MASTER THESIS

GOWRIMANI GOKAVARAPU

PERSISTENT HOMOLOGY

Abstract

The practical approach to define and visualize Persistent homology and its products is emphasized in this thesis. Persistent homology is a powerful notion originated from topology as an application of computational algebraic topology. This is widely used in analysis of large data to see the shape and features of the data sets, which have the data dealing with diverse domains such as Biology, Cheminformatics, Automatic classification of images, and Sensor and social network analysis. Since homology is homotopy invariant, the concept of homotopy theory can be used to detect the rich features of the data sets. Persistent homology is one of the foundation tools for TDA (Topological Data Analysis). Several ways to examine the features of underlying spaces of the point clouds are alpha complexes, Vietoris-Rips complex and direct usage of TDA packages. In addition, visualization of persistent diagrams and barcodes of persistent homology, are also part of TDA analysis.

Contents

Acknowledgements

I wish to express my sincere gratitude to my supervisor Professor Jesper Micheal Møller for his excellent guidance through out my M.Sc thesis work, particularly for his critical analysis. I indebted to him for all his co-operation and support for my work without any interruption. His contribution towards completion of my thesis is invaluable. My sincere thanks to him not only for his professional advice, but also for his moral support in my critical times. I would like to express my gratitude to Professor Alber Ruiz Cirera, Department of Mathematics at Universitat de Barcelona, for providing me the data required for my thesis work.

I am very much thankful to my family for their constant support and love during my entire studies. My special thanks to my husband Srikar, children Srimanth and Srinidhi, without their support I couldn't have finished my thesis on time.

Introduction

Persistent homology is a homology theory converted to a computational point of view. It analyses the shape and qualitative features of the data. In mathematical point of view, the data is in two different ways, such as set of points floated in a metric space $\mathbb{R}^n$, where n is dimension of a space represented as 'Point Clouds', and the points are considered as vectors. The second type of data is functional valued data, which represents its behavior to the corresponding function. The reason is to study the shape and qualitative features is, the more one can know the shape and geometric features of data the more chance to analyze and explore the science behind it. Note that qualitative features means global geometric features.

The quick overview to reach the persistent homology is, first considering the geometric object i.e.,data (the point cloud). Think of those data points as a union of balls, decompose those as a Voronoi diagram of the points, every center of the points assign the convex polyhedron with the points whose center is the nearest. Pick the nerve (complex which is constructed by edges where two Voronoi regions intersect, triangles where three of Voronoi regions and tetrahedron where three of Voronoi regions intersect), remove the convex pieces and the resulting complex is the alpha complex which is subset of Delaunay triangulation. Now is the time to find out the topological invariant homology in that complex. The final step is to check the persistency of that homology. The results give an idea whether they are homotopy equivalent to any other established shape, sometimes if lucky, they

are topologically equivalent to already established topological space. This justifies the shape and qualitative (geometric) features of that point cloud.

The basic foundation for this construction is Topology. The basic notions of Topology are explained in chapter 1. In chapter 2, the building blocks for the chain complexes are explained. The notion of homology is also introduced in this chapter. In chapter 3, geometrical notions such as Voronoi regions and Delaunay triangulation are explained. In chapter 4, the notion of alpha complex is introduced, and visualizations of alpha shape is also explained. Nerve and Vietoris Rips complexes are introduced to have an idea of which complex construction is recommended for practical approach of persistent homology. In chapter 5, the main objective of this thesis, introduction and products of Persistent homology are explained. The final section of this chapter analyzes point clouds and detects their topological features.

Chapter 1

Topology

To understand a topological space, here is quick introduction to address few notations such as metric spaces, open sets and closed sets. The definition of topology also give a more generalized version of the meaning of open and closed sets.

1.1 Metric spaces

Definition 1. *A metric space is a set X where it has a notation of distance. That is if $x, y \in X$, then $d(x, y)$ is the distance between x and y and the distance function must satisfy the following conditions.*

1. $d(x, y) \geqslant 0$ for all $x, y \in X$

2. $d(x, y) = 0$ if and only if $x = y$

3. $d(x, y) = d(y, x)$

4. $d(x, z) \leqslant d(x, y) + d(y, z)$ for $z \in X$

Example 1. *For any space X, let $d(x, y) = 0$ if $x = y$ and $d(x, y) = 1$, otherwise. This is called discrete metric.*

Example 2. *The Pythagorean theorem gives the most familiar notation of distance for points in $\mathbb{R}^n$. In particular, when given $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$, the distance f as*

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

1.1.1   Open Sets (In a metric Space)

Definition 2. *Let X be a metric space. A **ball B** of radius r around a point $x \in X$ is $B = \{y \in X | d(x, y) < r\}$*

Definition 3. *A subset $O \subseteq X$ is **open** if for every $x \in O$, there is a ball around x entirely contained in O*

For Example, Let $X = [0, 1]$. The interval $(0, 1/2)$ is open in X With an open set, it always has a chance to pick any point within the set, when taken an infinite dimensional step in any direction within the given space. That means for example, the interval $[0, 1/2)$ is not open in $\mathbb{R}$ where as same interval is open in $[0, 1]$

1.1.2   Closed Sets(In a metric space)

Definition 4. *A point z is a **limit point** for a set A if every open set containing z intersects A in a point other than z. Any*

set C is a closed set, if and only if it contains all of its limit points.

For example, Let $A = \mathbb{Z}$, a subset of $\mathbb{R}$. This is closed set because it does contain all of its limit points; no point is a limit point! A set that has no limit points is closed, by default, it contains all of its limit points.

From these definitions, it is obvious to see that, the empty set is open and whole space is open. Moreover the union of open sets of any collection of open sets is open, and the intersection of any finite number of sets is open. Similarly every union of closed sets is closed, and finite intersection of closed sets is closed.

1.1.3 Topological Space

Definition 5. *A **topological space** is a pair (X, τ) where X is a set and τ is a set of subsets of X satisfying following axioms, τ is called a topology*

1. *The empty set $\emptyset$ and the space X are both sets in the topology.*

2. *The union of any collection of sets in τ is contained in τ.*

3. *The intersection of any finitely many sets in τ is also contained in τ.*

Example 3. *Let X be a set and τ is a a topology on X, the sets in τ are called open. Therefore, if X does have metric (a notion*

*of distance), then $\tau = \{all\ open\ sets\ with\ the\ ball\}$ is indeed a topology, this is called **Euclidean Topology** .*

1.2   Continuity

Definition 6. *A function $f : X \to Y$ is continuous if and only if the pre image of any open set in Y is open in X, similarly the same applicable in closed set as well. That means the pre image of any closed set in Y is closed.*

Given a point x of X, a subset N of X is called a neighborhood of X, if one can find an open set O such that $x \in O \subset N$

A function $f : X \to Y$ is continuous, if for any neighborhood V of Y there is a neighborhood U of X such that $f(U) \subset V$. Note that a composition of 2 continuous function is also continuous.

1.2.1   Homeomorphism

Homeomorphism is the notion of equality in topology. A classic example in topology for Homeomorphism views doughnut and coffee cup are topologically same because one of the geometric objects can be stretched and bent continuously one from the other. Formally it is

Definition 7. *A homeomorphism is a function $f : X \to Y$ between two topological spaces X and Y that f is a continuous bijection, and it has a continuous inverse function f^{-1}.*

Two topological spaces X and Y are said to be homeomorphic, if there are continuous maps $f : X \rightarrow Y$ and $g : Y \rightarrow X$ such that,

$fog = I_Y$ and $gof = I_X$. Moreover, the maps f and g are inverses of each other.

The homeomorphism forms an equivalence relation of the class of all topological spaces.

- Reflexive : X is homeomorphic to X.

- Symmetry: X homeomorphic to Y, then Y homeomorphic to X.

- Transitivity: X homeomorphic to Y and Y homeomorphic to Z, then X homeomorphic to Z.

  The resulting equivalence classes are called homeomorphic classes

Two spaces X, Y are **homotopy equivalent**, if there exists a homotopy equivalence $X \rightarrow Y$ and is denoted by $X \simeq Y$. One should think of homotopy equivalent spaces as spaces, which can be deformed continuously one into other. Any homeomorphism $f : X \rightarrow Y$ is a homotopy equivalence with a homotopy inverse f^{-1}, but the converse need not necessarily be true.

For example the spaces S^1 and $\mathbb{R}^2 - \{0\}$ are homotopy equivalent spaces. The following three graphs are homotopy equivalent, but they are not homeomorphic [1].

In fact, they are homotopy equivalent to the disk with two points removed by the homotopies indicated by the following picture
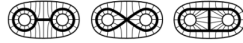
Definition 8. *A space is* **contractible***, if it is homotopy equivalent to one-point space.*

$\mathbb{R}^n$ is contractible.

1.2.2   Topological Invariants

Holes

A hole is a topological structure, which prevents the object from being continuously shrunk to a point. When a one dimensional shape in two dimensional space has a hole, it is not continuously shrinkable to a single point. That means a function mapping a space with a hole to a space without hole cannot be homeomorphism. They are also referred as voids. This is the key concept described in this thesis (called Betti numbers). The details of the $n-$dimensional holes is described in the following sections.

Note that, the tearing of the circle means that nearby points on the circle can be mapped to a very distant points in the line. This violates the continuity requirements of a homeomorphism.

Compactness

This is fundamental topological invariant. This explains the notion of a subset of Euclidean space being closed and bounded. An *open cover* of a

topological space X is a family of open subsets of X whose union is all of X. If F is an open cover of X. Let F_i is a subfamily of F and $\bigcup F_i = X$, then F_i is called the *subcover* of X. For example, Let F be the set of open balls of radius 1 whose centers have integer coordinates $\{x, y | x^2 + y^2 \leq 1\}$. The family of open balls covers the plane but if any one of the ball is removed, the family no longer covers the plane. Therefore this family of open balls has no proper subcover.

Definition 9. *A topological space X is compact if every open cover of X has a finite subcover.*

For example, the closed unit interval $[0, 1]$ is compact

Heine-Borel theorem states that, in $\mathbb{R}^n$ with the Euclidean topology, compact sets are precisely the closed and bounded sets of $\mathbb{R}^n$

Topogically Equivalent

Let X,Y be two topological spaces. Let $f : X \rightarrow Y$ be a bijective function, which is homeomorphism between X and Y. If f and f^{-1} are both continuous, then X and Y are said to be *topologically equivalent* or *homeomorphic*. That means they have same topology type. If a function $g : X \rightarrow Y$, whose restriction to the image $g(X) \subseteq Y$ is a homeomorphism. The triangle and the circle are topologically equivalent. Similarly the tetrahedron and the 2-dimensional sphere are topologically equivalent.

1.2.3   Surfaces

A surface is a topological space S such that every $s \in S$ has a neighborhood homeomorphic to $\mathbb{R}^2$. Which are also called 2-Manifolds. Manifolds can be explained in two different types: with or without boundary.

Definition 10. *A n-manifold(without boundary) is a topological space M such that every $x \in M$ has a neighborhood homeomorphic to $\mathbb{R}^n$*

One can get *manifolds with boundary* by removing open disks from manifolds without boundary. The boundary of n-dimension manifold is a $(n-1)$ Manifold. Here are examples for each of these.

1. The cylinder $\mathbf{S}^1 \times \mathbf{I}$ which is $\mathbf{S}^1 \times [0,1]$ (here, $\mathbf{I}$ denotes the $[0,1]$ in the usual topology) is a 2-manifold with boundary and its boundary consists of two closed curves. It can be connected from a square by gluing left edge to right edge.

Orientation

Which is a geometric notion that in two dimensions, allows one to say when a cycle goes around clockwise or counterclockwise, and in 3-dimensions when a figure is left handed or right handed. Technically, suppose for a closed manifold M without boundary with n-dimension, and for any point $x \in M$, a local orientation of M at $x \in M$ is a choice of generator μ_x which is n-dimensional loop.

The purpose of orientation definition is to see the spaces being glued with edges that give interesting new spaces and classify them according to it, whether they are *orientable surfaces* or not (non-orientable). As explained above, in cylinder $\mathbf{S} \times \mathbf{I}$ case, the opposite edges are being glued with same orientation. If it is done in reverse orientation, the resulting surface is then called *non-orientable* surface. *Mobius strip* is the surface obtained by gluing opposite edges in reverse orientation. For example, here are few constructions which give clear idea of what it means.

Torus $\mathbf{S} \times \mathbf{S}$ is obtained by gluing the top to the bottom edge,

then the boundary is removed and get compact surface, and glued the left edge and right edge with same orientation. In case, these side edges are being glued with reverse orientation, the resulting surface is the Klein bottle.

Triangulation

Triangulation is decomposition of a surface into triangles whenever any two triangles meet in a shared edge or shared vertex. By Euler characteristic of convex polyhedron definition [13] as the alternating sum of simplices (vertices, edges, faces, etc....), Triangulation does not cause any change in the value of Euler characteristic of the surface.

Connected Sum of Surfaces

The *connected sum* of any two manifolds is obtained by removing a disc in each manifold and gluing them together. The connected sum of two manifolds is again a manifold. The connected sum of two orientable manifold is again an orientable manifold, and the connected sum of non orientable manifold to either orientable manifold or non-orientable manifold is a non orientable manifold. This is denoted by '$\#$'. For example, the connected sum of two projective planes gives Klein bottle which is $\mathbf{K}^2 = \mathbb{R}^1 \# \mathbb{R}^1$

The two connected compact 2-manifolds (without boundary) are homeomorphic, if and only if they have the same Euler characteristic and they are both orientable and non-orientable. This is the classification theorem.

The orientability plays major role in computations when one is considering the coefficient group.

Chapter 2

Homology

Let X be topological space, one can define chain groups $C_p(X)$ of a topological space X for all $p \in \mathbb{N}$, and the boundary homomorphism $C_p(X) \to C_{p-1}((X)$

2.1 Free Abelian Group

Let X be a set, informally, the *free abelian group* with basis X is the formal linear combination of elements of X with coefficients in $\mathbb{Z}$ i.e. expressions

$$c = \sum c_x x$$

where all $c_x \in \mathbb{Z}$ and that $c_x = 0$ for all but finitely many $x \in X$. Two such expressions are equal if and only if the coefficients n_x agree for all $x \in X$. Addition is defined as, let $c^{'} = \sum c^{'}_x x$, then

$$c + c^{'} = \sum (c_x + c^{'}_x) x$$

Definition 11. *Let $\mathbb{Z}X$ be the group of formal linear combinations $\sum c_x x$ with addition defined as above*

Moreover, the set of functions $c : X \to \mathbb{Z}$ such that $\{x \in X | c(x) \neq 0\}$ is finite For each $x_0 \in X$ there is a corresponding element of $\mathbb{Z}X$, viz. the linear combination $\sum_{x \in X} c_x x$ with $c_{x_0} = 1$ and $c_x = 0$ for all $x \neq x_0$. The same notation can be used for $x_0 \in \mathbb{Z}X$ for this element associated to $x_0 \in X$.

Lemma 1. *Let A be an abelian group. The two homomorphisms $\phi, \psi : \mathbb{Z}X \to A$ for all $x \in X$*

For any function $f : X \to A$ there exists a homomorphism $\phi(x) = f(x)$ for all $x \in X$

Proof. Since any linear element of $\mathbb{Z}X$ is a $\mathbb{Z}-$linear combination of elements of X, and homomorphisms preserve linear combinations, two homomorphisms from $\mathbb{Z}X$ must agree if they agree X.

Given $f : X \to A$, define ϕ by $\phi(c) = \sum c_x f(x)$ if $c = \sum c_x x$. This is well defined because the numbers c_x are uniquely determined by c and satisfies $\phi(x) = f(x)$ for all $x \in X$. It is a homomorphism by the definition of addition in $\mathbb{Z}X$. $\square$

2.2   Standard simplices and their faces

A simplicial complex is a set of points (vertices), line segments(edges), triangles, and n-dimensional faces. Basic ingredients for the simplicial complex are vertices, all higher-dimensional faces can be determined as subsets of those vertices.

Definition 12. *For any integer $p \geqslant 0$, Let Δ^p be the space*

$$\Delta^p = \{(t_0, ..., t_p) \in \mathbb{R}^{p+1} | \sum_{i=0}^{p} t_i = 1, \forall i : t_i \geqslant 0\}$$

equipped with the subspace topology from $\mathbb{R}^{p+1}$ in its usual Euclidean topology.

For example, Δ^0 is a one-point space, Δ^1 is line segment or homeomorphic to an interval, Δ^2 is a solid triangle, Δ^3 is a tetrahedron shown in following figure 2.1. In general, the topological space Δ^p is called the standard p-simplex.

It can be written $e_0, ..., e_p \in \mathbb{R}^{p+1}$ for the standard basis, e.g $e_0 = (1, 0, ...0)$, $e_1 = (0, 1, ...0)$,etc. Then the point $t = (t_0, ...t_p) \in \Delta^p$ may be written as the linear combination $t = t_0 e_0 + ... + t_p e_p$.

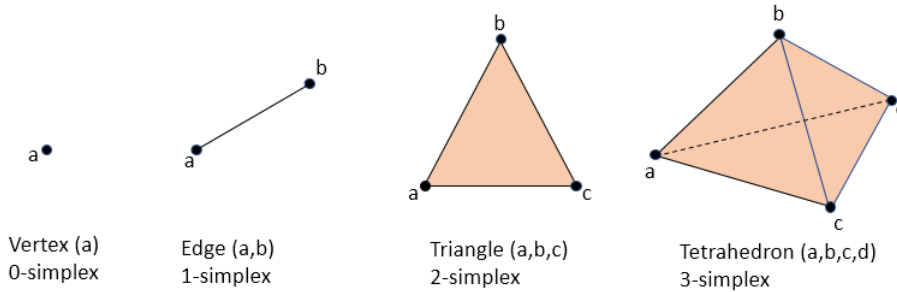| Vertex (a) | Edge (a,b) | Triangle (a,b,c) | Tetrahedron (a,b,c,d) |
| 0-simplex | 1-simplex | 2-simplex | 3-simplex |

Figure 2.1: Simplices

Definition 13. *For $0 \leq i \leq p$, let $\delta^i : \Delta^{p-1} \to \Delta^p$ be the map defined as,*

$$\delta^i(t_0, ..., t_{p-1}) = (t_0, ..., t_{i-1}, 0, t_i .... t_{p-1})$$

The image $\delta^i(\Delta^{p-1}) \subset \Delta^p$ is called the i-th face of Δ^p, and the union $\cup_{i=0}^{p} \delta^i(\Delta^{p-1})$ is called the boundary of Δ^p and denoted by $\partial \Delta^p$.

For example as shown in figure 2.1, the two faces of Δ^1 are the two end points of line segment, the three faces of Δ^2 are three edges in a triangle, and the four faces of Δ^3 are the triangles in the boundary of solid tetrahedron, etc.

For convenience of computations, the coefficient group is taken in this thesis for these operations $\mathbb{Z}_2$, whose elements are 0 and 1, and these are called modulo 2 coefficients.

2.2.1 Chain Complexes

Another way of defining triangulation of a topological space X is a simplicial complex K, whose underlying space is homeomorphic to space X. Let $\mathbf{K}$ be simplicial complex with underlying topological space $\mathbf{X}$ which contains p dimensional simplices, a p-chain is a formal linear sum of p-simplices in $\mathbf{K}$ like defined above $\sum c_i \sigma_i$, here c_i are the coefficients and σ_i are the p-simplices(continuous maps $\sigma : \Delta^p \to \mathbf{X}$). The two p-chains are added component wise like polynomials. Let $c = \sum a_i \sigma_i$ and $c' = \sum b_i \sigma_i$ be two chains, then $c + c' = \sum (a_i + b_i)\sigma_i$ is the addition operation and moreover considering the coefficient group is $\mathbb{Z}_2$, so $1 + 1 = 0$, the p-chains together with addition operation form the group of p-chains, denoted by $(\mathbf{C}_p, +)$. Associativity can be easily seen, with neutral element $0 = \sum 0\sigma_i$ and the inverse of a_i is $-a_i = a_i$ because $a_i + a_i = 0$. Moreover $\mathbb{Z}_2$ is abelian implies the group $(\mathbf{C}_p, +)$ is also abelian group. For convenience call the chain group as $\mathbf{C}_p$.

Boundaries and Cycles

The boundary of p-simplex is defined as the sum of $(p-1)$ - dimensional faces. They are $p+1$. The boundary of p-chain is the sum of boundaries

of its p-simplices. The boundary is $(p-1)$ chain. The boundary map is denoted by ∂ and the map between chain groups is shown below

$$\partial_p : \mathbf{C}_p \to \mathbf{C}_{p-1}$$

Which has homomorphism and hence is also called p - boundary homomorphism or boundary map. Since the relation is homomorphism with group operations, the maps and groups dimension wise arrangement give **Chain Complex** of **K** which is,

$$..... \xrightarrow{\partial_{p+2}} \mathbf{C}_{p+1} \xrightarrow{\partial_{p+1}} \mathbf{C}_p \xrightarrow{\partial_p} \mathbf{C}_{p-1} \xrightarrow{\partial_{p-1}} ....$$

A p - cycle is a p - chain with empty boundary and since ∂ commutes with addition, it is group of p-cycles and is denoted by $\mathbf{Z}_p$ or $\mathbf{Z}_p(\mathbf{K})$, which is a subgroup of p - chains, moreover this group is the kernel of the p-th boundary homomorphism. Similarly the group of p-boundary homomorphisms is also denoted by $\mathbf{B}_p$ or $\mathbf{B}_p(\mathbf{K})$ and which is also subgroup of the group of p-chains. Clearly this is image of p-boundary homomorphism.

A p-boundary is a p-chain which is boundary of $(p+1)$ -chain indeed, For $0 \leq p \leq \dim \mathbf{K}$ are non trivial groups. Since the boundary of a vertex is empty so it is clear that 0-chain maps to 0, moreover here the interesting points are the composition $\partial \circ \partial = 0$ is the trivial homomorphism and hence image of the homomorphism $\partial : \mathbf{C}_{p+1} \to \mathbf{C}_p$ is contained in the kernel of homomorphism $\partial : \mathbf{C}_p \to \mathbf{C}_{p-1}$. Since the p-boundaries form subgroups of p-cycles, now it is time to introduce quotients, that means one can partition each cycle group into classes of cycles that differ from each other by boundaries. This quotient group is called the p-th homology group.

2.2.2   Homology

The p-th homology group can be defined as the quotient of p-cycles mod out by p-boundaries

$$\mathbf{H}_p(\mathbf{K}) = \mathbf{Z}_p(\mathbf{K})/\mathbf{B}_p(\mathbf{K}) = \frac{ker(\partial : \mathbf{C}_p \to \mathbf{C}_{p-1})}{img(\partial : \mathbf{C}_{p+1} \to \mathbf{C}_p)}$$

Two chains $c, c' \in \mathbf{C}_p(\mathbf{K})$ are homologous, if $c - c' \in \mathbf{B}_p(\mathbf{K})$, and if there exist an $\alpha \in \mathbf{C}_{p+1}(\mathbf{K})$ with $\partial \alpha = c - c'$, this is an equivalence relation, and $\mathbf{H}_p(\mathbf{K})$ is the group of p-cycles modulo the relation of being homologous.

These are the homology groups of the spaces - circle, sphere and torus. with coefficients in $\mathbb{Z}$ are isomorphic to following groups

$$H_0(\mathbf{S}^1) \cong \mathbb{Z}, H_1(\mathbf{S}^1) \cong \mathbb{Z}, H_2(\mathbf{S}^1) \cong 0$$
$$H_0(\mathbf{S}^2) \cong \mathbb{Z}, H_1(\mathbf{S}^2) \cong 0, H_2(\mathbf{S}^2) \cong \mathbb{Z}.$$
$$H_0(\mathbf{S}^1 \times \mathbf{S}^1) \cong \mathbb{Z}, H_1(\mathbf{S}^1 \times \mathbf{S}^1) \cong \mathbb{Z} \oplus \mathbb{Z}, H_2(\mathbf{S}^1 \times \mathbf{S}^1) \cong \mathbb{Z}.$$

The homology groups of same spaces with coefficients in $\mathbb{Z}_2$ are isomorphic to indeed same combinations with $\mathbb{Z}_2$ group, instead of $\mathbb{Z}$. The benefit of considering the homology with $\mathbb{Z}/2\mathbb{Z}$ coefficients is, there is no need to worry about the orientation of the simplices. For computations and dealing with Gaussian elimination in matrices, considering this group is a good option.

Euler Characterstic

Let $\mathbf{A}$ be finitely generated abelian group, from fundamental theorem of algebra, one may find isomorphism

$$A \cong \mathbb{Z}^d \oplus T$$

where T is finite abelian group (torsion), the number of copies $\mathbb{Z}$, and d is the *rank* of A.

Definition 14. *Let* X *be a topological space such that $H_p(X)$ is finitely generated for all p such that there exists N with $H_p(X) = 0$ for all $p > N$ for such* X,

$$\chi(\mathrm{X}) = \sum_{i=0}^{\infty} (-1)^i rank(H_i(X))$$

The rank of the homology group is called the **Betti number**. If we compare the alternating sum of Betti numbers of corresponding simplices and Euler characteristic numbers, it is the same for common surfaces. Betti numbers are topological objects which were proved to be invariants by Poincăre and used by him to extend the Polyhedra formula to higher dimensional spaces. The numbers β_0, β_1, β_2 of any space represents the connected components, one-dimensional holes and cavities of that space respectively. The rank of the p-th homology group is the p-th Betti number of **k**. Moreover if substitute the dimension of quotient group [11] and from that the rank of the quotient group i.e.,

$$rank(\mathbf{H}_p) = rank(\mathbf{Z}_p) - rank(\mathbf{B}_p) \tag{2.1}$$

This rank is a measure of the difference between the p-th cycle group and p-th boundary group.

Chapter 3

Voronoi Diagram and Delaunay Triangulation

3.1 Convex Polygons

A set X is convex if two points $x, y \in X$ i.e., every point on the line segment connecting x and y belongs to X. For example, a disk in $\mathbb{R}^2$ is convex but a circle is not.

Note that the intersection of convex sets is convex.

Convex polygons can be constructed as the intersection of a finite number of half-planes or as the convex hull of a finite set of points.

3.2 Voronoi Diagrams

Let S be a finite set of points in $\mathbb{R}^2$, in order to distinguish them from other points in the plane, these elements are called ***sites***. It is known that the Euclidean distance between a point $x = (x_1, x_2)$ and a site $s = (s_1, s_2)$

is $||(x-s)||$ is $(x-s, x-s)^{1/2}$, the **Voronoi Region or Voronoi cell** of s is defined as,

$$V_s = \{x \in \mathbb{R}^2 | ||x-s|| \le ||x-t||, \forall t \in S\} \qquad (3.1)$$

The set of points those satisfy $||x-s|| \le ||x-t||$ is a closed half-plane, V_s is the intersection of finitely many half planes, hence it is a convex polygon. The Voronoi regions intersect at most, along their boundaries. The Voronoi regions covers the entire plane. The set of all these Voronoi regions is the **Voronoi Diagram** of S (Figure 3.1).
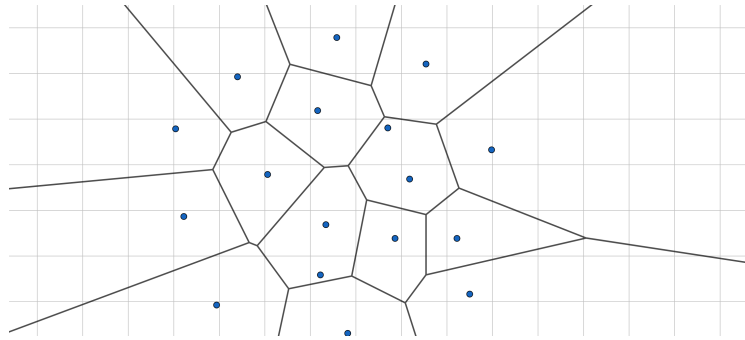

Figure 3.1: Voronoi Diagram for finite set S of points $\mathbb{R}^2$

The Voronoi ball of any site s w.r.t S is defined as the intersection of the Voronoi region with the closed ball of radius say r around that point.

In easy words, imagine the concept by taking arbitrary finite set of points in a plane and draw the circles around each point with some radius r. Start increasing the radius of circle, at some point, they all intersect and it looks, the whole plane contains union of circles. But in Voronoi diagrams, they are not merging with neighboring circles. Instead the circles with neighboring sites in the plane start colliding at a single point (like horizontal 8 figure). As they expand, they grow into a line where they squish together. The lines are drawn in the circle boundaries. If this happens with all the

finite sites, it looks like irregular patterns as shown in Figure 3.1. Moreover, the border of each Voronoi region(cell) is always at the same distance to the two nearest sites, and whenever three lines meet, there obtained a vertex and that is equally distanced to the three nearest sites.

3.3   Delaunay Triangulation

Suppose S is a Voronoi region in $\mathbb{R}^2$, the Delaunay triangulation is obtained by connecting two sites by a straight edge, whenever the corresponding Voronoi regions shared by an edge (the boundary of Voronoi region). The intersection of four or more Voronoi regions is empty, and the intersection of three Voronoi regions form a triangle in Delaunay Triangulation, is shown in Figure 3.2.
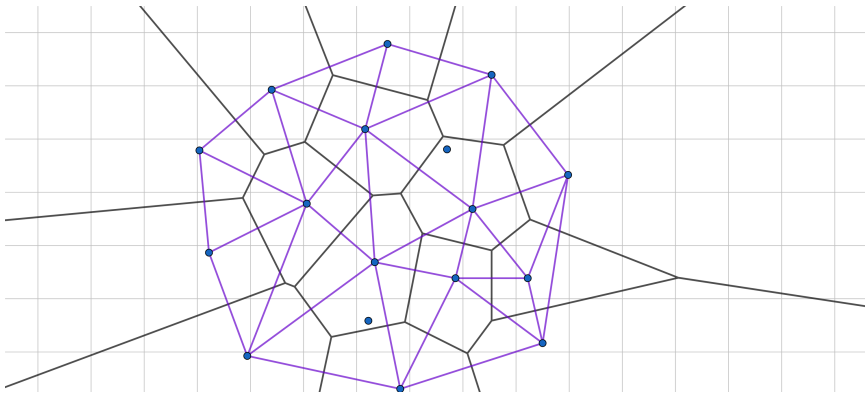

Figure 3.2: Voronoi Diagram and corresponding Delaunay triangulation of finite set S of points $\mathbb{R}^2$

The **star** of a site in the Delaunay triangulation is defined as the collection of edges and triangles that share s. The **link** of s is the collection of sites and edges in the boundary of the star that doesn't contain s. The

triangles in the Delaunay triangulation decompose the convex hull of S.

Chapter 4

Complex Constructions

The main reason to study the concept of alpha shapes is to develop concrete explanation of **shape** of a finite point set.

4.0.1 Jarvis Construction

Given a set of points in plane, the convex hull of a set is the smallest convex polygon (Figure 4.1), that contains all of the points in that plane.

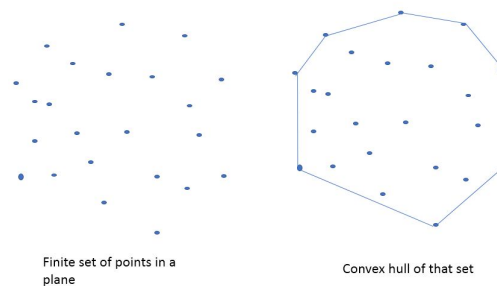Finite set of points in a
plane

Convex hull of that set

Figure 4.1: Convex hull: A bound around the plane

Jarvis March algorithm solves the convex hull of set of points such that all the points are covered and the boundary is minimal. The basic idea starts with choosing the left most point and pick the next point as left most point of starting points, and repeat the process until it reaches the initial point. Left most point can be determined by cross product.

The time complexity of this algorithm is $O(n, h)$, n is number of points in the set and h is the number of points in the boundary. If h is small, the algorithm is very efficient. If not, it would be $O(n^2)$, that means it has to check every point until it gets the boundary. Space complexity is something that might be holding all the collinear points.

Non convex can be drawn in generalized way, that is instead of rotating line, one can draw line segment by decreasing length (fixed length). Since the line segment should be drawn with fixed length, it wouldn't get nice shape unless the sequence of points(sites) in the plane are nicely distributed.

4.1   Alpha Shapes

The α-**shape** [2] of a finite set of points for arbitrary real α. This notion is generalization of one common definition of the convex hull. The main objective is to construct efficient algorithm for the α-shapes of a point set for several α's. There is a close connection between α-hull and α-shapes with Delaunay Triangulation that will be explained in following section.

Some relevant definitions are as following.

Let S be a set of n (n being positive integer) points in the plane. The convex hull of S may be defined as the intersection of all closed half planes that contain all the points of S.

Definition 15. *For an arbitrary sufficiently small real value α,*

the α-hull of S is the intersection of all closed discs with radius $1/\alpha$ that contain all the points of S.

In order to get an intersection of discs, there exists at least one disc of the chosen size that contain all the points. The smallest possible value for $1/\alpha$ is exactly equal to radius of the smallest enclosing circle.

Definition 16. *For negative real α's, the α-hull is defined as the intersection of the complements of discs where radii of these discs $-1/\alpha$ that contains all the points in S.*

Let us assume the intersection of no discs and discs with large positive α is equal to the entire plane, general disc radius be $1/\alpha$ for $\alpha > 0$, the complement of a disc of radius $-1/\alpha$, if $\alpha < 0$ and a half plane if $\alpha = 0$, then the family of α-hull ranging from $-\infty$ to ∞. The family contains entire plane for sufficiently large α's, the smallest enclosing circle of S, when its radius equals to $1/\alpha$ and convex hull of S for $\alpha = 0$ and S itself, when α ls sufficiently small.

Definition 17. *A point a in a set S is called as α-**extreme** in S, if there exists a closed generalized disc of radius $1/\alpha$ such that a lies on its boundary and it contains all the points of S. For two such a, b as extreme points, there exists a closed generalized disc of radius $1/\alpha$ with both points on its boundary, which contain all the points of S, then a and b are said to be α-**neighbors**.*

Definition 18. *Let S be a set of points and a real valued α, the α-**shape** of S is the straight line graph, whose vertices are the α-extreme points and edges are connections between the respective α-neighbors.*

If any space or point cloud data is provided, it can then be decomposed into Voronoi regions. The dual space of that decomposition is Delaunay triangulation. In fact α-complex is Delaunay triangulation and through this complex, the filtration of the complex is defined. Here, s denotes site in Voronoi region.

4.1.1   Union of disks

Let x be a point and be center of disk constructed around each point with radius α in the space of points, and note x is the center of an empty disk of radius α, if and only if it is further than α. Then the union of all the disks of radius α centered at x is the entire region covered by the disks.

$$\mathbf{U}_s(\alpha) = \bigcup_{s \in S} D_s(\alpha) \tag{4.1}$$

4.1.2   Voronoi Decomposition

It is a formulation of α-shape as the union of simplices in the α-complex. i.e., overlaying of the union of disks with the Voronoi diagram, decomposing the union into convex regions.

$$R_s(\alpha) = V_s \cap D_s(\alpha) \tag{4.2}$$

The intersection of convex set is again a convex set, so

$$\mathbf{U}_s(\alpha) = \bigcup_{s \in S} R_s(\alpha) \tag{4.3}$$

The region R_s covers the entire union without overlapping. The common intersection of regions are shared edges and vertices only. So the Delaunay triangulation is constructed as follows. Construct a α-complex between two sites if the regions are intersecting in common edge, and draw a triangle

among the regions of the respective sites if all of them intersect in common point. Note that, if three regions of three sites are intersecting with common vertex, a triangle can be drawn as shown in Figure 4.3. This is the convex of Nerve theorem. The construction is explained in the following Figure 4.3 which give an idea [3]
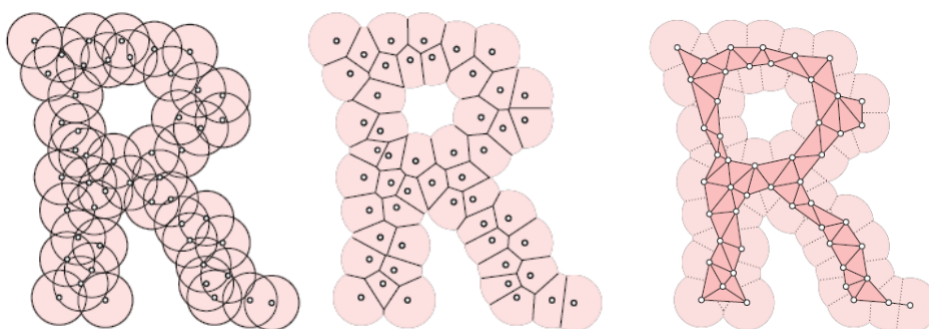

Figure 4.2: *Left*-The union of disks with radius α, *middle*- The Voronoi decomposition of the union, *right*- The α-complex is superimposed on the union of disks, convex regions of Voronoi decomposition
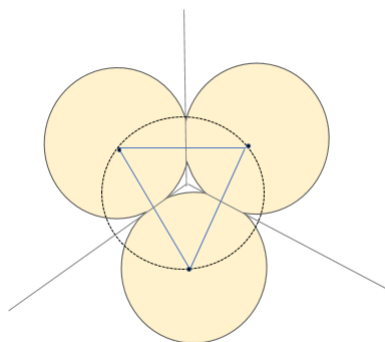

Figure 4.3: The intersection of three Voronoi regions

4.2   Visualization of Alpha shapes of a point cloud data

In this study, a sample point cloud set with 400 points is taken, and the α-shapes with different values of α are constructed as shown below (Figure 4.4).
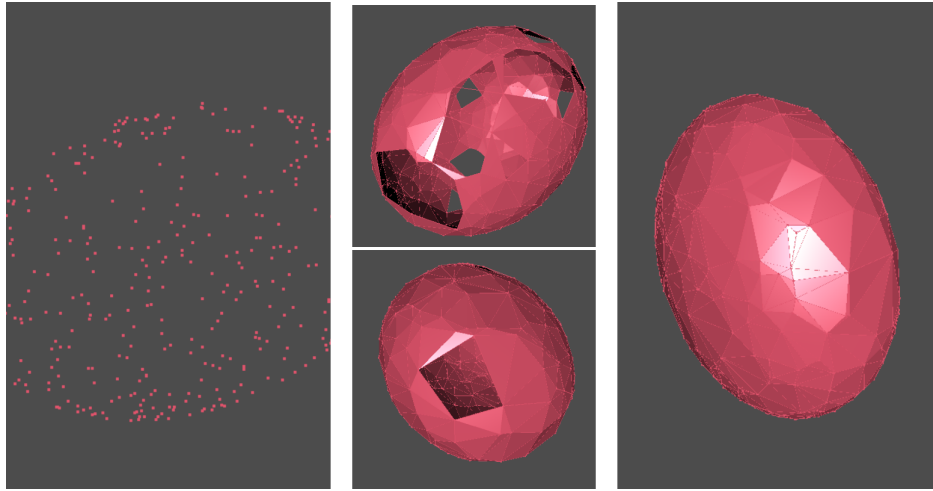

Figure 4.4: Left is a sample point cloud with 400 points, rest of the figures are α-shapes of the point cloud at various α's $0.2, 0.25$ and 0.5 respectively

In the centre-top sub-figure of Figure 4.4 the α-shape has holes and has big tunnel ([2]). In the centre-bottom sub-figure of Figure 4.4 the α-shape has most of the holes disappeared, except one. But it still has enclosed volume which is homotopy equivalent to 2-dimensional sphere with removed 1-dimensional disk. The right side sub-figure of Figure 4.4 is complete convex α-shape with enclosed volume (seen in middle portion of the figure).

Another example of point cloud α-shapes construction is shown in Figure 4.5. In this case, 450-points are distributed in 3-dimensional space. It is identified as circular distribution via α-value of 0.15, and by increasing the value of α to 0.25, 0.5 and 1.0, the respective α-shapes are constructed and

analyzed. For α-value 0.5, the α-shape of point cloud has holes, and they begin to merge. When α-value is 1.0, it turns out to be a giant α-complex.
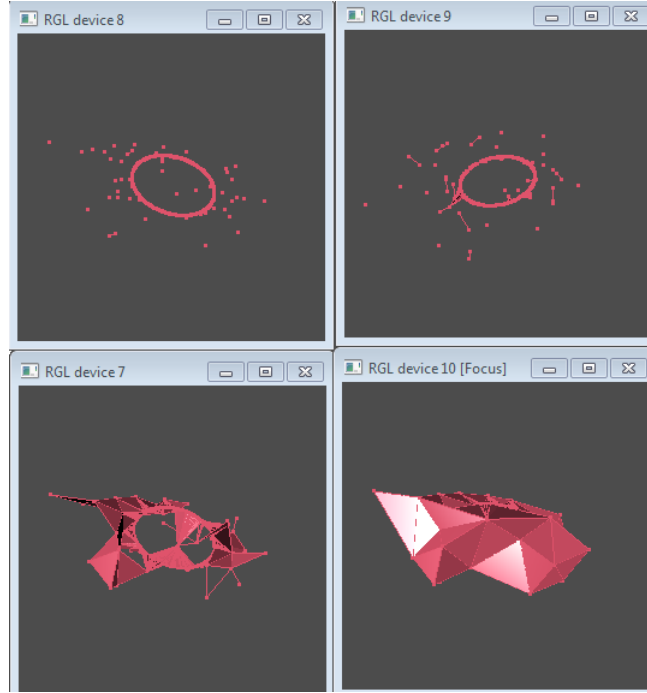

Figure 4.5: The alpha shapes of the another sample of point clouds, the alpha complex constructed with α-values

In this thesis, the α-complex construction is done in **R**(http://cran.r-project.org/). **R** is a programming language used for statistical computing and graphics. R-programming has certain nice properties and it has many packages for computing and visualizations. It is very easy to make inter-active plots via**R**-programming. Since it is a scripting language, it is easy to use, but **R** is slower in execution when compared with **C** or **C++**. The **R**-packages used in the current analysis are "geometry", "rgl" and "al-phashape3d" for α-shape construction. It has to be noted that the input data should be in the required class. For example, the data used here is in the form of list, it is then converted into double class in order to construct α-shape on to it.

```r
############### Installing Required Packages ###############
########### Here we are using "alphashape3d" ################
install.packages("geometry")
install.packages("rgl")
install.packages("alphashape3d")


############# Loading Package ###########################
library("geometry")
library("rgl")
library("alphashape3d")
############################################################
Sample_Point_cloud <- read.csv(file.choose(), header = T)
## This is "data frame" class, we are making this into matrix form
###### constructing alpha shape for this sample ############3
sample_matrix_PC <- data.matrix(Sample_Point_cloud[1:400,1:3])
alphashape3d_Sample_PC <- ashape3d(sample_matrix_PC, alpha = 0.5)
    #plotting with 4 different values of alpha
######### plotting the alpha shape ####################
plot(alphashape3d_Sample_PC)
####### sample of circular alpha shape ########################
sample_circ_data <- data.matrix(PC3)
alphashape3d_circ <- ashape3d(sample_circ_data, alpha = 0.5)
alpha <- c(0.15,0.25,0.5,1)
alphashape3d_circ <- ashape3d(alphashape3d_circ, alpha = alpha)
plot(alphashape3d_circ, indexAlpha = 1:4)
Device 7  : alpha = 0.5
Device 8  : alpha = 0.15
Device 9  : alpha = 0.25
Device 10  : alpha = 1
```

4.3 Complex constructions on Point clouds

Homology can be computed for spaces if they are equipped with a triangulation which is a homeomorphism to a simplicial complex. A point cloud is a geometric object which is a set of points, floating around randomly in a space. Topologically it is a discrete space i.e., they are isolated from each other. In order to obtain the information from the point cloud, one can construct complex on it, and computations can be done on it.

Abstract Simplicial Complex: An Abstract Simplicial Complex of a space is a system of sub collection $\mathbf{A}$, the finite sets are called faces (vertices, edges and faces). Every sub-face of a face is again face of that complex i.e. if $\alpha \in \mathbf{A}$ and $\beta \subseteq \alpha$ imply $\beta \in \mathbf{A}$.

4.3.1 Čech Complex

Let $\mathbf{X}$ be a topological space and let $u = \left\{ U_\alpha \right\}_{\alpha \in \mathbf{A}}$ of $\mathbf{X}$ indexed by a set $\mathbf{A}$

Definition 19. *The Čech complex of* u, *$C(u)$, is the simplicial complex whose vertex set is $\mathbf{A}$, and where a subset $\{\alpha_0, \alpha_1, ...\alpha_k\}$ is a simplex, if and only if $\mathbf{U}_{\alpha 0} \cap \mathbf{U}_{\alpha 1} \cap ...\mathbf{U}_{\alpha k} \neq \emptyset$*

The Čech complex of covering of u is homotopy equivalent to $\mathbf{X}$, so the homology of covering space is isomorphic to homology of $\mathbf{X}$ i.e. if all the sets of the form $U_{\alpha 0} \cap U_{\alpha 1} \cap ...U_{\alpha k}$ are either empty or contractible, then $C(u)$ is homotopy equivalent to X.

If $\mathbf{S}$ is a finite subset of a metric space, it can be written as $\mathbf{C}_{(B_\epsilon)}$, where B is the collection of metric balls $\{B_\epsilon(s) | s \in \mathbf{S}\}$ in the case of Euclidean data. Here, the union of balls $\mathbf{S}_\epsilon$, which is homotopy equivalent

to $\mathbf{X}$, where $\mathbf{S}$ is sampled from a space $\mathbf{X} \subseteq \mathbb{R}^n$. Induced homology is also isomorphic to homology of $\mathbf{X}$

4.3.2   Vietoris-Rips Complex

The Vietoris-Rips complex characterizes the topology of a point set. This complex is popular in Topological Data Analysis and its construction can easily be extended to higher dimensions. This complex needs distance between points in a space, which is an abstract simplicial complex defined on a finite metric space.

Definition 20. *Let* X *be a metric space with radius* r, *for any finite subset* S *of* X, *the Vietoris-Rips Complex* $\mathrm{VR}_r(\mathrm{X})$ *is all finite subsets of diameter at most of* r. *The set* $\{\sigma \subseteq \mathrm{X} | diameter(\sigma) \leq r\}$ *where* $\sigma \in \mathrm{S}$ *is finite.*

4.3.3   Nerve

Let $\mathbf{X}$ be a finite collection of sets. The *nerve* of $\mathbf{X}$ is the system of sub collections of $\mathbf{X}$ whose sets have non-empty common intersection.

$$\mathbf{NrvX} = \{\emptyset \neq V \subseteq \mathbf{X} | \cap V \neq \emptyset\}$$

Theorem 1. *Nerve Theorm: If* V *is a finite collection of* $\boldsymbol{X}$ *with all non-empty intersections of subcollections of* V *is contractible, then* $\boldsymbol{Nrv}(V)$ *is homotopic to the union of elements of* V

If the cover of the sets is sufficiently nice, then the Nerve Theorem says that the nerve of the cover and the underlying space X have the same homology group. For example, let S be a finite set points in a metric

space X, and for every $\epsilon > 0$ the space S_ϵ as the union $\bigcup B(x, \epsilon)$ where $B(x, \epsilon)$ denotes the closed ball with radius ϵ centered at x. It follows that $\{B(x, \epsilon) | x \in S\}$ is a cover of S_ϵ, and nerve of this cover is the Čech complex on S at scale ϵ. If the space X is Euclidean space, then the Nerve theorem guarantees that the simplicial complex $\check{C}_\epsilon(S)$ recovers the homology of S_ϵ.

In order to preserve one dimensional homology group, the Čech complex can be used. If the diameter of ball increases, the hole disappears, so when one perform the persistent homology, there is not a lot of difference because eventually the Čech complex does embed inside a Vietoris-Rips complex when increasing the diameter. So one can conclude that there is no big difference of which kind of homology is used for computing persistent homology [8].

Betti numbers (Figure 4.6) provides a signature of the underlying topology. If unknown data cloud is given, using these Betti numbers one can conclude the shape of data by computing persistent homology and compare whether it has same homotopic type that of standard topological space such as circle, torus, Klein or sphere etc.

From a computational point of view, the Rips complex is less expensive than the corresponding Čech complex, even though the Rips complex has more simplices. It is a flag complex, and is maximal among all simplicial complexes with the given 1-skeleton(topological graph) that completely determines the complex. The Rips complex can be stored as a graph and reconstructed instead of storing the entire boundary operator needed for Čech complex.

In order to convert a point cloud data set into a global complexes whether it is Rips or Čech, it requires a choice of parameter ϵ. If the chosen ϵ is sufficiently small, the complex is a discrete set. If it is sufficiently large, the complex is a single giant complex.
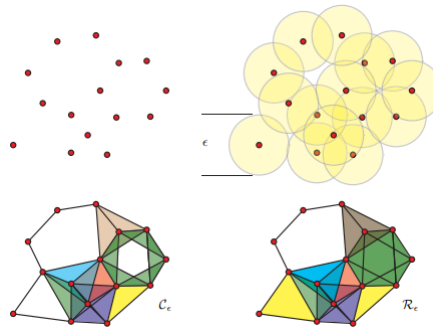
Figure 4.6: A fixed point set can be completed to a Čech complex C_ϵ or to a Rips complex R_ϵ based on a parameter ϵ

[10], This Čech complex has the homotopy type of the $\epsilon/2$ cover $(S^1 \vee S^1 \vee S^1)$, where as the Rips Complex has a homotopy type $(S^2 \vee S^1)$.

Chapter 5

Persistent Homology

5.0.1   Introduction

Persistent homology is a technique that has been developed over last 20 years. Persistent homology can be used to measure the scale or resolution of a topological space. Persistent homology is also a method that is used in topological data analysis (TDA) to study qualitative features of data that persist across multiple scales. TDA is a field that lies at the intersection of data analysis, algebraic topology, computational geometry, statistics, and other related areas. When it is time to handle large amount of data, the complexity arises in such form of noise in the data, higher dimension, and incomplete, and these are more significant challenges. The clustering techniques [5], machine learning from Computer Science, and uncertainty qualifications along with mathematical and statistical models are often useful for data analysis. TDA is a mathematical approach to help people to manage complicated point cloud data.

The main goal of TDA is to use ideas and results from geometry and topology to develop tools for studying qualitative features of the data. To achieve this, one needs precise definitions of qualitative features, tools

to compute them in practice and some acceptance about the robustness of these features. One mathematical way to address these features is a method in TDA called *Persistent homology*. This method is appealing for applications since it is based on *Algebraic topology*, which gives good understanding of theoretical framework and commutable via *Liner algebra*, and is robust with respect to small perturbations in input data.

The data sets that are studied with persistent homology includes finite metric spaces, digital images, level sets of real-valued functions, and networks. Which is also used in virtual chemical compounds testings, study the cancer cells data as well. In TDA literature, Finite-metric spaces are considered as point-cloud data sets. From Topological point of view, these types of data does not have any interesting information. Persistent homology is an algebraic method for discerning topological features (connected components, holes and enclosed volumes etc..) of data by building simplicial complex with different techniques (Vietoris-Rips complex, čech complex and a few more) in order to achieve Delaunay Triangulation. The output of persistent homology is the data of finite collection of intervals (birth and death of the homology). These finite collection of intervals are called barcodes. They can be visualized in two ways, viz. persistent diagrams and barcodes plot.

5.1   Persistentency

5.1.1   Filtered Complex

Let K be a finite simplicial complex and let $K_1 \subset K_2 \subset ...... \subset K_l = K$ be a finite sequence of nested subcomplexes of K. The simplicial complex K with a sequence of such subcomplexes is called a *filtered simplicial complex*. One can apply homology to each of the subcomplexes. For all p(dimension), the inclusion maps $K_i \rightarrow K_j$ induce $\mathbb{Z}_2$-linear maps $f_{i,j}$:

$H_p(K_i) \rightarrow H_p(K_j)$ for all $i, j \in \{1, ...l\}$ with $i \leq j$. By functionality it follows that

$$f_{k,j} \circ f_{i,k} = f_{i,j}$$

let $K_1 \subset K_2 \subset ...... \subset K_l = K$ be a filtered simplicial complex. The p-th persistent homology of K is the pair $\left(\{H_p(K_i)\}_{1 \leq i \leq l}, \{f_{i,j}\}_{1 \leq i \leq j \leq l}\right)$ for all $i, j \in \{1, ...l\}$ with $i \leq j$, the linear maps $f_{i,j} : H_p(K_i) \rightarrow H_p(K_j)$ are the maps induced by the inclusion maps $K_i \rightarrow K_j$ [7]
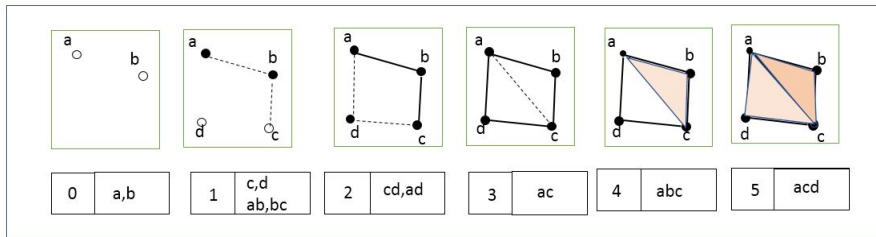
Figure 5.1: Example for increasing sequence of simplicial complex

The p-th persistent homology of a filtered simplicial complex give more refined information than just homology of the single subcomplexes.

In Figure 5.1, the filtration followed by the incremental algorithm [9] for Betti numbers is shown. For each sequence of filtration, by adding a simplex, either chain extension or cycle is created. If the added simplex creates a cycle, then one of the Betti number in that sequence($\beta_1 \beta_2 \beta_3$) decreases by 1, otherwise it increases by one, there by creating a chain extension.

The complex construction contains six increasing finite sequences of simplicial complexes $\mathbf{C}^0 \subset \mathbf{C}^1 \subset \mathbf{C}^2 \subset \mathbf{C}^3 \subset \mathbf{C}^4 \subset \mathbf{C}^5$, is starting with two vertices a, b at sequence 0, then in sequence 1, two vertices and two edges are added. In each sequence, the dotted line represents creation

of new edges. The light pink area in sequence 4 and sequence 5 represents
the formation of new triangle by boundary edges. In sequence 5, the dark
pink triangle represents the solid face triangle. The solid face triangle is
the boundary of tetrahedron, a 3-dimensional face. Moreover a,b are in
$\mathbf{C}_0^0 \subseteq \mathbf{C}_0^1 \subseteq \mathbf{C}_0^2 \subseteq \mathbf{C}_0^3 \subseteq \mathbf{C}_0^4 \subseteq \mathbf{C}_0^5$, and 2-dimensional simplex 'abc'
is in $\mathbf{C}_2^4 \subseteq \mathbf{C}_2^5$, the 'ab','ac','cd' and 'ad' represent edges. In the entire
sequences, the superscript of chains denotes the time of filtration level n,
and subscript denotes the dimension of simplices. So by the definition of
homology discussed earlier, consider the chain complex

$$..... \xrightarrow{\partial_3} \mathbf{C}_2 \xrightarrow{\partial_2} \mathbf{C}_1 \xrightarrow{\partial_1} \mathbf{C}_0 \xrightarrow{\partial_0} 0$$

The homology group H_0 is the (kernel of ∂_0)$/$ (image of ∂_1). Con-
sider the chain complex below here, the vertical map is the boundary map
of chain complex at n-dimension where one can find the homology and the
horizontal maps are inclusion maps where those maps embedded in next se-
quence of chain groups in the process of filtration. So that one can find the
homology at time n, for example in above filtration in the figure 5.1, we can
find homology at time 0 to 5

The following giant chain complex is the nested sequence of chain
subcomplexes with boundary homomorphisms in each dimension. In this
chain complex $C_k{}^p$ is not only inclusion map for $C_k{}^{p+1}$, but also the inclu-
sion map for all values greater than k and p. The induced homology of the
chain complex is: $\varphi_k^{i,j} : H_k(C_k^i) \rightarrow H_k(C_k^j)$

$$
\begin{array}{ccccccc}
\Big\downarrow \partial_3 & & \Big\downarrow \partial_3 & & \Big\downarrow \partial_3 & \\
C_2{}^0 & \xrightarrow{f^0} & C_2{}^1 & \xrightarrow{f^1} & C_2{}^2 & \xrightarrow{f^2} & \ldots \\
\Big\downarrow \partial_2 & & \Big\downarrow \partial_2 & & \Big\downarrow \partial_2 & \\
C_1{}^0 & \xrightarrow{f^0} & C_1{}^1 & \xrightarrow{f^1} & C_1{}^2 & \xrightarrow{f^2} & \ldots \\
\Big\downarrow \partial_1 & & \Big\downarrow \partial_1 & & \Big\downarrow \partial_1 & \\
C_0{}^0 & \xrightarrow{f^0} & C_0{}^1 & \xrightarrow{f^1} & C_0{}^2 & \xrightarrow{f^2} & \ldots \\
\Big\downarrow \partial_0 & & \Big\downarrow \partial_0 & & \Big\downarrow \partial_0 & \\
0 & & 0 & & 0 &
\end{array}
$$

This gives information about the homology at dimension k, and how long the homology is going to be persistent at time i to j. For example, consider the Figure 5.1, $H_0^{0,1}$ computes the zero dimensional cycles created at time 0 and checks whether they disappear at time 1, i.e. at time $0 + 1$ the zero dimensional cycles at time 0 are going to be boundaries at time 1. $\{< a, b > |(a + b)\}$ is the only generator for $H_0^{0,1}$. The interval of cycles created at time 0 and destroyed at time 1. Similarly the persistent homology can be computed at all time intervals.

$$
H_k^{i,p} = Z_K^{i,p}/(B^{i+p} \cap Z_k^i)
$$

which is $p-$persistent kth homology group.

Topological Noise and Features

If precise number of holes or voids for a good fraction of the complex in the underlying topological space is expected, sometimes during the filtration process it creates p-cycles which are not persistent in a long time and does not give any interesting information. The $n-$dimensional holes with short lifetimes are considered to be *topological noise*, where as the $n-$dimensional holes with a long lifetime are considered to be *topological features*.

Birth and Death

In the sequence of homomorphisms connecting the the homology groups of the complexes on the filtration [3], for $i \leq j$, K_i is a subcomplex of K_j which has the injective and inclusion map $f^{i,j} : K_i \hookrightarrow K_j$, is inclusion map since it is within same complexes. In case of p-cycles the inclusion map $f_p^{i,j} : Z_p(K_i) \hookrightarrow Z_p(K_i)$, it induces map on homology

$$\varphi_p^{i,j} : H_p(K_i) \to H_p(K_j)$$

Which is not inclusion map, since $\varphi_k^{i,j}$ does not depend on the choice of representative. It sends p-cycle in K_i to K_j.

In the Figure 5.1, The 1-dimensional homology group is generated by a 2-simplex $< abc >$, which is created in this complex. It is the **birth** of that homology group. It fills in the next sequence of complex in the filtration, that means the hole is filled. There is no more 1-dimensional homology group generated by $< abc >$, hence it dies in that sequence. That it is called **death** of that homology group.

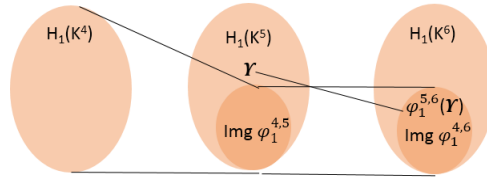Figure 5.2: The generator $< abc >= \gamma$ is created at K_5 and dies when entering the K_6

In Figure 5.2, $\gamma \in H_1(K_5)$ is born at K_5 and it dies at K_6 and $\varphi_1^{5,6}(\gamma)$ is the image of $\varphi_1^{4,6}$. So the index of persistency of the group H_1 is only one [3]

If γ surrounds a hole in K_i, and fills up at the time when it reaches K_j, then $\varphi_p^{i,j}$ sends γ to 0, which belongs to $H_p(K_j)$. The image of $\varphi_p^{i,j}$ is called the **Persistent homology group**, and it contains all p-dimensional homology classes that have generators already in K_i. The corresponding **Persistent Betti number** is called the rank of the persistent homology group.

Barcodes

Pairing up the births and deaths of n-dimensional homology group for $n \in \mathbb{N}$, one can get a set of intervals or bars (connecting line segment of death and birth). The set is called as **barcode** of the homology group. The following Figure 5.4 illustrates the barcode of the filtration. The construction of the simplicial complex can be followed with incremental algorithm. In each sequence, the Betti number changes by ± 1, depending on birth or death of a simplex. Short bars in the barcode diagram represent the topological noise. The long bars represent the persistent homology in the corresponding dimension.
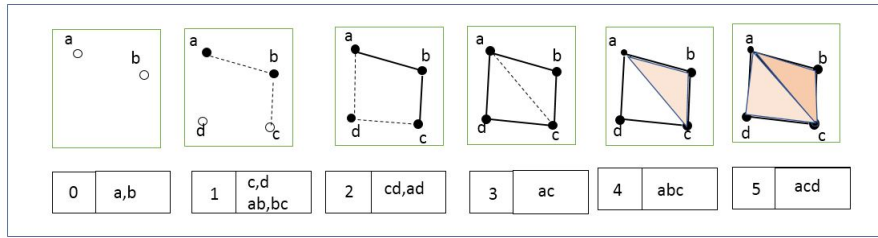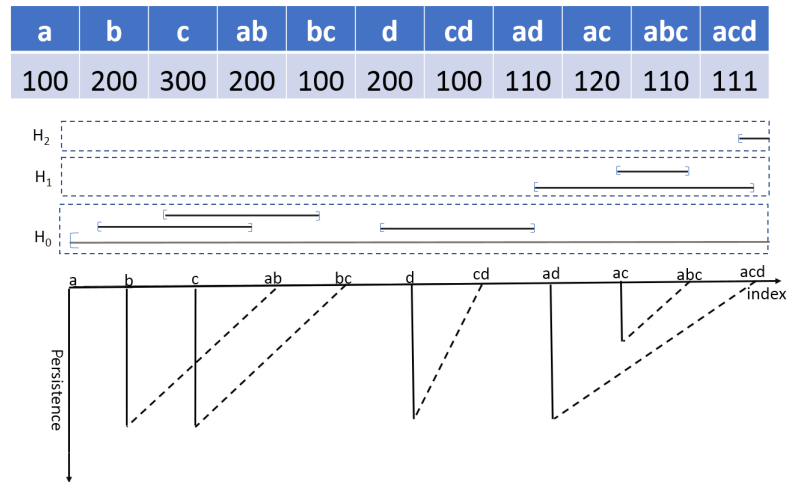
Figure 5.3: simplicial complex filtration

Figure 5.4: Table: Filteration sequence, Betti numbers, bar codes and Persisent diagrams

In the table (Figure 5.4): Top-row represent the construction of simplex and Betti numbers $\beta_0, \beta_1, \beta_2$ in the sequence of corresponding simplices. The middle-row represent the barcode of the filtration. The bottom-row with vertical bars represent creation of simplices. The diagonal dotted bars represent the death of that simplex paired up with birth.

In the Figure 5.4, the bars represent the persistency of Betti number for the corresponding cycle. As already explained (Figure 5.3), the bar denotes the persistency of β_1, that is born with 'ad' and dies with 'acd'. The

bottom part of Figure 5.4 is a ***Persistent diagram*** of the simplex.

Computation of persistent homology

For computing persitent homology, the matrix method can be used. The Gaussian elimination procedure is used to compute the persistent homology. As explained in the previous sections, the construction follows the incremental algorithm [7].

5.2 Persistent homology analysis in point clouds using TDA, R-programming

In this section, concepts and visualization of the plots such as persistent diagrams and barcodes is shown. The analysis of time lapse for the Vietoris Rips complex is also explained with the help of Topological Data Analysis (TDA) package, which is used in programming language $\mathbf{R}$(http://cran.r-project.org/).

TDA package in $\mathbf{R}$ bridges between $\mathbf{C}{+}{+}$ libraries (GUDHI, Dionysus, PHAT) and $\mathbf{R}$. The TDA package provides an $\mathbf{R}$ interface for the efficient algorithm of the $\mathbf{C}{+}{+}$ libraries (GUDHI, Dionysus, PHAT). By using this package, short execution time and quick development can be achieved. This also provides tools for Topological Data Analysis. The persistent homology of the Rips filtration is implemented in this thesis.

"TDA stats" package is also one of the channels for computing persistent homology. Which also conducts statistical interface on persistent homology calculations and visualize it by using persistent plots and barcode plots.

The package "alphashape3d" in **R**, implements the 3-dimensional alpha shapes for the reconstruction of 3-dimensional sets from a point cloud. It computes the alpha-shape and provides the users with functions to compute the volume of the alpha-shape, and identifies the connected components and produce three dimensional graphical visualizations of the set.

Before installing R package TDA, four additional packages are needed to be installed: parallel, FNN, igraph, and scales. The 'parallel' is included when installing R. But the FNN, igraph, and scales have to be installed additionally.

5.2.1   Persistent homology analysis on Point Clouds

A point cloud is a set of points in floating around n-dimensional space. In programming world, it is a data set or a data frame which has 'n' number of variables, both quantitative and qualitative variables. It is an array of points arranged in columns and rows(matrices).

Now, 3 point clouds are taken in the Euclidean space and the corresponding data is visualized, each with a set of 450 data points in three variables X_1, X_2, X_3. All these are real numbers ranging from -4 to 5 distributed in a plane shown in the Figure 5.5. The implementation in RStudio is shown below. The "rgl" package provides 3D view of point cloud data.

PC1, PC2, PC3 are short forms used in this thesis for the point cloud 1, point cloud 2 and point cloud 3 respectively.

```
############### Installing Required Packages ###############
install.packages('TDA', dependencies=TRUE,
    repos='http://cran.rstudio.com/')
install.packages("TDAstats")
install.packages('rgl', dependencies=TRUE,
    repos='http://cran.rstudio.com/')
```

```r
install.packages('FNN', dependencies=TRUE,
    repos='http://cran.rstudio.com/')


############## Loading R package TDA ###########################
library(package = "TDA")
library(TDAstats)
library(rgl)     #installed required packages as well


################### Import files from Excel(First point cloud) ###
PointCloud1 <- PC1
PC1 <- read.csv(file.choose(), header = T)
plot3d(PC1)
#### same procedure for rest of 2 PointClouds}


####################################################################
PointCloud2 <- PC2
PC2 <- read.csv(file.choose(), header = T)
plot3d(PC2)


####################################################################
PointCloud3 <- PC3
PC3 <- read.csv(file.choose(), header = T)
plot3d(PC3)
```
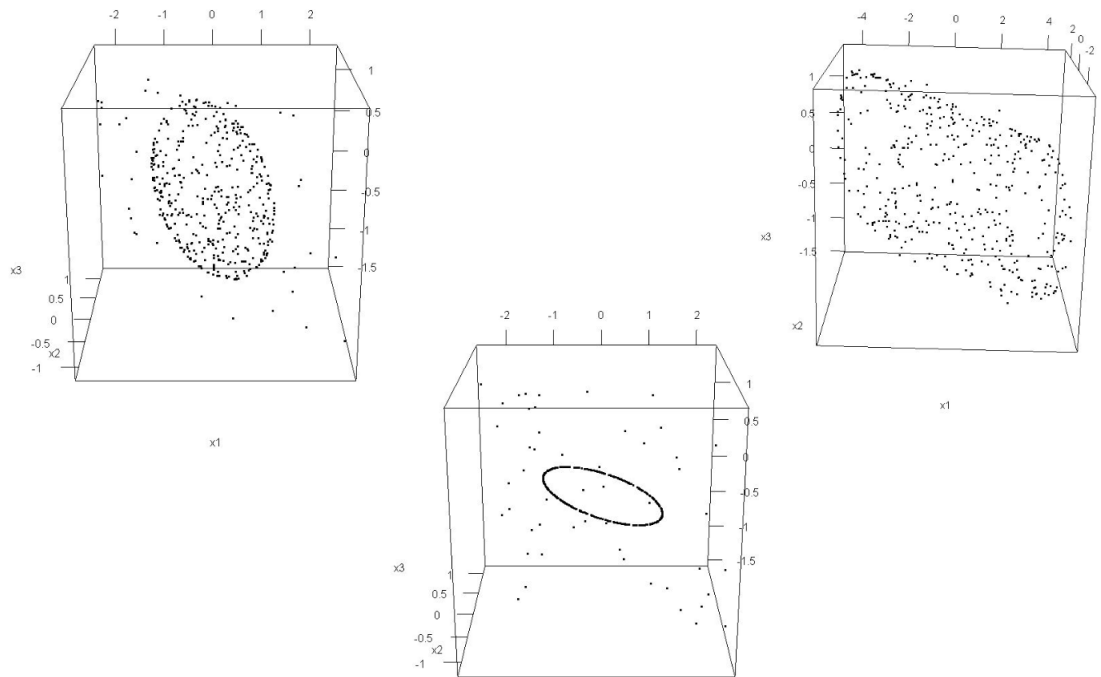
Figure 5.5: 3 different PointClouds

The 3 point clouds are decomposed into Voronoi regions, i.e.letting the points of point clouds as a union of balls in 3-dimensional Euclidean space. They are shown in Figure 5.6 for PC1, Figure 5.7 for PC2, and Figure 5.8 for PC3 respectively. The left-hand side subplots of these Figures are Voronoi decomposition of Voronoi regions with original point clouds. The right-hand side subplots are decomposition of Voronoi regions as a result of removing 50-data points in each point cloud. The nerve from triangulation of these Voronoi diagrams are clearly homotopy equivalent to sphere, torus and circle. The implementation is expecting to find the topological features of original point clouds.

```r
library(rgl)
plot3d(PC1, type='s', size=3, col = "pink")
#same for rest of two
```
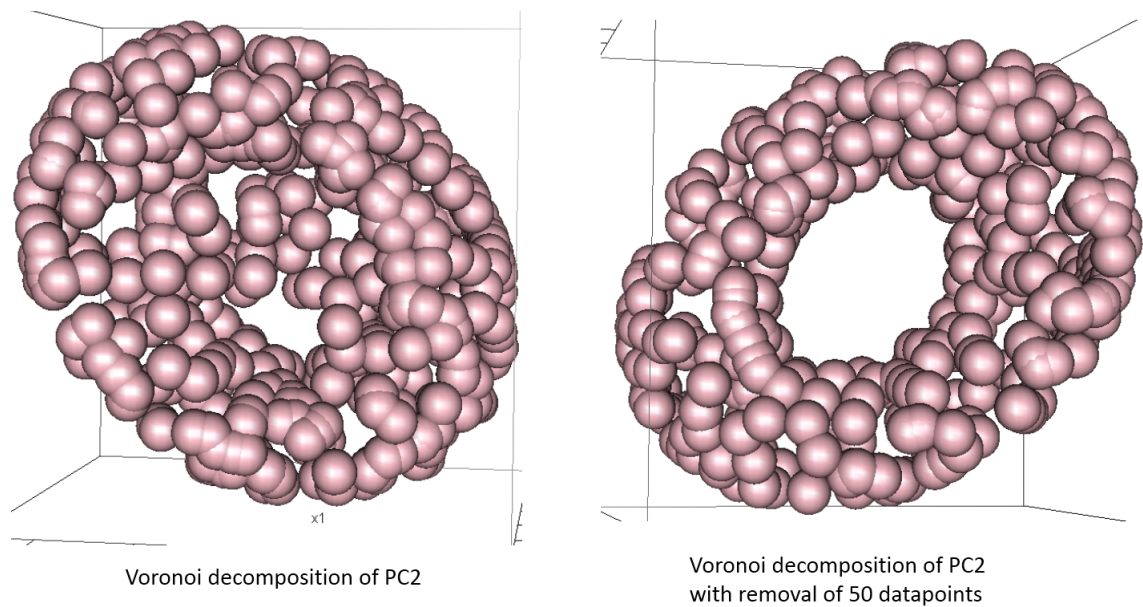
Voronoi decomposition of PC1

Voronoi decomposition of PC1
some 50 points are removed
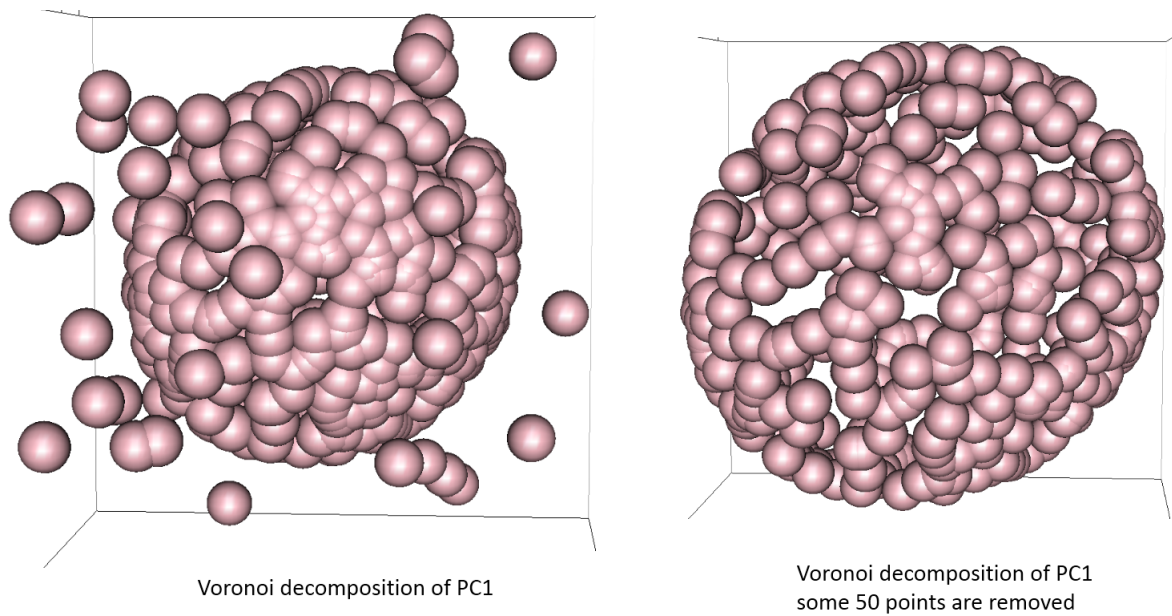
Figure 5.6: Voronoi decomposition of data points of PC1

Voronoi decomposition of PC2

Voronoi decomposition of PC2
with removal of 50 datapoints

Figure 5.7: Voronoi decomposition of data points of PC2

Voronoi decomposition of PC3

Voronoi decomposition of PC3
with removal of 50 data points
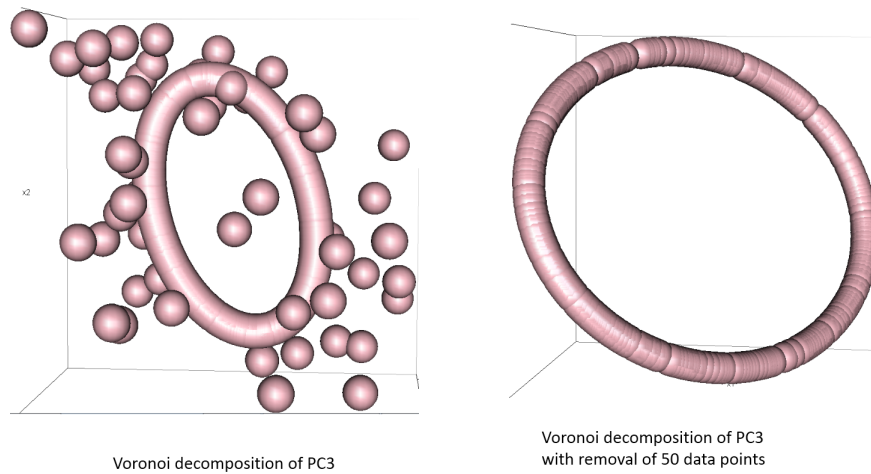
Figure 5.8: Voronoi decomposition of data points of PC3

The duality of these Voronoi diagrams, the Delaunay triangulation is acquired by picking the nerve of convex polyhedron whose centers is the nearest (Figure 4.2 explains the general case of points in $2-$dimensional plane). The resulting complex is the Delaunay complex, Figure 5.9 is the delaunay triangulation of corresponding point clouds.

```
######### Delaunay triangualtion of 3 point clouds ###############
library(alphahull)

del1 <- delvor(PC1)
plot(del1)

del2 <- delvor(PC1)
plot(del2)

del3 <- delvor(PC1)
plot(del3)
```
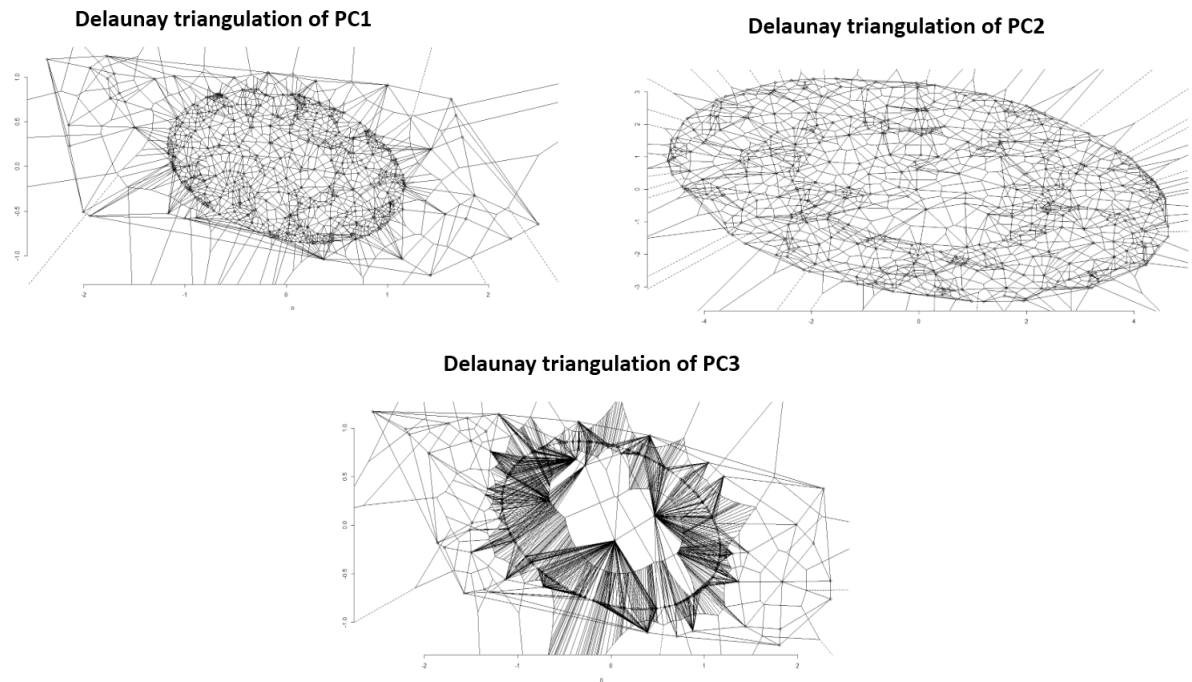
Figure 5.9: Delaunay Triangulation of respective point clouds

Note that the alpha complex can be subset of Delaunay triangulation. The alpha shape of the point clouds is illustrated below. The respective $\alpha-$shapes of the PC1,PC2 and PC3 are shown in Figure 5.10.

```
##### Alpha hulls of 3 point clouds ###############
library(alphashape3d)
AShape3d_PC1 <- ashape3d(PC1_Dat_Matr, alpha = 0.75)
plot(AShape3d_PC1, byComponents = TRUE, indexAlpha = 1)
Device 2  : alpha = 0.75
AShape3d_PC2 <- ashape3d(PC2_Dat_Matr, alpha = 0.75)
plot(AShape3d_PC2, byComponents = TRUE, indexAlpha = 1 )
Device 1  : alpha = 0.75
AShape3d_PC3 <- ashape3d(PC3_Dat_Matr, alpha = 0.75)
 plot(AShape3d_PC3, byComponents = TRUE, indexAlpha = 1)
Device 3  : alpha = 0.75
```
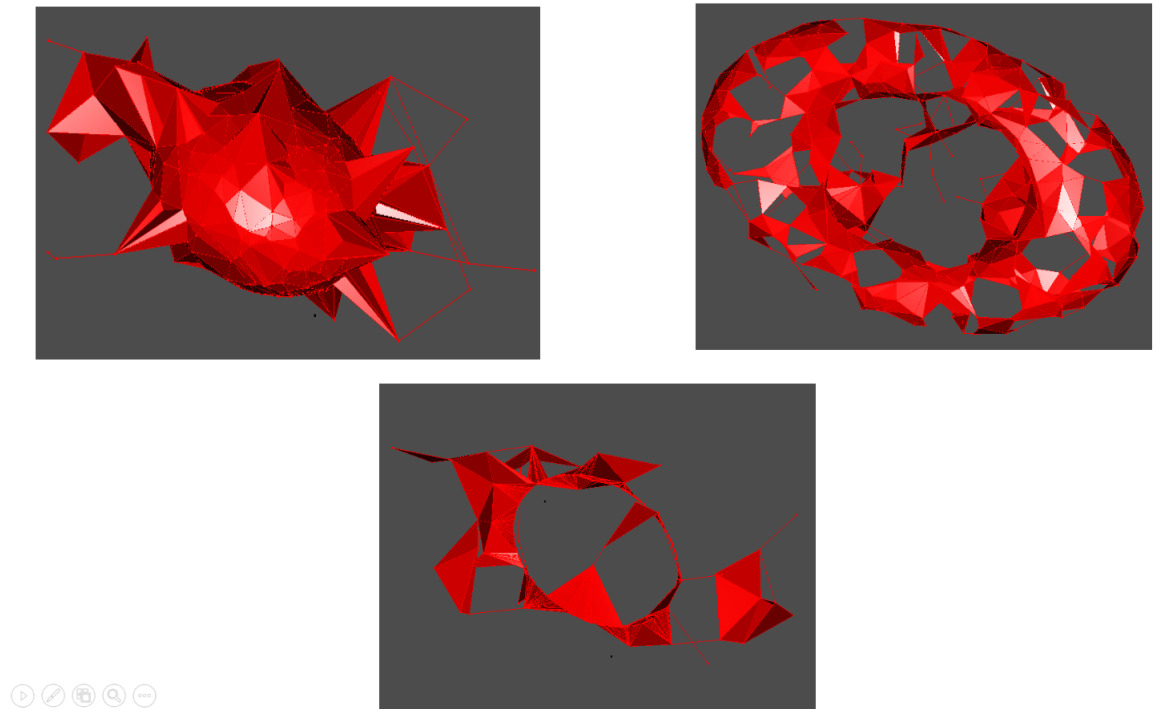
Figure 5.10: $\alpha-$shapes for 3 different PointClouds

The $\alpha-$shapes of the respective point clouds with proxy parameter α with different values (detailed implementation described in section 4.2). The volume of alpha shape is also determined. The Figure 5.11 shows the Alpha shape of the second point cloud. For the rest of 2 point clouds, the respective alpha shapes for different α shown in Figures 4.4 and 4.5

```r
########### Alpha Shapes ###################################
#########Alpha complexs for point cloud samples ############3
PC1_Dat_Matr <- data.matrix(PC1)#Making data frame into data matrix
AShape3d_PC1 <- ashape3d(PC1_Dat_Matr, alpha = 0.75) #Checking
    with various alphas
plot(AShape3d_PC1)
```

```r
##### Computation of volumes of the connected components of the
    alpha ahapes#####
AShape3d_PC1 <- ashape3d(AShape3d_PC1, alpha = alpha)
plot(AShape3d_PC1, indexalpha = 1:3)
volume_ashape3d(AShape3d_PC1, byComponents = TRUE)
#[1] 8.499966
plot(AShape3d_obj1, byComponents = TRUE, indexAlpha = 2)
### Same produre followed with rest of the data clouds###
######## Volume of PC2 Alpha shape ##############
volume_ashape3d(AShape3d_PC2)
#[1] 19.09232
######## Volume of PC2 Alpha shape ##############
volume_ashape3d(AShape3d_PC3)
#[1] 6.243968
```
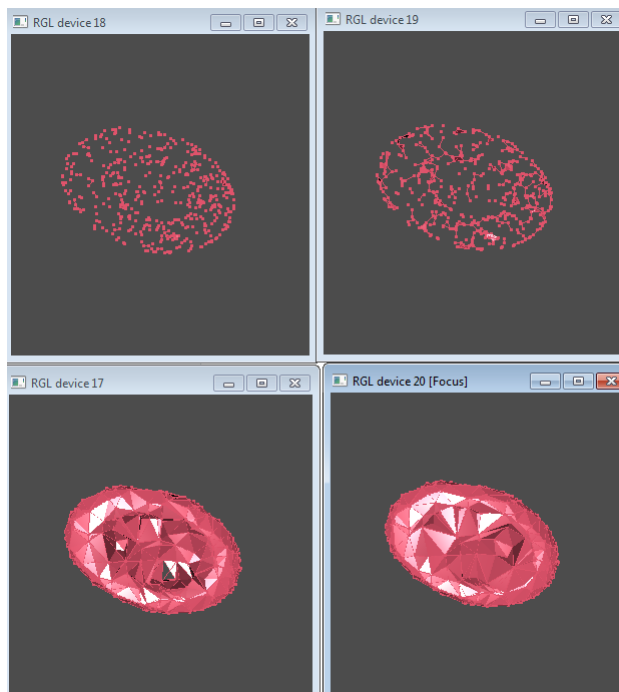

Figure 5.11: Alpha Shapes of Point Cloud 2 for various α's

One of the applications of alpha shape is, in molecular shape characterization, the novel properties are exhibited and then compared against the established shape descriptors [12]. In a Cheminformatics research, it is of significant importance to characterize, analyze and predict properties that describe the shape of molecules.

Now by using TDA package, Rips complex is determined and the time lapse for its construction is also observed. It is also observed how long the homology groups are persistent in that Rips complex. In order to visualize the persistency diagram and barcodes of these three homology groups at one, The package "TDAstats" can be useful. The $\mathbb{Z}_2$ coefficient group is taken, so there is no need to worry about the orientation of the simplices obtained from point clouds.

Point cloud 1: The implementation for the time lapse of the Rips complex construction for point cloud 1 with maximum dimension 1 is described in the following. The generated complex is of size 15185192, and it took 12:52 minutes. The ϵ (diameter) value is selected so that the Rips complex is neither complex of disjoint components nor giant complex. The same is repeated until maximum dimension 2. However, the process of generating Rips complex was never completed with the available resources (laptop). The resultant persistent homology of PC1 from construction of Rips complex on it of dimension 0 and 1 is shown in Figure 5.12

```
\begin{lstlisting}
##########################Rips complex#####################
RipsDiagPC1 <- ripsDiag(PC1, maxdimension = 1, maxscale = 5,
    library = "GUDHI", printProgress = FALSE)
plot(RipsDiagPC1[["diagram"]])
#RipsDiagPC1 <- ripsDiag(PC1, maxdimension = 2, maxscale = 5,
    library = "GUDHI", printProgress = FALSE)
#can not be completed
RipsDiagPC1$diagram[450:570,1:3] # No long intervals #noise
```

```
RipsDiagPC1 <- ripsDiag(PC1, maxdimension = 1, maxscale = 5,
    library = "Dionysus", printProgress = TRUE)
# Generated complex of size: 15185192

0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
**************************************************
# Persistence timer: Elapsed time [ 772.387000 ] seconds
```
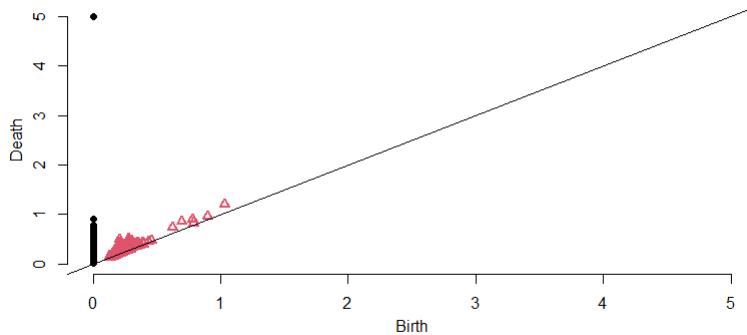

Figure 5.12: Persistent homology of Rips complex constructed on PC1 upto dimension 1

The implementation of the Rips complex construction on PC1 to see the persistent homology upto dimension 2 is described as following

```
######## Homology of PointCloud1 ####################
library(TDAstats)
PC1_Hom <- calculate_homology(PC1, dim = 2)
view(PC1_Hom) #table containes filtration of sequence of 572
    subcomplexes
plot_persist(PC1_Hom)
plot_barcode(PC1_Hom)
```
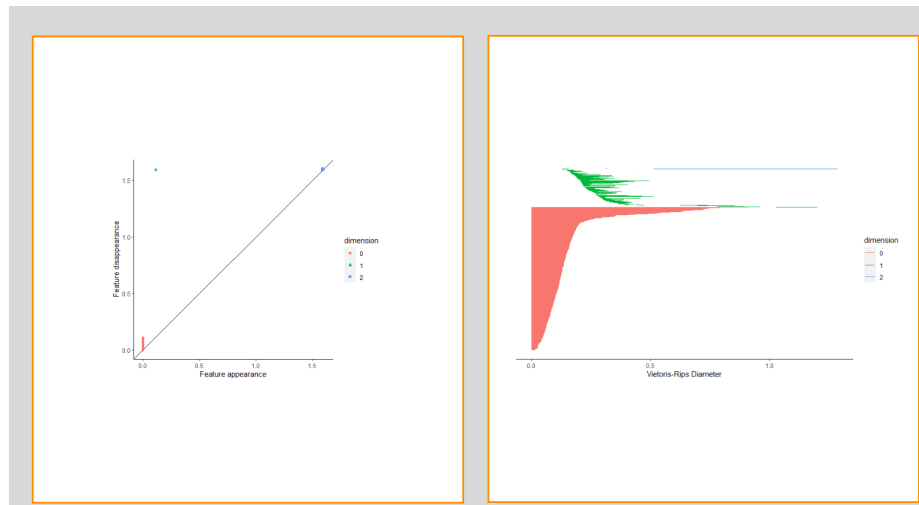
Figure 5.13: left-Persistent Diagram, Right-Barcode plots of point cloud 1

The point cloud 1 results in Figure 5.13 explains that $H_0(PC1)$ is persistent through out the whole filtration. The final obtained complex has only one component with generator in $H_2(PC1)$, and that is persistent and never dies. The right-hand side subplot in Figure 5.13 has a few bars that represent $H_1(PC1)$, and they are not consistent and lives very short time. They are considered as *topological noise* during the filtration. The left-hand side subplot in Figure 5.13 explains that there is no persistent homology $H_1^{450,572}(PC1)$. So the Rips complex obtained from PC1 has only $H_0(PC1)$ and $H_2(PC1)$ persistency, and is isomorphic to group generated by only one copy of generators in group with $\mathbb{Z}_2$ coefficients. So this point cloud has a persistent homology that is isomorphic to homology of the two dimensional sphere generated by the group with $\mathbb{Z}_2$ coefficients. The Rips complex obtained from PC1 is homotopy equivalent to two dimensional sphere.

Point cloud 2: The implementation for the time lapse of the Rips complex construction for point cloud 2 with maximum dimension 1 is described in the following. The generated complex is of size 5640057, and it took around 3:56 minutes. The same implementation is repeated for max-

imum dimension 2. However, the process of generating Rips complex was never completed with the available resources (laptop).The resultant persistent homology of PC2 from construction of Rips complex on it of dimension 0 and 1 is shown in Figure 5.14. Scaling parameter ϵ is taken for getting appropriate Rips complex on it.

```
RipsDiagPC2 <- ripsDiag(PC2, maxdimension = 1, maxscale = 5,
    library = "Dionysus", location = TRUE, printProgress = TRUE)
#plot(RipsDiagPC2[["diagram"]]
# Generated complex of size: 5640057


0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
**************************************************
# Persistence timer: Elapsed time [ 236.377000 ] seconds


#RipsDiagPC2 <- ripsDiag(PC2, maxdimension = 2, maxscale = 5,
    library = "GUDHI", printProgress = FALSE)
#The process can not be completed.
```
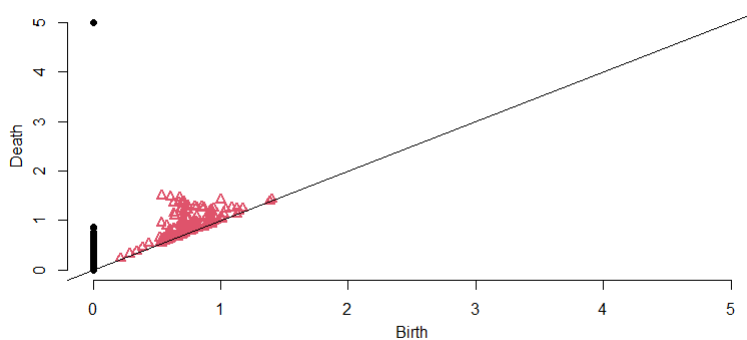

Figure 5.14: The resultant persistent homology on Rips complex constructed from PC2 at dimension 0 and 1

The implementation of the Rips complex construction in PC2 to see the persistent homology upto dimension 2 is described as following

```
######## Homology of PointCloud2 #####################
PC2_Hom <- calculate_homology(PC2, dim = 2)
view(PC2_Hom) #table containes filtration process of 638 complexes
plot_persist(PC2_Hom)
plot_barcode(PC2_Hom)
```
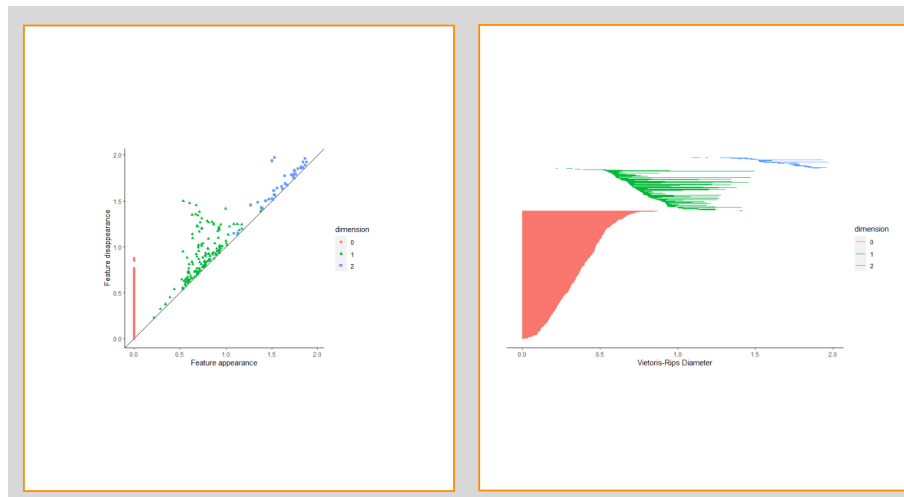

Figure 5.15: Left-The persistent diagram of PC2 and right-The barcodes diagram of PC2

The barcode diagram, persistent diagram of second point cloud are shown in Figure 5.15. The resultant homology groups are $H_0(PC2)$ $H_1(PC2)$ and $H_2(PC2)$. The left-hand side subplot in Figure 5.15 is the persist diagram. The right-hand side subplot is barcode. It can seen in the barcode diagram that several bars are persistently long enough to represents the $H_1(PC2)$ group. The two green points on the plot (viz.top most) are persistent. as well as their birth and death in the form of intervals are [0.4, 1.5] and [0.5,1.4] (clearly seen the red scatters in Figure 5.14 also). So the persistency of $H_1(PC2)$ is generated by two 1-dimensional cycles, the two

copies of the quotient group $\mathbb{Z}_2$. This homology groups is isomorphic to first homology group of the torus generated by cycle in group with $\mathbb{Z}_2$ coefficients. On the barcode diagram, there is one long blue bar, which is persistent and it represents the $H_2(PC2)$ in the filtration process.

When the filtration table is observed (in R interface), which contain sequence of 638 subcomplexes of filtration. There are two intervals which are persistent from filtration level 450 to filtration level 598 for a long time with dimension 1. It shows that $H_1(PC2)$ is generated by two generators. These are two copies of $\mathbb{Z}_2$ group. The direct sum of these two copies generate $H_1(PC2)$. Hence the persistent homology groups of this point cloud PC2 are isomorphic to homology of torus generated by $\mathbb{Z}_2$ group up to dimension of 2 (without orientation into consideration).

Point cloud 3:The implementation for the time lapse of the Rips complex construction for point cloud 3 with maximum dimension 1 is described in the following. The generated complex is of size 15184305, and it took 8:43 minutes. The same is repeated until maximum dimension 2. However, the process of generating Rips complex was never completed with the available resources. The ϵ value is selected like what is done for PC1.

```
RipsDiagPC3 <- ripsDiag(PC3, maxdimension = 1, maxscale = 5,
    library = "Dionysus", location = TRUE, printProgress = TRUE)
plot(RipsDiagPC3[["diagram"]])
# Generated complex of size: 15184305
0%   10   20   30   40   50   60   70   80   90   100%
|----|----|----|----|----|----|----|----|----|----|
**************************************************
# Persistence timer: Elapsed time [ 523.779000 ] seconds
#RipsDiagPC3 <- ripsDiag(PC3, maxdimension = 2, maxscale = 5,
    library = "GUDHI", printProgress = FALSE)
#RipsDiagPC3 <- ripsDiag(PC3, maxdimension = 2, maxscale = 5,
    library = "Dionysus", printProgress = FALSE)
```

```
#They can not be completed with existing resources
```

The resultant persistent homology of the PC3 from construction of Rips complex on it of dimension 0 and 1 is shown in Figure 5.16
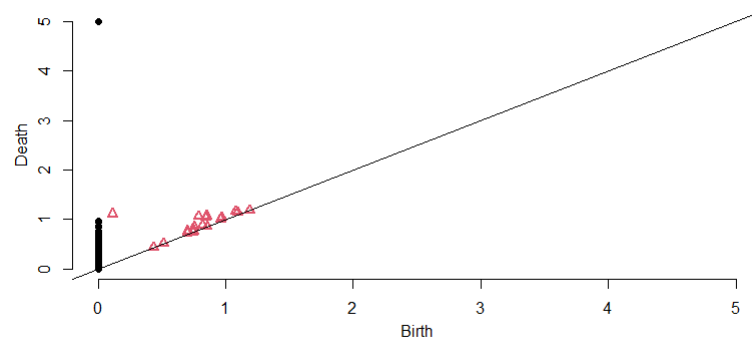

Figure 5.16: Black data denotes the persistency of connected components and the triangle points denotes 1−dimesional homology groups

The implementation of the Rips complex construction in PC3 to see the persistent homology upto dimension 2 is described as following

```
######## Homology of PointCloud3 #####################
PC3_Hom <- calculate_homology(PC3, dim = 2)
view(PC3_Hom) #table containes filtration process of 469 complexes
plot_persist(PC3_Hom)
plot_barcode(PC3_Hom)
```
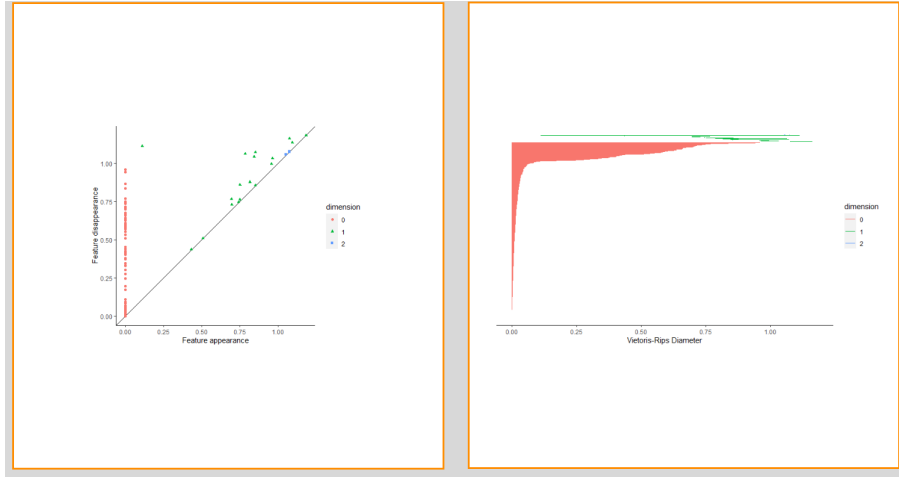
Figure 5.17: The persistent diagram and the barcode plot for PC3

The plots for point cloud 3 are shown in Figure 5.17. The left-hand side subplot in Figure 5.17 is the persist diagram. The right-hand side subplot is barcode. The resultant homology groups are only $H_0(PC3)$ and $H_1(PC3)$. They are persistent in very long time and there is no $H_2(PC3)$ group. There are no blue bars that represent the persistency of $H_1(PC3)$). In barcode diagram in figure 5.17, the longest red bars area and longest green bar explains that there exists one generator each for $0-$dimension and $1-$dimension. The corresponding homology groups persists in a long intervals. It concludes that the persistent homology group exists for this point cloud only in those dimensions but not in dimension 2. The persistent $H_0(PC3)$ and $H_1(PC3)$ are generated different copies each. These are representative classes of $\mathbb{Z}_2$ group. Therefore the persistent homology of PC3 is is isomorphic to the homology groups of circle with $\mathbb{Z}_2$ coefficients.

The implementation is repeated with several values of scaling parameter for all 3 point clouds. The resulting Rips complexes does not change their topological features, but with some topological noise in the PC1. For PC3, all the values of ϵ, the resultant persistent barcodes are consistently same. If it is the case with Rips complex constructed on point clouds is

expected to be other than these, the approximation of ϵ has to be checked to get the features of that point cloud.

In fact Homotopy equivalence is richer topological invariant than homology. It is concluded that the alpha complex and Rips complex generated by these three point clouds are topologically equivalent to the sphere, torus and circle respectively. See the Figure 4.4, Figure 5.11 and Figure 4.5.

Bibliography

[1] Allen Hatcher's book Algebraic Topology

[2] Herbert Edelsbrunner, David G.Kirkpatrick and Raimund Seidel. On the shape of point sets in the plane

[3] Herbert Herbert Edelsbrunner, A short course in Computationl Geometry and Topology

[4] Jesper.M.Møller The fundamental Group and Covering Spaces

[5] SE Schaeffer(2007) Graph clustering Comput Sci Rev 1:27-64

[6] Gameiro M, Hiraoka Y, Izumi S, Kramár M, Mischaikow K, Nanda V (2015) A topological measurement of protein compressibility. Jpn J Ind Appl Math 32:1-17

[7] Afra Zomorodian and Gunnar Carlsson, Discrete and Computational Geometry

[8] Herbert Edelsbrunner(auth.) - A short course in Computational Geometry and Topology (§ page 76-78)

[9] Herbert Edelsbrunner(auth.) - A short course in Computational Geometry and Topology (§ page 84)

[10] ROBERT GHRIST, BARCODES: THE PERSISTENT TOPOLOGY OF DATA

[11] Halmos.P.R., Finite dimensional vector Spaces

[12] J. Anthony Wilson, Andreas Bender, Taner Kaya, and Paul A. Clemons,
     Alpha shape applied to molecular shape characterization exhibit novel
     properties compared to established shape descriptors 2011, Aug 19.

[13] Weisstein, Eric W. "Euler characteristic". MathWorld.