filtration** of K is a totally ordered set of subcomplexes of K so that

$$\emptyset = K_{-1} \subset K_0 \subset K_1 \subset ... \subset K_n = K.$$

We note that each K_i is itself a simplicial complex hence we can create an increasing (by embedding) sequence of chain complexes.

$$C_\bullet(K_0) \hookrightarrow C_\bullet(K_1) \hookrightarrow ... \hookrightarrow C_\bullet(K_n).$$

The embedding is given by the map induced by the inclusion of $C_k(K_i)$ into $C_k(K_{i+1})$ for each k and i. From this we see that this induces a linear map η_k^i : $H_k(K_i, R) \to H_k(K_{i+1}, R)$.

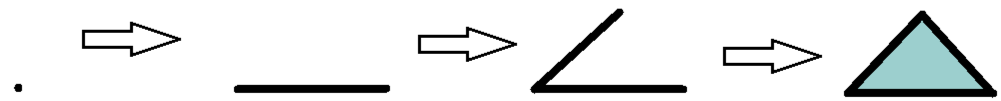


Figure 10. This is a Filtration.

Definition 2.58. Given a filtration of a simplicial complex K and a commutative ring with unit R, the *kth* **persistence module** $\mathcal{H}_k$ of K over R is the family of the kth homology modules $H_k(K_i, R)$ together with the induced linear maps between them, $\eta_k^i : H_k(K_i, R) \to H_k(K_{i+1}, R)$.

As the name suggests, persistence modules have a module structure. In fact, they can be given a graded module structure over the polynomial ring $R[x]$. This will allow us to apply a standard decomposition theorem that us to define diagrams and barcodes, which are topological summaries of data from which we draw conclusions.

Definition 2.59. A **graded ring** R is a ring that decomposes as a direct sum of abelian groups $R = \bigoplus_{i \in \mathbb{Z}_{\geq 0}} R^i$ so that $x \in R^i$ and $y \in R^j$ implies $xy \in R^{i+j}$. Any element $x \in R^i$ is said to be **homogeneous of degree** i. Finally, if $I \subset R$ is a two sided ideal and $I = \bigoplus_{i \in \mathbb{Z}_{\geq 0}} I \cap R^i$, then I is called a **graded ideal**.

Definition 2.60. A left **graded module** is a left module M over a graded ring R such that $M = \bigoplus_{i \in \mathbb{Z}_{\geq 0}} M^i$ and $R^i M^j \subset M^{i+j}$. M is **non-negatively graded** if $M^i = 0$ whenever $i < 0$.

As mentioned before, a persistence module can be given a graded module structure over the polynomial ring $R[x]$. That is, $\mathcal{H}_k = \bigoplus_{i=0}^{\infty} H_k^i$ where the action of x is given by $x \cdot \sum_{i=0}^{\infty} m^i = \sum_{i=0}^{\infty} \eta_k^i(m^i)$ for $m^i \in H_k^i$. This is no large leap as we think of the action intuitively as a shift up by one unit. This means that the action is the link that connects homologies across different complexes in the filtration. The following theorem is a generalization of Theorem 5 on page 463 in [DF04].

Theorem 2.61. *[Structure for Theorem for Finitely Generated Graded Modules over a PID] Let R be a graded Principal Ideal Domain (P.I.D.) and let M be a finitely generated graded R-module. Then M is isomorphic to the direct sum of finitely many cyclic modules $M \cong \bigoplus_{i=1}^{m} \Sigma^{b_i} R/(a_i) \oplus \bigoplus_{j=1}^{r} \Sigma^{c_j} R$ where Σ^k denotes a k-shift upward in grading and where $b_i, c_i \in \mathbb{N}$ and $a_i \in R$ with a_i divides $a_i + 1$ for all $1 \leq i \leq m-1$. The isomorphism is unique up to reordering.*

Proof. The proof of this is nearly identical to the proof of Theorem 5 on page 463 of [DF04] which we now go through.

Since M is finitely generated and graded we can find homogeneous elements $\{x_1, ..., x_n\}$ that generate M. Recall that homogeneous means for each i we have $x_i \in M^j$ for some j. Now let R^n denote the free R-module with basis of homogeneous elements $\{b_1, ..., b_n\}$ of the same grade as the x_i. Define the map $\pi : R^n \to M$ by $\pi(b_i) = x_i$. It is clear that π is a surjective homomorphism; hence we apply the first isomorphism theorem to obtain $M \cong R^n / \ker \pi$. By Theorem 4 page 462 of [DF04] there exists a basis $\{y_1, ..., y_n\}$ of homogeneous elements of R^n so that $\{a_1 y_1, ..., a_m y_m\}$

is a basis for $\ker \pi$ for some elements $a_1, ..., a_m$ with $a_1|a_2|\cdots|a_m$. Hence we see that we have

$$M \cong R^n / \ker \pi = \left(\bigoplus_{i=1}^{n} Ry_i\right) / \left(\bigoplus_{i=1}^{m} Ra_iy_i\right).$$

Now we define a map

$$\phi : \bigoplus_{i=1}^{n} Ry_i \to \bigoplus_{i=1}^{m} \Sigma^{deg(y_i)} R/(a_i) \oplus \bigoplus_{j=1}^{n-m} \Sigma^{deg(y_j)} R$$

where Σ^k denotes a k-shift upward in grading and where

$$(\alpha_1 y_1, ..., \alpha_n y_n) \mapsto (\alpha_1 \text{mod}(a_1), ..., \alpha_m \text{mod}(a_m), \alpha_{m+1}, ..., \alpha_n).$$

It is clear that $\ker \phi = \{(\alpha_1 y_1, ..., \alpha_m y_m, 0, ..., 0) \mid a_i | \alpha_i \ \forall \ i\}$, but this is exactly $\ker \pi$. Hence we have then

$$M \cong \bigoplus_{i=1}^{m} \Sigma^{deg(y_i)} R/(a_i) \oplus \bigoplus_{j=m+1}^{n} \Sigma^{deg(y_j)} R.$$

Now by calling $r = n - m$, $deg(y_i) = b_i$ for $1 \leq i \leq m$, and $deg(y_j) = c_i$ for $m + 1 \leq j \leq n$ we get exactly the form we seek. Uniqueness follows from the divisibility property of the $a_i's$ □

The following theorem is just a direct application of the graded structure theorem to persistence modules.

Theorem 2.62 (Structure Theorem for Persistence Modules)**.** *Suppose $\mathcal{H}_k$ is a persistence module over the polynomial ring $R[x]$ as above. Then,*

$$\mathcal{H}_k = \bigoplus_{i=1}^{n} (x^{a_i}) \oplus \bigoplus_{j=1}^{m} (x^{b_j})/(x^{c_j})$$

where $a_i, b_j, c_j \in \mathbb{R}_+$ are non -negative real numbers for all i, j.

Really what we are saying is that given a persistence module and a nice polynomial ring we are able to decompose the persistence module into free and torsion portions. The left side (free portion) will come in at step a_i in the filtration and will persist for all future parameters whereas the right portion (torsion elements) corresponds to the homology generators that come in to existence at b_j and die at $b_j + c_j$.

2.6 Persistence Diagrams and Barcodes

We would like to know that if we have two (possibly) different samples from the same space, the persistence diagrams are "close." We will define a distance between diagrams and prove that diagrams are stable under small perturbations with respect to this distance. This entire section follows the development in [CSEH07] which was the first paper to provide a proof of the stability theorem.

Definition 2.63. Formally, a **multi-set** $\mathcal{A}$ is the graph of a function $\mu : A \to \mathbb{N} \cup \{+\infty\}$ where A is a set. Elements of $\mathcal{A}$ are of the form $(a, \mu(a))$. We call $\mu(a)$ the **multiplicity** of $a \in A$.

Informally, a multi-set is simply a collection of objects that are allowed to appear multiple times. For example the collection $\{1, 1, 2, 3, 4, 4, 4, 5\}$ can be written formally as the multi-set $\{(1, 2), (2, 1), (3, 1), (4, 3), (5, 1)\}$.

Definition 2.64. Let $\mathbb{F}$ be a field and suppose $\mathcal{H}_k$ is a persistence module over $\mathbb{F}[x]$ with its decomposition

$$\mathcal{H}_k = \left(\bigoplus_i x^{a_i} \right) \oplus \left(\bigoplus_j x^{b_j}/x^{c_j} \right).$$

We define the **persistence barcode**, or just **barcode**, $\mathcal{B}_k$ to be a multi-set of intervals in $\bar{\mathbb{R}}_+$ with elements of the form $[a_i, \infty]$ and $[b_j, b_j + c_j]$. Where $a_i, b_j, c_j \in \mathbb{R}_+$.

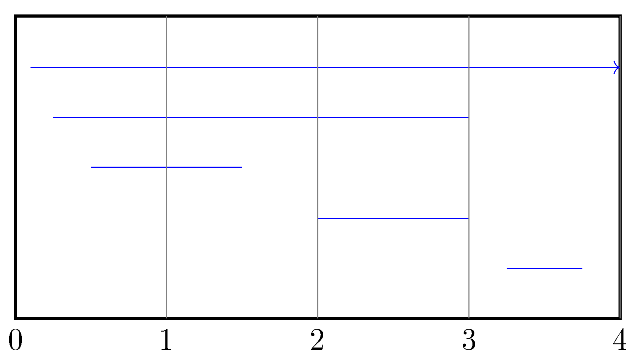


Figure 11. A Barcode

Definition 2.65. Let $\mathbb{F}$ be a field and suppose $\mathcal{H}_k$ is a persistence module over $\mathbb{F}[x]$ with its decomposition as above. We define the **persistence diagram** D_k to be a multi-set of points in $\bar{\mathbb{R}}^2_+$ of the form (a_i, ∞) and $(b_j, b_j + c_j)$, where $a_i, b_j, c_j, \in \mathbb{R}_+$, union the diagonal $\Delta = \{(x, x) : x \geq 0\}$ counted with infinite multiplicity.

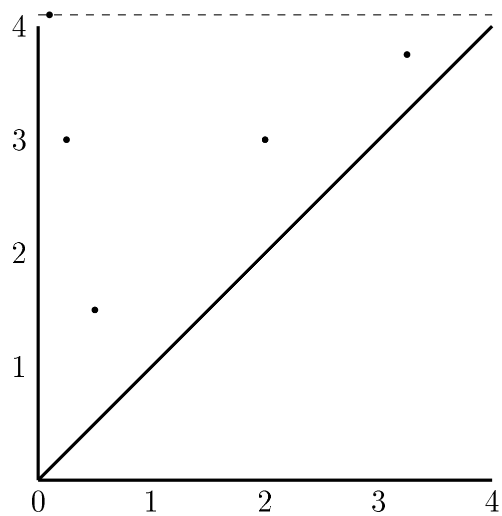


Figure 12. A Persistence Diagram

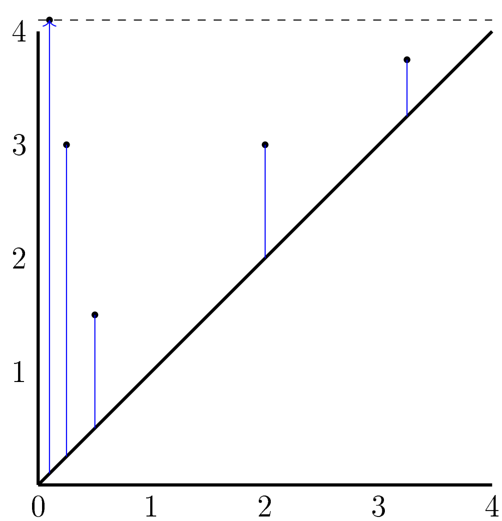


Figure 13. The Barcode Encoded into the Persistence Diagram.

We are headed towards the stability of persistence diagrams where we see the filtration as arising from a function. The following development will be in terms of singular homology, but a classical result in algebraic topology is that singular and simplicial homology are isomorphic on triangulable spaces [Rot08]. In our case, the nice spaces are triangulable. Similar to the previous section, a continuous map between topological space $f : \mathbb{X} \to \mathbb{Y}$ induces linear maps $f_k : H_k(\mathbb{X}) \to H_k(\mathbb{Y})$. Also if $f : \mathbb{X} \to \mathbb{Y}$ and $g : \mathbb{Y} \to \mathbb{Z}$ are continuous functions then $(g \circ f)_k = g_k \circ f_k$. For our purposes we shall consider the case where $\mathbb{X}$ is a subspace of $\mathbb{Y}$ and f as the inclusion map.

Definition 2.66. Let $\mathbb{X}$ be a topological space and $f : X \to \mathbb{R}$ be a function. The **sub-level set** of f at some real number a is the set $f^{-1}((-\infty, a])$.

Definition 2.67. Let $\mathbb{X}$ be a topological space and $f : \mathbb{X} \to \mathbb{R}$ be a real function on $\mathbb{X}$. A **homological critical value** of f is a real number a so that there is an integer k so that for every sufficiently small $\epsilon > 0$ the map $H_k(f^{-1}(-\infty, a - \epsilon]) \to H_k(f^{-1}(-\infty, a + \epsilon])$ induced by inclusion is not an isomorphism.

Put simply, a homological critical value is exactly the point where the generators of the homology of a sub-level set change.

Definition 2.68. A function $f : \mathbb{X} \to \mathbb{R}$ is said to be **tame** if it has finitely many homological critical values and the homology modules $H_k(f^{-1}(-\infty, a])$ are finite dimensional for every $k \in \mathbb{Z}_{\geq 0}$ and $a \in \mathbb{R}$.

To reduce the cumbersome notation, fix an integer k. Let $F_x = H_k(f^{-1}(-\infty, x])$. For $x \leq y$ let $f_x^y : F_x \to F_y$ be the map induced by the inclusion $f^{-1}(-\infty, x] \subset f^{-1}(-\infty, y]$. We will write F_x^y to mean $\text{im} f_x^y$. As a convention we set $F_x^y = \{0\}$ whenever x or y is infinite.

Definition 2.69. Let $f : \mathbb{X} \to \mathbb{R}$ be a tame function and $x \leq y \in \bar{\mathbb{R}}$. We call F_x^y a **persistent homology group** and we call $\beta_x^y = \text{rank}\, F_x^y$ a **persistent Betti number**.

Lemma 2.70 (Critical Value Lemma)**.** *Suppose the closed interval* $[x, y]$ *contains no homological critical value of some function* f*. Then* f_x^y *is an isomorphism for each* $k \in \mathbb{Z}_{\geq 0}$*.*

Proof. Let $m_0 = (x + y)/2$. Then $f_x^y = f_{m_0}^y \circ f_x^{m_0}$. If f_x^y is not an isomorphism then either $f_{m_0}^y$ or $f_x^{m_0}$ is not an isomorphism. Without loss of generality suppose it is the latter. Let $M_0 = [x, m_0]$. Now take $m_1 = (x + m_0)/2$. As before either $f_{m_1}^{m_0}$ or $f_x^{m_1}$ is not an isomorphism. Continue in this manner to obtain a countable decreasing sequence of intervals whose intersection is a point. That point is a homological critical value by definition contradicting our hypothesis. □

From here on we shall fix our dimension for homology to be $k \geq 0$. Suppose $f : \mathbb{X} \to \mathbb{R}$ is tame with $(a_i)_{i=1}^n$ its homological critical values. Let $(b_i)_{i=0}^n$ be so that $b_i - 1 < a_i < b_i$. Let $b_{-1} = a_0 = -\infty$ and $a_{n+1} = b_{n+1} = \infty$. Note that here we obtain a filtration $F_{b_0} \hookrightarrow F_{b_1} \hookrightarrow ... \hookrightarrow F_{b_n}$, which yields a persistence module.

Definition 2.71. Take integers $0 \leq i, j \leq n + 1$ and define the **multiplicity** of the pair (a_i, a_j) to be

$$\mu_i^j = \beta_{b_{i-1}}^{b_j} - \beta_{b_i}^{b_j} + \beta_{b_i}^{b_{j-1}} - \beta_{b_{i-1}}^{b_{j-1}}$$

Definition 2.72. The **persistence diagram** arising from f, $D(f)$ is the multiset of points $(b_i, b_j) \in \mathbb{R} \times \mathbb{R}$ counted with multiplicity μ_i^j, union all of the points on the diagonal, Δ with infinite multiplicity.

Notice that this definition is the same as the previous definition of a persistence diagram. We will use this definition as it is more convenient. We wish to include the diagonal points for the following reason. Given two persistence diagrams, we would like to be able to pair each point on the diagram in an "optimal way" which will be explained later. Given that there might not be a pairing without the diagonal (as the number of off diagonal points could differ in each diagram) we must have it to achieve the goal.

Definition 2.73. Given a multi-set A we define the total multiplicity of A to be the sum of the multiplicities of elements in A and is denoted $\#(A)$.

For example the total multiplicity of the persistence diagram without the diagonal (because with the diagonal the total multiplicity would be trivial) is $\#(D(f) - \Delta) = \sum_{i<j} \mu_i^j$. We will refer to this number as the **size** of the persistence diagram. Now for a bit of notation. We will let $Q_a^b = [-\infty, a] \times [b, \infty]$. That is the upper left quadrant determined by the point (a, b).

Lemma 2.74 (k-Triangle Lemma)**.** *Let f be a tame function and suppose $x < y$ are not homological critical values of f. Then the total multiplicity of the persistence diagram in the resulting upper left quadrant is $\#(D(f) \cap Q_x^y) = \beta_x^y$.*

Proof. Since $x < y$ suppose without loss of generality that $x = b_i$ and $y = b_{j-1}$. For ease of notation we will let $\beta_i^j = \beta_{b_i}^{b_j}$. Then by definition we have

$$\begin{aligned}
\#(D(f) \cap Q_x^y) &= \sum_{k=-1}^{i} \sum_{\ell=j}^{n+1} \mu_k^\ell \\
&= \sum_{k=-1}^{i} \sum_{\ell=j}^{n+1} (\beta_{k-1}^\ell - \beta_k^\ell + \beta_k^{\ell-1} - \beta_{k-1}^{\ell-1} \\
&= \beta_{-1}^{n+1} - \beta_i^{n+1} + \beta_i^{j-1} - \beta_{-1}^{j-1}.
\end{aligned}$$

Notice that the first, second, and fourth term are all 0 by convention. Hence, we are left with $\beta_i^{j-1} = \beta_x^y$ as desired. $\square$

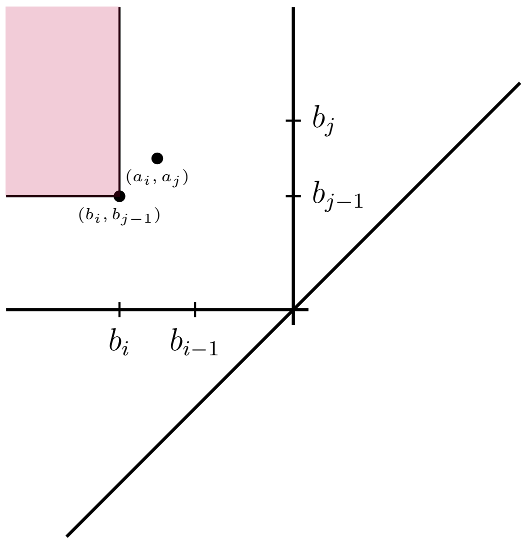
The multiplicity μ_i^j can also be written as a difference of differences. That is $\mu_i^j = (\beta_{b_i}^{b_{j-1}} - \beta_{b_i}^{b_j}) - (\beta_{b_{i-1}}^{b_{j-1}} - \beta_{b_{i-1}}^{b_j})$. The first term, $\beta_{b_i}^{b_{j-1}}$, can be interpreted as the number of independent homology classes, or features, in $F_{b_{j-1}}$ that are born before F_{b_i}. Then the first difference, $\beta_{b_i}^{b_{j-1}} - \beta_{b_i}^{b_j}$, counts the number of features in $F_{b_{j-1}}$, born before F_{b_i}, that die before F_{b_j}. In the same way, the second difference, $\beta_{b_{i-1}}^{b_{j-1}} - \beta_{b_{i-1}}^{b_j}$ counts the features in $F_{b_{j-1}}$ born before $F_{b_{i-1}}$, that die before F_{b_j}. Then we conclude that the multiplicity μ_i^j counts the features born between $F_{b_{i-1}}$ and F_{b_i} that die between $F_{b_{j-1}}$ and F_{b_j}, see Figure 14.

Definition 2.75. Let X and Y be multisets of points in metric space. Define the **Hausdorff distance** (also known as the Pompeiu-Hausdorff distance) between X and Y is

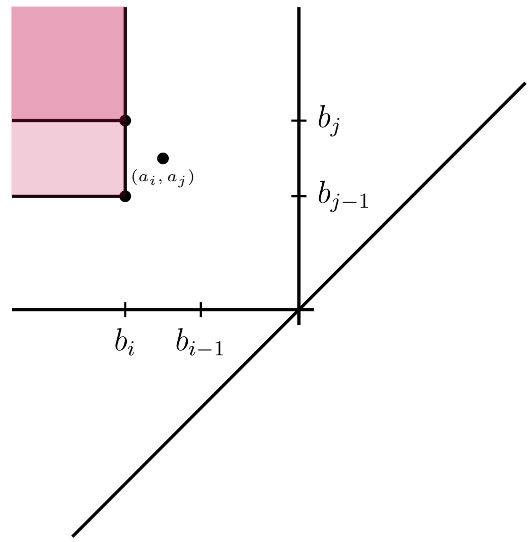
$$d_H(X,Y) = \max\{\sup_{x\in X}\inf_{y\in Y} d(x,y), \sup_{y\in Y}\inf_{x\in X} d(x,y)\}.$$

Definition 2.76. Let X and Y be multisets of points of the same total multiplicity. Let $\Gamma = \{\gamma : X \to Y \mid \gamma \text{ is a bijection}\}$. We define the **bottleneck distance** to be

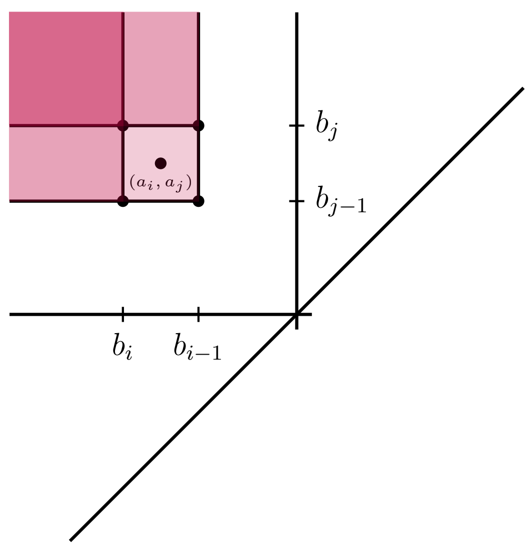
$$d_B(X,Y) = \inf_{\gamma\in\Gamma}\sup_{x\in X} \|x - \gamma(x)\|_\infty.$$



(a) The Shaded Region Represents β_i^{j-1}



(b) The Lighter Region Represents $\beta_i^{j-1} - \beta_i^j$.



(c) This is μ_i^j

Figure 14. A Visual Representation of Multiplicity

Note that for persistence diagrams, Γ is not empty as every diagram contains infinitely many diagonals Δ. We are closing in on the stability theorem in [CSEH07]. We will first need to cover a few more lemmata. Recall that if f is a tame function on a topological space X, $F_x = H_k(f^{-1}((-\infty, x]))$. In the same way for a tame function g on X, $G_x = H_k(g^{-1}((-\infty, x]))$. Also, $f_x^y : F_x \to F_y$, $g_x^y : G_x \to G_y$, $F_x^y = \mathrm{im} f_x^y$, and $G_x^y = \mathrm{im} g_x^y$. Finally, let $\epsilon = \|f - g\|_\infty$, $Q = Q_b^c$ and $Q_\epsilon = Q_{b-\epsilon}^{c+\epsilon}$ for $b < c$. What we will show next is that if we have two tame functions, then the diagrams are somehow interleaved, see Figure 15. This will be made more precise in the theorem.

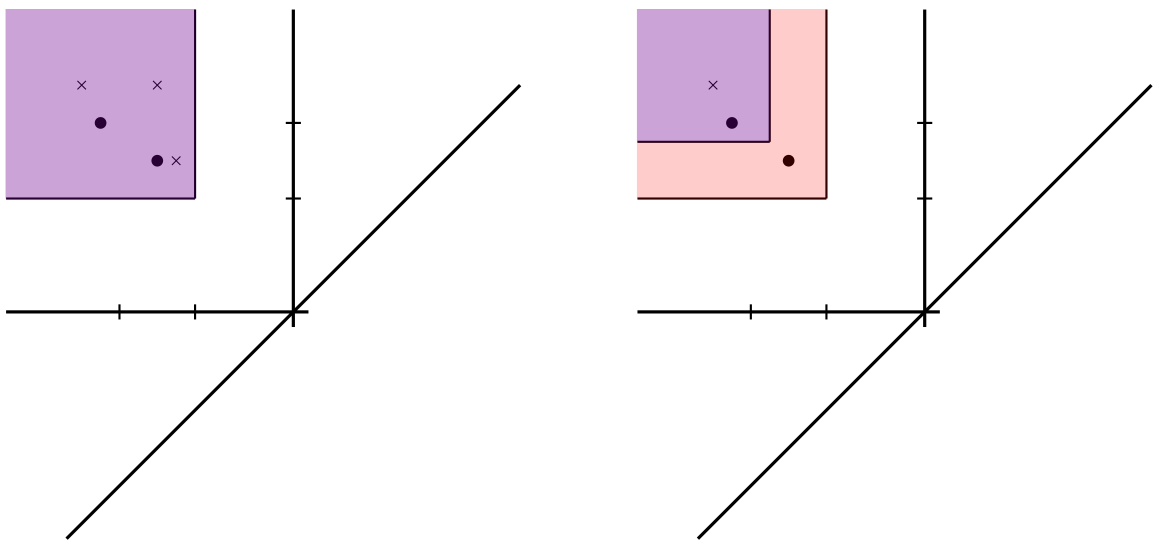


Figure 15. A Visual Representation of the Quadrant Lemma

Lemma 2.77 (Quadrant Lemma)**.** *Let $\mathbb{X}$ be a topological space. Let $f : \mathbb{X} \to \mathbb{R}$ and $g : \mathbb{X} \to \mathbb{R}$ be two tame functions. Then $\#(D(f) \cap Q_\epsilon) \leq \#(D(g) \cap Q)$.*

Proof. Since $\epsilon = \|f - g\|_\infty$ we have $f^{-1}((-\infty, x]) \subset g^{-1}((-\infty, x + \epsilon])$. Let $\phi_x : F_x \to G_{x+\epsilon}$ be the induced inclusion map. Similarly $g^{-1}((-\infty, x]) \subset f^{-1}((-\infty, x + \epsilon])$ Let $\psi_x : G_x \to F_{x+\epsilon}$. We have then the following two diagrams.

$$
\begin{array}{ccc}
F_{b-\epsilon} & \xrightarrow{f_{b-\epsilon}^{c+\epsilon}} & F_{c+\epsilon} \\
\downarrow{\scriptstyle \phi_{b-\epsilon}} & & {\scriptstyle \psi_c}\uparrow \\
G_b & \xrightarrow{g_b^c} & G_c
\end{array}
\qquad\qquad
\begin{array}{ccc}
F_{b+\epsilon} & \xrightarrow{f_{b+\epsilon}^{c+\epsilon}} & F_{c+\epsilon} \\
{\scriptstyle \psi_b}\uparrow & & {\scriptstyle \psi_c}\uparrow \\
G_b & \xrightarrow{g_b^c} & G_c
\end{array}
$$

Since the inclusion maps commute, we have that the induced maps commute also. Hence from the left diagram we get $f_{b-\epsilon}^{c+\epsilon} = \psi_c \circ g_b^c \circ \phi_{b-\epsilon}$ Suppose $\xi \in F_{b-\epsilon}^{c+\epsilon}$. Then by definition there is some $\eta \in F_{b-\epsilon}$ so that $\xi = f_{b-\epsilon}^{c+\epsilon}(\eta)$. Hence with $\zeta = g_b^c(\phi_{b-\epsilon}(\eta))$ we have $\xi = \psi_c(\zeta)$. This means that $F_{b-\epsilon}^{c+\epsilon} \subset \psi(G_b^c)$. From the second diagram we have $\psi_c(G_b^c) = \psi_c \circ g_b^c(G_b) = f_{b+\epsilon}^{c+\epsilon} \circ \psi_b(G_b) \subset F_{b+\epsilon}^{c+\epsilon}$. Putting these together we find $F_{b-\epsilon}^{c+\epsilon} \subset \psi_c(G_b^c) \subset F_{b+\epsilon}^{c+\epsilon}$.

Interpreting this tells us that $\dim F_{b-\epsilon}^{c+\epsilon} \leq \dim G_b^c$. By applying the k-Triangle Lemma we have that this inequality applies to the total multiplicities and therefore we obtain our desired result that $\#(D(f) \cap Q_\epsilon) \leq \#(D(g) \cap Q)$. We note that if $b, b-\epsilon, c$, or $c+\epsilon$ happen to be homological critical values we can simply introduce a sufficiently small δ and repeat the argument. □

Lemma 2.78 (Box Lemma)**.** *Let $a < b < c < d \in \bar{\mathbb{R}}$ and let f and g be tame functions. Let $R = [a,b] \times [c,d]$ and let $R_\epsilon = [a+\epsilon, b-\epsilon] \times [c+\epsilon, d-\epsilon]$. Then $\#(D(f) \cap R_\epsilon) \leq \#(D(g) \cap R)$.*

We will omit this proof as it is quite lengthy with the same flavor as the proof of the Quadrant Lemma. We will, however, give a picture of how this lemma works.

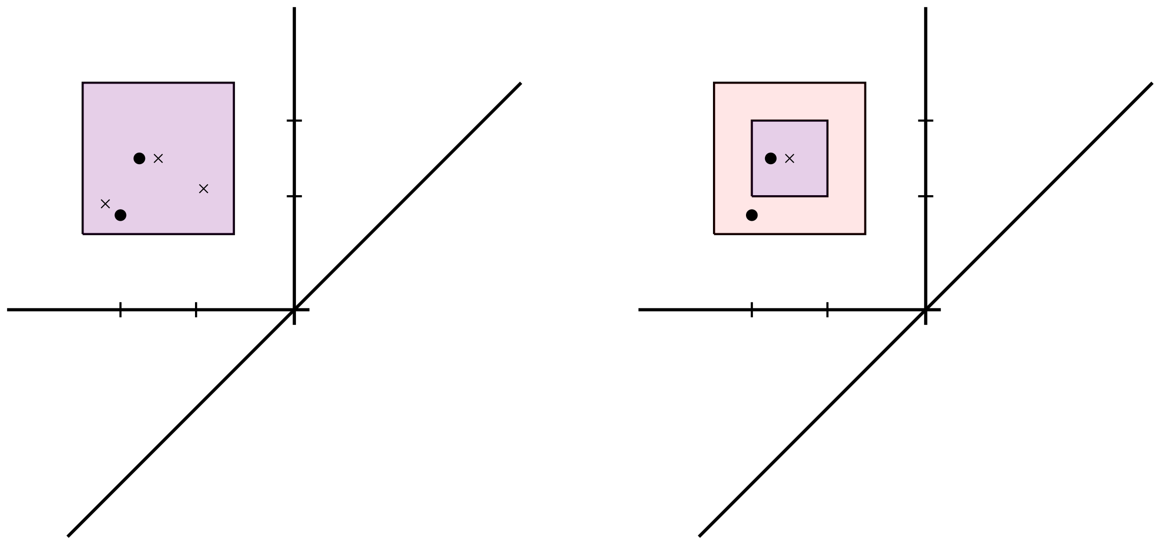


Figure 16. An Illustration of the Box Lemma

From the Box lemma we have $d_H(D(f), D(g)) \leq \|f-g\|_\infty$. Before the next lemma we need a definition.

Definition 2.79. Let $f, g : \mathbb{X} \to \mathbb{R}$ be tame and let $\delta_f = \min\{\min\{\|p-q\|_\infty \mid p, q \in D(f) - \Delta\}, \min\{\|p-q\|_\infty \mid p \in D(f) - \Delta\ q \in \Delta\}\}$. In the case where there are not two off-diagonal points in $D(f)$ we set the first minimum to ∞. We call f and g **very close** if $\|f-g\|_\infty < \delta_f/2$.

Intuitively, δ_f measures how much room there is between the closest points in the diagram of f. Then if g is so that the distance between g and f is less δ_f, then g must be very close or even almost the same function as f.

Lemma 2.80 (Easy Bijection Lemma)**.** *Let $f, g : \mathbb{X} \to \mathbb{R}$ be very close and tame functions. Then $d_B(D(f), D(g)) \leq \|f-g\|_\infty$.*

Proof. Let μ denote the multiplicity of the point p in $D(f)-\Delta$. Let Γ_ϵ be the square with center p and radius $\epsilon = \|f-g\|_\infty$. Applying the Box Lemma gives us that

$\mu \leq \#(D(g) \cap \Gamma_\epsilon) \leq \#(D(f) \cap \Gamma_{2\epsilon})$. By definition of very close we have $2\epsilon \leq \delta_f$ which means that p is the only point in $D(f) \cap \Gamma_{2\epsilon}$. But this means $\#(D(g) \cap \Gamma_\epsilon) = \mu$. This means we can map all points of $D(g) \cap \Gamma_\epsilon$ to p bijectively. We apply this to all off diagonal points in $D(f)$. If anything in $D(g)$ remains, it is exactly those points farther than ϵ away from $D(f) - \Delta$. However since $d_H(D(f), D(g)) \leq \epsilon$ we see that these remaining points are less than ϵ away from Δ. Sending these points to the diagonal we obtain a bijection which moves points by at most ϵ completing this proof. □

For our final lemma before we prove the Bottleneck Stability Theorem, let $\hat{f}$ and $\hat{g}$ be piecewise-linear functions defined on a simplicial complex K. Let $h_\lambda = (1-\lambda)\hat{f} + \lambda\hat{g}$ where $0 \leq \lambda \leq 1$. The collection of all h_λ forms a linear interpolation between $\hat{f}$ and $\hat{g}$.

Lemma 2.81 (Interpolation Lemma)**.** *In the notation above,* $d_B(D(\hat{f}), D(\hat{g})) \leq \|\hat{f} - \hat{g}\|_\infty$.

Proof. Let $c = \|f - g\|_\infty$. Note that for each λ, h_λ is tame. Also, $\delta(\lambda) = \delta_{h_\lambda}$ is positive. Then the set C of open intervals $C_\lambda = (\lambda - \delta(\lambda)/4c, \lambda + \delta(\lambda)/4c)$ forms an open cover of $[0, 1]$. Since $[0, 1]$ is compact we may take not only a finite subcover of C, but a minimal subcover C' of C. So let $\lambda_1 < \lambda_2 < ... < \lambda_n$ be the centers of the intervals in C'. Since C' is minimal we know that $C_{\lambda_i} \cap C_{\lambda_{i+1}} \neq \emptyset$. Thus $\lambda_{i+1} - \lambda_i \leq (\delta(\lambda_i) + \delta(\lambda_{i+1}))/4c \leq \max\{\delta(\lambda_i), \delta(\lambda_{i+1})\}/2c$. Now, by definition of c we have $\|h_{\lambda_i} - h_{\lambda_{i+1}}\|_\infty = c(\lambda_{i+1} - \lambda_i)$. That is to say $\|h_{\lambda_i} - h_{\lambda_{i+1}}\|_\infty \leq \max\{\delta(\lambda_i), \delta(\lambda_{i+1})\}/2c$. But this means h_{λ_i} and $h_{\lambda_{i+1}}$ are very close and hence we apply the Easy Bijection Lemma to get $d_B(D(h_{\lambda_i}), D(h_{\lambda_{i+1}}) \leq \|h_{\lambda_i} - h_{\lambda_{i+1}}\|_\infty$ for each $1 \leq i \leq n-1$. Putting $\lambda_0 = 0$ and $\lambda_{n+1} = 1$ we have this holding for $0 \leq i \leq n$ Then the triangle inequality gives us

$$d_B(D(\hat{f}), D(\hat{g})) \leq \sum_{i=0}^{n} d_B(D(h_{\lambda_i}), D(h_{\lambda_{i+1}}) \leq \sum_{i=0}^{n} \|h_{\lambda_i} - h_{\lambda_{i+1}}\|_\infty.$$

But the latter sum is bounded above by $\|\hat{f} - \hat{g}\|_\infty$ since the h_λ sample the interpolation between $\hat{f}$ and $\hat{g}$, which concludes the proof. □

With all of our combined technical results we now state and prove the Bottleneck Stability theorem.

Theorem 2.82 (Bottleneck Stability)**.** *Let $\mathbb{X}$ be a triangulable space with continuous tame functions $f, g : \mathbb{X} \to \mathbb{R}$. Then $d_B(D(f), D(g)) \leq \|f - g\|_\infty$.*

Proof. Since $\mathbb{X}$ is triangulable there is a finite simplicial complex L and homeomorphism $\Phi : L \to \mathbb{X}$. Note Φ can be chosen so that $f \circ \Phi$ is tame and has the same diagram as f. Since f and g are continuous and L is compact there is a subdivision K of L so that

$$|f \circ \Phi(u) - f \circ \Phi(v)| \leq \delta \text{ and } |g \circ \Phi(u) - g \circ \Phi(v)| \leq \delta$$

where u and v are vertices of a common simplex in K and δ is sufficiently small. Now let $\hat{f}, \hat{g} : \operatorname{Sd} K \to \mathbb{R}$ be the piecewise linear interpolations of $f \circ \Phi$ and $g \circ \Phi$ on K. Then by the construction of K, $\|\hat{f} - f \circ \Phi\|_\infty \leq \delta$ and $\|\hat{g} - g \circ \Phi\|_\infty \leq \delta$. Now by the Interpolation Lemma we have $d_B(D(\hat{f}), D(\hat{g})) \leq \|\hat{f} - \hat{g}\|_\infty \leq \|f - g\|_\infty + 2\delta$. By supposing $\delta \leq \min\{\delta_f, \delta_g\}$ we obtain from the Easy Bijection Lemma, $d_B(D(f), D(\hat{f})) = d_B(D(f \circ \Phi), D(\hat{f}) \leq \delta$ and $d_B(D(g), D(\hat{g})) = d_B(D(g \circ \Phi), D(\hat{g}) \leq \delta$. Finally, by putting it all together we have

$$d_B(D(f), D(g)) \leq d_B(D(f), D(\hat{f})) + d_B(D(\hat{f}), D(\hat{g})) + d_B(D(g), D(\hat{g})) \leq \|f - g\|_\infty + 4\delta$$

By allowing δ to tend to 0 we conclude $d_B(D(f), D(g)) \leq \|f - g\|_\infty$. □

This theorem validates the method of persistent homology. It guarantees that any two good samples of a space will have similar looking diagrams. In other words, if one were to repeat an experiment, then one should recover the shape of the data that was recovered in the original experiment.

CHAPTER III

RESULTS

3.1 A Multi-scale Rips Lemma

In this chapter we present our results. We will often consider a data cloud in $\mathbb{R}^d$. Recall that the Rips and Čech complexes were defined at a scale ϵ. We will extend these definitions to be defined for multiple radii in the form of a function $\mathbf{r} : X \to (0, \infty)$. The following definition is an obvious extension of the classical definition made simply by replacing the scale with a function.

Definition 3.1. Let $X \subset \mathbb{R}^d$ be finite with n elements. Let $\mathbf{r} : X \to (0, \infty)$ be a function. Define the **multi-scale Čech complex** at scale $\mathbf{r}$ to be the set

$$\check{C}_{\mathbf{r}}(X) = \{\sigma \neq \emptyset \subset X \mid \cap_{x_i \in \sigma} B_{r_i}(x_i)\}$$

where $r_i = \mathbf{r}(x_i)$.

Clearly, by taking $\mathbf{r}(x) = \epsilon$ for all $x \in X$ we obtain $\check{C}_\epsilon(X)$ the classical Čech complex at scale ϵ. Now recall that the Rips condition requires $d(x_i, x_j) \leq 2r = r + r$. So, if we take the function $\mathbf{r}$ the Rips condition then simply turns into $d(x_i, x_j) \leq r_i + r_j$. We use the notation with the $x_i's$ for convenience, since we can place a total ordering on any finite set X.

Definition 3.2. Let X be a dubset of a metric space (M, d) and $\mathbf{r} : X \to (0, \infty)$ be as above. Define the **multi-scale Rips complex** to be

$$R_{\mathbf{r}}(X) = \{\sigma \subset X \mid \forall x_i, x_j \in \sigma \; d(x_i, x_j) \leq r_i + r_j\}.$$

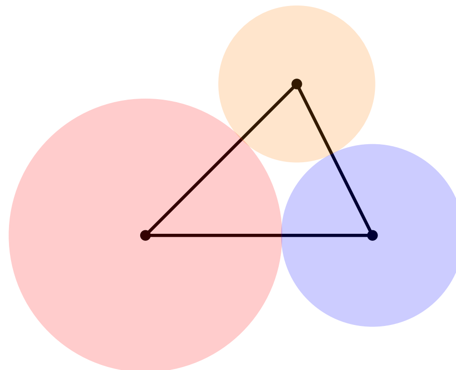


Figure 17. The Multi-Scale Čech Complex

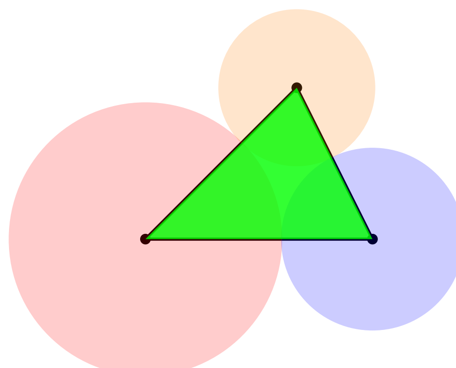


Figure 18. The Multi-Scale Rips Complex

Theorem 3.3. *Suppose $X \subset \mathbb{R}^d$ is finite with N elements. Further suppose that $\mathbf{r} : X \to (0, \infty)$ and $\epsilon, \epsilon' > 0$ so that $\epsilon \geq \epsilon' \cdot \sqrt{\frac{2d}{d+1}}$. Then $R_{\epsilon'\mathbf{r}} \subset \check{C}_{\epsilon\mathbf{r}} \subset R_{\epsilon\mathbf{r}}$*

Proof. The second containment $\check{C}_{\varepsilon\mathbf{r}}(X) \subseteq R_{\varepsilon\mathbf{r}}(X)$ follows from the fact that the multi-scale Rips complex is the flag complex of the Čech complex. To show that $R_{\varepsilon\mathbf{r}}(X) \subset \check{C}_{\epsilon\mathbf{r}}$, we take an element of the Rips complex at scale ε'. That is suppose there is some finite collection $\{x_k\}_{k=0}^{\ell} \subseteq \mathbb{R}^d$ so that $\|x_i - x_j\|_2 \leq \varepsilon'(\mathbf{r}(x_i) - \mathbf{r}(x_j))$ whenever $i \neq j$. Define a function $f : \mathbb{R}^d \to \mathbb{R}$ by

$$f(y) = \max_{0 \le j \le \ell} \left\{ \frac{\|x_j - y\|_2}{\mathbf{r}(x_j)} \right\}.$$

Clearly, f is continuous and $|f| \to \infty$ as $\|y\|_2 \to \infty$. Thus f attains a minimum on some compact set containing $\text{conv}(\{x_k\}_{k=0}^{\ell})$. It follows that f attains an absolute minimum, say y_0, on $\mathbb{R}^d$. By a reordering of vertices if needed, we may assume $f(y_0) = \frac{1}{\mathbf{r}(x_j)}\|x_j - y_0\|_2^2$ for some subcollection $\{x_j\}_{j=0}^{n} \subseteq \{x_k\}_{k=0}^{\ell}$ and $f(y_0) > \frac{1}{\mathbf{r}(x_j)}\|x_j - y_0\|_2^2$ for $\{x_j\}_{j=n+1}^{\ell}$. Let $g(y) = \max_{0 \le j \le n}\{\frac{1}{\mathbf{r}(x_j)}\|x_j - y\|_2\}$ and $h(y) = \max_{n+1 \le j \le \ell}\{\frac{1}{\mathbf{r}(x_j)}\|x_j - y\|_2\}$.

Now we wish to show that $y_0 \in \text{conv}(\{x_j\}_{j=0}^{n})$. To this end, we apply Farkas' Lemma [HUL04]: either $y_0 \in \text{conv}(\{x_j\}_{j=0}^{n})$ or there is a $v \in \mathbb{R}^d$ such that $v \cdot x_j \ge 0$ for all $0 \le j \le n$ and $v \cdot y_0 < 0$. Thus we need only show that there is no $v \in \mathbb{R}^d$ so that $v \cdot (x_j - y_0) > 0$ for $0 \le j \le n$. By way of contradiction, suppose otherwise. Since

$$\|x_j - (y_0 + \lambda v)\|_2^2 = \|x_i - y_0\|_2^2 - 2\lambda v \cdot (x_j - y_0) + \lambda^2 \|v\|_2^2$$

for each $0 \le j \le n$, it follows that $g(y_0 - \lambda v) < f(y_0)$ for all $\lambda \in (0, \lambda_1)$ where $\lambda_1 = \min_{0 \le j \le n} 2v \cdot (x_j - y_0)/\|v\|_2^2$. Since $h(y)$ is continuous and $h(y_0) < f(y_0)$, there exists a λ_2 so that $h(y_0 + \lambda v) < f(y_0)$ for $\lambda \in [0, \lambda_2)$. Thus there exists a $\lambda > 0$ such that $f(y_0 + \lambda v) = \max\{g(y_0 + \lambda v), h(y_0 + \lambda v)\} < f(y_0)$, a contradiction to the minimality of y_0.

By Carathéodory's theorem [GWZ96] and reordering of vertices if necessary, there exists some subcollection of vertices $\{x_i\}_{i=0}^{m}$ where $0 < i \le \min\{d, n\}$. It is not possible that $i = 0$. If so, then $y_0 = x$ and $f(y_0) = \frac{1}{\mathbf{r}(x_0)}\|x_0 - y_0\|_2 = 0$ and f is identically zero. Since σ contains a vertex $x_1 \neq x_0$, it follows that $f(y_0) = f(x_0) > \frac{1}{\mathbf{r}(x_1)}\|x_1 - x_0\|_2 > 0$, a contradiction.

By way of notation, let $\hat{x}_j = x_j - y_0$. Note that

$$\|\hat{x}_j\|_2^2 = \mathbf{r}(x_j)^2 f(y_0)^2. \tag{3.1}$$

Take $a_0, a_1, \ldots, a_m \in \mathbb{R}_{\geq 0}$ so that $\sum_{i=0}^m a_i = 1$ and $y_0 = \sum_{i=1}^m a_i x_i$. Then $\sum_{i=0}^m a_i \hat{x}_i$. By relabeling, we may assume that $a_0\mathbf{r}(x_0) \geq \mathbf{r}(x_i)a_i$ when $i > 0$. Then we obtain $\hat{x}_0 = \sum_{i=1}^m \frac{a_i}{a_0}\hat{x}_i$, and so

$$\mathbf{r}(x_0)^2 f(y_0)^2 = \|\hat{x}_0\|_2^2 = -\sum_{i=1}^m \frac{a_i}{a_0}\hat{x}_0\hat{x}_i.$$

Among the indices $1, 2, \ldots, m$, there is some ι such that

$$\frac{1}{d}\mathbf{r}(x_0)^2 f(y_0)^2 \leq \frac{1}{m}\mathbf{r}(x_0) f(y_0)^2 \leq -\frac{a_\iota}{a_0}\hat{x}_0\hat{x}_\iota. \tag{3.2}$$

Putting (3.1) and (3.2) together, we find

$$\begin{aligned} f(y_0)^2\left(\mathbf{r}(x_0)^2 + \frac{2a_0\mathbf{r}(x_0)^2}{a_\iota d} + \mathbf{r}(x_\iota)^2\right) &\leq \|\hat{x}_0\|_2^2 - 2\hat{x}_0\hat{x}_\iota + \|\hat{x}_\iota\|_2^2 \\ &= \|\hat{x}_0 - \hat{x}_\iota\|_2^2 \\ &= \|x_0 - x_\iota\|_2^2 \\ &\leq (\varepsilon'(\mathbf{r}(x_0) + \mathbf{r}(x_\iota)))^2. \end{aligned}$$

We will now show that

$$\frac{(\mathbf{r}(x_0)^2 + \mathbf{r}(x_\iota)^2)^2}{\mathbf{r}(x_0)^2 + \frac{2a_0\mathbf{r}(x_0)^2}{a_\iota d} + \mathbf{r}(x_\iota)^2} \leq \frac{2d}{d+1}.$$

It suffices to show $(d - 1 + 4\frac{a_0}{a_\iota})\mathbf{r}(x_0)^2 - 2(d+1)\mathbf{r}(x_0)\mathbf{r}(x_\iota) + (d-1)\mathbf{r}(x_\iota)^2 \geq 0$. Since $\frac{a_0}{a_\iota} \geq \frac{\mathbf{r}(x_\iota)}{\mathbf{r}(x_0)}$ we get

$$\begin{aligned}
(d - 1 + 4\frac{a_0}{a_\iota})\mathbf{r}(x_0)^2 - 2(d+1)\mathbf{r}(x_0)\mathbf{r}(x_\iota) + (d-1)\mathbf{r}(x_\iota)^2 \\
&\geq \left(d - 1 + 4\frac{\mathbf{r}(x_\iota)}{\mathbf{r}(x_0)}\right)\mathbf{r}(x_0)^2 - 2(d+1)\mathbf{r}(x_0)\mathbf{r}(x_\iota) + (d-1)\mathbf{r}(x_\iota)^2 \\
&= (d-1)(\mathbf{r}(x_0) - \mathbf{r}(x_\iota))^2 \\
&\geq 0
\end{aligned}$$

as desired. Our assumption that $\varepsilon \geq \varepsilon'\sqrt{2d/(d+1)}$ implies $f(y_0) \leq \varepsilon$ and thus

$$y_0 \in \bigcap_{i=0}^{m} \bar{B}_{\varepsilon\mathbf{r}(x_i)}(x_i).$$

Therefore $\sigma \in \check{C}_{\varepsilon\mathbf{r}}(X)$ and we are done. □

3.2 Stability

We would like to be able to fit the multi-scale Rips complexes into the stability framework presented in the previous section. Hence, we must define a function whose sub-level sets form the Rips complex and is continuous and tame. This is done through the entry function.

Definition 3.4. Let $X \subset \mathbb{R}^d$ be finite. Let $\mathbf{r} : X \to (0, \infty)$. Finally let $\mathcal{X}$ denote a compact set in $\mathbb{R}^d$ containing X. We define the **entry function** $f_{X,\mathbf{r}} : \mathcal{X} \to \mathbb{R}$ by

$$f_{X,\mathbf{r}}(y) = \min_{x\in X}\left\{\frac{\mathrm{d}(x,y)}{\mathbf{r}(x)}\right\}.$$

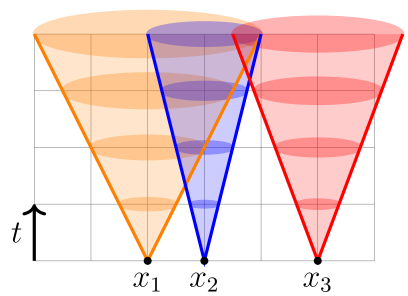


Figure 19. The Entry Function Traces the Bottom of the Cones.

For a bit of intuition imagine a point set in the plane. Begin growing cones above the points with corresponding radius ratios. As the cones grow they will intersect. The entry function traces the bottom of these cones. Now we'd like to show the sub-level sets indeed have the same homology as the Čech complex.

Proposition 3.5. *Suppose $X \subset \mathbb{R}^d$ is finite of size N. Let $\mathbf{r} : X \to (0,\infty)$. Then $H_k(f_{X,\mathbf{r}}^{-1}(-\infty,a]) \cong H_k(\check{C}_{a\mathbf{r}}(X)$.*

Proof. It will suffice to show $f_{X,\mathbf{r}}^{-1}(-\infty,a] = \bigcup_{x_i\in X} B_{a\mathbf{r}(x_i)}(x_i)$ since the Nerve Theorem will grant us the isomorphism between the homology spaces. We proceed with the standard argument. Suppose $x \in f_{X,\mathbf{r}}^{-1}(-\infty,a)$. Then $a \geq f_{X,\mathbf{r}}(x) = \min_{x_i\in X}\frac{d(x,x_i)}{\mathbf{r}(x_i)} = \frac{d(x,x_j)}{\mathbf{r}(x_j)}$ for some x_j. Hence $ar_j \geq d(x,x_j)$ That is to say $x \in B_{a\mathbf{r}(x_j)}(x_j)$. Thus $f_{X,\mathbf{r}}^{-1}(-\infty,a) \subset \bigcup_{x_i\in X} B_{a\mathbf{r}(x_i)}(x_i)$. Next let x be in the union. Then there is some $x_i \in X$ for which $d(x,x_i) \leq a\mathbf{r}(x_i)$. Hence $\frac{d(x,x_i)}{\mathbf{r}(x_i)} \leq a$. Therefore $f(x) \leq a$, or $x \in f_{X,\mathbf{r}}^{-1}(-\infty,a]$. $\square$

Lemma 3.6. *For a finite set $X \subset \mathbb{R}^d$, a function $\mathbf{r} : X \to (0,\infty)$ and a compact set $\mathcal{X}$ containing X, the entry function $f_{X,\mathbf{r}} : \mathcal{X} \to \mathbb{R}$ is continuous and tame.*

Proof. Due to the finiteness of our point set, the function is a minimum over skewed distances. Since we are looking at a minimum of continuous functions, we conclude that the entry function is continuous.

As we are assuming that there are only finitely many points in X, there are only finitely many local maximums and minimums. Hence there are finitely many homological critical values. Thus the entry function is tame. □

With this lemma we have satisfied the hypotheses of the stability theorem, which we state fully here.

Theorem 3.7. *Let $X, X' \subset \mathbb{R}^d$ of size N. Let $\mathbf{r}, \mathbf{r}' : X \to (0, \infty)$ and $\mathcal{X}$ be a compact set containing $X \cup X'$. Then the following three things hold,*

(i) $d_B(D(f_{X,\mathbf{r}}, D(f'_{X,\mathbf{r}})) \leq \|f_{X,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty$;

(ii) $d_B(D(f_{X,\mathbf{r}}, D(f_{X',\mathbf{r}})) \leq \|f_{X,\mathbf{r}} - f_{X',\mathbf{r}}\|_\infty$;

(iii) $d_B(D(f_{X,\mathbf{r}}, D(f'_{X',\mathbf{r}})) \leq \|f_{X,\mathbf{r}} - f_{X',\mathbf{r}'}\|_\infty$.

This theorem follows directly from stability result [CSEH07]. But what we'd like to do is strengthen this result by getting a bound on the distance between the entry functions. We accomplish this with the next three theorems.

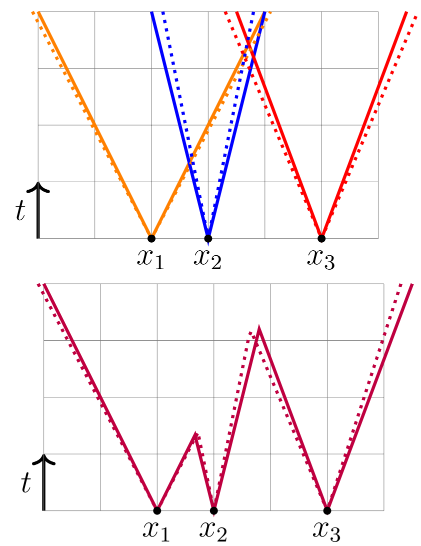


Figure 20. Perturbation of $\mathbf{r}$

Theorem 3.8 (Weight Stability)**.** *Let $X \subseteq \mathbb{R}^d$ be finite. Let $\mathbf{r}, \mathbf{r}' : X \to (0, \infty)$ be functions and $\mathcal{X}$ be a compact set containing X. Then for every $\varepsilon > 0$*

$$\|f_{X,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty < \varepsilon \text{ whenever } \|\mathbf{r} - \mathbf{r}'\|_\infty < \delta = \frac{\varepsilon \min\{\mathbf{r}(x)\mathbf{r}'(x) \,|\, x \in X\}}{\text{diam(X)}}.$$

Proof. It suffices to show that

$$\|f_{X,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty \leq \frac{\|\mathbf{r} - \mathbf{r}'\|_\infty \text{diam}(X)}{\min\{\mathbf{r}(x)\mathbf{r}'(x) \,|\, x \in X\}}.$$

To begin,

$$\|f_{X,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty = \max_{y \in \mathcal{X}} \{|f_{X,\mathbf{r}}(y) - f_{X,\mathbf{r}'}(y)|\}$$

since $f_{X,\mathbf{r}}$ and $f_{X,\mathbf{r}'}$ are continuous functions with compact domain, it follows that there exists some $y_0 \in \mathcal{X}$ such that

$$\|f_{X,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty = |f_{X,\mathbf{r}}(y_0) - f_{X,\mathbf{r}'}(y_0)|.$$

Then we have

$$\|f_{X,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty = \left| \min_{x \in X} \left\{ \frac{\|x - y_0\|_2}{\mathbf{r}(x)} \right\} - \min_{x \in X} \left\{ \frac{\|x - y_0\|_2}{\mathbf{r}'(x)} \right\} \right|.$$

Since X is a finite set, there exist $x_j, x_k \in X$ so that

$$\left| \min_{x \in X} \left\{ \frac{\|x - y_0\|_2}{\mathbf{r}(x)} \right\} - \min_{x \in X} \left\{ \frac{\|x - y_0\|_2}{\mathbf{r}'(x)} \right\} \right| = \left| \frac{\|x_j - y_0\|_2}{\mathbf{r}(x_j)} - \frac{\|x_k - y_0\|_2}{\mathbf{r}'(x_k)} \right|.$$

It is either the case that $\|x_j - y_0\|_2/\mathbf{r}(x_j) = \|x_k - y_0\|_2/\mathbf{r}'(x_k)$ or, without loss of generality, $\|x_j - y_0\|_2/\mathbf{r}(x_j) > \|x_k - y_0\|_2/\mathbf{r}'(x_k)$. If $\|x_j - y_0\|_2/\mathbf{r}(x_j) = \|x_k - y_0\|_2/\mathbf{r}'(x_k)$ then $\|f_{X,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty = 0$ and we are done. Now, suppose $\|x_j - y_0\|_2/\mathbf{r}(x_j) > \|x_k - y_0\|_2/\mathbf{r}'(x_k)$. Since $\|x - y_0\|_2/\mathbf{r}(x) \geq \|x_j - y_0\|_2/\mathbf{r}(x_j)$ for all $x \in X$, it must hold that

$$\frac{\|x_j - y_0\|_2}{\mathbf{r}(x_j)} - \frac{\|x_k - y_0\|_2}{\mathbf{r}'(x_k)} \leq \frac{\|x_k - y_0\|_2}{\mathbf{r}(x_k)} - \frac{\|x_k - y_0\|_2}{\mathbf{r}'(x_k)}.$$

Therefore,

$$\begin{aligned}
\|f_{X,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty &= \left| \frac{\|x_j - y_0\|_2}{\mathbf{r}(x_j)} - \frac{\|x_k - y_0\|_2}{\mathbf{r}'(x_k)} \right| \\
&\leq \left| \frac{\|x_k - y_0\|_2}{\mathbf{r}(x_k)} - \frac{\|x_k - y_0\|_2}{\mathbf{r}'(x_k)} \right| \\
&= \left| \frac{[\mathbf{r}'(x_k) - \mathbf{r}(x_k)]\|x_k - y_0\|_2}{\mathbf{r}(x_k)\mathbf{r}'(x_k)} \right|.
\end{aligned}$$

Finally,

$$\begin{aligned}
\|f_{X,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty &\leq \left| \frac{[\mathbf{r}'(x_k) - \mathbf{r}(x_k)]\|x_k - y_0\|_2}{\mathbf{r}(x_k)\mathbf{r}'(x_k)} \right| \\
&\leq \frac{\|\mathbf{r} - \mathbf{r}'\|_\infty \operatorname{diam}(X)}{\min\{\mathbf{r}(x)\mathbf{r}'(x) \mid x \in X\}},
\end{aligned}$$

as desired. □

Corollary 3.9. *If* $\mathbf{r}, \mathbf{r}' : X \to [1, \infty)$, *then for every* $\varepsilon > 0$

$$\|f_{x,\mathbf{r}} - f_{X,\mathbf{r}'}\|_\infty < \varepsilon \text{ whenever } \|\mathbf{r} - \mathbf{r}'\|_\infty < \delta = \frac{\varepsilon}{\operatorname{diam}(X)}.$$

Before we go through the next theorem we will say a word on the idea. Imagine a sensor network with moving or movable sensors. As the sensors move, the sensing range doesn't change. To model this we require a bijection η between X and X'.

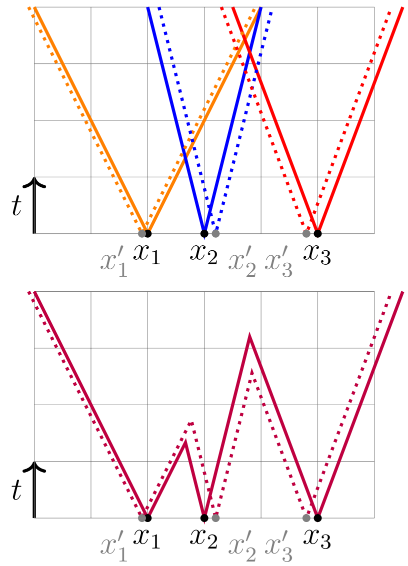


Figure 21. Perturbation of the Points

Theorem 3.10 (Point Perturbation Stability)**.** *Suppose $X, X' \subseteq \mathbb{R}^d$ of common size N, $\mathbf{r} : X \to (0, \infty)$, and suppose that $\eta : X' \to X$ is a set bijection. Furthermore, suppose that $\mathcal{X}$ is a compact set containing both X and X'. Let $f_{X',\mathbf{r}\circ\eta}$ be the entry function on X' induced by $\mathbf{r} \circ \eta$. Then for every $\varepsilon > 0$*

$$\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty < \varepsilon \textit{ whenever } \max_{x \in X'}\{\|x - \eta(x)\|_2\} < \delta = \varepsilon \min_{x \in X}\{\mathbf{r}(x)\}.$$

Proof. We will proceed by showing

$$\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty \leq \frac{\max\left\{\|x - \eta(x)\|_2 \mid x \in X'\right\}}{\min\left\{(\mathbf{r} \circ \eta)(x) \mid x \in X'\right\}}.$$

Since $f_{X,\mathbf{r}}$ and $f_{X',\mathbf{r}'}$ are continuous functions with compact domain,

$$\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty = \max\left\{|f_{X,\mathbf{r}}(x) - f_{X',\mathbf{r}\circ\eta}(x)| \mid x \in \operatorname{conv}(X \cup X')\right\}.$$

It follows that there exists some $y_0 \in \mathcal{X}$ so that

$$\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty = \left| \min_{x \in X} \left\{ \frac{\|x - y_0\|_2}{\mathbf{r}(x)} \right\} - \min_{x \in X'} \left\{ \frac{\|x - y_0\|_2}{(\mathbf{r} \circ \eta)(x)} \right\} \right|.$$

The finiteness of X and X' implies the existence of $x_j \in X$ and $x_k \in X'$ such that

$$\left| \min_{x \in X} \left\{ \frac{\|x - y_0\|_2}{\mathbf{r}(x)} \right\} - \min_{x \in X'} \left\{ \frac{\|x - y_0\|_2}{(\mathbf{r} \circ \eta)(x)} \right\} \right| = \left| \frac{\|x_j - y_0\|_2}{\mathbf{r}(x_j)} - \frac{\|x_k - y_0\|_2}{(\mathbf{r} \circ \eta)(x_k)} \right|.$$

Now it is either the case that $\|x_j - y_0\|_2/\mathbf{r}(x_j) = \|x_k - y_0\|_2/(\mathbf{r} \circ \eta)(x_k)$ or, without loss of generality, $\|x_j - y_0\|_2/\mathbf{r}(x_j) > \|x_k - y_0\|_2/(\mathbf{r} \circ \eta)(x_k)$. In the first case, we have that $\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty = 0$ and we are done. To continue, suppose that $\|x_j - y_0\|_2/\mathbf{r}(x_j) > \|x_k - y_0\|_2/(\mathbf{r} \circ \eta)(x_k)$. Since $\|x_j - y_0\|_2/\mathbf{r}(x_j) \leq \|x - y_0\|_2/\mathbf{r}(x)$ for all $x \in X$, it must be the case that $\|x_j - y_0\|_2/\mathbf{r}(x_j) \leq \|\eta(x_k) - y_0\|_2/(\mathbf{r} \circ \eta)(\eta(x_k))$. Therefore

$$\frac{\|x_j - y_0\|_2}{\mathbf{r}(x_j)} - \frac{\|x_k - y_0\|_2}{(\mathbf{r} \circ \eta)(x_k)} \leq \frac{\|\eta(x_k) - y_0\|_2}{(\mathbf{r} \circ \eta)(x_k)} - \frac{\|x_k - y_0\|_2}{(\mathbf{r} \circ \eta)(x_k)}.$$

This implies

$$\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty = \left| \frac{\|x_j - y_0\|_2}{\mathbf{r}(x_j)} - \frac{\|x_k - y_0\|_2}{(\mathbf{r} \circ \eta)(x_k)} \right| \leq \left| \frac{\|\eta(x_k) - y_0\|_2}{(\mathbf{r} \circ \eta)(x_k)} - \frac{\|x_k - y_0\|_2}{(\mathbf{r} \circ \eta)(x_k)} \right|$$

$$= \frac{|\|\eta(x_k) - y_0\|_2 - \|x_k - y_0\|_2|}{(\mathbf{r} \circ \eta)(x_k)} \leq \frac{\|\eta(x_k) - x_k\|_2}{(\mathbf{r} \circ \eta)(x_k)} \leq \frac{\max\{\|x - \eta(x)\|_2 \mid x \in X'\}}{\min\{(\mathbf{r} \circ \eta)(x) \mid x \in X'\}}.$$

□

Corollary 3.11. *If* $\mathbf{r} : X \to [1, \infty)$, *then*

$$\|f_{X,\mathbf{r}} - f_{X,\mathbf{r}\circ\eta}\|_\infty < \max_{x \in X}\{\|x - \eta(x)\|_2\}.$$

Proof. From the proof of Theorem 3.10, $\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty < \frac{\max\{\|x-\eta(x)\|_2 | x \in X'\}}{\min\{(\mathbf{r}\circ\eta)(x) | x\in X'\}}$. But now, $\min\{(\mathbf{r}\circ\eta)(x) \mid x \in X'\} \geq 1$. Hence, $\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty < \frac{\max\{\|x-\eta(x)\|_2 | x \in X'\}}{\min\{(\mathbf{r}\circ\eta)(x) | x\in X'\}} \leq \max\{\|x - \eta(x)\|_2 \mid x \in X'\}$. □

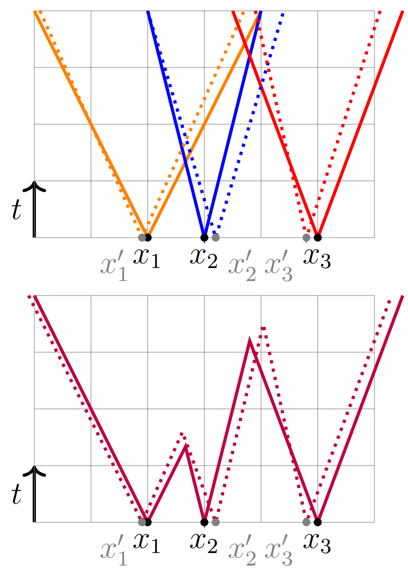


Figure 22. Perturbation of the Points and Radii

Theorem 3.12 (Combined Stability)**.** *Suppose* $X, X' \subseteq \mathbb{R}^d$ *of common size* N, $\mathbf{r} : X \to (0, \infty)$ *and* $\mathbf{r}' : X \to (0, \infty)$ *are functions, and* $\eta : X' \to X$ *is a set bijection. Let* $m_1 = \min_{x\in X'}\{\mathbf{r}(x)\}$ *and* $m_2 = \frac{\min\{(\mathbf{r}\circ\eta)(x)\mathbf{r}'(x) | x \in X'\}}{\operatorname{diam}(X)}$. *Then for every* $\varepsilon > 0$ *we have* $\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}'}\|_\infty < \varepsilon$ *whenever* $\max_{x\in X'}\{\|x - \eta(x)\|_2\} + \|(\mathbf{r}\circ\eta) - \mathbf{r}'\|_\infty < \delta = \epsilon \min\{m_1, m_2\}$.

Proof. Let $\varepsilon > 0$,$m_1 = \min_{x \in X'}\{\mathbf{r}(x)\}$ and $m_2 = \frac{\min\{(\mathbf{r}\circ\eta)(x)\mathbf{r}'(x) | x \in X'\}}{\mathrm{diam}(X)}$. By Theorem 3.8,

$$\|f_{X',\mathbf{r}\circ\eta} - f_{X',\mathbf{r}'}\|_\infty < \frac{\varepsilon}{2} \text{ whenever } \|(\mathbf{r}\circ\eta) - \mathbf{r}'\|_\infty < \frac{\varepsilon \min\{(\mathbf{r}\circ\eta)(x)\mathbf{r}'(x) \mid x \in X'\}}{2\mathrm{diam}(X')}.$$

Also,

$$\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty < \frac{\varepsilon}{2} \text{ whenever } \max_{x \in X'}\{\|x - \eta(x)\|_2\} < \frac{\varepsilon}{2}\min_{x \in X}\{\mathbf{r}(x)\}$$

by Theorem 3.10. Therefore, if we require

$$\max_{x \in X'}\{\|x - \eta(x)\|_2\} + \|(\mathbf{r}\circ\eta) - \mathbf{r}'\|_\infty < 2\min\{\varepsilon m_1, \varepsilon m_2\},$$

then we have

$$\begin{aligned}\|f_{X,\mathbf{r}} - f_{X',\mathbf{r}'}\|_\infty &\leq \|f_{X,\mathbf{r}} - f_{X',\mathbf{r}\circ\eta}\|_\infty + \|f_{X',\mathbf{r}\circ\eta} - f_{X',\mathbf{r}'}\|_\infty \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\end{aligned}$$

and we are done. □

3.3 The Outlier Problem

The goal of this section is to develop a way to interpolate between persistence on a set with a lot of noise and persistence on a set with the noise removed. But first we are going to see that if we add in finitely many points with radius 0 to a finite set, then homology for $n \geq 1$ remains unchanged.

Lemma 3.13. *Let $X, Y \subset \mathbb{R}^d$ be finite. Let $\mathbf{r} : X \cup Y \to [0, \infty)$ be so that $\mathbf{r}(x) \neq 0$ for each $x \in X$ and $\mathbf{r}(y) = 0$ for each $y \in Y$. Then $H_n(\check{C}_{\mathbf{r}}(X \cup Y)) \cong H_n(\check{C}_{\mathbf{r}}(X))$ for $n \geq 1$.*

Proof. It will suffice to show the conclusion holds when Y consists of just one point y. Now either $y \in \bigcup_{x \in X} B_{\mathbf{r}}(x)$ or not. Suppose first that y is in the union. First note that $\bigcup_{x \in X} B_{\mathbf{r}}(x) = \bigcup_{x \in X} B_{\mathbf{r}}(x) \cup \{y\}$. Then by applying the Nerve Lemma twice we obtain the following string of homology equivalences $H_n(\check{C}_{\mathbf{r}}(X \cup \{y\})) \cong H_n(\bigcup_{x \in X} B_{\mathbf{r}}(x) \cup \{y\}) \cong H_n(\bigcup_{x \in X} B_{\mathbf{r}}(x)) \cong H_n(\check{C}_{\mathbf{r}}(X))$. Now if y is not in the union, then we have a disjoint union $\check{C}_{\mathbf{r}}(X) \sqcup \{y\}$. Then $H_n(\check{C}_{\mathbf{r}}(X)) \sqcup \{y\}) = H_n(\check{C}_{\mathbf{r}}(X)) \bigoplus H_n(\{y\}) = H_n(\check{C}_{\mathbf{r}}(X))$ as long as $n \geq 1$. It is clear that we can repeat this argument for any additional points in Y hence we are done. □

Note that the addition of points can change 0 homology if the points added land outside the cover of the balls. This lemma is indeed necessary to the next proof as our stability theorems require we have two sets of the same size.

Theorem 3.14. *Let $X, Y \subset \mathbb{R}^d$ be disjoint and finite of size N and M respectively. Call $Z = X \cup Y$ Let $\mathbf{r} : X \to [0, \infty)$ be so that $\mathbf{r}(x) = 1$ for each $x \in X$ and $\mathbf{r}(y) = 0$ for each $y \in Y$. Let $\mathbf{s}(x) = 1$ for every $x \in Z$ Then for every $\varepsilon > 0$ and $n \geq 1$ there exists $\delta > 0$ so that $\lambda < \delta$ implies $d_B(D_n(f_{Z,\mathbf{r}_\lambda}), D_n(f_{X,\mathbf{s}})) \leq \varepsilon$.*

Proof. By the previous lemma, points with 0 radius do not affect homology for $n > 0$, thus we have $D_n(f_{X,\mathbf{s}}) = D_n(f_{Z,\mathbf{r}_0})$. Hence, $d_B(D_n(f_{Z,\mathbf{r}_\lambda}), D_n(f_{X,\mathbf{s}})) = d_B(D_n(f_{Z,\mathbf{r}_\lambda}), D_n(f_{Z,\mathbf{r}_0}))$. Note that $\|r_\lambda - r_0\|_\infty = \lambda$ By applying radius stability, we can find a δ so that for any $\lambda < \delta$ we have $d_B(D_n(f_{Z,\mathbf{r}_\lambda}), D_n(f_{X,\mathbf{s}})) = d_B(D_n(f_{Z,\mathbf{r}_\lambda}), D_n(f_{Z,\mathbf{r}_0})) \leq \varepsilon$. □

So what we see is that if $\{\lambda_m\}$ is any sequence converging to 0 then $D_n(f_{Z,\mathbf{r}_{\lambda_\mathbf{m}}}) \to D_n(f_{X,\mathbf{s}})$. Moreover, we know that given a reliable method to locate noise, we can more precisely detect the underlying structure without having to completely throw away points.

3.4 The Coverage Problem

In this section our goal is to use the idea of multiple-radii to reduce the cost of coverage. We will be considering a compact, simply-connected region of interest.

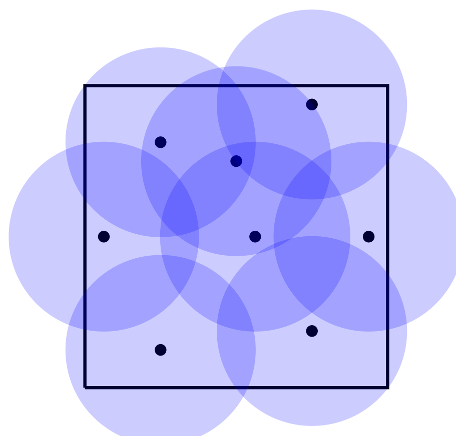


Figure 23. Coverage of Region of Interest

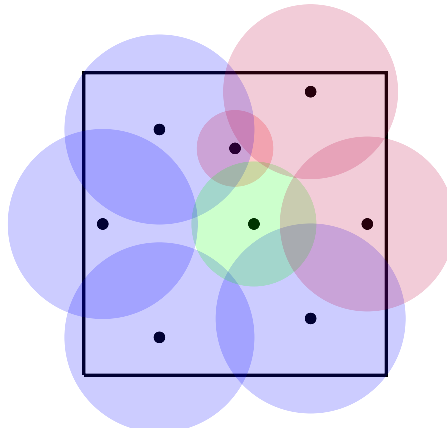


Figure 24. Reduced Cost Coverage of Region of Interest

Definition 3.15. Given a system X of n sensors and weights $\mathbf{r} : X \to (0, \infty)$ we define the **total cost** of the sensors running at power $\mathbf{r}$ to be $C(X, \mathbf{r}) = \sum_{x \in X} \mathbf{r}(x)^2$.

Definition 3.16. Give a system X of sensors and weights $\mathbf{r} : X \to (0, \infty)$, and a compact simply connected region of interest D in some metric space we say $(X, \mathbf{r})$ **covers** D if for each $y \in D$ there is some $x \in X$ so that $d(x, y) \leq \mathbf{r}(x)$.

Notice that we can say that the time in which $(X, \mathbf{r})$ covers D is exactly $\max\{f_{X,\mathbf{r}}(y) \mid y \in D\}$. Hence by letting $u_{\mathbf{r}} = \max\{f_{X,\mathbf{r}}(y) \mid y \in D\}$ we see the cost of coverage is $C(X, \mathbf{r}) = \sum_{i=1}^{n} u_{\mathbf{r}}^2 r_i^2$.

Definition 3.17. Suppose $X \subset \mathbb{R}^d$ is finite. Then the **Voronoi cell** of a point x in X is the set of all points in $\mathbb{R}^d$ for which x is closest, $V_x = \{u \in \mathbb{R}^d \mid d(x, u) \leq d(y, u) \; \forall \; x \in X\}$. The collection of all V_x is called the **Voronoi diagram**.

Definition 3.18. The **Delaunay complex** of X is the nerve of the Voronoi diagram, $DC = \{\sigma \in X \mid \bigcap_{x \in \sigma} V_x \neq \emptyset\}$. We refer often to the **Delaunay triangulation** which is just the 1-skeleton of the Delaunay complex, $DT = DC^{(1)}$.

Proposition 3.19. *If $(X, \mathbf{r})$ covers D where $\mathbf{r}(x) = \mathbf{r}(y)$ for every $x, y \in X$, then the edges of the Delaunay complex are edges in the nerve of the cover.*

Proof. Suppose $[xy]$ is an edge in the Delaunay complex. Then the intersection of the Voronoi cells of x and y is not empty. In fact the intersection of these cells with the edge is nonempty. Suppose w is in that intersection. Since we have coverage, there must be point $z \in X$ so that $d(z, w) \leq \mathbf{r}(z)$. But by definition of the Voronoi cell of x, and since we are assuming $\mathbf{r}(x) = \mathbf{r}(y)$, we must have $d(x, w) \leq \mathbf{r}(x)$. For the same reason, w is also in the Voronoi cell of y we must have $d(w, y) \leq \mathbf{r}(y)$. Hence by the triangle inequality we have $d(x, y) \leq \mathbf{r}(x) + \mathbf{r}(y)$ as desired. □

Our coverage algorithm makes use of two pre-existing algorithms. The first is to compute the Delaunay triangulation given a finite set. The second is a quadratic programming algorithm used to minimize our cost subject to some constraints. Quadratic programming (quadprog) is the problem of finding a vector x that minimizes a quadratic function, possibly subject to linear constraints. That is $\min_x\{\frac{1}{2}x^T Hx + c^T x\}$ subject to $Ax \leq b$. We will now see that we can apply this directly to our situation. Let X be a set of n sensor locations. Since X is finite, we put a total order on it. Let $r_i = \mathbf{r}(x_i)$ and $\vec{r} = [r_1, ..., r_n]$ be the associated vector. Then we are minimizing $\sum_{i=1}^n r_i^2$. Hence to fit this into the quadprog equation we let $H = 2I$ where I is the $n \times n$ identity and we see that indeed $\frac{1}{2}\vec{r}^T H\vec{r} = \vec{r}^T \cdot \vec{r} = \sum_{i=1}^n r_i^2$. Now, for our constraints suppose we have the set E of edges of the Delaunay triangulation of X. We know X is finite hence E is finite. Suppose then we have a total ordering on E. then we take b to be the vector whose entries are the negative norms of the vertices in the edges of the Delaunay triangulation. So if $z = [xy] \in E$ let us write $\|z\|$ to mean $-\|y - x\|_2$. Then the i-th entry of b is $b_i = \|z_i\|$. Let A have rows and columns numbered by the ordering on X. Then define A to be the matrix where the i-th row contains a -1 for each of the vertices in b_i. Then $A\vec{r} \leq b$ implies $d(x_i, x_j) \leq r_i + r_j$ for each edge in the Delaunay triangulation. We will refer to applying the quadprog function as quadprog(H, f, A, b).

Algorithm 3.20. *This algorithm will be used to reduce cover cost of a sensor network.*

Data: *sensorLocations*

Result: *sensorPowers*

begin

Compute Delaunay triangulation on sensorLocations

Edges = edges in Delaunay triangulation

b = $\|z_i\|$ *for* $z_i \in$ *Edges*

A = *Corresponding matrix*

H = $2 \cdot I$

f = *0*

sensorPowers = *quadprog*(H, f, A, b)

end

We immediately see the following theorem as a consequence of this algorithm and the preceding proposition.

Theorem 3.21. *Let X be a set of sensor locations with region of interest D. Let $\mathbf{r} : X \to (0, \infty)$ be the radius obtained from the algorithm. Let $\mathbf{s} : X \to (0, \infty)$ be constant so that $(X, \mathbf{s})$ covers D. Then $C(X, \mathbf{r}) \leq C(X, \mathbf{s})$.*

CHAPTER IV

SUMMARY AND FUTURE DIRECTIONS

In this paper we have generalized the notion of a single radii to multiple radii by defining two classical complexes and generalizing the classical relationship between them. In doing so we have presented three notions of stability and a notion interpolation between two methods of data analysis. Furthermore, we have shown that it is possible to reduce cost of covering a region. In the future we hope to develop an algorithm which assigns the best possible weights to detect the right feature without having to remove noise. We also hope to optimize the cost of covering compact domains with sensor placement and radii assignment.

REFERENCES

[CBK09] Moo K Chung, Peter Bubenik, and Peter T Kim, *Persistence diagrams of cortical surface data*, Information Processing in Medical Imaging, Springer, 2009, pp. 386–397.

[CSEH07] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer, *Stability of persistence diagrams*, Discrete & Computational Geometry **37** (2007), no. 1, 103–120.

[DF04] D.S. Dummit and R.M. Foote, *Abstract algebra*, Wiley, 2004.

[DSG07] Vin De Silva and Robert Ghrist, *Coverage in sensor networks via persistent homology*, Algebraic & Geometric Topology **7** (2007), no. 1, 339–358.

[EH10a] Herbert Edelsbrunner and John Harer, *Computational topology: an introduction*, American Mathematical Soc., 2010.

[EH10b] ———, *Computational topology: an introduction*, American Mathematical Soc., 2010.

[ELZ02] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian, *Topological persistence and simplification*, Discrete and Computational Geometry **28** (2002), no. 4, 511–533.

[Fro90] Patrizio Frosini, *A distance for similarity classes of submanifolds of a euclidean space*, Bulletin of the Australian Mathematical Society **42** (1990), 407–415.

[Gol14] Daniel Goldfarb, *An application of topological data analysis to hockey analytics*.

[GWZ96] PM Gruber, JM Wills, and GM Ziegler, *Handbook of convex geometry*, Jahresbericht der Deutschen Mathematiker Vereinigung **98** (1996), no. 4, 40–40.

[HUL04] J.B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of convex analysis*, Grundlehren Text Editions, Springer Berlin Heidelberg, 2004.

[LKC+12] Hyekyoung Lee, Hyejin Kang, Moo K Chung, Bung-Nyun Kim, and Dong Soo Lee, *Persistent brain network homology from the perspective of dendrogram*, Medical Imaging, IEEE Transactions on **31** (2012), no. 12, 2267–2277.

[Rob99] Vanessa Robins, *Towards computing homology from finite approximations*, Topology proceedings **24** (1999), no. 1, 503–532.

[Rot98] J. Rotman, *An introduction to algebraic topology*, Graduate Texts in Mathematics, Springer New York, 1998.

[Rot08] ———, *An introduction to homological algebra*, Universitext, Springer New York, 2008.