Introduction to Topological Data Analysis

a project under *Prof. Kashyap Rajeevsarathy*

as part of Summer Internship Program - 2024 at IISER Bhopal

Venkat Trivikram
ISI Bangalore

Contents

Chapter 1

Some defnitions

1.1 Topological spaces

Given a (raw)set of some elements, kneading them by defining some properties and making them into a *study*-able structure has been a practice that mathematicians been doing since ages. Owing to that, this is yet again another structure, a *topological space*. We first see what a *topology* is.

For any set X, we have the power set $\mathcal{P}(X)$ that is the set of all the $2^{|X|}$ subsets of X. A topology $\mathcal{T}$ is a *nice* enough subset of $\mathcal{P}(X)$. We define the *nice*-ness as follows.

Definition 1.1.1 (Topology)**.** A **topology** on a set X is a collection $\mathcal{T}$ of subsets of X having the following properties

- The sets $\emptyset$ and X are in $\mathcal{T}$
- The union of the elements of any subcollection of $\mathcal{T}$ is in $\mathcal{T}$.
- The intersection of the elements of any finite subcollection of $\mathcal{T}$ is in $\mathcal{T}$.

The set X together with the collection $\mathcal{T}$, is called a **topological space**. We abuse the notation and call X itself as a topological space when $\mathcal{T}$ is evident. Elements of $\mathcal{T}$ are called **open sets** and a **closed set** is that whose complement is an open set.

Next thing is, whenever we have a new structures, we try to get more information out of it by considering functions between them. But, we at least seek some *nice*-ness to the functions. We at least want the function between two topological spaces to preserve the respective topologies, don't we? Here's such least nice kind of map [1]

Definition 1.1.2 (Continuous map)**.** Let, $(X, \mathcal{T})$, $(Y, \mathcal{T}')$ be two topological spaces. A map $f : X \to Y$ is said to be **continuous** if for any open set $\mathcal{U}$ in Y, $f^{-1}(\mathcal{U})$ an open set of X.

So, how exactly is it preserving the respective topologies? Here's how. We know that topologies are characterised by the open sets. So, preserving the open sets, would preserve the topologies. Suppose we have a continuous map as above. Further suppose $\mathcal{U}$ and $\mathcal{V}$ are open in Y, then $f^{-1}(\mathcal{U})$ and $f^{-1}(\mathcal{V})$ are open in X. Now observe that,

- as $\emptyset$ is open in Y, $f^{-1}(\emptyset) = \emptyset$ is open in X
- as Y is open in Y, $f^{-1}(Y) = X$ is open in X

[1] we might call functions as maps also. it's short and cute. moreover function reminds us of duty/work whereas map is more close to earth, nature ;)

- $f^{-1}(\mathcal{U} \cup \mathcal{V}) = f^{-1}(\mathcal{U}) \cup f^{-1}(\mathcal{V})$ which is open since topologies are closed under unions
- $f^{-1}(\mathcal{U} \cap \mathcal{V}) = f^{-1}(\mathcal{U}) \cap f^{-1}(\mathcal{V})$ which is open since topologies are closed under intersections

You see that! We exploited the intrinsic set theoretic properties of the inverse functions that behaves well with unions and intersections and thus it is true for arbitrary unions and finite intersections too. Hence, it preserves the open sets, and thus the topology (cf. the definition above).

Now, you might think, why can't we define like,

$$\text{“ for every open set } \mathcal{U} \text{ of } X,\ f(\mathcal{U}) \text{ should be an open set in } Y\text{ ”}$$

Such maps are called **open** maps. But, these are certainly not helpful. Set theoretically, image of function f doesn't behave that well as f^{-1} does. To start with, the intersections: for any two sets A, B in the domain of f,

$$f(A \cap B) \subseteq f(A) \cap f(B)$$

equality is not true in general. Thus, there are open maps which are not continuous, see this.

Definition 1.1.3 (Subspace)**.** Consider a topological space $(X, \mathcal{T})$. We can define a topology on a subset $S \subseteq X$ as follows. A subset of S is open in the subspace topology if and only if it is the intersection of S with an open set of X. The **subspace topology** is given by

$$\mathcal{T}_S = \{S \cap \mathcal{U} \,|\, \mathcal{U} \in \mathcal{T}\}$$

Remarks.

1. By the definition of continuity, we can see that if S is a subspace of X then the inclusion map $i : S \hookrightarrow X$ is continuous.
2. Since, complement of open is closed, we can just replace the word *closed* with *open* in the above definition and we are good to go.

We now see an useful result concerning subsets and continuous functions.

Theorem 1.1.1 (Pasting lemma)**.** Let, A and B be two closed(or open) subsets of a topological space X such that $X = A \cup B$. Consider another topological space Y. If $f : X \to Y$ is continuous when restricted to both A and B, then f is continuous.

Proof. Please refer to Theorem 18.3 in [10]. □

We now see an *extremely* nice map.

Definition 1.1.4 (Homeomorphism)**.** A **homeomorphism** between two topological spaces X and Y is a bijective continuous map $f : X \to Y$ such that the inverse f^{-1} is also continuous.

This notion leads to comparison of two topological spaces. The continuity of f and f^{-1} tells us the open sets are presevered both in the forward and backward direction. Moroeover, f is a bijection too. So, if there exists a homeomorphism between two topological spaces, then they are considered to be *same*, **homeomorphic**.

1.2 Euclidean spaces

Look at the collection of all real *open* intervals

$$\mathcal{I} = \{(a, b) \mid a, b \in \mathbb{R}\}.$$

Now consider the collection

$$\mathcal{T} = \left\{ \bigcup_{\alpha} I_\alpha \mid I_\alpha \in \mathcal{I} \right\}$$

This forms a topology on the set of real numbers $\mathbb{R}$. The $\mathcal{T}$ is call **usual topology**. Now, this is a topological space. It is a one dimensional space called **real 1-space**. We now extend this to n dimensions.
Consider the set of all n-tuples of real numbers,

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \mid x_i \in \mathbb{R}\, \forall\, i\}$$

The usual cartesian distance between any two points $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ in $\mathbb{R}^n$ is given by

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

We have the extended notion of open interval to n dimensions as follows. An open ball of radius r in $\mathbb{R}^n$ is defined as

$$B(x; r) = \{y \in \mathbb{R}^n \mid d(x, y) < r\}$$

An subset $\mathcal{U}$ of $\mathbb{R}^n$ is said to be open iff for every $x \in \mathcal{U}$ there exists $r > 0$ such that $B(x; r) \subseteq \mathcal{U}$. This collection of open sets is a topology on $\mathbb{R}^n$ called the **standard topology**. $\mathbb{R}^n$ with the standard topology is called **real n-space** or **euclidean spaces** in general.

Definition 1.2.1 (Affinely independent)**.** An ordered set of points

$$\{v_0, v_1, \ldots, v_k\} \subset \mathbb{R}^n$$

are said to be **affinely independent** if

$$\{v_1 - v_0, v_2 - v_0, \ldots, v_k - v_0\}$$

is a linear independent subset of $\mathbb{R}^n$.

Consider a set of points $v_0, v_1, \ldots, v_k$ in $\mathbb{R}^n$.

Definition 1.2.2 (Convex combination)**.** A **convex combination** of the points above is a point

$$x = t_0 v_0 + t_1 v_1 + \cdots + t_k v_k$$

such that, $\sum\limits_{i=1}^{n} t_i = 1, t_i \geq 0, \forall\, i$

Definition 1.2.3 (Convex set)**.** The **convex set** $[v_0, v_1, \ldots, v_k]$, spanned by the points $v_0, v_1, \ldots, v_k$, is the set of all convex combinations of the points $v_0, v_1, \ldots, v_k$.

Chapter 2

Simplicial Complexes

2.1 Defining the structure

In order to study the data, giving a structure to it is essential. Since, we think that we well enough understand the structure of euclidean spaces, we try to somehow (fit) embed the data into the euclidean spaces. Now this is the question of understanding sub-structures v.i.z., sub-spaces of the euclidean spaces. Yet, again we at least try to understand nice-enough sub-spaces of the euclidean spaces. What are nice subspaces we know? If we think of one dimension, a line segment. In two dimiensions, there are triangles, circles etc. In three dimensions there are tetrahedrons, which are extensions of triangles. These are well understood structures. We now try to extend this notion of triangles to arbitrary dimensions. This leads to defining **simplexes**.

Definition 2.1.1 (k-simplex)**.** Consider $k + 1$ many affinely independent points $\{v_0, v_1, \ldots, v_k\}$ in $\mathbb{R}^n$. Then $[v_0, v_1, \ldots, v_k]$, the convex set spanned by the set $\{v_0, v_1, \ldots, v_k\}$, is called the ***k*-simplex** with vertices $v_0, v_1, \ldots, v_k$.

Definition 2.1.2 (Face)**.** The **faces** of the simplex $[v_0, v_1, \ldots, v_k]$ are the simplices[1] spanned by the subsets of $\{v_0, v_1, \ldots, v_k\}$.

Remark. A ***m*-face** of a simplex $[v_0, v_1, \ldots, v_k]$ is a m-simplex spanned by $(m + 1)$ many vertices from $\{v_0, v_1, \ldots, v_k\}$.

Definition 2.1.3 (Simplicial complex)**.** A **simplicial complex** K in $\mathbb{R}^n$ is a collection of simplices such that

1. any face of a simplex of K is a simplex of K
2. the intersection of any two simplices of K is either empty or a common face of both.

Definition 2.1.4 (Underlying space)**.** Consider any simplicial complex K in $\mathbb{R}^n$. We define $|K|$, the **underlying space** of K to be the union of simplices of K.

Remark. Since, each simplex is a subspace, $|K|$ is also a subspace of $\mathbb{R}^n$. The topology of $|K|$ is the subspace topology induced from the standard topology in $\mathbb{R}^n$

Definition 2.1.5 (Abstract simplicial complex)**.** Let, $V = \{v_1, v_2, \ldots, v_k\}$ be any finite set. An abstract simplicial complex $\tilde{K}$ with vertex set V is a set of finite subsets of V satisfying the two conditions

1. The elements of V belong to $\tilde{K}$

[1] plural for simplex

2. if $\tau \in \tilde{K}$ and $\sigma \subseteq \tau$, then $\sigma \in \tilde{K}$

As mentioned earlier, our aim is to realize the data as an embedding in $\mathbb{R}^n$. So, we have to somehow realize the abstract simplicial complex also as a subspace in $\mathbb{R}^n$. We do that as follows.

Definition 2.1.6 (Geometric realization)**.** Let $\tilde{K}$ be a (abstract)simplicial complex. Further let $\phi : V \to \mathbb{R}^n$ be a map that sends vertices of $\tilde{K}$ to $\mathbb{R}^n$. Then the **geometric realization** of $\tilde{K}$ with respect to ϕ is the union

$$|\tilde{K}|_\phi = \bigcup_{\sigma \in \tilde{K}} |\sigma|_\phi$$

where for each $\sigma = \{v_{i_1}, v_{i_2}, \cdots v_{i_j}\} \in \tilde{K}$, the set $|\sigma|_\phi \subset \mathbb{R}^n$ is the simplex spanned by the points $\phi(v_{i_1}), \cdots, \phi(v_{i_j})$.

Remark. Since, ϕ was arbitrary, we might consider ϕ to be a map that takes each vertex to the same point in $\mathbb{R}^n$. But such maps are of no use in order to understand the information from simplicial complex. Thus, considering ϕ to be *injective* would be interesting. We call a map $\phi : V \to \mathbb{R}^n$ as **affine embedding** if ϕ is injective and $\phi(V)$ is an affinely independent subset of $\mathbb{R}^n$.

Earlier, immediately after defining underlying space of simplicial complex, we talked about its topology. Here too, one would like to talk about the same. But, here the (underlying space) realization is depending upon the map ϕ. So, we expect each map ϕ would give different realizations in $\mathbb{R}^n$. But our expectation is faltered in the case of affine embeddings. This is given by the following proposition.

Proposition 2.1.1. For any two affine embeddings $\phi, \psi : V \to \mathbb{R}^n$, there is a homeomorphism

$$h : |\tilde{K}|_\phi \to |\tilde{K}|_\psi$$

between the corresponding geometric realizations.

Before proving this, let us do the following lemma.

Lemma 2.1.1. If, K_1 and K_2 are two simplicial complexes, and $f : K_1 \to K_2$ is given by a bijection $f : V(K_1) \to V(K_2)$ such that $[v_0, v_1, \cdots, v_r]$ is an r-simplex of K_1 if and only if $[f(v_0), f(v_1), \cdots, f(v_r)]$ is a r-simplex of K_2. Then, this induces a homeomorphism between the respective underling spaces as follows,

$$\varphi : |K_1| \to |K_2|,\ \varphi\left(\sum_{i=0}^{r} t_i v_i\right) = \sum_{i=0}^{r} t_i f(v_i)$$

Proof. Since,

1. f is bijective on the vertices
2. we have a simplex in K_1 if and only if the corresponding image of it via f is a simplex in K_2
3. the underlying space is just the union of all simplices in the complex

it clearly follows that the map φ is a bijection. We will now show that the map φ is continuous. As φ defined is a linear map for each r-simplex and as linear maps are continuous, we conclude that φ when restricted to each simplex is a continuous map. We observe that, since each r-simplex is a closed subset of $\mathbb{R}^n$ and thus closed subsets of the $|K_1|$ in subspace topology. So, our φ is continuous on all (simplexes)closed sets whose union is the $|K_1|$. We now appeal to the pasting lemma and conclude that φ is continuous on $|K_1|$. Similarly, we have the inverse map

$$\varphi^{-1} : |K_2| \to |K_1|,\ \varphi\left(\sum_{i=0}^{r} t_i f(v_i)\right) = \sum_{i=0}^{r} t_i v_i,$$

which is continuous for the exact same reasons that we concluded for the continuity of φ. Thus, we conclude that φ is continuous bijective map whose inverse is continuous, thus a homeomorphism. ☐

We now prove the proposition using the above lemma.

Proof. (of Proposition. 2.1.1) Consider the vertex set, $V = \{v_0, v_1, \ldots, v_k\}$ of $\tilde{K}$. Look at the following sets

$$\phi(V) = \{\phi(v_0), \phi(v_1), \cdots, \phi(v_k)\}$$

and

$$\psi(V) = \{\psi(v_0), \psi(v_1), \cdots, \psi(v_k)\}.$$

Since, ϕ and ψ are affine embeddings,

1. (*injectivity*) $\phi : V \to \phi(V)$ and $\psi : V \to \psi(V)$ are bijections.
2. (*affine independence*) we can talk about the simplexes formed by $\phi(V)$ and $\psi(V)$

So, we look at the simplicial complexes with the vertex sets $\phi(V)$ and $\psi(V)$, call them K_ϕ and K_ψ respectively. Now, (by definition) the geometric realizations $|\tilde{K}|_\phi$ and $|\tilde{K}|_\psi$ are nothing but the underlying spaces of K_ϕ and K_ψ respectively. Set,

$$\varphi := \psi \circ \phi^{-1} : \phi(V) \to \psi(V),\ \varphi(\phi(v_i)) = \psi(v_i),\ \forall\, i\, \in\, \{0, 1, \cdots, k\}$$

This is clearly a bijection. Moreover, $\{\phi(v_{i_1}), \phi(v_{i_2}), \cdots, \phi(v_{i_l})\}$ is a simplex in K_1 if and only if $\{v_{i_1}, v_{i_2}, \cdots, v_{i_l}\}$ is a simplex in $\tilde{K}$ if and only if $\{\psi(v_{i_1}), \psi(v_{i_2}), \cdots, \psi(v_{i_l})\}$ is a simplex in K_2. Thus, we have a simplex in K_1 if and only if the corresponding image of it via φ is a simplex in K_2. Thus, we now can appeal to the lemma 2.1.1 and conclude that the underlying spaces $|K_1|$ and $|K_2|$ are homeomorphic. Thus, from our earlier observation $|\tilde{K}|_\phi$ and $|\tilde{K}|_\psi$ are homeomorphic. □

2.2 New structures from old

Following the same fundamental ideology as we had in the chapter 1, immediately after having new structures, we try to extract more information about them by considering the maps between them. The following is one such attempt.

Definition 2.2.1 (Simplicial maps)**.** Consider two (abstract)simplicial complexes[2] K and L. A **simplicial map** $f : K \to L$ is an assignment to vertices of K to vertices of L such that it sends simplices to simplices. In other words, if $\sigma = \{v_{i_1}, v_{i_2}, \cdots v_{i_j}\}$ is a simplex in K then $f(\sigma) = \{f(v_{i_1}), f(v_{i_2}), \cdots f(v_{i_j})\}$ is a simplex in L.

Remark. The type of map between two simplicial complexes that we stated in the hypothesis of the lemma 2.1.1 is an example of bijective simplicial map as it maps vertex set bijectively and takes simplices to simplices. Moreover, that map is called an **isomorphism** between simplicial complexes.

After defining the continuous map, we saw the notion of subspace. We are essentially getting a new topological space arising from the existing one. We apply the same idea here. We try to build new simplicial complexes from the existing ones. This leads to the following notions in the next subsection.

2.2.1 Barycentric subdivision

Since, we were talking of extending the notion of *triangles* to the higher dimensions, we should also extend the notion of the center of triangle, i.e., the *centroid*, to higher dimensions. We do that as follows.

[2] since, usual simplicial complex can also be looked as an abstract simplicial complex, we look at general cases itself. from now we represent the letters without tilde, it's clean without that :)

Definition 2.2.2 (Barycenter)**.** If $\sigma = [v_0, v_1, \cdots, v_k]$, then the **barycenter** of σ is defined to be the point

$$\hat{\sigma} = \frac{1}{k+1} \sum_{i=0}^{k} v_i$$

Definition 2.2.3 (Barycentric subdivision)**.** The **barycentric subdivision** of an affine n-simplex Σ^n, denoted by $Sd\,\Sigma^n$, is a family of affine n-simplexes defined inductively for $n \geq 0$

1. $Sd\,\Sigma^0 = \Sigma^0$
2. if $\sigma_0, \sigma_1, \cdots, \sigma_{n+1}$ are the n-faces of Σ^{n+1} and if $\hat{\sigma}$ is the barycenter of Σ^{n+1}, then $Sd\,\Sigma^{n+1}$ consists of all the $(n+1)$-simplexes spanned by $\hat{\sigma}$ and n-simplexes in $Sd\,\sigma_i$ for $i = 0, 1, \cdots, n+1$.

2.3 Triangulation

We now are in the place to believe that simplexes and thus simplicial complexes are relatively nicer spaces to understand and work with. Now, we can ask ourselves a question that given any topological space X can have a *same* structure as X in terms of a simplicial complex? This leads to the following notion.

Definition 2.3.1 (Triangulation)**.** A triangulation τ of a topological space X is a simplicial complex K whose underlying space $|K|$ is homeomorphic to X.

So, our earlier question now changes to "does every topological space admits a triangulation?". This is very non-trivial question and we know that very specific spaces (mostly the spaces which we work with) admits one. For example, there's a big theorem that "every compact surface admits a triangulation". One can read a graph theoretic proof of this theorem at this paper by Carsten Thomassen.

We famously know the *Euler's formula* that for any convex polyhedron or spherical polyhedras,

$$V - E + F = 2$$

where, V, E and F are number of vertices, edges and faces of the polyhedron respectively. Since, vertices are nothing but 0-simplexes, edges are 1-simplexes and faces are 2-simplexes, we generalise the euler's formula in the following way.

Definition 2.3.2 (Euler characteristic)**.** Let, K be a simplicial complex whose largest simplex is of dimension m. For each $q \geq 0$, let α_q be the number of q-simplexes in K. The Euler characteristic of K is given by,

$$\chi(K) = \sum_{q=0}^{m} (-1)^q \alpha_q$$

Since, any compact surface K admits a triangulation, we can talk about Euler characteristic of this surface $\chi(K)$, which would be the Euler characteristic of the underlying simplicial complex structure of K. But, a surface admits many triangulations, so is this a valid definition? We shall come to this in the later parts after building some theory.

2.4 Orientation

In this section we shall give an additional structure to a simplicial complex in the expectation of getting more information about the space.

Definition 2.4.1 (Orientation)**.** An orientation of a simplicial complex K with vertex set V is an injective function $o : V \to \mathbb{N}$ which assigns unique natural numbers to vertices.

Thus, now given an orientation on K, we shall represent its simplices as an ordered tuple v.i.z., $(v_0, v_1, \cdots, v_k)$ is an *oriented* k-simplex where $o(v_0) < o(v_1) < \cdots < o(v_k)$.

Definition 2.4.2. Let, K be an oriented simplicial complex and let $\sigma = (v_0, \cdots, v_k)$ be an oriented k-simplex in K. For each i in $\{0, 1, \cdots, k\}$ the ***i*-th face**[3] of σ is the $(k-1)$-dimensional simplex

$$\sigma_{-i} = (v_0, v_1, \cdots, v_{i-1}, \cancel{v_i}, v_{i+1}, \cdots, v_k).$$

Since we have a orientation, we can talk about start and ending points of a paths along the simplexes. Thus, we can talk about *cycles* which is a path that ends at its starting point. We can also talk about *boundaries* of a simplex which are nothing but the simplicial complex of one dimension less that bounds a simplex. For example, we have a oriented 2-simplex, (v_0, v_1, v_2). The boundary of this would be $(v_0, v_1) + (v_1, v_2) - (v_0, v_2)$. Here, by $+$ and $-$ I mean to say that we trasverse through the simplex along the way and away from the way respectively. This leads to the following definition.

Definition 2.4.3 (Algebraic boundary)**.** Let, σ be an oriented k-simplex. The **algebraic boundary** of σ is the linear combination

$$\partial_k \sigma = \sum_{i=0}^{k} (-1)^i \sigma_{-i} \sum_{i=0}^{k} (-1)^i \sigma_{-i}$$

which sends each basis k-chain σ to the $(k-1)$-chain. where, σ_{-i} denotes the i-th face as defined above.

So, with this notation, we have $\partial_2(v_0, v_1, v_2) = (v_0, v_1) + (v_1, v_2) - (v_0, v_2)$. We compute ∂_1 of this. This is valid as the boundary map we defined can just be extended linearly on to any combination of simplices. So, we have

$$\begin{aligned}
\partial_1 \partial_2 (v_0, v_1, v_2) = \partial_1((v_0, v_1) + (v_1, v_2) - (v_0, v_2)) &= \partial_1(v_0, v_1) + \partial_1(v_1, v_2) - \partial_1(v_0, v_2) \\
&= v_1 - v_0 + v_2 - v_1 - v_2 + v_0 \\
&= 0
\end{aligned}$$

This suggests us to say that the boundaries of the boundaries is 0. But is this true in general. Yes, it is and this is proposed as follows.

Proposition 2.4.1. For any oriented k-simplex σ $k \geq 0$, we have

$$\partial_{k-1} \circ \partial_k \sigma = 0.$$

Proof.

$$\begin{aligned}
\partial_{k-1} \circ \partial_k \sigma &= \partial_{k-1} \left(\sum_{i=0}^{k} (-1)^i \sigma_{-i} \right) \\
&= \sum_{i=0}^{k} (-1)^i \partial_{k-1}(\sigma_{-i}) \\
&= \sum_{i=0}^{k} (-1)^i \partial_{k-1}(v_0, \cdots, \cancel{v_i}, \cdots, v_k) \\
&= \sum_{i=0}^{k} (-1)^{i+j} \left(\sum_{j=0}^{i} (v_0, \cdots, \cancel{v_j}, \cdots, \cancel{v_i}, \cdots, v_k) + \sum_{j=i+1}^{k-1} (v_0, \cdots, \cancel{v_i}, \cdots, \cancel{v_j}, \cdots, v_k) \right)
\end{aligned}$$

[3]one should not get confuse with the m-face that we defined in 2.1

$$= \sum_{i=0}^{k}\sum_{j=0}^{i}(-1)^{i+j}(v_0, \cdots, \cancel{v_j}, \cdots, \cancel{v_i}, \cdots, v_k) + \sum_{i=0}^{k}\sum_{j=i+1}^{k-1}(-1)^{i+j}(v_0, \cdots, \cancel{v_i}, \cdots, \cancel{v_j}, \cdots, v_k)$$

Consider a typical summand $(v_0, \cdots, \cancel{v_a}, \cdots, \cancel{v_b}, \cdots, v_k)$. The coefficient of this in the sum would be,

$$(-1)^{a+b} + (-1)^{a+b-1} = (-1)^{a+b-1}(-1+1) = 0.$$

□

If boundary map evaluates to 0 then it must be a cycle! Thus, every boundary is a cycle. Now we observe that, suppose only a boundary of a simpex exists but not the interior of a simplex, then it can be seen as a hole. So, it is essentially not present as a boundary of a simplex as the object which makes it a boundary is not present in the simplicial complex. That is a cycle which is not a boundary(of a any simplex). Presence of a hole in the simplicial complex is an essential information of the structure of space. So, looking at the *cylces* which are not *boundaries* seems to be giving nice information of the simplcical complex. We expand this idea in the next chapter.

Chapter 3

Simplicial (Co)Homology

Earlier we were considered *sums* v.i.z., linear combinations of the simplices and the coefficient 1 and -1 gave the information about the orientaion of the path. Here, it we considered the field of three elements $\{-1, 0, 1\}$. But, we wish to extend this idea in general to any field expecting to get more information. Let K be a simplicial complex and $\mathbb{F}$ be a field.

Definition 3.0.1. For each dimension $k \geq 0$, the k-th chain group of K is the vector space $C_k(K)$ over F generated by treating the k-simplices of K as a basis.

So, an element γ in $C_k(K)$ is called k-chain of K. It can be expressed as a linear combination of the form

$$\gamma = \sum_{\sigma \text{ over } k\text{-simplices}} \gamma_\sigma \cdot \sigma,$$

where γ_σ are the elements of the field $\mathbb{F}$. Here, these chains are mere formal sums, this doesn't apriori have any meaning[1]. Earlier we have defined a boundary map to a simplex. We now linearly extend the map and we have the following.

Definition 3.0.2 (Boundary operator)**.** For each dimension $k \geq 0$, the k-th **boundary operator** of K is the $\mathbb{F}$-linear map

$$\partial_k(K) : C_k(K) \to C_{k-1}(K),\ \partial_k(\sigma) = \sum_{i=0}^{k} (-1)^i \sigma_{-i}$$

which sends each basis k-chain σ to the $(k-1)$-chain.

Remark. Thus, for any k-chain γ, we have

$$\partial_k(\gamma) = \partial_k\left(\sum_{\sigma \text{ over } k\text{-simplices}} \gamma_\sigma \cdot \sigma\right) = \sum_{\sigma \text{ over } k\text{-simplices}} \gamma_\sigma \cdot \partial_k(\sigma)$$

Now, we can restate the Proposition 2.4.1 as follows.

Corollary 3.0.1. For every $k \geq 0$, the composite

$$\partial_k(K) \circ \partial^K_{k+1} : C_{k+1}(K) \to C_{k-1}(K)$$

is a zero map.

[1]"Not everything has to mean something. Some things just are" -*Charles de Lint*

Remark. Since, these are linear maps, we can talk about the kernels[2]. Thus, the above corollary means that the image of ∂^K_{k+1} lies in the kernel of ∂^K_k.

Thus, now given any oriented simplcial complex K, we came up with a collection of finite $\mathbb{F}$-vector spaces and $\mathbb{F}$-linear maps between them. We thus have a *nice* sequence as follows,

$$\cdots \xrightarrow{\partial^K_{k+1}} C_k(K) \xrightarrow{\partial^K_k} C_{k-1}(K) \longrightarrow \cdots \longrightarrow C_1(K) \xrightarrow{\partial^K_1} C_0(K) \longrightarrow 0$$

what makes it *nice* is that composition of any two adjacent maps is a zero map. Apparently, this kind of sequences shows up left and right in the theory. Thus, these have a name.

Definition 3.0.3 (Chain complex)**.** A **chain complex** $(C_\bullet, d_\bullet)$ over the field $\mathbb{F}$ is a collection of F-vector spaces C_k and $\mathbb{F}$-linear maps $d_k : C_k \to C_{k-1}$ which satisfy the condition $d_k \circ d_{k+1} = 0$ for all $k \geq 0$.

Remark. The chain complexes $(C_\bullet(K), \partial^K_\bullet)$ that arise from a simplicial complex K is called a **simplicial chain complexes**.

As observed earlier, im $\partial^K_{k+1} \subseteq$ ker $\partial^K_k, \forall\, k \geq 0$. The chains in im ∂^K_{k+1} are nothing but the boundaries and the chains in ker ∂^K_k are nothing but the cycles. Thus, unlike earlier we can now formally say that $k+1$-boundaries are k-cycles. We mentioned that finding the k-cycles which are not $k+1$-boundaries seems to be of our interest. Since, ker ∂^K_k and im ∂^K_{k+1} are subspaces, we have to get a structure where im ∂^K_{k+1} as single element and all other sets of ker ∂^K_k as different elements. This can be achieved by quotienting im ∂^K_{k+1} out of ker ∂^K_k. We formally state this as follows.

3.1 Homology

Fix a chain complex $(C_\bullet, d_\bullet)$ over the field $\mathbb{F}$.

Definition 3.1.1 (Homology group)**.** For each $k \geq 0$, the k-th **homology group**[3] of $(C_\bullet, d_\bullet)$ is the quotient vector space

$$H_k(C_\bullet, d_\bullet) = \text{ker } d_k/\text{im } d_{k+1}$$

We call the elements in ker d_k as k-cycles and the elements in im d_{k+1} as k-boundaries.

Remark. When we have a simplicial chain complex i.e., $(C_\bullet, d_\bullet) = (C^K_\bullet, d^K_\bullet)$, then homology groups associated to it are called **simplicial homology** groups of K. We denote the homology groups of a simplicial complex K with coefficients from $\mathbb{F}$ as $H_k(K; \mathbb{F}), \forall\, k \geq 0$.

When we have a finite dimensional vector space V, then the main characteristic of it is the dimension of the vector space as any two finite dimensional vector spaces with same dimension are isomorphic. These homology groups are finite quotient spaces. Thus we can talk about its dimension and that appears to be very useful in the literature. It is named as follows.

Definition 3.1.2 (Betti numbers)**.** The dimension of the k-th simplicial homology group is called the k-th **betti number** denoted by $\beta_k(K)$

$$\beta_k(K; \mathbb{F}) = \dim H_k(K; \mathbb{F}).$$

There is s remarkable relation between the Euler characteristic and the betti numbers. That is given as follows.

[2]here is a shameless plug-in of my blurbs; one of them on why *kernels* are called *kernels*

[3]one might think why are we calling this as group. but we know that any vector space is an additive abelian group and this homology theory can be extended to coefficients in the chain complex being non-fields where we have to talk about the *free abelian groups*

Theorem 3.1.1 (Euler-Poincaré)**.** For a simplicial complex K of dimension n, we have

$$\chi(K) = \sum_{k=0}^{n} (-1)^k \beta_k(K)$$

This shows that the homology groups determine the Euler characteristic! Fantastic right!.

3.2 Categories

Consider two composable continuous functions between topological spaces. The composition is also a continuous function. Infact, as we argued earlier in the Chapter 1.1 it was defined to be such that it preserves the structure. Same in the case with other famous structures as vector spaces and groups. Take linear maps between vector spaces and their composition is also a linear map i.e., a map preserving the vector space structure. Take group homomorphisms, it's the same case again. Thus, given any structure, we can get more information about it by considering the structure preserving functions. This (structure, structure preserving functions) gives rise to an additional global structure called *Categories*. We define them formally now.

Definition 3.2.1 (Categories)**.** A **category** $\mathcal{C}$ consists of

1. a collection $\mathcal{C}_0$, whose elements are called *objects*
2. for every pair of objects x, y in $\mathcal{C}$ a set $\mathcal{C}(x, y)$ of *morphisms* from x to y, where a typical element of it is $f : x \to y$
3. for each triple x, y, z of objects a composition law

$$\mathcal{C}(x, y) \times \mathcal{C}(y, z) \to \mathcal{C}(x, z)$$

sending $f : x \to y$ and $g : y \to z$ to $g \circ f : z \to z$ such that

(a) (*identity*) for each x in $\mathcal{C}_0$, there exists an identity morphism $g \circ id_x = g$ and $id_x \circ h = h$ for any object y and morphims $g : y \to x, h : x \to y$

(b) (*associative*) given any triple of morphisms

$$x \xrightarrow{f} y \xrightarrow{g} z \xrightarrow{h} w$$

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

Examples. We have the categories like

- `Top` whose objects are topolocial spaces and morphisms are continuous maps
- $\texttt{Vect}_{\mathbb{F}}$ whose objects are $\mathbb{F}$-vector spaces and morphisms are $\mathbb{F}$-linear maps
- `Grp` whose ojects are groups and morphims are group homomorphims

Currently the structures of interest are the simplicial complexes. Moreover the simplicial maps are the structure preserving maps and composition of two simplicial map is a simplicial map. Thus we have a category,

`Simp`[4]; whose objects are simplcial complexes and the morphims are simplicial maps

[4] yes..very suggestive nomenclature, easy to *simp* over `Simp`

We have encountered this situation above many times. What do we do after defining a new structure? Yes! We define the structure preserving maps between the structures.

Definition 3.2.2 (Functors)**.** A **functor** $F : \mathcal{C} \to \mathcal{C}'$ assigns

1. to each object x in $\mathcal{C}_0$ an object Fx in $\mathcal{C}'_0$
2. to each morphism $f : x \to y$ in $\mathcal{C}$ a morphism $Ff : Fx \to Fy$ in $\mathcal{C}'$ such that
   (a) we have $F\,id_x = id_{Fx}$ for each x in $\mathcal{C}_0$ and
   (b) for any pair of morphisms f in $\mathcal{C}(x, y)$ and g in $\mathcal{C}(y, z)$, we have
   $$F(g \circ f) = Fg \circ Ff$$

So, functors are the functions between categories. Let us recall the two known categories `Simp` and $\mathtt{Vect}_{\mathbb{F}}$. We have simplicial complexes in the former and the corresponding homology groups in the later category. So, do the homology "induces" a functor? This is good question to ask. The even more good thing is the answer is true!

Theorem 3.2.1. For each $k \geq 0$, the map
$$K \mapsto H_k(K; \mathbb{F})$$
constitutes a functor from `Simp`, the category of simplicial complexes to $\mathtt{Vect}_{\mathbb{F}}$, the category of $\mathbb{F}$-vector spaces.

To get into the above theorem, we need to build some theory. We start with defining what chain maps are.

3.2.1 Functoriality

Definition 3.2.3 (Chain maps)**.** A **chain map** $\phi_\bullet$ from a chain complex $(C_\bullet, d_\bullet)$ to a chain complex $(C'_\bullet, d'_\bullet)$ is defined to ba sequence of $\mathbb{F}$-linear maps $\{\phi_k : C_k \to C'_k\}_{k\geq 0}$ such that
$$d'_k \circ \phi_k = \phi_{k-1} \circ d_k,\ \forall\, k \geq 0.$$
v.i.z, each square commutes in the following diagram.
$$\begin{tikzcd}
\cdots \arrow[r, "d_{k+2}"] & C_{k+1} \arrow[r, "d_{k+1}"] \arrow[d, "\phi_{k+1}"'] & C_k \arrow[r, "d_k"] \arrow[d, "\phi_k"] & C_{k-1} \arrow[r, "d_{k-1}"] \arrow[d, "\phi_{k-1}"] & \cdots \\
\cdots \arrow[r, "d'_{k+2}"] & C'_{k+1} \arrow[r, "d'_{k+1}"] & C'_k \arrow[r, "d'_k"] & C'_{k-1} \arrow[r, "d'_{k-1}"] & \cdots
\end{tikzcd}$$

Now, fix simplicial complexes K, L; a simplicial map $f : K \to L$ and a field $\mathbb{F}$. We then have simplicial chain complexes
$$(C_\bullet(K), \partial^K_\bullet),\ (C_\bullet(L), \partial^L_\bullet).$$
We define a map between them as follows.

Definition 3.2.4. For each $k \geq 0$, let $C_k f : C_k(K) \to C_k(L)$ be the $\mathbb{F}$-linear map between chain groups defined by the following action on each basis k-simplex σ of K
$$C_k f(\sigma) = \begin{cases} f(\sigma), & \text{if} \dim f(\sigma) = k \\ 0, & \text{otherwise} \end{cases}$$

So, we have a sequence of maps as follows

$$
\begin{tikzcd}
\cdots \arrow[r, "\partial_{k+2}^K"] & C_{k+1}(K) \arrow[r, "\partial_{k+1}^K"] \arrow[d, "C_{k+1}f"'] & C_k(K) \arrow[r, "\partial_k^K"] \arrow[d, "C_k f"] & C_{k-1}(K) \arrow[r, "\partial_{k-1}^K"] \arrow[d, "C_{k-1}f"] & \cdots \\
\cdots \arrow[r, "\partial_{k+2}^L"] & C_{k+1}(L) \arrow[r, "\partial_{k+1}^L"] & C_k(L) \arrow[r, "\partial_k^L"] & C_{k-1}(L) \arrow[r, "\partial_{k-1}^L"] & \cdots
\end{tikzcd}
$$

If we show that each square commutes, then this is a chain map. Moreover the map we defined is indeed a chain map. We show that now.

Proposition 3.2.1. For each $k \geq 0$, and k-simplex σ in K, we have

$$\partial_k^L \circ C_k f(\sigma) = C_{k-1} f \circ \partial_k^K(\sigma)$$

Proof. Since, there are two cases in the defnition of the function. We prove this in two cases.

- (Case 1:) If, $\dim f(\sigma) = k$ then $C_k f(\sigma) = f(\sigma)$. Then we have,

$$\partial_k^L \circ C_k f(\sigma) = \partial_k^L \circ f(\sigma) = \sum_{i=0}^{k} (-1)^i f(\sigma)_{-i} \tag{3.1}$$

  Moreover, since $\dim f(\sigma) = k$, f is injective and thus f is injective on the faces σ_{-i} of σ. Thus,

$$C_{k-1} f(\sigma_{-i}) = f(\sigma_{-i}) = f(\sigma)_{-i}$$

  the last equality follows from the fact that f is a simplicial map. Now observe,

$$C_{k-1} f \circ \partial_k^K(\sigma) = C_{k-1} f \left(\sum_{i=0}^{k} (-1)^i \sigma_{-i} \right) = \sum_{i=0}^{k} (-1)^i f(\sigma)_{-i} \tag{3.2}$$

  equation (3.2) and (3.2) shows the required.

- (Case 2:) If $\dim f(\sigma) < k$ then $C_k f(\sigma) = 0$. Thus, $\partial_k^L \circ C_k f(\sigma) = 0$. So it suffices to show

$$C_{k-1} f \circ \partial_k^K(\sigma) = 0.$$

  To show this, we first give orientations o_K and o_L on K and L so that f is orientation preserving i.e., if $o_K(v) < o_L(v')$ then $o_K(f(v)) < o_L(f(v'))$. Consider oriented $\sigma = (v_0, \cdots, v_k)$. Moreover since $\dim f(\sigma) < k$, there exists a vertex v_p such that $f(v_p) = f(v_{p+1})$. Thus, f fails to be injective on the vertices of every face σ_{-i} possibly except σ_p and $\sigma_{-(p+1)}$. Now,

$$
\begin{aligned}
C_{k-1} f \circ \partial_k^K(\sigma) = C_{k-1} f \left(\sum_{i=0}^{k} (-1)^i \sigma_{-i} \right) &= \sum_{i=0}^{k} (-1)^i C_{k-1} f(\sigma_{-i}) \\
&= (-1)^p f(\sigma_{-p}) + (-1)^{p+1} f(\sigma_{-(p+1)}) \\
&= (-1)^p (f(\sigma_{-p}) - f(\sigma_{-(p+1)})) \\
&= (-1)^p (f(\sigma)_{-p} - f(\sigma)_{-(p+1)})
\end{aligned}
$$

  $f(\sigma)_{-p}$ and $f(\sigma)_{-(p+1)}$ are the simplices formed by removing $f(v_p)$ and $f(v_{p+1})$ respectively. But, as $f(v_p) = f(v_{p+1})$, $f(\sigma)_{-p} = f(\sigma)_{-(p+1)}$. Thus, the above equation evaluates to 0.

□

Thus, we conclude that the simplicial maps
$$f : K \to L$$
induce chain maps
$$C_\bullet f : (C_\bullet(K), \partial^K_\bullet) \to (C_\bullet(L), \partial^L_\bullet).$$
Now, we prove some more propositions to finally arrive at the functoriality.

Proposition 3.2.2. Let $\phi_\bullet : (C_\bullet, d_\bullet) \to (C'_\bullet, d'_\bullet)$ be a chain map. For each $k \geq 0$, there exists a well defined $\mathbb{F}$-linear map $H_k\phi : H_k(C_\bullet, d_\bullet) \to H_k(C'_\bullet, d'_\bullet)$

Proof. We recall that $H_k(C_\bullet, d_\bullet) = \ker d_k / \text{im } d_{k+1}$. So, we have to get a map from $\ker d_k/\text{im } d_{k+1} \to \ker d'_k/\text{im } d'_{k+1}$. Thus, it is enough to show that under the map ϕ_k the $\ker d_k$ maps to $\ker d'_k$ and the im d_{k+1} maps to im d'_{k+1}. Consider a k-cycle α v.i.z., $d_k(\alpha) = 0$. Since, it's a chain map we have
$$d'_k \circ \phi_k(\alpha) = \phi_{k-1} \circ d_k(\alpha) = 0$$
thus, $\phi_k(\alpha) \in \ker d'_k$. Now, suppose $\beta \in \text{im } d_{k+1}$. Then there exists a $\gamma \in C_{k+1}$ such that $d_{k+1}(\gamma) = \beta$. Thus, again from chain map commutativity we have
$$\phi_k(\beta) = \phi_k \circ d_{k+1}(\gamma) = d'_{k+1} \circ \phi_{k+1}(\beta)$$
thus, $\phi_k(\beta) \in \text{im } d'_{k+1}$. Thus, done. □

Similarly one shall prove that

Proposition 3.2.3. For $k \geq 0$, given chain maps $\phi_\bullet : (C_\bullet, d_\bullet) \to (C'_\bullet, d'_\bullet)$ and $\psi_\bullet : (C'_\bullet, d'_\bullet) \to (C''_\bullet, d''_\bullet)$ we have
$$H_k(\psi \circ \phi) = H_k\psi \circ H_k\phi$$
Thus, combining all these we shall conclude that our map in theorem 3.2.1 is indeed a functor.

3.3 Triangulation returns

Though a digression, this small discussion on triangulation is essential. Though every compact surface admits a triangulation, finding the (simplicial) triangulation is hard. Let us discuss an example.

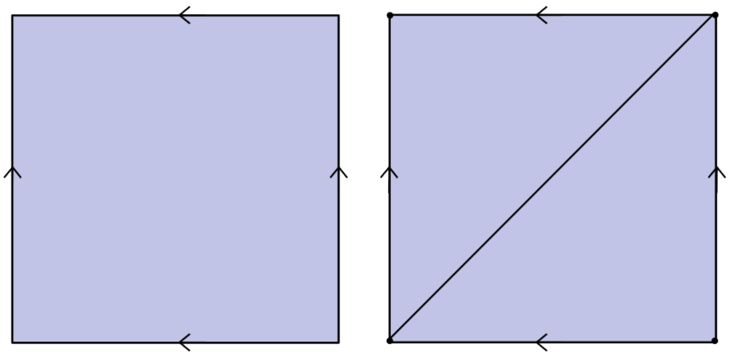

Figure 3.1: on left: identification of torus; on right: we have something like a triangulation, is it really a triangulation?

We consider the torus being identified as shown. Our interest is the triangulation of the torus. So, we triangulate the square first and then identify it to get a triangulation of torus. So, the natural way that we would think is the triangulation that is shown in the above figure. But, unfortunately that is a (simplicial) triangulation. We give the reasons now. In a triangulation, a triangle is uniquely determined by three vertices. So, if you give me three vertices, if it exists, there should be an unique triangle corresponding to them. But, here, all the four vertices are being identified to a single vertex. Thus, given any three vertices, it doesn't guarantee us to get a unique triangle, it can be any of the two as in the figure. Moreover, the simplices must meet at simplices for a simplicial complex. Since, our triangulation by definition was a simplicial complex, this fails in this case.

3.3.1 How would one get a triangulation?

So, how would one get a triangulation? One straightforward way (though not easy to prove) is to start with a *good* enough triangulation. The one where any two triangles meet at a face, something as shown in the earlier figure. I shall give the explicit meaning of the word *good* soon. Then, performing barycentric subdivisions twice would give us a simplicial complex and thus a triangulation. The *good* triangulations here are the Δ-complexes. One can see Δ-complexes defined in page number 103 of [5]. Further, the exercise 2.1.23 of the same tells us the result that

"the second barycentric subdivision of a Δ-complex is a simplicial complex"

But, second barycentric subdivision would contain many triangles and thus working with that wouldn't be preferable. In a hands on way we have a following triangulation of torus which has around 18 triangles, 9 vertices, 27 edges after identification.

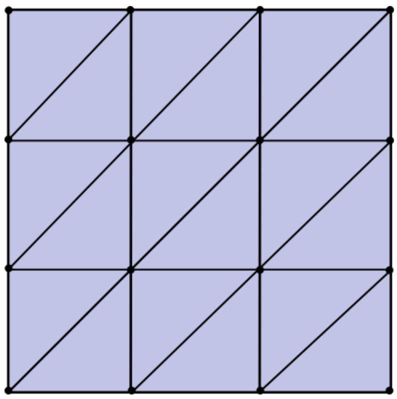

Figure 3.2: A Triangulation of Torus

This brings us to the question does there exist any *minimal* triangulation? Here, minimal in the sense of minimal number of faces/triangles. It turns out that we have a nice result in the exercise 8.2 of [8] which states the following.

Theorem 3.3.1. For any triangulation of a compact surface K, show that

$$3f = 2e \tag{3.3}$$

$$e = 3(v - \chi(K)) \tag{3.4}$$

$$v \geq \frac{1}{2}(7 + \sqrt{49 - 24\chi(K)}) \tag{3.5}$$

where, v, e, f are the number of vertices, edges and faces respectively and $\chi(K)$ is the Euler characteristic of K.

Remark. Earlier I mentioned about the clarity about uniqueness of the object $\chi(K)$. Since, K admits many triangulations one would think this definition is not consistent. But, here comes the power of homology. Here's an result from Chapter 7 of [13]

Result (Alexander-Veblen). Let, X be a polyhedron with triangulations K and K'. Then, $H_k(K) \cong H_k(K')$ for all $k \geq 0$.

Now, since we know that Euler characteristics are characterised by the homology groups uniquely, we conclude that for any two triangulations of a compact surface, the Euler characteristic is invariant.

The proof of the theorem 3.3.1 is straightforward. We shall absorb the idea of the proof given here and do as follows.

Proof. We count the edges. Since, each face consists of 3 edges, the number of edges from faces is $3f$. Now, we observe any two faces meet at an edge. Thus, we are counting an each edge twice. Thus, we have the first equation 3.3

$$3f = 2e.$$

Moreover, $\chi(K) = v - e + f$ this combined with first equation 3.3 gives the second equation (3.4)

$$e = 3(v - \chi(K)).$$

Suppose, we have n vertices then maximum number of edges that we can draw between them is $\binom{n}{2}$. Thus, we have the inequality

$$e \leq \frac{v(v-1)}{2}$$

This combined with the equation (3.4) gives following,

$$3(v - \chi(K)) \leq \frac{v(v-1)}{2} \implies 6v - 6\chi(K) \leq v^2 - v \implies v^2 - 7v + 6\chi(K) \geq 0$$
$$\implies v \geq \frac{1}{2}(7 + \sqrt{49 - 24\chi(K)})$$

□

Since we already have a triangulation of torus with 18 triangles, 9 vertices, 27 edges as shown above. We can say that Euler characteristic of the torus is $9 - 18 + 27 = 0$. Thus, we substitute $\chi(K) = 0$ in the above inequality to get

$$v \geq \frac{1}{2}(7 + \sqrt{49 - 24(0)}) = \frac{1}{2}(14) = 7 \implies \boxed{v \geq 7}$$

Thus, a triangulation of torus should minimum have seven vertices in it. But, this leads to a question "is the bound tight?" In other words, is the minimal triangulation achieved? Turns out that this is actually true modulo some cases.

1. G. Ringel in 1955 showed that the inequality in (3.5) is tight for all non-orientable surfaces except for two cases
   - Klien Bottle in which the minimal vertices is showed to be 8 as opposed to 7 from the inequality
   - non-orientable surface of genus 3 where minimal vertices is showed to be 9 as opposed to 7 from the inequality
2. In 1980, Jungerman and Ringel in the article [6] showed that the inequality in (3.5) is tight for all orientable surfaces S_p of genus p except for $p = 2$ for which minimal number of vertices is 10 as opposed to 8.

Thus there exists a minimal triangulation with 7 vertices for torus and it is given in the figure 3.3.

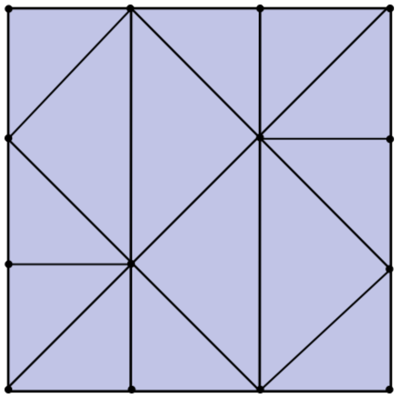

Figure 3.3: Minimal Triangulation of Torus

3.4 Cohomology

We recall the notion of dual of a vector space. Consider a $\mathbb{F}$-vector space V. The dual V^* of V is the set of all F-linear maps from $V \to \mathbb{F}$. Is this anything more than set? Yes, it is. It is also an $\mathbb{F}$-vector space. We can dualize a linear map as follows. For vector spaces V, W and for every linear map $A : V \to W$, the dual $A^* : W^* \to V^*$ which sends $q \mapsto q \circ A$. We have a chain complex which consists of vector spaces and linear maps between them. Now, we shall dualize them as follows. Consider a chain complex over $\mathbb{F}$

$$\cdots \xrightarrow{d_{k+1}} C_k \xrightarrow{d_k} C_{k-1} \longrightarrow \cdots \longrightarrow C_1 \xrightarrow{d_1} C_0 \longrightarrow 0$$

and we dualize everything to get

$$\cdots \xleftarrow{d_{k+1}^*} C_k^* \xleftarrow{d_k^*} C_{k-1}^* \longleftarrow \cdots \longleftarrow C_1^* \xleftarrow{d_1^*} C_0^* \longleftarrow 0$$

For chain maps we had the property that composition of two adjacent maps evaluates to zero. Does that property get induced in our new chain? Let us find out. For $k \geq 0$, consider linear map $\alpha : C_k \to \mathbb{F}$. This is an element in C_k^*. Thus, observe that following

$$d_{k+2}^* \circ d_{k+1}^*(\alpha) = d_{k+2}^* \circ (\alpha \circ d_{k+1}) = \alpha \circ d_{k+1} \circ d_{k+2}$$

The last two equalities are from the definition of dual. Now, this evalueates to zero for any α as $d_{k+1} \circ d_{k+2} \equiv 0$. Thus, $d_{k+2}^* \circ d_{k+1}^* \equiv 0$. So, the property of the "chain" is preserved. Asusual even this appears in the literature often and thus we name it.

Definition 3.4.1 (Cochain complexes)**.** A **cochain complex** $(C^\bullet, d^\bullet)$ over $\mathbb{F}$ is a sequence of vector spaces and linear maps of the form

$$0 \longrightarrow C^0 \xrightarrow{d^0} C^1 \xrightarrow{d^1} \cdots \longrightarrow C^k \xrightarrow{d^k} C^{k+1} \xrightarrow{d^{k+1}} \cdots$$

such that $d^{k-1} \circ d^k = 0$ for all $k \geq 1$.

Here, C^k is called k-th cochain group and each d^k is called k-th coboundary map. Again, we have a sequence of vector spaces and maps between them thus we can talk about kernels. As observed above, im $d^{k-1} \subseteq$ ker d^k. Thus, same funda, we just quotient it out.

Definition 3.4.2 (Cohomology)**.** For each $k \geq 0$, the k-th **cohomology group** of $(C^\bullet, d^\bullet)$ is defined to be the quotient space

$$H^k(C^\bullet, d^\bullet) = \ker d^k / \mathrm{im}\, d^{k-1}$$

Elements of $\ker d^k$ are calle k-cocyles and that of $\mathrm{im}\, d^{k-1}$ are called k-coboundaries.Now, we come to the realm of simplicial complexes. Let K be a simplicial complex and $C_k(K)$ be k-chain groups for all $k \geq 0$. The corresponding duals will be $C_k(K)^*$ denoted by $C^k(K)$ for sanity and duals of boundaries are defined as follows.

Definition 3.4.3 (Simplicial Coboundary operator)**.** The k-th simplicial coboundary operator $\partial_K^k : C^k(K) \to C^{k+1}(K)$ is the dual of the boundary operator $\partial_k^K : C_{k+1}(K) \to C_k(K)$. Thus it satisfies, $\partial_K^k(\alpha^*) = \alpha^* \circ \partial_{k+1}^K$ for each α^* in $C^k(K)$.

Remark. Thus, consider any cochain $\zeta \in C^k(K)$ and an oriented $(k+1)$-simplex $\sigma = (v_0, v_1, \cdots, v_{k+1})$. Then $\partial_K^k \zeta$ is a $(k+1)$-cochain that acts on σ. It acts as follows

$$\partial_K^k \zeta(\sigma) = \sum_{i=0}^{k} (-1)^i \zeta(\sigma_{-i})$$

The k-th cohomology group of the simplicial cochain complex $(C^\bullet(K), \partial_K^\bullet)$ is called the k-th simplicial cohomology group of K and denoted by $H^k(K; \mathbb{F})$.

We observe that ∂_K^k is the transpose of ∂_{k+1}^K thus their ranks are same as linear maps. Exploit that fact and a simple application of rank-nullity theorem will give us the following.

Proposition 3.4.1. Let $(C^\bullet, \partial^\bullet)$ be the dual chain complex of $(C_\bullet, \partial_\bullet)$ over a field $\mathbb{F}$ whose $\dim C_k$ is finite for all k. Then,

$$\dim H_k(C_\bullet^K, \partial_\bullet) = \dim H^k(C^\bullet, \partial^\bullet)$$

Moreover, similarly as we had the chain map inducing the functor between `Simp` and $\mathtt{Vect}_{\mathbb{F}}$. Here too with the same machinery as we done in the simplicial homology case we can conclude that the cochain maps induce a functor between `Simp` and $\mathtt{Vect}_{\mathbb{F}}$ where it takes a simplicial complex K to its $H^k(K; \mathbb{F})$ for all k. There are two very useful products that we can impose on the cochain maps. Since, cochains acts on chains and will be taking are on the field, the multiplication comes from that. The natural possibilities of product are take two cochains and produce a new cochain; or take one cochain and a chain produce another chain. These two products are sort of dual notions to each other. We now see these products.

3.5 Cup and Cap products

Fix an oriented simplicial complex K, so that each k-simplex σ can be uniquely written as an increasing list of vertices $(v_0, \cdots, v_k)$.

Definition 3.5.1 (Front and back faces)**.** For each $i \geq 0$, the i-th **front face** of σ is the i-dimensional simplex $\sigma_{\leq i} = (v_0, v_1, \cdots, v_i)$ and the i-th **back face** of σ is the $(k-i)$-dimensional simplex $\sigma_{\geq i} = (v_i, \cdots, v_k)$.

Definition 3.5.2 (Cup product)**.** Fix a simplicial complex K. Let, $\xi \in C^k(K)$, k-cochain and $\eta \in C^l(K)$, l-cochain. The **cup product** of ξ and η is a $(k+l)$-cochain $\xi \smile \eta \in C^{k+l}(K)$ defined by an action of $(k+l)$-dimensional simplex σ as

$$\xi \smile \eta(\sigma) = \xi(\sigma_{\leq k}) \cdot \eta(\sigma_{\geq k})$$

Remark. Since, ξ is a k-cochain, it acts on k-chains and thus it takes the front faces. Whereas, eta is a l-cochain and it acts on l-chains and thus it takes the back faces which are $k + l - k = l$-dimensional simplices.

Definition 3.5.3 (Cap product)**.** The **cap product** of an i-cochain ξ with a k-chain

$$\gamma = \sum_{\sigma} \gamma_\sigma \sigma$$

is a new $(k-i)$-chain $\xi \frown \eta$ defined as,

$$\xi \frown \eta = \sum_{\sigma} \gamma_\sigma \cdot \xi(\sigma_{\leq i}) \cdot \sigma_{\geq i}$$

Proposition 3.5.1. For any $\xi \in C^k(K;\mathbb{F})$ and $\eta \in C^l(K;\mathbb{F})$ we have

$$\partial_K^{k+l}(\xi \smile \eta) = \left(\partial_K^k(\xi) \smile \eta\right) + (-1)^k \cdot \left(\xi \smile \partial_K^l(\eta)\right)$$

Proposition 3.5.2. For each simpicial complex K and dimensions $k, l \geq 0$, the cup product map $\smile : C^k(K;\mathbb{F}) \times C^l(K;\mathbb{F}) \to C^{k+l}(K;\mathbb{F})$ induces a well-defined bilinear map of cohomology groups.

Proposition 3.5.3. For each $\xi \in C^i(K)$ and $\gamma \in C_k(K)$, we have

$$\partial_{k-i}^K(\xi \frown \gamma) = (-1)^i \cdot \left((\xi \frown \partial_k^K(\gamma)) - (\partial_K^i(\xi) \frown \gamma)\right)$$

Proposition 3.5.4. For each simpicial complex K and dimensions $i \leq k$, the cap product map $\smile : C^i(K;\mathbb{F}) \times C_k(K;\mathbb{F}) \to C_{k-i}(K;\mathbb{F})$ induces a well-defined bilinear map of cohomology groups.

We now try to prove this proposition.

Proof. Since,coboundaries and boundaries are the identity elements in the respective groups. It is enough to check three cases. Cap product of cocycle and a cycle; cocycle and a boundary; coboundary and a cycle.

- (cocycle $\frown$ cycle = cycle) Let ξ and γ are i-cocyle and k-cycle respectively, we have $\partial_K^i(\xi) = 0 = \partial_k^K(\gamma)$. then from the proposition 3.5.3 we have,

  $$\partial_{k-i}^K(\xi \frown \gamma) = (-1)^i \cdot ((\xi \frown 0) - (0 \frown \gamma)) = 0$$

  thus, cocycle $\frown$ cycle is a cycle.

- (cocycle $\frown$ boundary = boundary) Let ξ be i-cocycle. Then, $\partial_K^i(\xi) = 0$. Further let β be a k-boundary. Then there exists $\alpha \in C_{k+1}(K)$ such that $\beta = \partial_{k+1}(\alpha)$. Consider the proposition 3.5.3 for ξ and α. We then have,

  $$\begin{aligned} \partial_{k+1-i}(\xi \frown \alpha) &= (-1)^i((\xi \frown \partial_k^K(\alpha)) - (\partial_K^i(\xi) \frown \alpha)) \\ \partial_{k+1-i}(\xi \frown \alpha) &= (-1)^i((\xi \frown \beta) - (0 \frown \alpha)) \\ \xi \frown \beta &= (-1)^i \cdot \partial_{k+1-i}(\xi \frown \alpha) \end{aligned}$$

  So, modulo the orientation, we have got a $(k+1-i)$-chain $\xi \frown \alpha$ such that its output under the boundary map is $\xi \frown \beta$. Thus, cocycle $\frown$ boundary = boundary.

- (coboundary $\frown$ cycle = boundary) Let, ξ be i-coboundary and γ be a k-cycle. Then there exists η such that $\partial^{i-1}(\eta) = \xi$. Now, consider the proposition 3.5.3 for η and γ. We then have,

  $$\begin{aligned} \partial_{k-i+1}(\eta \frown \gamma) &= (-1)^i((\eta \frown \partial_k^K(\gamma)) - (\partial_K^{i-1}(\eta) \frown \gamma)) \\ \partial_{k-i+1}(\eta \frown \gamma) &= (-1)^i((\eta \frown 0) - (\xi \frown \gamma)) \\ \xi \frown \gamma &= (-1)^i \cdot \partial_{k-i+1}(\eta \frown \gamma) \end{aligned}$$

  So, modulo the orientation, we have got a $(k-i+1)$-chain $\eta \frown \gamma$ such that its output under the boundary map is $\xi \frown \gamma$. Thus, coboundary $\frown$ cycle = boundary.

Thus, this induces a bilinear maps of cohomology groups and it's often used by the same symbol a cup product

$$\frown: H^i(K;\mathbb{F}) \times H_k(K;\mathbb{F}) \to H_{k-i}(K;\mathbb{F}).$$

□

3.5.1 Uses of cap products

There's an astounding result that if we know just homology groups then we can know the cohomology groups. That is achieved by cap products. Though proving that result in the realm of simplicial complex is hard, we just state the result. Before that let us do some theory building. Let M be a simplicial complex whose geometric realisation $|M|$ is a compact, connected n-dimensional manifold. We now give a new notion orientable.

Definition 3.5.4 (Orientable)**.** We say M is **orientable** over field $\mathbb{F}$ of there exists a function

$$\omega : \{n\text{-simplices of } M\} \to \{\pm 1\}$$

such that, the chain

$$[M] = \sum_{\sigma:\dim\sigma=n} \omega(\sigma)\cdot\sigma$$

is an n-cycle in $C_n(M;\mathbb{F})$.

Theorem 3.5.1 (Poincaré Duality)**.** Let M be orientable over $\mathbb{F}$. For each i in $\{0,1,\cdots,n\}$, the linear map

$$D_i : H^i(M;\mathbb{F}) \to H_{n-i}(M;\mathbb{F})$$

given by $D_i(\xi) = \xi \frown [M]$ is an isomorphism of $\mathbb{F}$-vector spaces.

Using this theorem we have a useful corollory.

Corollary 3.5.1. M be simplicial complex with the same adjectives as above, then the following hold:

1. The betti numbers $\beta_0(M), \beta_1(M), \cdots, \beta_n(M)$ are palindromic v.i.z.,

$$\beta_k = \beta_{n-k},\ \forall\, k$$

2. If n is odd, then $\chi(M) = 0$
3. If n is even($= 2i$), then

$$\chi(M) \text{ is odd} \iff \beta_i(M) \text{ is odd.}$$

Proof. 1. we observe that,

$$\begin{aligned}
\beta_k(M) &= \dim H_k(M) && \text{(definition)}\\
&= \dim H^{n-k}(M) && \text{(thm. 3.5.1)}\\
&= \dim H_{n-k}(M) && \text{(prop. 3.4.1)}\\
&= \beta_{n-k}(M) && \text{(definition)}
\end{aligned}$$

2. we have from theorem 3.1.1 that, $\chi(M) = \sum_{k=0}^{n}(-1)^k\beta_k(M)$ So, if n is odd, then β_k and β_{n-k} will appear in opposite signs from the above. Thus $\chi(M) = 0$.

3. we observe that

$$\begin{aligned}\chi(M) = \sum_{k=0}^{n}(-1)^k\beta_k(M) &= \sum_{k=0}^{i-1}(-1)^k\beta_k(M) + \sum_{k=i+1}^{2i}(-1)^k\beta_k(M) + \beta_i(M)\\ &= \sum_{k=0}^{i-1}(-1)^k\beta_k(M) + \sum_{n-k=0}^{i-1}(-1)^{n-k}\beta_k(M) + \beta_i(M)\\ &= 2(\cdots) + \beta_i(M)\end{aligned}$$

thus, $\chi(M)$ is an odd number if and only if β_i is odd.

□

Chapter 4

Persistent Homology

We started our study in the earlier chapters with trying to understand the given data points better by giving it structure. We realised that presence of "holes" in the structure is an important information. We developed simplicial homology groups which were essentiallly giving information about the "holes" in the simplicial complex. Now, given a raw point cloud data, can we somehow get a structure out of it? a simplicial complex maybe? such that we could use our homological methods and find information about the holes which would essentially give the shape of the data. So, our first aim is to get a simplicial complex structure out of a point cloud data. That's the next section.

4.1 Filtrations

Definition 4.1.1 (Filtration)**.** Let K be a simplicial complex, a **filtration** of K(of length n) is a nested sequence of subcomplexes of the form

$$F_1K \subseteq F_2K \subseteq \cdots \subseteq F_{n-1}K \subseteq F_nK = K$$

We now build a filtration on top of finite sets of points in two ways; each has their own merits and de-merits.

Definition 4.1.2 (Čech filtration)**.** Let M be a finite subset of a metric space (Z, d). The **Čech filtration** of M with respect to Z is the increasing family of simplicial complexes $C_{\epsilon\epsilon\geq 0}$ with the rule that: a subset $\{x_0, x_1, \cdots, x_k\} \subset M$ forms a k-dimensional simplex in $C_\epsilon(M)$ if an only if there exists some z in Z satisfying $d(z, x_i) \leq \epsilon$ for all i.

Definition 4.1.3 (Vietoris-Rips filtration)**.** Let (M, d) be a finite metric space. The **Vietoris-Rips filtration** of M is an increasing family of simplicial complexes $VR_\epsilon(M)$ indexed by the real numbers $\epsilon \geq 0$ with the rule that :a subset $\{x_0, x_1, \cdots, x_k\} \subset M$ forms a k-dimensional simplex in $VR_\epsilon(M)$ if and only if the pairwise distances satisfy $d(x_i, x_j) \leq \epsilon$ for all i, j.

4.1.1 Why Persistent Homology?

So, the fundamental idea is that as the radius increases at each point, we will have simplicial complexes. So, initially everything will be disconnected. Slowly, the connectivity increases, and in between some holes will be formed and be destroyed, that is, those holes will be filled as the radius increases. So, simplicial homology can only count the holes for a fixed ϵ and gives information about the holes at that radius. But those holes might be filled immediately for a very subtle increase in the radius. This type of holes are redundant for us. But this information is not given by simplicial homology. A hole is a hole for simplicial homology, there is no additional adjective such as *good* hole or *bad* hole. Thus, an introduction of new notion was required, which gives us information about which holes last for long time v.i.z. persists for long time. That's *Persistent Homology*.

So, consider any filtration

$$\phi = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_{n-1} \subseteq K_n = K$$

For every $i \leq j$ we have an inclusion map from the underlying space of K_i to that of K_j and therefore an induced homomorphism,

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

for each dimension p. The filtration thus corresponds to a sequence of homology groups connected by homomorphisms,

$$0 = H_p(K_0) \longrightarrow H_p(K_1) \longrightarrow \cdots \longrightarrow H_p(K_n) = H_p(K)$$

for each dimension p. As we go from K_{i-1} to K_i, we gain new homology classes, and we lose some when they become trivial or merge with each other. As mentioned earlier, we collect the classes that are born at or before a given threshold and die after another threshold in groups.

Definition 4.1.4 (Persistent homology group)**.** The p-th persistent homology groups are images of the homomorphisms induced by inclusion, $H_p^{i,j} = \text{im } f_p^{i,j}$ for $0 \leq i \leq j \leq n$.

Remark. We note that, $H_p^{i,i} = H_p(K_i)$. The persistent homology groups contain the homology classes of K_i that are still alive at K_j in other words,

$$H_p^{i,i} = \ker d_p^{K_i} / \text{im } d_p^{K_j} \cap \ker d_p^{K_i}$$

But, this raises the question that, the Vietoris-Rips filtrations or the Čech filtrations are the continuous radii processes. So, we have a collection of homologies indexed by $\mathbb{R}$. Then how do we make sense of the persistent homology? So, we build our theory abstractly and impose some finite type conditions and we see that amazingly information of every such continuous type is uniquely given by finitely many vector spaces. This is leads to defining persistence modules.

4.2 Persistence modules

Definition 4.2.1 (Persistence module)**.** A persistence module is a pair (V, π) where V is a collection $\{V_t\}_{t\in\mathbb{R}}$ of finite dimensional $\mathbb{F}$-vector spaces and π is a collection $\{\pi_{s,t}\}_{s<t\in\mathbb{R}}$ of $\mathbb{F}$-linear maps $\pi_{s,t} : V_s \to V_t$ such that the following diagram commutes

$$\begin{tikzcd} V_s \arrow[rr, "\pi_{s,t}"] \arrow[dr, "\pi_{s,r}"'] & & V_t \arrow[dl, "\pi_{t,r}"] \\ & V_r & \end{tikzcd}$$

We also seek some finite-type conditions which are as follows,

1. there exists a finite set S such that $t \in \mathbb{R} \setminus S$, there exists a neighborhood $\mathcal{U}$ of t such that $\pi_{r,s}$ is isomorphism for all $r < s \in \mathcal{U}$
2. (*semi-continuity*) for all $t \in \mathbb{R}$, for all $t \leq s$ and $t - s$ sufficiently small the map $\pi_{s,t}$ is isomorphism

these 1 and 2 double imply the following criterion which we take to be the finite-type criterion from now on

$$\forall\, r \in \mathbb{R},\ \exists\, s \in S \text{ such that } s < r \text{ and } \pi_{s,t} \text{ is isomorphism } \forall\, t \in (s, r].$$

Suppose (V, π) and (W, θ) are two persistent modules then a morphism between them is

$$\varphi : (V, \pi) \to (W, \theta)$$

a family of linear maps $\{\varphi_t : V_t \to W_t\}_{t\in\mathbb{R}}$ such that the following diagram commutes for every $s < t \in \mathbb{R}$

$$\begin{array}{ccc} V_s & \xrightarrow{\varphi_s} & W_s \\ {\scriptstyle \pi_{s,t}}\downarrow & & \downarrow{\scriptstyle \theta_{s,t}} \\ V_t & \xrightarrow{\varphi_t} & W_t \end{array}$$

Thus, with the persistent modules (V, π) as objects and ϕ as the morphisms we have a category of persistent modules called `Pers`. We now see a specific finite-type persistent module. Let I be an interval of the form $(a, b]$ or (a, ∞)

Definition 4.2.2 (Interval module)**.** We define **interval module** $\mathbb{F}(I)$ by letting

$$\mathbb{F}(I)_t = \begin{cases} \mathbb{F}, & \forall\, t \in I \\ 0, & \text{otherwise} \end{cases}$$
$$\pi_{s,t} = \begin{cases} id_{\mathbb{F}}, & \text{if } s, t \in I \\ 0, & \text{otherwise} \end{cases}$$

The remarkable theorem is any persistent module has an unique representation in terms of the interval modules. This is given by the following

Definition 4.2.3 (Structure theorem)**.** Let (V, π) be finite type persistent module. Then there exists a finite collection $\{(I_i, m_i)\}_{i=1}^N$ of intervals I_i and their multiplicities m_i such that

$$V = \bigoplus_{i=1}^{N} \mathbb{F}(I_i)^{m_i}$$

Remark. This finite collection/multiset $\{(I_i, m_i)\}_{i=1}^N$ uniquely determines the persistence module thus this have a name called **Barcodes**.

Proving the structure theorem requires a lot of theory building. We shall now try to build some theory which would only partially help us in the proof but nevertheless we do so. Primary question should be why persistent modules are called modules? Algebraically the notion of module is not at all near to the so-called persistent module that we are seeing here. But here's the catch, each finite-type persistent module can be looked as a $\mathbb{F}[t]$ graded module. We shall now see some graded rings and modules.

4.2.1 Structure theorem

Definition 4.2.4 (Group ring)**.** Let G be a group and R be a ring. The group ring of G over R denoted by $R[G]$ is the set of formal sums

$$\sum_{g\in G} r_g g$$

where, $r_g \in R, \forall\, g \in G$ and $r_g = 0$ for all but finitely many g.

Remark. The operations are the usual coefficient-wise addition and the multiplication is the usual multiplication of formal sums commutative w.r.t elements of G.

Definition 4.2.5 (Graded ring)**.** Fix a group G. Consider a ring R. We say R is G-**graded**, if there exists a family

$$\{R_g,\, g \in G\}$$

of additive subgroups R_g of R such that

$$R = \bigoplus_{g \in G} R_g$$

and for every $g, h \in G$,

$$R_g R_h \subseteq R_{gh}.$$

Definition 4.2.6 (Graded module)**.** Let ring R be G-**graded**. A G-**graded** R-**module** is a module over the ring R such that

$$M = \bigoplus_{g \in G} M_g$$

where every M_g is an additive subgroup of M and for every $g, h \in G$,

$$R_g M_h \subseteq M_{gh}.$$

The category `RING` has rings as objects and morphisms are the ring homomorphims. By fixing a group G, we have a category G-`Gr`, whose objects are G-graded rings and morphism between two G-graded rings R and S is a ring homomorphism,

$$\varphi : R \to S$$

such that

$$\varphi(R_g) \subseteq S_g \,\forall\, g \in G.$$

Consider a G-graded ring R. We have a category R-`Gr-Mod` whose objects are R-graded modules and morphism between two R-garded modules M and N is a module homomorphism,

$$\varphi : M \to N$$

such that

$$\varphi(M_g) \subseteq N_g \,\forall\, g \in G.$$

Remark. idea is to have the morphims preserving the degree[1].

We look at the category of all G-graded modules over $R[G]$ called $R[G]$-`Gr-Mod`. In particular, our interest is $\mathbb{F}[\mathbb{R}]$-`Gr-Mod`. We now see that following lemma.

Lemma 4.2.1. The subcategory of finite type persistence modules of `Pers` is equivalent to the sub category of finitely generated $\mathbb{R}$-graded $\mathbb{F}[\mathbb{R}]$-modules of $\mathbb{F}[\mathbb{R}]$-`Gr-Mod`.

We need to give a forward and a backward map to establish the equivalence. In the name of proof we shall give the maps but showing the maps are actually functor that take finite type `Pers` modules to finitely generated $\mathbb{F}[\mathbb{R}]$-`Gr-Mod` module and vice versa requires a lot more theory to be build. Since, we just know enough theory to understand the maps, we just define the maps.

Proof.

[1]degree of any M_g is the corresponding group element g

- **(Forward map)** For, $(\mathcal{M}, \pi) \in \texttt{Pers}$, we define the forward map α as,

$$\alpha(\mathcal{M}) = \bigoplus_{r \in \mathbb{R}} M_r$$

and the multiplication is given by $x \cdot m_r = \pi_{r,x+r}(m_r)$.

- **(Backward map)** We define the backward map β as

$$\beta\left(\bigoplus_{r\in\mathbb{R}} M_r\right) = (\{M_r\}_{r\in\mathbb{R}}, \{\pi_{r_1,r_2}\}_{r_1<r_2\in\mathbb{R}})$$

where,

$$\pi_{r_1,r_2}(m_{r_1}) = (r_2 - r_1) \cdot m_{r_1}$$

$\square$

So, since finitely generated =, we can consider the category on the right hand side as $\mathbb{F}[\mathbb{N}]$-$\texttt{Gr-Mod}$, which is indeed same as $\mathbb{F}[t]$-$\texttt{Gr-Mod}$ as $\mathbb{F}[t] = \bigoplus_{i\geq 0} \mathbb{F}t^i$. So, every finite type persistent moule is equivalent to a $\mathbb{F}[t]$ graded module. Moreover, $\mathbb{F}[t]$ is a PID. Thus, from the structure theorem for graded modules over graded principal ideal domains[2], we have the representation as

$$\left(\bigoplus_{i=1}^{n} \Sigma^{\alpha_i}\mathbb{F}[t]\right) \oplus \left(\bigoplus_{i=1}^{m} \Sigma^{\gamma_j}\mathbb{F}[t]/(t^{\beta_j})\right)$$

Here, Σ^{α_i} is the shift operator shifting the grade by α_i. We associate a graded $\mathbb{F}[t]$-module to a set S of intervals via a bijection Q. We define $Q(i,j) = \Sigma^i\mathbb{F}[t]/(t^{j-i})$ for an interval $(i,j) \in S$ and $Q(i,\infty) = \Sigma^i\mathbb{F}[t]$. Since, the representation in the structure theorem is unique we have the corresponding intervals with the multiplicities as given by the bijection mentioned above, which consists of the Barcode. Thus, given a persistent module the corresponding barcode uniquely determines it.

Let us zoom back a bit and try to understand what's going on. We have a set of data, from which we get a continuous persistence modules V which consists of the homologies. Since, we impose finite type conditions, we have a unique representation of the persistence module in terms of a multiset of intervals called barcode $\mathcal{B}(V)$. So this assignment $V \mapsto \mathcal{B}(V)$ is an injection. So, the information about the shape of data is completely given by $\mathcal{B}(V)$. Now, if there is just a slight perturbation in the data, we expect that there also should be just a slight perturbation in the corresponding barcodes. So the "distance" between persistence modules V and W should be same as the "distance" between $\mathcal{B}(V)$ and $\mathcal{B}(W)$. But, to make sense of this properly we need to define the notion of distances in both the spaces. We do that in the next section.

4.2.2 Stability theorem

Definition 4.2.7 (δ-Shift)**.** For a persistence module (V,π) and $\delta \in \mathbb{R}$, define a persistence module $(V[\delta], \pi[\delta])$ called the δ-shift of V by taking $(V[\delta])_t = V_{t+\delta}$ and $(\pi[\delta])_{s,t} = \pi_{s+\delta,t+\delta}$

Definition 4.2.8 (δ-shift operator)**.** The map $\Phi^\delta : (V,\pi) \to (V[\delta], \pi[\delta])$ defined by $\Phi^\delta_t = \pi_{t,t+s}$ is a morphism of persistence modules called δ-shift morphism.

Remark. If, $F : V \to W$ is a morphism then corresponding morphism between their δ-shifts is denoted by $F[\delta] : V[\delta] \to W[\delta]$.

[2]apparently it is again a very involved result and a elaborative comment on this is written by Clara Löh in the article [7]

Definition 4.2.9 (δ-interleaved)**.** Given a $\delta > 0$, we say two persistence modules are δ-interleaved if there exists two morphisms

$$F : V \to W[\delta] \text{ and } G : W \to V[\delta]$$

such that the following diagrams commute

$$\begin{tikzcd} V \arrow[r, "F"] \arrow[dr, "\Phi_V^{2\delta}"'] & W[\delta] \arrow[d, "G[\delta]"] \\ & V[2\delta] \end{tikzcd} \qquad \begin{tikzcd} W \arrow[r, "G"] \arrow[dr, "\Phi_W^{2\delta}"'] & V[\delta] \arrow[d, "G[\delta]"] \\ & W[2\delta] \end{tikzcd}$$

We call F and G as δ-interleaving morphisms.

Definition 4.2.10 (Interleaving distance)**.** For two persistence modules (V, π) and (W, θ), we define the **interleaving distance** between them to be

$$d_{int}(V, W) = \inf\{\delta > 0 | (V, \pi) \text{ and } (W, \theta) \text{ are } \delta\text{-interleaved}\}$$

This is the notion of distance in the space of persistence modules. We now introduce a distance in the space of barcodes.

Notation. Given an interval $I = (a, b]$ we denote the interval obtained by expanding δ on both sides as $I^\delta = (a - \delta, b + \delta)$.

Let, $\mathcal{B}$ be a barcode. For $\epsilon > 0$, denote $\mathcal{B}_\epsilon$ as the set of all bars[3] from $\mathcal{B}$ of length greater than ϵ. Thus, by considering $\mathcal{B}_\epsilon$ we are neglecting short bars[4].

Definition 4.2.11 (Matching)**.** A matching between two multisets X and Y is a bijection $\mu : X' \to Y'$ where, $X' \subset X$ and $Y' \subset Y$.

Definition 4.2.12 (δ-matching)**.** A δ-matching between two barcodes $\mathcal{B}$ and $\mathcal{C}$ is a matching $\mu : \mathcal{B}' \to \mathcal{C}'$ such that

1. $\mathcal{B}_{2\delta} \subset \mathcal{B}'$
2. $\mathcal{C}_{2\delta} \subset \mathcal{C}'$
3. if $\mu(I) = J$, then $I \subset J^{-\delta}$ and $J \subset I^{-\delta}$

Definition 4.2.13 (Bottleneck distance)**.** The bottleneck distance between two barcodes $\mathcal{B}$ and $\mathcal{C}$) is defined to be

$$d_{bot}(\mathcal{B}, \mathcal{C}) = \inf\{\delta > 0 \,|\, \exists\, \delta\text{-matching between } \mathcal{B} \text{ and } \mathcal{C}\}$$

Theorem 4.2.2. [Stability Theorem] The map $V \mapsto \mathcal{B}(V)$ is an isometry v.i.z., for any two V, W persistence modules we have

$$d_{int}(V, W) = d_{bot}(\mathcal{B}(V), \mathcal{B}(W))$$

Any equality is two inequalities. The $\geq$ is apparently a big theorem called the algebraic stability theorem. More about is given in [12] and [1]. We shall prove the $\leq$ inequality here.

Theorem 4.2.3. Let V and W be persistence modules. If there exists a δ-matching between their barcodes then V and W are δ-interleaving. In particular,

$$d_{int}(V, W) \leq d_{bot}(\mathcal{B}(V), \mathcal{B}(W))$$

[3] bars are the intervals in the barcode

[4] neglecting the short bars corresponds to neglecting the holes which doesn't last long in the process of the filtration, we see how exactly, later

Proof. By structure theorem we have,

$$V = \bigoplus_{I \in \mathcal{B}(V)} \mathbb{F}(I),\, W = \bigoplus_{I \in \mathcal{B}(W)} \mathbb{F}(I)$$

For, $\mathcal{B}(V)' \subset \mathcal{B}(V)$ and $\mathcal{B}(W)' \subseteq \mathcal{B}(W)$, let $\mu : \mathcal{B}(V)' \to \mathcal{B}(W)'$ be a δ-matching. In order to construct a δ-interleaving between V and W we shall use the matched intervals and neglect the unmatched intervals which are relatively small.

Notation. Denote,

$$V_Y = \bigoplus_{I \in \mathcal{B}(V)'} \mathbb{F}(I),\, W_Y = \bigoplus_{J \in \mathcal{B}(W)'} \mathbb{F}(I)$$

$$V_N = \bigoplus_{I \in \mathcal{B}(V) \setminus \mathcal{B}(V)'} \mathbb{F}(I),\, W_Y = \bigoplus_{J \in \mathcal{B}(W) \setminus \mathcal{B}(W)'} \mathbb{F}(I)$$

Clearly,

$$V = V_Y \bigoplus V_N,\, W = W_Y \bigoplus W_N.$$

Moreover, for any matched pair I, J, we know that $I \subseteq J^{-\delta}$ and $J \subseteq I^{\delta}$. Then we have the following claim.

Claim 4.2.4. I, J are δ-matched then $\mathbb{F}(I)$ and $\mathbb{F}(J)$ are δ-interleaved.

Observe, for $\mathbb{F}[J][\delta]$

$$(\mathbb{F}(J)[\delta])_t = (\mathbb{F}(J))_{t+\delta} = \begin{cases} id, & \text{if } t+\delta \in J \\ 0, & \text{otherwise} \end{cases} = \mathbb{F}(J - \delta)$$

Since, $I \subseteq J^{-\delta}$, we have $I \subseteq J - \delta$. Thus, we have natural inclusion morphisms between $\mathbb{F}[I]$ and $\mathbb{F}(J-\delta)$. Similarly, since, $J \subseteq I^{-\delta}$, we have $J \subseteq I - \delta$ and thus we have morphisms between $\mathbb{F}(J)$ and $\mathbb{F}(I-\delta)$. So, we now have a pair of δ-interleaving morphisms

$$f_I : \mathbb{F}(I) \to \mathbb{F}(J)[\delta], \qquad g_J : \mathbb{F}(J) \to \mathbb{F}(I)[\delta]$$

These pair induce a pair of δ-interleaving morphisms

$$f_Y : V_Y \to W_Y[\delta], \qquad g_Y : W_Y \to V_Y[\delta]$$

Since, the intervals that are not matched by μ are of length less than 2δ, V_N and W_N are δ-interleaved with the trivial module. So, consider

$$f : V \to W,\, f|_{V_Y} \equiv f_Y,\, f|_{V_N} \equiv 0$$

and

$$g : W \to V,\, g|_{W_Y} \equiv g_Y,\, g|_{W_N} \equiv 0$$

these gives δ-interleaving morphisms between V and W. $\square$

This with the celebrated algebraic stability gives us the isometry between the space of persistence modules and the corresponding space of barcodes.

4.3 Persistence diagrams

To get the information about the persistence module, we have the unique barcode corresponding to it, which is a multiset consisting of intervals. To evaluate the closeness of the data, we have to calculate the bottleneck distance. For that we need to produce matchings between intervals which if you think is tedious to at least visualise. So, we need to come up with a visualisable diagram, a plot rather, which computer understands better, without loss of any information from the barcode. These diagrams are called **Persistence diagrams**.

Definition 4.3.1 (Persistence diagram)**.** Consider a persistence module V and its corresponding barcode $\mathcal{B}(V)$. For every interval $(p_i, q_i) \in \mathcal{B}(V)$, plot a point at $(p_i, q_i) \in \mathbb{R}^2$[5]. This plot of the intervals is called the **Presistence diagram**.

Remark. Consider a point (p_i, q_i) in the plot. Here, p_i represents the birth time/radius of a hole and q_i represents death time/radius of a hole. Initially the data is disconnected, in our study disconnected means that there exists a 0-dimensional hole. So, the points which are near to the diagonal $y = x$ tells us that the duration of the holes was very less. Usually there shall be more accumulated points near the diagonal as many points gets connected in short span of time. To know about the global structure of the data, we want the holes which persists long enough. So, unlike statisticians we look at the outliers in the persistence diagram.

This new plot gives the exact information as a barcode. Moreover, working over $\mathbb{R}^2$, is uncomplicated and computer understands this better. We now see the notion of bottleneck distance in the persistence diagrams.

4.3.1 Bottleneck distance in persistence diagrams

Consider two persistence diagrams $A = \{a_1, a_2, \cdots, a_m\}$ and $B = \{b_1, b_2, \cdots, b_m\}$ where, $a_i = (p_i, q_i)$ and $b_j = (p'_j, q'_j)$ for all i and j.

Definition 4.3.2 (Partial match)**.** A **partial match**[6] between A and B os a bijective map between the subsets of A and B

Definition 4.3.3 (Cost)**.** Consider a partial match $\mathcal{M}$ between A and B. Let, domain of $\mathcal{M}$ be A' and range of $\mathcal{M}$ be B'. The **cost** of $\mathcal{M}$ is defined as

$$c(\mathcal{M}) = \max\left\{\sup_{a_i \in A'} \{||a_i - \mathcal{M}(a_i)||_\infty\};\ \sup_{(p_i,q_i)\in A\setminus A'} \left\{\frac{1}{2}(q_i - p_i)\right\};\ \sup_{(p'_i,q'_i)\in B\setminus B'} \left\{\frac{1}{2}(q'_i - p'_i)\right\}\right\}$$

Remark. Here the norm $||\cdot||_\infty$ is the supremum norm. Moreover, for the points with are not matched, without loss of generality we match it to a corresponding perpendicular point on the diagonal $y = x$. Thus, we have the norm from the non matched point (p_i, q_i) to the corresponding point on perpendicular as $\frac{1}{2}(q_i - p_i)$.

Definition 4.3.4 (Bottleneck distance)**.** The **bottleneck distance** between two persistence diagrams A and B is defined as the infimum of the cost of all partial matchings between A and B.

$$d_{bot}(A, B) = \inf_{\mathcal{M} \text{ is a matching}} c(\mathcal{M})$$

We now shift to computer science and see some code regarding the usefulness of bottleneck distance. Codes for the following section is available over the colab notebook. We use `ripser` package which produces Vietoris-Rips filtration in the backend to find the homologies. The core codes are mainly taken from the `GitHub` repository [9] by Elizabeth Munch.

[5]we plot with the multiplicity as in, when we do a matching between these points, we consider a match from the points with multiplicity
[6]cf.4.2.11

4.3.2 Computing d_{bot} over `python`

Here, we first plot a point cloud of 200 randomly generated points in the shape of annulus.

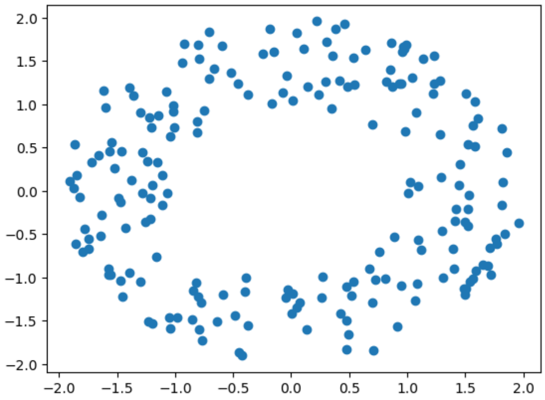

Figure 4.1: Point cloud in the shape of annulus

As mentioned in the remark 4.3, since annulus initially is disconnected, there exists a 0-dimensional hole and since, it has a 1-dimensional hole, we expect one point as outlier in the 1-dimensional persistence diagram. We now see what `ripser` has to say about this.

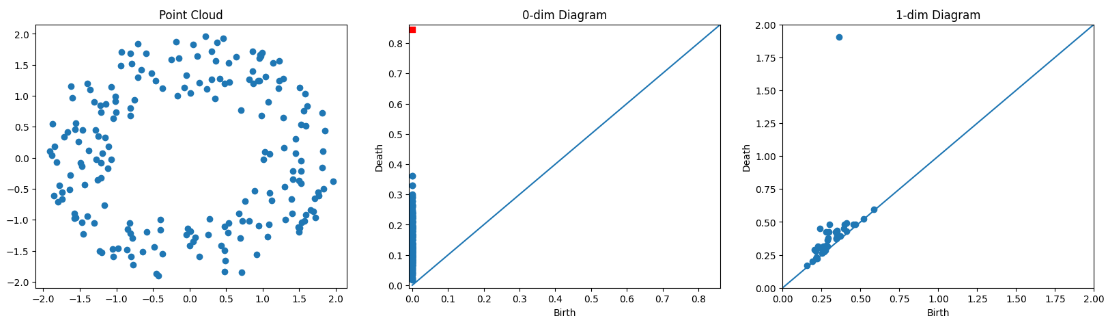

Figure 4.2: Generated annulus with the corresponding persistence diagrams

Observations

- As expected, in the 0-dimensional diagram, the increase sequence of points closer to the $x = 0$ represents that the connectivity is increasing quickly and finally after a point the whole plot is connected and remains connected all over; that is represented by the red dot in the plot.
- As expected, in the 1-dimensional diagram, there is a dense points along the line $y = x$ which tells that the one dimensional holes i.e, the triangles were formed in the filtration and filled immediately. But, since globally there is one hole for the annulus, we have a one outlier point in the plot. The point is around $(0.3, 1.9)$, which tells that the hole created at 0.3 radius and lasted till 1.9 which relatively persisted longer than the other holes. Thus, this is consistent with our expectation.

We now look at the persistence diagram of a point cloud of 200 randomly generated points in the shape of a double annulus. Since, there are two 1-dimensional holes for double annulus, we have two points as outliers in

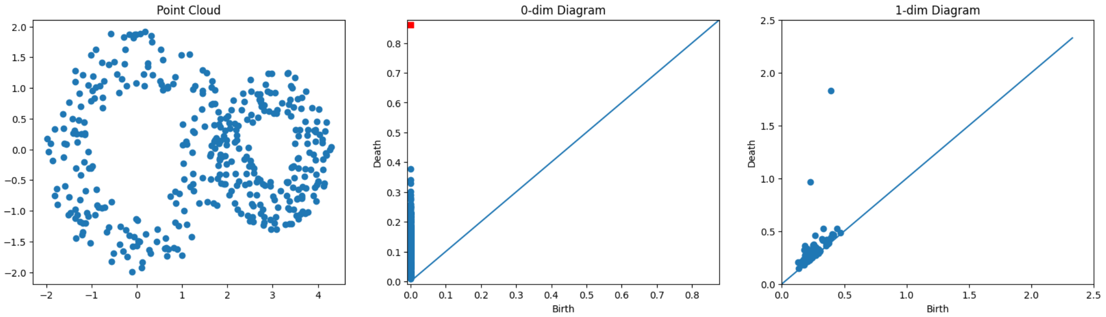

Figure 4.3: Generated double-annulus with the corresponding persistence diagrams

the 1-dimensional diagram.

We consider two single annulus and a double annulus. Bottleneck distance is a measure of how close one data is to other. It is a comparative measure. Thus, we expect that the bottleneck distance between the two annulus should be less than that of between an annulus and a double annulus. We generate point clouds of 800 randomly generated points in the shape of two single annuli and a double annulus.

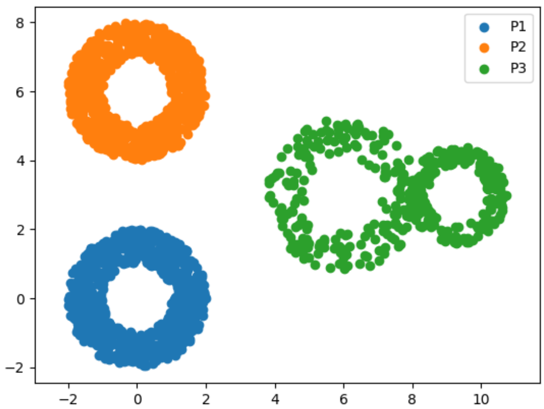

Figure 4.4: Generated two single annuli and a double annulus

We first ask computer to output the matching diagram and the bottleneck distance between the two single annuli. The figure 4.5 gives the matching diagram of two single annuli and the bottleneck distance generated is $\sim$ `0.036`[7]. Whereas, the observations from the figure 4.6 includes

1. the matching diagram of a single annulus and a double annulus, (thus we have two orange dots corresponding to double annulus and a single blue dot corresponding to single annulus)
2. the red color line segment represents the bottleneck distance which is $\sim$ `0.661`

This distance is $\sim$ 21 times larger than the distance between two single annulus. Thus this is consistent with our expectation.

[7] kindly see the colab notebook for the bottleneck distance output

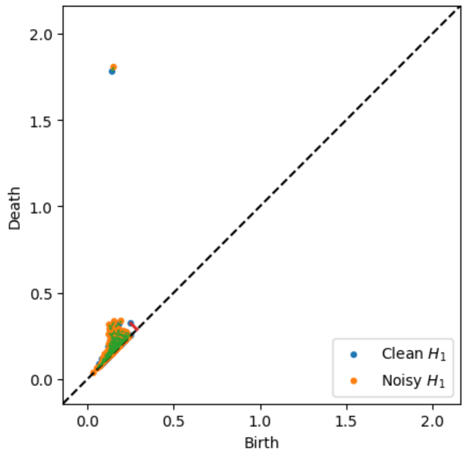

Figure 4.5: Matchings between two single annuli

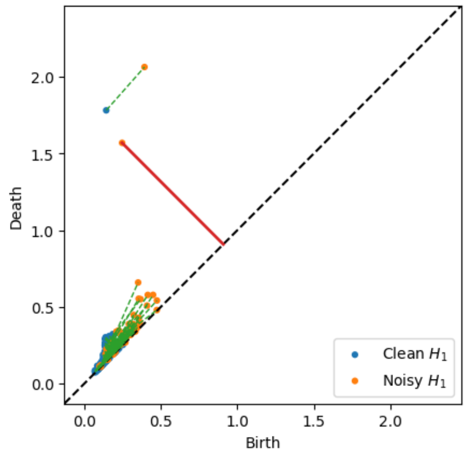

Figure 4.6: Matchings between annulus and double annulus

4.3.3 Comparing english alphabets with d_{bot}

Suppose we have a data in the shapes of a english letter say **R**. We know that the shape are like the alphabets, but the computer doesn't. To make the computer understand we need to compare the shape of data with the actual letters. Here, we use the bottleneck distance. Suppose we upload our handwritten letter **R** and handwritten **O** into the computer. We expect that the bottleneck distance between our uploaded letter **R** and the data in the shape of **R** should be less compared to that of our **O** and the data in the shape of **R**. This is the basic idea for this section. We see the results as follows.

To get a data loosely in the shape of **R**, we add some noise i.e., blur-ness to the letter **R** and plot the points in the shape of the blurred **R**.

Figure 4.7: Generated blurred image of **R**

We get the point cloud in the shape of this blurred **R** and calculate its persistence diagrams as in figure 4.8.

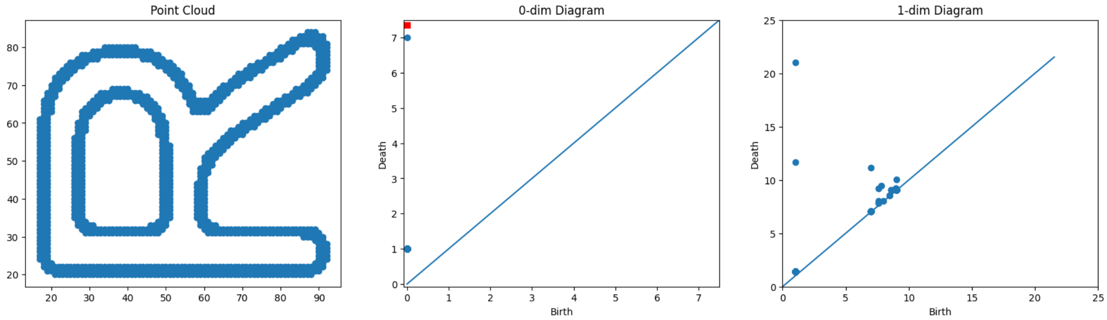

Figure 4.8: Generated point cloud and corresponding persistence diagrams

Now, we upload handwritten letters **R** and **O** to the computer[8].

Figure 4.9: Handwritten letter **R**

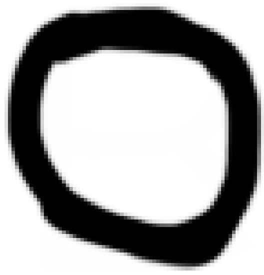

Figure 4.10: Handwritten letter **O**

With this uploaded letters, we get the data points in terms of them and compute their persistence diagrams. That is given as follows.

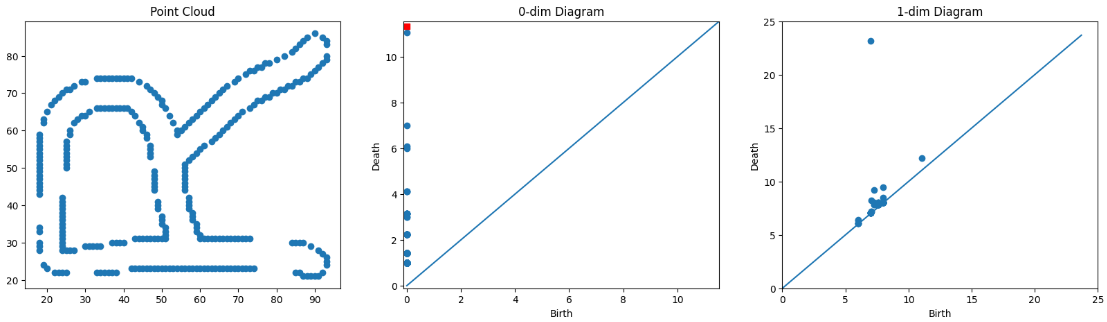

Figure 4.11: Generated points of written **R** and corresponding persistence diagram

[8]when you use the code, make sure you upload the handwritten images with file names `R.png` and `O.png`

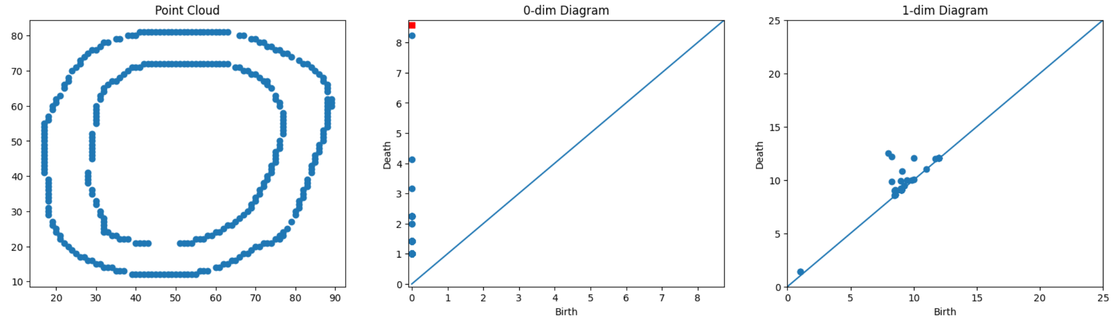

Figure 4.12: Generated points of written **O** and corresponding persistence diagram

We now compute the matchings and bottleneck distance between the generated data in shape of **R** and handwritten **R** as shown in figure 4.13. The bottleneck distance turns out to be ~ 6. From figure 4.14 we observe that we have matchings between the generated data in shape of **R** and handwritten **O** where the bottleneck distance turns out to be ~ 19.41 which is almost three times larger than that of with **R**. Thus, the generated **R** is more closer to the hand-written **R** than that of hand-written **O** as expected.

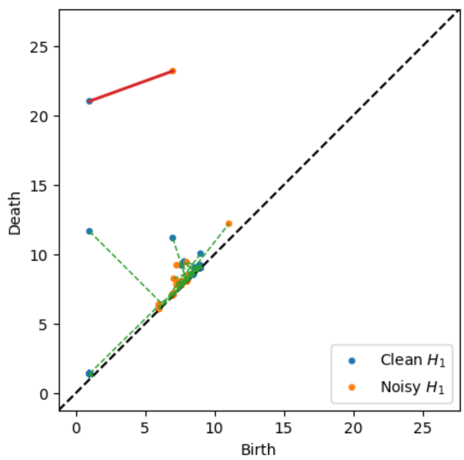

Figure 4.13: Matchings of generated **R** and hand-drawn **R**

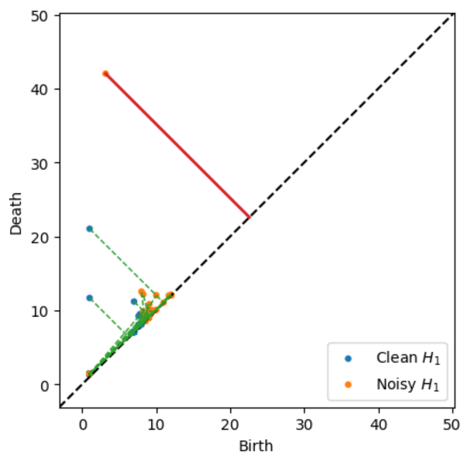

Figure 4.14: Matchings of generated **R** and hand-drawn **O**

Though this is just a first use of bottleneck distance, turns out to be a great tool in comparative study of the structure of data. We end our discussion here.

Bibliography

[1] Ulrich Bauer and Michael Lesnick. "Induced matchings and the algebraic stability of persistence barcodes". en. In: *Journal of Computational Geometry* (2015), Vol. 6 No. 2 (2015): Special issue of Selected Papers from SoCG 2014. DOI: 10.20382/JOCG.V6I2A9. URL: https://jocg.org/index.php/jocg/article/view/2983.

[2] J.D. Boissonnat, F. Chazal, and M. Yvinec. *Geometric and Topological Inference*. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2018. ISBN: 9781108410892.

[3] René Corbet and Michael Kerber. "The representation theorem of persistence revisited and generalized". In: *Journal of Applied and Computational Topology* 2.1–2 (July 2018), pp. 1–31. ISSN: 2367-1734. DOI: 10.1007/s41468-018-0015-3. URL: http://dx.doi.org/10.1007/s41468-018-0015-3.

[4] Herbert Edelsbrunner and John Harer. *Computational Topology - an Introduction.* American Mathematical Society, 2010, pp. I–XII, 1–241. ISBN: 978-0-8218-4925-5.

[5] A. Hatcher. *Algebraic Topology*. Algebraic Topology. Cambridge University Press, 2002. ISBN: 9780521795401.

[6] M. Jungerman and G. Ringel. "Minimal triangulations on orientable surfaces". In: *Acta Mathematica* 145.none (1980), pp. 121–154. DOI: 10.1007/BF02414187.

[7] Clara Loeh. *A comment on the structure of graded modules over graded principal ideal domains in the context of persistent homology*. 2023. arXiv: 2301.11756 [math.AC]. URL: https://arxiv.org/abs/2301.11756.

[8] William S. Massey. *Algebraic topology : an introduction*. Vol. 56. Graduate Texts in Mathematics. New York: Springer, 1967. ISBN: 3540902716 9783540902713 0387902716 9780387902715.

[9] Elizabeth Munch. *TDA-Python-Workshop-JMM21*. Jan. 2021. URL: https://github.com/lizliz/TDA-Python-Workshop-JMM21.

[10] J.R. Munkres. *Topology*. Featured Titles for Topology. Prentice Hall, Incorporated. ISBN: 9780131816299.

[11] V Nanda. "Lecture notes in Computational Algebraic Topology". https://people.maths.ox.ac.uk/nanda/cat/TDANotes.pdf. Mar. 2021.

[12] Leonid Polterovich et al. *Topological Persistence in Geometry and Analysis*. 2021. arXiv: 1904.04044 [math.AT]. URL: https://arxiv.org/abs/1904.04044.

[13] J. Rotman. *An Introduction to Algebraic Topology*. Graduate Texts in Mathematics. Springer New York, 1998. ISBN: 9780387966786.

[14] J. Tierny. *Topological Data Analysis for Scientific Visualization*. Mathematics and Visualization. Springer International Publishing. ISBN: 9783319715070.

[15] Afra Zomorodian and Gunnar Carlsson. "Computing Persistent Homology". In: *Discrete and Computational Geometry* 33 (Feb. 2005), pp. 249–274. DOI: 10.1007/s00454-004-1146-y.