UNIVERSITY OF GHANA, LEGON
DEPARTMENT OF MATHEMATICS

APPLYING PERSISTENT HOMOLOGY TO TOPOLOGICAL DATA ANALYSIS (TDA)

BY

PEARL SANDOH KUURIDONG
10738817

December, 2022

A project submitted to the Department of Mathematics, University of Ghana, in partial fulfillment of the requirements for a Bachelor of Science degree.

Declaration

I declare that this project is my original work. Except where due acknowledgment has been provided, this project contains no material previously published by any other individual. To the best of my knowledge, there is no content in this project that has been accepted as part of any other academic degree requirements at the University of Ghana or any other university.
This project was carried out in the Department of Mathematics, University of Ghana in partial fulfillment of the Bachelor of Science degree under the supervision of Dr. Ralph Agyei Twum.

Name of Student: Ms. Pearl Sandoh Kuuridong

Signature:______________________________

Date:______________________________

Name of Supervisor: Dr. Ralph Agyei Twum

Signature:______________________________

Date:______________________________

Abstract

Although orthodox methods for data analysis like predictive analysis have many advantages, they have a serious drawback; these methods do not work with data whose dimension cannot be visualized. Given data from a five-dimensional space, for instance, renders these techniques and methods inefficient.
A field of mathematics known as Topological Data Analysis (TDA), which is a mix of concepts from algebraic topology and computational geometry, allows for analyses to be done on higher dimensional data. The most basic idea is that data in any space has a shape, and within that underlying shape is relevant information to be extracted for use. One of the most common tools in TDA is *persistent homology*, a mathematical tool that measures the topological features of the data that remain while changing a particular variable of the shape of the data. This tool has been very useful in the success of TDA since it allows for topological features of higher-dimensional data to be represented in a lower-dimensional form.

This essay gives the reader an understanding of what persistent homology is and how it is applied in TDA.

Dedication

I dedicate this paper to my family; The Kuuridong Family and to my nieces and nephew; Afia Mwinmaalo Nyarko Ansah, Adwoa Mwinnuo Aggrey Ansah, Akosua Mwinso Konadu Ansah, and Phinley-Ethan Mwinsoma Sando Kuuridong.

Acknowledgments

I wish to thank God almighty first of all, for life for without it I would not have been able to do this work. I also want to thank Him for the ability to work on, and complete this project. In addition, I want to say a very big thank to my supervisor, Dr. Ralph Agyei Twum for his guidance and direction throughout this project.

I want to thank, in a very special way, my colleagues at the Department of Mathematics, University of Ghana most especially Desmond Kofi Boateng, for his support throughout this four-year journey, most especially during this final year.
I would also like to appreciate researchers in this field whose work informed me enough to put this essay together.

Finally, my utmost appreciation goes to my family and friends, most especially my parents, Mr. Augustine Sabatu Kuuridong and Mrs. Dora Kob Kuuridong, and my cousin, Maalvladedon Ganet Some, for their constant prayers, support and advice throughout my academic journey.

Contents

Chapter 1

Introduction

It is often said that we are in the digital age, and this claim is not without reason. The efforts by individuals and organizations to make life easier as well as connect people have turned the world into what we call a global village. What this means is that people are connected like never before, and can complete multiple tasks with the single click of a button. Data is made available to these organizations, through interactions with their products, to aid their decision-making. The advancements in technology in recent years mean that the size and complexity of this data have increased.

Extracting information from this big data has become challenging; traditional methods for data analysis like simple linear regression, are inefficient in the analysis of big data. Recent developments in mathematics are giving insight into these traditional methods, the result of which is new, efficient methods for data analysis that allow people to better decipher big data.

Topological Data Analysis (TDA) is a field of mathematics that lies at the intersection of algebraic topology, data analysis, computational geometry, computer science, and statistics (Otter et al. 2017[27]). Popularized by the likes of Gunnar Carlsson with his 2009 article, Topology and Data [3], TDA really started coming out as a field with the pioneering works of Edelsbrunner et al. in 2002 with the article Topological Persistence and Simplification [9]. TDA aims at developing tools to study qualitative features of data. The idea is to use concepts and results from topology and geometry to develop these tools.

Persistent homology (PH) remains the most popular tool developed by TDA, but it is not the only one. The Mapper algorithm, which combines clustering, and graph network techniques with dimension reduction and cellular sheaves (which are data structures that assign local data to cells), and linear transformations between cells of incident ascending dimension are a few of the other tools the theories in TDA have successfully developed. The fact that PH is based on algebraic topology, whose rich theory makes the study of the qualitative features of complex structures possible, and can be computed using algebra are among the reasons persistent homology is an especially popular tool.

As a fairly recent tool that has been useful in the analysis of big data, people are curious about the extensibility of PH. As a result, various research efforts have been undertaken to determine how PH can be extended. We discuss one of these research efforts in section 4.2; the statistical interpretation of topological summaries. The main question to be answered in this area is the definition of the average of a collection of persistence diagrams− a visual representation of the result of the calculation of persistent homology. Current research has been able to define a space of persistence diagrams on which a distance function is defined. This allows us to perform standard statistical calculations on this space of persistence diagrams. Other areas of research include multidimensional persistence, which is concerned with filtering data points along more than one axis, and the efficient creation of cell complexes. Cell complexes are a model of the underlying shape of data. The goal of this second area of research is to find alternate complexes of smaller sizes than the ones that are currently in use, using geometric approximation techniques.

This essay is divided into five chapters. Chapter 1, which is this chapter, introduces the reader to persistent homology and TDA in general. In chapters 2 and 3, we discuss concepts in topology and algebraic topology that are key to acquiring a basic understanding of persistent homology. We discuss concepts like topological spaces, continuous maps between topological spaces, and some topological properties in chapter 2. Chapter 3 discusses concepts like simplices and simplicial complexes, homotopy, and homology, which are central to the study of algebraic topology. The latter is the concept on which persistent homology is built. Chapter 4 is the heart of this essay. We discuss persistent homology in this chapter and talk about the pipeline for the computation of persistent homology. We also discuss an area of active research in the field; the statistical interpretation of topological summaries and analyze some data using the methods discussed in the chapter. We conclude this essay in chapter 5 with a recap of the main ideas discussed in the essay.

Chapter 2

Basic Concepts in Topology

2.1 Topology

General topology, also known as point-set topology, is the area of mathematics that deals with the study of shapes. Unlike in geometry where the shapes are seen as rigid in space, topologists study these shapes as though they were made of rubber and so could be continuously deformed. In the case where one object can be continuously deformed into another, we consider them as the same object possessing similar properties. In topology we are more concerned with the properties an object has rather than what it may look like. In this chapter we look at some of the most basic concepts underlying the study of Topology.

2.1.1 Definition

Let X be a non-empty set. We define a topology on the set X to be a collection τ of subsets of X such that the following are satisfied:

1. $X, \emptyset \in \tau$.
2. Arbitrary unions of elements in τ are in τ. That is, given $A_i \in \tau$, where i belongs to the index set I, $\bigcup\limits_{i \in I} A_i \in \tau$.
3. Finite intersections of elements of τ are in τ. That is, given $A_1, A_2, ...A_n \in \tau$, $\bigcap\limits_{i=1}^{n} A_i \in \tau$.

We call the pair (X, τ) a topological space. For convenience sake however, we represent the topological space (X, τ) by its underlying set X.

A subset U of X is said to be open if $U \in \tau$.

A subset V of X is said to be closed if $X \backslash V \in \tau$.

A subset of X can be both open and closed. In that case, it is called a clopen subset.

2.1.2 Examples

1. Let X be a non-empty set. Let $\mathcal{P}(X)$ denote the collection of all subsets of X. Then, $\mathcal{P}(X)$ is a topology on the set X, known as the *discrete topology* on X.

2. Let X be a non-empty set. The collection $\tau = \{X, \emptyset\}$ is a topology on X, known as the *indiscrete or trivial* topology.

3. Let $X = \mathbb{R}$. Define $\tau = \{S \subseteq \mathbb{R} \mid \forall x \in S, \exists a, b \in \mathbb{R}$ with $a < b$ such that $a < x < b$, $(a, b) \subset S\}$. τ is known as the *usual topology or the Euclidean topology* on $\mathbb{R}$.

4. Let X be a non-empty set. The collection $\tau = \{X, \emptyset, U \mid U \subset X, X\backslash U$ is finite$\}$ is a topology on X. It is known as the *finite-closed* or *cofinite topology* on X.

5. Let X be a non-empty set. The collection $\tau = \{X, \emptyset, U \mid U \subset X, X\backslash U$ is countable $\}$ is the *countable-closed* topology on X.

2.1.3 Basis and Subbasis

Let (X, τ) be a topological space. A collection β, of open subsets of X is known as a basis for X if $\forall\, U \in \tau$, $U = \bigcup\limits_{B \in \beta} B_i$.

Let (X, τ) be a topological space. A collection, δ, of open subsets of X is a *subbasis* for X if for $i \in \mathbb{N}$, $\bigcap\limits_{i=1}^{n} A_i$, where $A_i \in \delta\ \forall i$, is a basis element for the topology τ on X.

Given a topology and a basis for the topology, we say that the basis generates the topology.

A topology can have more than one basis defined on it; that is to say that a basis for a topology is not unique.

Every topology is a basis for itself.

2.2 Metric Spaces

Given a non-empty set X we define a *metric* on X as a map $d : X \times X \to \mathbb{R}^+$ such that the following conditions are satisfied:
$\forall x, y, z \in X$,

1. $d(x, y) \geq 0$

2. $d(x, x) = 0$

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

4. $d(x, y) \leq \ d(x, z) + d(z, y)$

The pair (X, d) is defined as a *metric space.* For convenience however, we represent the metric space (X, d) by its underlying set X.

2.2.1 Examples

1. The space $(\mathbb{R}^n \ , d)$, with $d : \mathbb{R}^n \to \mathbb{R}^+$ defined as $d(\mathbf{x},\ \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$ $\forall \mathbf{x},\ \mathbf{y} \in X$. with $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, $\mathbf{y} = (y_1, y_2, \cdots, y_n)$ is known as the *Euclidean metric space.*

2. Let X be the set of all sequences of complex numbers such that each sequence is bounded. We define a metric $d : X \times X \to \mathbb{R}^+$ by $d(\mathbf{x},\mathbf{y}) = \sup\limits_{i \in \mathbb{N}} |x_i - y_i| \quad \forall \mathbf{x},\ \mathbf{y} \in X$ with $\mathbf{x} = (x_1, x_2, \cdots)$ and $\mathbf{y} = (y_1, y_2, \cdots)$. The space (X, d) is known as the *l^∞ space.*

3. Let X be the set of all real-valued functions which are defined and continuous on a closed and bounded interval, A. Define a map $d : X \times X \to \mathbb{R}^+$ by $d(f, g) = \max\limits_{x \in A} |f(x) - g(x)| \quad \forall f,\ g \in X$. Then d is a metric on X and the space (X, d) is known as the *function space of bounded continuous functions on A.*

4. Let X be the set of all n-tuples whose entries are either 0 or 1. Define a map $d : X \times X \to \mathbb{R}^+$ by $d(x, y) = m$, where m is the number of places where $x_i \neq y_i$. The map d is a metric on the set X and it is known as the *Hamming distance* on X.

5. Let $p \in \mathbb{N}$, with $p \geq 1$. Let X be the set of sequences of numbers such that $\forall x \in X$ with $x = (x_1, x_2, \cdots)$, $\sum\limits_{i=1}^{\infty} |x_i|^p$ converges. Define a metric $d : X \times X \to \mathbb{R}^+$ by

$$d(\mathbf{x},\ \mathbf{y}) = (\sum_{i=1}^{\infty} |x_i - y_i|^p)^{\frac{1}{p}}$$

$\forall \mathbf{x},\mathbf{y} \in X$ with $\mathbf{x} = (x_1, x_2, \cdots)$ and $\mathbf{y} = (y_1, y_2, \cdots)$. This map d is a metric on X and the space (X, d) is known as the *l^p space.* The space l^2 is known as the *Hilbert Sequence Space.*

2.2.2 Metric-Induced Topology

An open ball centered at a point $a \in X$ with radius r denoted by $B_r(a)$ is the set defined as $B_r(a) = \{x \in X \mid d(x, a) < r\}$.

A closed ball centered at a point $a \in X$ with radius r denoted by $\bar{B}_r(a)$ is the set defined as $\overline{B}_r(a) = \{x \in X \mid d(x, a) \leq r\}$.

Let (X, d) be a metric space. Let A be a subset of X such that, for all $x \in A$, there exist $r \in \mathbb{R}^+$ with $B_r(x) \subset A$. Suppose that $\tau = \{\emptyset, X, A\}$. Then the collection τ is a topology on X. It is known as the *metric-induced topology* on X. It is induced by the metric d on X.

Every open ball is an open set and every closed ball is a closed set in the topology induced by the metric d.

Every metric space (X, d) is a topological space as the metric, d, induces a topology, τ_d on X. τ_d is generated by all the open balls in X.

2.3 Limit Points, Continuous Functions, and Connectedness

2.3.1 Limit Points, Closure, and Denseness

Let (X, τ) be a topological space and $A \subseteq X$. A point $x \in X$ is said to be a *limit point* of A if for each $U \in \tau$ such that $x \in U, \exists a \in U$ such that $a \in A$ and $a \neq x$. The collection of all the limit points of A denoted by A' is known as the limit point set of A. That is, for every neighbourhood of a point $x \in X$, no matter how small, always contains another point of A other than x.

Let A be a subset of a topological space, (X, τ). The *closure* of A, denoted by $\bar{A}$, is defined as the union of the set A together with all of its limit points. That is, $\bar{A} = A \cup A'$.

The closure of a subset A of topological space (X, τ) is a closed set. It is the smallest closed set containing A, that is $\bar{A}$ is the intersection of all the closed subsets of X containing A.

Every subset of a topological space is closed if and only if it contains all of its limit points.

A subset A of a topological space (X, τ) is said to be *dense* if its closure $\bar{A} = X$. That is, for every open subset $U \subseteq X$ there exists an element $a \in A$ such that $a \in U$.

2.3.2 Examples

1. Let (X, τ) be the indiscrete topological space. Then the limit point set of every subset $A \subseteq X$ is the whole space X. That is, $\bar{A} = X$, so that every subset of X is dense.

2. Let (X, τ) be the discrete topological space. The limit point set of every subset $A \subseteq X$ is the empty set. That is, $A' = \emptyset$. Hence $\bar{A} = A$. That is, every subset of X is closed.

3. Let $X = \mathbb{N}$ and let τ be the finite-closed topology on X. The set of limit points for every finite subset $A \subseteq X$ is the empty set. For any infinite subset $B \subseteq X$, $B' = \mathbb{N}$. Hence every open set in the finite-closed topology is dense.

4. Let X be a finite set and let τ be the finite-closed topology on X. Then the limit point set for every subset $A \subseteq X$ is X the empty set and so no subset of a finite set is dense under the finite-closed topology.

2.3.3 Continuous Functions and Homeomorphisms

Let (X, τ) and (Y, τ') be topological spaces. A map $f : X \to Y$ is *continuous* if and only if the pre-image of every open subset $V \subseteq Y$ is open in X. That is, $\forall\, V \in \tau', f^{-1}(V) \in \tau$.

Let X and Y be topological spaces. A homeomorphism between X and Y is defined as a map $f : X \to Y$ such that:

1. f is invertible

2. f is continuous

3. The inverse of f, f^{-1}, is continuous

X and Y are said to be homeomorphic, denoted by $X \cong Y$, if there exists a homeomorphism between them.

Although every homeomorphism is a continuous map, the converse of this statement is not necessarily true.

Homeomorphisms preserve topological properties between spaces.

2.3.4 Examples

1. Every constant function between topological spaces is continuous.

2. Let τ be the Euclidean topology on $\mathbb{R}$. Then every map $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = ax$, for $a \in \mathbb{R}$ is continuous.

3. The identity map from a topological space to itself is a homeomorphism.

4. Given a homeomorphism f between topological spaces X and Y then f^{-1} is also a homeomorphism.

5. Let X, Y, Z be topological spaces. Let $f : X \to Y$ and $g : Y \to Z$ be homeomorphisms. Then the map $g \circ f : X \to Z$ is also a homeomorphism.

Examples 3 to 5 imply that the relation *"homeomorphic to"* is an equivalence relation.

2.3.5 Connected Topological Spaces

Let X be a topological space. X is said to be *connected* if the only clopen subsets of X are X and the empty set, $\emptyset$.

If a topological space (X, τ) is not connected, then there exists a disconnection $A, B \subseteq X$. That is, there exists $A, B \in \tau$ such that $A \cup B = X$ and $A \cap B = \emptyset$.

A topological space (X, τ) is said to be *path-connected* if for every distinct pair of points, $a, b \in X$, there exists a continuous function $f : [0, 1] \to X$ such that $f(0) = a, f(1) = b$.

Every path-connected space is connected.

2.4 Compactness and Separation Axioms

2.4.1 Compactness

Let (X, τ) be a topological space. An arbitrary collection, $\mathcal{A}$ of subsets of X is said to be a *cover* for a subset $M \subseteq X$ if $M \subset \bigcup_{A_i \in \mathcal{A}} A_i$. If for each $A \in \mathcal{A}, A \in \tau$ then $\mathcal{A}$ is called an open covering for M.

Let X be a topological space. A subset $M \subseteq X$ is said to be compact if for every open cover of M, there exists a finite subcover of M. That is to say, if $\mathcal{A}$ is an open cover for M and there exists a finite non-empty subset of $\mathcal{A}$ which is also an open cover for M, then the set M is said to be compact.

2.4.2 Examples

1. Let (X, τ) be the indiscrete topological space. Then every subset of X is compact.

2. Let X be a topological space and M, a finite subset of X. Then M is a compact set.

3. Let X, Y be topological spaces, and let $f : X \to Y$ be a continuous map. The image, $f(M)$, of a compact subset $M \subseteq X$ is also a compact subset of Y.

4. Every closed interval $[a, b] \subseteq \mathbb{R}$ is compact.

2.4.3 Separation Axioms

Let (X, τ) be a topological space. X is said to be a T_0-space if for every pair of distinct points $x, y \in X, \exists U \in \tau$ such that $x \in U$ and $y \notin U$ or $y \in U$ and $x \notin U$.

A topological space (X, τ) is said to be T_1 if for every pair of distinct points $x, y \in X$, there exists $U, V \in \tau$ such that $x \in U$ but $y \notin U$ and $y \in V$ but $x \notin V$. A T_1 space is also known as a *Fréchet* space.

A topological space (X, τ) is said to be T_2 if for every pair of distinct points $x, y \in X$, there exists $U, V \in \tau$ such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$. A T_2 space is also known as a *Hausdorff* space.

Chapter 3

Basic Concepts in Algebraic Topology

One of the most basic problems of topology is determining whether or not any two given topological spaces are topologically equivalent; whether they are homeomorphic. It is fairly simpler to show that two given spaces are homeomorphic than to show that they are not. To show that any two given spaces are homeomorphic is just a matter of finding a homeomorphism between them. However, to show that some spaces are not homeomorphic, we would have to show that no homeomorphism exists between them, a task which can be tiresome. However, since topological properties remain invariant under homeomorphisms we can instead argue that there is a topological property that one space has that the other does not have, to prove that they are indeed not homeomorphic.
Algebraic topology is a field of mathematics that seeks to construct certain topological invariants to help answer the question of whether or not any two given topological spaces are homeomorphic to each other.

3.1 Simplices and Simplicial Complexes

3.1.1 Simplices

Consider the Euclidean space $\mathbb{R}^n$. A set of vectors $\{a_0, a_1, ..., a_n\} \subset \mathbb{R}^n$ is said to be *geometrically independent* if for every choice of scalars t_i, the equations $\sum_{i=0}^{n} t_i = 0$ and $\sum_{i=0}^{n} t_i a_i = \mathbf{0}$ imply that that each $t_i = 0$.

Consider the Euclidean space $\mathbb{R}^{n+1}$ and a geometrically independent subset of vectors $\{a_0, a_1, ..., a_n\}$ of $\mathbb{R}^{n+1}$. An *n-simplex*, σ is the set of points spanned by $\{a_0, a_1, ..., a_n\}$ such that $\sum_{i=0}^{n} t_i = 1$ for some choice of scalars t_i. That is, for every point $x \in \sigma, x = \sum_{i=0}^{n} t_i a_i$ and $\sum_{i=0}^{n} t_i = 1$.
The vectors $\{a_0, a_1, ..., a_n\}$ are known as the *vertices* of σ.

Assuming the set $\{a_0, a_1, ..., a_n\}$ is the canonical basis for the space $\mathbb{R}^{n+1}$, then the n-simplex formed is known as the *standard* n-simplex and is denoted by Δ^n.

Let σ be a simplex. We call σ an *oriented* n-simplex if it is an n-simplex and the order of the vertices is fixed.

3.1.2 Examples

1. A point in space is a simplex. It is spanned by a single point in $\mathbb{R}$ and is known as the *0-simplex.* Δ^0 is spanned by $1 \in \mathbb{R}$.

2. A line segment is a simplex. It is spanned by two distinct points in $\mathbb{R}^2$ and is known as the *1-simplex.* Δ^1 is spanned by $(1,0),(0,1) \in \mathbb{R}^2$.

3. A triangle is a simplex, known as the *2-simplex.* It is spanned by 3 geometrically independent points in $\mathbb{R}^3$. In $\mathbb{R}^3$, Δ^2 is spanned by $(0,0,1),(1,0,0),(0,1,0)$. It is known as the *equilateral triangle.*

4. A tetrahedron is a simplex, known as the *3-simplex.* It is spanned by 4 geometrically independent points in $\mathbb{R}^4$. Δ^3 is known as the *regular tetrahedron* and the vertices are $(1,0,0,0),(0,1,0,0),(0,0,1,0),(0,0,0,1)$.

3.1.3 Simplicial Complexes

Let σ be a simplex. A *face* δ of σ is a subset $\delta \subseteq \sigma$ such that δ is also a simplex whose vertices are vertices of σ. An *edge* γ of σ is a line segment that joins any two vertices in σ.
A simplicial complex is a collection K of simplices such that:

1. If $\sigma \in K$, then $\delta \in K$ where δ is a face of σ.

2. If σ_1, $\sigma_2 \in K$ then $\sigma_1 \cap \sigma_2 \in K$.

3.1.4 Examples

Let (X,d) be a finite metric space and $\varepsilon > 0$. Let $r = \frac{\varepsilon}{2}$. For each nonempty subset $A \subseteq X$ form an open ball $B_r(a)$ centered at a for each $a \in A$.

1. We form a simplicial complex K in the following way:

   - For each point $x_0 \in X$, let x_0 be a vertex for at least one simplex $\sigma \in K$.

   - If $\bigcap\limits_{a \in A} B_r(a) \neq \emptyset$ then $A \in K$.

   The simplicial complex K is known as the *Čech Complex.*

2. We form a simplicial complex H in the following way:

   - For each point $x_0 \in X$, let x_0 be a vertex for at least one simplex $\sigma \in H$.

- For each simplex $\sigma \subset X, \sigma \in H$ if $d(x, y) \leq \varepsilon$, where x, y are vertices of σ.

The simplicial complex H is known as the *Vietoris-Rips Complex.*

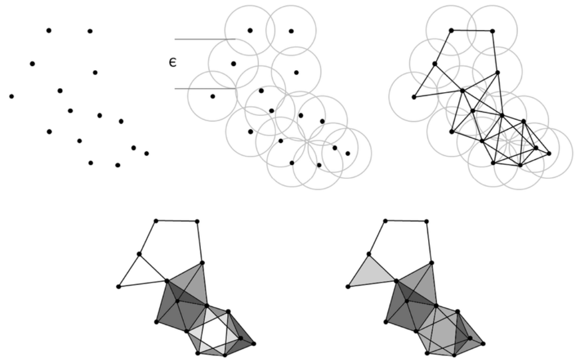


Figure 3.1: Formation of Rips Complex and Čech Complex.
source: Snášel et al. [28]

3.2 Homotopy and the Fundamental Group

The Fundamental group of a space X, denoted by $\pi_1(X)$, is a topological invariant that is constructed from continuous maps in the space. Maps in X that can be continuously deformed into each other are said to be homotopic and are considered to be the same in the fundamental group of X. The elements of the fundamental group are equivalence classes known as path-homotopy classes. The fundamental group aids us, to some extent, in finding out whether two spaces are homeomorphic or not. If any two spaces are topologically equivalent then they must have isomorphic fundamental groups. The converse of this statement is not true in general.

3.2.1 Homotopy

Let X and Y be topological spaces and let $f, g : X \to Y$ be continuous maps. f is said to be *homotopic* to g if there exists a continuous map $F : X \times [0, 1] \to Y$ such that

1. $F(x, 0) = f(x)$

2. $F(x, 1) = g(x)$

In that case the map F is said to be a homotopy between f and g and we write $f \simeq g$ to mean f is homotopic to g.

If f is homotopic to a constant map, we say f is *nulhomotopic.*

Let X be a topological space. A *path* f in X is a continuous map $f : [0, 1] \to X$ where $f(0)$ is known as the *starting point* or *initial point* of the path and $f(1)$ is called the *endpoint* of the path.

Let X be a topological space and let f, g be paths in X. Suppose $f(0) = g(0)$,

$f(1) = g(1)$. f is said to be *path-homotopic* to g if there exists a homotopy $F : [0,1] \times [0,1] \to X$ between f and g such that

1. $F(0,t) = f(0)$

2. $F(1,t) = f(1)$

In that case F is a path-homotopy between f and g and we write $f \simeq_p g$ to mean f is path-homotopic to g. If $f \simeq_p g$ then $f \simeq g$.

3.2.2 Examples

1. Let X be a topological space and f be a path in X. Then there exists a path-homotopy $F : [0,1]\times[0,1] \to X$ between f and itself defined by $F(x,t) = f(x)$.

2. Let X be a topological space and let f, g be paths in X. Suppose that F is a path-homotopy between f and g. The map $G : [0,1] \times [0,1] \to X$ defined by $G(x,t) = F(x, 1-t)$ also a path-homotopy between g and f.

3. Let X be a topological space and let f, g, h be paths in X. Suppose that $F : [0,1] \times [0,1] \to X$ is a path-homotopy between f, g and $G : [0,1] \times [0,1] \to X$ is a path homotopy between g, h. Then the map $H : [0,1] \times [0,1] \to X$ defined by

$$H(x,t) = \begin{cases} F(x,\ 2t), & t \in [0, \frac{1}{2}] \\ G(x,\ 2t-1), & t \in [\frac{1}{2}, 1] \end{cases}$$

   is a path-homotopy between f and h.

4. Consider the topological space $\mathbb{R}^2$ and let f, g be paths in $\mathbb{R}^2$ with the same initial point and endpoint. Then the map $F : [0,1] \times [0,1] \to \mathbb{R}^2$ defined by $F(x,t) = (1-t)f(x) + tg(x)$ is a path-homotopy between f and g. It is known as the *Linear Homotopy* or the *Straight-Line Homotopy*.

Note that examples 1 to 3 mean that *"path-homotopy"*, and by extension *"homotopy"*, is an equivalence relation.

3.2.3 The Fundamental Group

Let X be a topological space. A path f in X is known as a *loop* if $f(0) = f(1)$.

Let X be a topological space and f a loop in X. The path $\bar{f} : [0,1] \to X$ defined by $\bar{f}(t) = f(1-t)$ is also a loop in X with $\bar{f}(0) = f(1)$ and $\bar{f}(1) = f(0)$. It is known as the *reverse* of f.

Let X be a topological space. Let f, g be paths in X such that $f(1) = g(0)$. We define a product of paths $*$, by

$$(f * g)(s) = \begin{cases} f(2s), & s \in [0, \frac{1}{2}] \\ g(2s-1), & s \in [\frac{1}{2}, 1] \end{cases}$$

Note that the above operation is not defined for all paths. To be more specific the the product of paths is only defined for paths f, g such that $f(1) = g(0)$.

Let X be a topological space and $x_0 \in X$ be a point. Let $\mathcal{A}$ denote the set of path-homotopy classes of loops at $x_0 \in X$. For any path-homotopy class $[f] \in \mathcal{A}$, $[f](s) = f(s)$ and $[f] * [g] = [f * g]$ for all loops f, g in X. The pair $(\mathcal{A}, *)$ forms a group, where $*$ is the product of paths defined above.

We prove this claim.

Let $[f], [g] \in \mathcal{A}$.

$$([f] * [g])(s) = (f * g)(s) = \begin{cases} f(2s), & s \in [0, \frac{1}{2}] \\ g(2s-1), & s \in [\frac{1}{2}, 1] \end{cases}$$

$$(f * g)(0) = f(2 \times 0) = f(0) = f(1) = x_0 = g(1) = g(0) = g((2 \times 1) - 1) = (f * g)(1)$$

Hence $\mathcal{A}$ is closed under the product of paths.

To prove the associativity of $*$ is not straight forward so we define the following map that helps us prove this property:

Let X be a topological space. A *reparametrization* of a path f in X is the map $f \circ \rho : [0, 1] \to X$, where $\rho : [0, 1] \to [0, 1]$ is any continuous map such that $\rho(0) = 0, \rho(1) = 1$.

$f \circ \rho \simeq f$ by the straight-line homotopy.

We now prove the associativity of $*$:

Let $[f], [g], [h] \in \mathcal{A}$.

$$\begin{aligned}
([f] * ([g] * [h]))(s) &= (f * (g * h))(s) \\
&= \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ (g * h)(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ g(4s-2) & s \in [\frac{1}{2}, \frac{3}{4}] \\ h(4s-3) & s \in [\frac{3}{4}, 1] \end{cases}
\end{aligned}$$

$$\begin{aligned}
((f * g) * h)(s) &= \begin{cases} (f * g)(2s) & s \in [0, \frac{1}{2}] \\ h(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} f(4s) & s \in [0, \frac{1}{4}] \\ g(4s-1) & s \in [\frac{1}{4}, \frac{1}{2}] \\ h(2s-1) & s \in [\frac{1}{2}, 1] \end{cases}
\end{aligned}$$

Consider the function $\rho : [0, 1] \to [0, 1]$ defined by

$$\rho(s) = \begin{cases} \frac{1}{2}s & s \in [0, \frac{1}{2}] \\ s - \frac{1}{4} & s \in [\frac{1}{2}, \frac{3}{4}] \\ 2s - 1 & s \in [\frac{3}{4}, 1] \end{cases}$$

$$\begin{aligned}
(((f * g) * h) \circ \rho)(s) &= ((f * g) * h)(\rho(s)) \\
&= \begin{cases} (f * g)(2\rho(s)) & \rho(s) \in [0, \frac{1}{2}] \\ h(2\rho(s) - 1) & \rho(s) \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} f(4\rho(s)) & \rho(s) \in [0, \frac{1}{4}] \\ g(4\rho(s) - 1) & \rho(s) \in [\frac{1}{4}, \frac{1}{2}] \\ h(2\rho(s) - 1) & \rho(s) \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} f(4 \cdot \frac{1}{2}(s)) & s \in [0, \frac{1}{2}] \\ g(4(s - \frac{1}{4}) - 1) & s \in [\frac{1}{2}, \frac{3}{4}] \\ h(2(2s-1) - 1) & s \in [\frac{3}{4}, 1] \end{cases} \\
&= \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ g(4s-2) & s \in [\frac{1}{2}, \frac{3}{4}] \\ h(4s-3) & s \in [\frac{3}{4}, 1] \end{cases} = f * (g * h)(s)
\end{aligned}$$

That is, $f * (g * h)$ is a reparametrization of $(f * g) * h$ hence $f * (g * h) \simeq (f * g) * h$. Therefore $*$ is associative.

Let $[f] \in \mathcal{A}$ and consider the path $c : [0,1] \to X$ defined by $c(t) = x_0$ for all $t \in [0,1]$ then $c \in \mathcal{A}$ since $c(0) = c(1) = x_0$ and

$$\begin{aligned}
([f] * c)(s) &= (f * c)(s) \\
&= \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ c(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ x_0 & s \in [\frac{1}{2}, 1] \end{cases} \\
&= f(s) = [f](s).
\end{aligned}$$

Notice also that

$$\begin{aligned}
(c * [f])(s) &= (c * f)(s) \\
&= \begin{cases} c(2s) & s \in [0, \frac{1}{2}] \\ f(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} x_0 & s \in [0, \frac{1}{2}] \\ f(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= f(s) = [f](s)
\end{aligned}$$

that is, there exists a path $c \in \mathcal{A}$ defined by $c(t) = x_0$ for all $t \in [0,1]$, which acts as the identity element in $\mathcal{A}$.

For paths that are not loops, this "identity" element is not unique. That is to say that, the left and right identity elements do not coincide.

Consider the following example:

Let α be a path from x_0 to x_1 in X. Let e_{x_0} be the path defined by $e_{x_0}(s) = x_0$ for all $s \in [0,1]$.

$$\begin{aligned}
(e_{x_0} * \alpha)(s) &= \begin{cases} e_{x_0}(2s) & s \in [0, \frac{1}{2}] \\ \alpha(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} x_0 & s \in [0, \frac{1}{2}] \\ \alpha(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \alpha(s).
\end{aligned}$$

Hence we conclude that e_{x_0} is the left identity element.

Let e_{x_1} be the path defined by $e_{x_1}(s) = x_1$ for all $s \in [0,1]$.

$$
\begin{aligned}
(\alpha * e_{x_1})(s) &= \begin{cases} \alpha(2s) & s \in [0, \frac{1}{2}] \\ e_{x_1}(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} \alpha(2s) & s \in [0, \frac{1}{2}] \\ x_1 & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \alpha(s)
\end{aligned}
$$

Therefore we conclude that e_{x_1} is the right identity element.

Now since $e_{x_1} \neq e_{x_0}$ then the identity element is not unique for paths in general. Let $[f] \in \mathcal{A}$. Consider the following:

$$
\begin{aligned}
([f] * [\bar{f}])(s) &= (f * \bar{f})(s) \\
&= \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ \bar{f}(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} f(2s) & s \in [0, \frac{1}{2}] \\ f(-2s) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= x_0.
\end{aligned}
$$

Notice also that

$$
\begin{aligned}
([\bar{f}] * [f])(s) &= (\bar{f} * f)(s) \\
&= \begin{cases} \bar{f}(2s) & s \in [0, \frac{1}{2}] \\ f(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= \begin{cases} f(1-2s) & s \in [0, \frac{1}{2}] \\ f(2s-1) & s \in [\frac{1}{2}, 1] \end{cases} \\
&= x_0.
\end{aligned}
$$

So we conclude that for every $[f] \in \mathcal{A}$, $[\bar{f}]$ which is the reverse of $[f]$, is the inverse of $[f]$ under the operation $*$.

The pair $(\mathcal{A}, *)$ is known as *The First Homotopy Group* or *The Fundamental Group* of X with respect to the base point x_0 and it is denoted by $\pi_1(X, x_0)$.

3.2.4 The Fundamental Group of $\mathbb{R}$ and S^1

In this section we calculate the fundamental group of some spaces; specifically $\mathbb{R}$ the real line, and S^1, the unit circle. Before we calculate the fundamental groups of $\mathbb{R}$ and S^1, let us consider the following:

Let X be a path-connected topological space. The fundamental group of X is independent of the base point chosen. That is, for any distinct pair of points $a, b \in X$, $\pi_1(X, a) \cong \pi_1(X, b)$.

We prove this claim:
Let X be a path-connected space and a, b be distinct points in X. Define a map $\rho : \pi_1(X, a) \to \pi_1(X, b)$ by $\rho([f]) = [\bar{g} * [f] * g]$, where $[f]$ is a path-homotopy class of loops at a in X and g is a fixed path from b to a.

Let $[f], [h]$ be path-homotopy classes of loops at a in X.
$\rho([f * h]) = [\bar{g} * (f * h) * g] = [\bar{g} * (f * g * \bar{g} * h) * g] = [(\bar{g} * f * g) * (\bar{g} * h * g)] = [\bar{g} * f * g] * [\bar{g} * h * g] = \rho([f]) * \rho([h])$.
Also, ρ is invertible. The map $\rho^{-1} : \pi_1(X, b) \to \pi_1(X, a)$ defined by $\rho^{-1}([k]) = [g * k * \bar{g}]$ is the inverse of ρ. Since $(\rho \circ \rho^{-1})([k]) = \rho(g * k * \bar{g}) = [\bar{g} * (g * k * \bar{g}) * g] = [k]$ and $(\rho^{-1} \circ \rho)([f]) = \rho^{-1}([\bar{g} * f * g]) = [g * (\bar{g} * f * g) * \bar{g}] = [f]$. That is, $\rho \circ \rho^{-1}$ acts as the identity element in $\pi_1(X, b)$ and $\rho^{-1} \circ \rho$ acts as the identity element in $\pi_1(X, a)$ so ρ is invertible and hence bijective.

Consider the real line $\mathbb{R}$. Since $\mathbb{R}$ is path-connected, then $\pi_1(\mathbb{R}, x_0)$ is independent of the basepoint x_0. Consider $0 \in \mathbb{R}$ and let $[f]$ be a path-homotopy class of loops at 0. The map $F : \mathbb{R} \times [0, 1] \to \mathbb{R}$ defined by $F(s, t) = (1 - t)f(s)$ is a homotopy between $[f]$ and the constant map $c : [0, 1] \to X$, defined by $c(x) = 0$ for $x \in [0, 1]$. Since $F(s, 0) = f(s)$ and $F(s, 1) = 0$. Hence $\pi_1(\mathbb{R}, x_0) = \{c\}$. That is, the fundamental group of $\mathbb{R}$ is trivial.

Before we calculate the fundamental group of the unit circle, let us first look at the following concepts that would aid us in calculating $\pi_1(S^1)$:

Let X and $\tilde{X}$ be topological spaces and $p : \tilde{X} \to X$ be a map. The pair $(\tilde{X}, p)$ is called a *covering space* for X if the following conditions are satisfied:

1. For each $x \in X$, there exists an open neighborhood $U \subset X$ of x such that $p^{-1}(U) = \bigcup_{i \in I} A_i$ where the A_i's are disjoint, open sets and I is some indexing set.

2. For all i, the map $p : A_i \to U \subset X$ is a homeomorphism.

The set U is said to be *evenly covered*.

Consider the following sequence of maps between topological spaces:

$$\begin{tikzcd} & Z \arrow[d, "\beta"] \\ X \arrow[ur, dashed, "\gamma"] \arrow[r, "\alpha"'] & Y \end{tikzcd}$$

A *lift* is a map $\gamma : X \to Z$ such that $\beta \circ \gamma = \alpha$.

Given a topological space X and a covering space $(\tilde{X}, p)$, the following facts about covering spaces are useful in calculating $\pi_1(S^1)$:

1. For every path f in X with initial point x_0 and every point $\tilde{x}_0 \in p^{-1}(x_0)$ there is a unique lift $\tilde{f} : I \to \tilde{X}$ with initial point $\tilde{x}_0$.

2. For each homotopy of paths $F : [0,1] \times [0,1] \to X$ with $F(0,0) = x_0$ and each $\tilde{x}_0 \in p^{-1}(x_0)$ there is a unique lifted homotopy of paths $\tilde{F} : [0,1] \times [0,1] \to \tilde{X}$ with $F(0,0) = \tilde{x}_0$.

The intuition behind the fundamental group of the unit circle is that every loop can be wound any number of times starting and ending at a particular point to form the unit circle. Also, since S^1 is path-connected then the first homotopy group of the unit circle is independent of the basepoint chosen. Hence, we claim that the fundamental group of the unit circle $\pi_1(S^1)$ is isomorphic to $\mathbb{Z}$. That is, $\pi_1(S^1)$ is an infinite cyclic group.

We prove this claim:

Let $(\mathbb{R}, p)$ be a covering space for S^1, where $p : \mathbb{R} \to S^1$ is the map defined by $p(s) = (\cos 2\pi s, \sin 2\pi s)$. Let ω_n be a sequence of loops at $(1,0)$ in S^1 defined by $\omega_n = (\cos 2\pi ns, \sin 2\pi ns)$, for $n \in \mathbb{Z}$ and let f be an arbitrary loop at $(1,0)$ in S^1. We are trying to show that every loop in S^1 based at $(1,0)$ is homotopic to ω_n for each $n \in \mathbb{Z}$.

From the facts about covering spaces, there exists a lift of f given by $\tilde{f} : [0,1] \to \mathbb{R}$ with $\tilde{f}(0) = 0$ and $\tilde{f}(1) = n \in \mathbb{Z}$. Also, there exists a lift of ω_n given by $\tilde{\omega}_n$ with $\tilde{\omega}_n(0) = 0$, $\tilde{\omega}_n(1) = n \in \mathbb{Z} \subset \mathbb{R}$. Then $\tilde{f} \simeq \tilde{\omega}_n$ by the linear homotopy f_t and so $f \simeq \omega_n$ with homotopy $F : [0,1] \times [0,1] \to S^1$ defined by $F(s,t) = (f_t \circ p)(s)$. Hence $[f] \simeq [\omega_n]$.

We also claim that n is uniquely determined by the homotopy class of each loop in S^1. That is, if $[f] \simeq [\omega_n]$ and $[f] \simeq [\omega_m]$ then $m = n$.
Let $[f]$ be a homotopy class of loops in S^1 based at $(1,0)$. suppose that $[f] \simeq [\omega_n]$ and $[f] \simeq [\omega_m]$ so that $[\omega_m] \simeq [\omega_n]$. Let F be the homotopy between $[\omega_m]$ and $[\omega_n]$ with $F(0,t) = \omega_m$ and $F(1,t) = \omega_n$. Again, by the facts about covering spaces, there exists a lift of F given by $\tilde{F}$ such that $\tilde{F}(0,t) = 0$ and $\tilde{F}(0,t) = \tilde{\omega}_m$ and $\tilde{F}(1,t) = \tilde{\omega}_m$. Since $\tilde{F}$ is a homotopy of paths then $\tilde{F}(1,t)$ is independent of t hence $m = n$ since $\tilde{F}(1,0) = m$ and $\tilde{F}(1,1) = n$.

3.3 Homology and Homology Groups

While the fundamental group of a topological space X helps us in answering the question of whether or not X is homeomorphic to some other space Y, it does not carry us very far when working with spaces of higher dimension. Although higher homotopy groups exist, the calculation of those are difficult. Loosely speaking, homology groups, denoted by $H_n(X)$ where $n \in \mathbb{N}$, are considered the abelian analog to homotopy groups. They give information about the number of n-dimensional holes in X.

3.3.1 Simplicial Homology

Let X be an abelian group. X is said to be a *free abelian group* if X has a basis, when considered as a module over $\mathbb{Z}$.

Let K be a simplicial complex and let $C_i(K)$ be the free abelian group generated by all the i simplices of K. The sequence

$$\cdots \longrightarrow C_{i+1}(K) \xrightarrow{f_{i+1}} C_i(K) \xrightarrow{f_i} C_{i-1}(K) \longrightarrow \cdots$$

is called a *chain complex* if $f_i \circ f_{i+1} = 0$ for all i where the f_i's are group homomorphisms. The homomorphism f_i for all i, is called a *boundary map*.

Denote by $ker(f_i)$ the kernel of f_i and let $Im(f_i)$ denote the image of f_i. Since for any $i \in I$, $ker(f_i)$ is a subgroup of X_i and $Im(f_{i+1})$ is also a subgroup of X_i, then the group $ker(f_i)/Im(f_{i+1})$ is defined. This group is known as the *ith Homology Group* of K.

The group $Im(f_i)$ is called the *i-th boundary* and the group $ker(f_i)$ is called the *i-th cycle.*

The dimension of $H_i(K)$ is known as the *ith Betti number* of K. It is denoted by $\beta_i(K)$. $\beta_i(K)$ measures the number of i-dimensional holes in a simplicial complex however, Betti numbers and homology groups are not restricted to simplicial complexes alone.

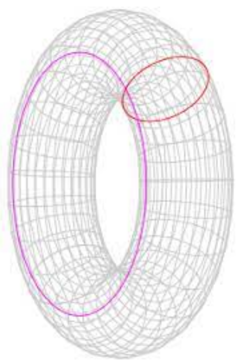


Figure 3.2: The number of 1-dimensional holes in a torus.
source: Wikipedia [18]

The zero dimensional Betti number tells us the number of connected components in a space. So for a sphere in a 3 dimensional space, $\beta_0 = 1$. The one dimensional Betti number tells us about the number of one-dimensional holes in the space. The one dimensional Betti number represents the "tunnels" in the space. A hollow cylinder with both ends open is what "tunnels" represents. The two dimensional Betti number tells us about the 3 dimensional holes (voids) in the space. The table below summarizes the ith Betti number for some spaces, where $i \in \{0, 1, 2, 3\}$.

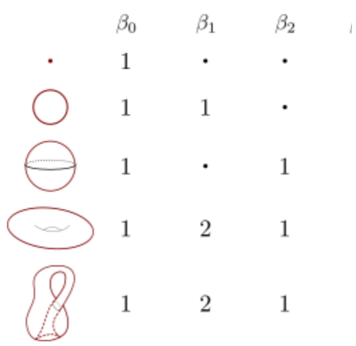


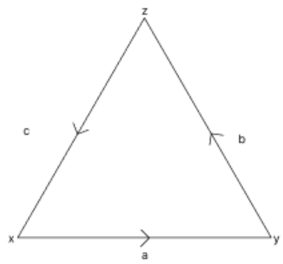
Figure 3.3: source: Elizabeth Munch[23]

3.3.2 Differences in Homology for Different Simplicial Complexes

This section seeks to draw the reader's attention to the differences in homology groups that may arise as a result of the choice of simplicial complex while introducing the reader to some homology computations. We do this by first calculating the homology groups of the unit circle S^1 and the disc. We then consider some points in the Euclidean space $\mathbb{R}^2$ and then draw the Čech complex and Rips complex with these points as our vertex set. For the Čech complex, the resulting simplicial complex is the boundary of the 2-simplex, which is the set of 1-simplices and 0-simplices in the 2-simplex. And for the Rips complex the resulting simplicial complex is the 2-simplex. These two simplicial complexes are the triangulations of the circle and the disc.
We use triangle throughout the remainder of this section to refer to the boundary of a triangle and 2-simplex to denote the boundary of a triangle together with its face.

Let us first consider the circle. We notice that the circle is homeomorphic to the triangle. So the homology groups of the triangle will be the same as the circle.

Consider the following oriented triangle with vertices $\{x, y, z\}$ and edges $\{a, b, c\}$.



Let us call this oriented triangle X. Notice that X is a topological space with the discrete topology $\tau_x = \{x, y, z, a, b, c\}$. We build a chain complex on X. Since X is the boundary of the 2-simplex, then for $n \geq 2, C_n = \{0\}$.

The chain complex is given by:

$$C_2(X) \xrightarrow{f_2} C_1(X) \xrightarrow{f_1} C_0(X) \xrightarrow{f_0} 0$$

where f_0 is the zero map, $f_1(a) = y - x$, $f_1(b) = z - y$, $f_1(c) = x - z$ and f_2 is the inclusion map. Each f_n for $n = 0, 1, 2$ is a homomorphism.

$C_0(X) = \langle x, y, z \rangle$ so $C_0(X) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$.
$C_1(X) = \langle a, b, c \rangle$ so $C_1(X) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$.
$C_2(X) \cong \{0\}$ since the 2-simplex is not in X.

To calculate the homology groups we need to find the kernel and image of each map. Let $ker(f_n)$ denote the kernel of f_n and $Im(f_n)$ denote the image of f_n.

Since f_0 is the zero map then $ker(f_0) = C_0(X) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$ and $Im(f_0) = \{0\}$.

Similarly, since $C_2(X) = \{0\}$ then $ker(f_2) = C_2(X)$ and $Im(f_2) = \{0\}$.

$Im(f_1) = \langle y - x, z - y, x - z \rangle$ since $y - x, z - y, x - z$ are the images of the generators of $C_1(X), a, b, c$ respectively. Also notice that $x - z = -(y - x) - (z - y)$ hence $Im(f_1) = \langle y - x, z - y \rangle \cong \mathbb{Z} \oplus \mathbb{Z}$.

To calculate $ker(f_1)$, we find the elements in $C_1(X)$ whose image in $C_0(X)$ is 0. Since every element in $C_1(X)$ can be written as a linear combination a, b, c, let $v \in C_1(X)$ with $v = la + mb + nc$ where l, n, m are scalars and $f_1(v) = 0$. Then $f_1(v) = f_1(la + mb + nc) = 0$. Since f_1 is a homomorphism,

$$\begin{aligned} f_1(la + mb + nc) &= lf_1(a) + mf_1(b) + nf_1(c) \\ &= l(y - x) + m(z - y) + n(x - z) \\ &= (n - l)x + (l - m)y + (m - n)z = 0. \end{aligned}$$

We notice that since x, y, z are geometrically independent this is only possible when $n - l = l - m = m - n = 0$. That is $m = n = l$. So $ker(f_1) = \langle a + b + c \rangle \cong \mathbb{Z}$.

We now compute the homology groups of X.

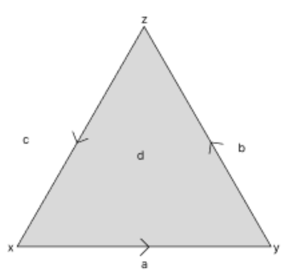
$H_0(X) = ker(f_0)/Im(f_1) = C_0/\langle y - x, z - y \rangle \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}/\mathbb{Z} \oplus \mathbb{Z} \cong \mathbb{Z}$.

$H_1(X) = ker(f_1)/Im(f_2) = \langle a + b + c \rangle/\{0\} \cong \mathbb{Z}/\{0\} = \mathbb{Z}$.

$H_2(X) = \{0\}$ since $C_2(X) = \{0\}$.

So we have that $H_n(X) = \begin{cases} \mathbb{Z} & n = 0, 1 \\ 0 & n \geq 2 \end{cases}$.
What this tells us is that the first and second homology groups are isomorphic to the cyclic group with one generator, hence $\beta_0 = \beta_1 = 1$. This is to be expected since the circle has one connected component and one tunnel. In addition, this stresses that fact that the homology groups are the abelian analog to homotopy groups since $\pi_1(S^1) \cong \mathbb{Z}$.

We now calculate the homology groups of the disc. Like in the first case, notice how the disc is homeomorphic to the 2-simplex. So their homology groups will be isomorphic. Consider the oriented 2-simplex below:



Like in the case of the circle, the vertices are $\{x, y, z\}$ and the edges are $\{a, b, c\}$. In this case there is a face, the 2-simplex d and this time we call this oriented 2-simplex Y. It is easy to see that Y is a topological space with discrete topology $\tau_y = \{x, y, z, a, b, c, d\}$. We build a chain complex on Y. Since Y is the 2-simplex then for $n \geq 3, C_n = \{0\}$. The chain complex is given by:

$$C_2(Y) \xrightarrow{g_2} C_1(Y) \xrightarrow{g_1} C_0(Y) \xrightarrow{g_0} 0$$

where g_0 is the zero map, $g_1(a) = y - x, g_1(b) = z - y, g_1(c) = x - z$, $g_2(d) = a + b + c$. Each g_n for $n = 0, 1, 2$ is a homomorphism.

$C_0(Y) = \langle x, y, z \rangle$ so $C_0(Y) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$
$C_1(Y) = \langle a, b, c \rangle$ so $C_1(Y) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$
$C_2(Y) = \langle d \rangle$ so $C_2(Y) \cong \mathbb{Z}$

To calculate the homology groups we need to find the kernel and image of each g_n. Let $ker(g_n)$ denote the kernel of g_n and $Im(f_n)$ denote the image of g_n.

Since g_0 is the zero map then $ker(g_0) = C_0(Y) \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$ and $Im(f_0) = \{0\}$.

Since $g_1 = f_1$ above, then $ker(g_1) = \langle a+b+c \rangle \cong \mathbb{Z}$ and $Im(g_1) = \langle y-x, z-y \rangle \cong \mathbb{Z} \oplus \mathbb{Z}$.

$Im(g_2) = \langle a + b + c \rangle$.
For $ker(g_2)$ choose an arbitrary $v \in C_2(Y)$ such that $g_2(v) = 0$. Since $v \in C_2(Y)$ then $v = ld$ for some scalar l.

$$\begin{aligned} g_2(v) &= g_2(ld) \\ &= lg_2(d) \\ &= l(a + b + c) = 0. \end{aligned}$$

This is only possible when $l = 0$ hence $v = 0$ and so $ker(g_2) = \{0\}$.

We now calculate the homology groups of Y:

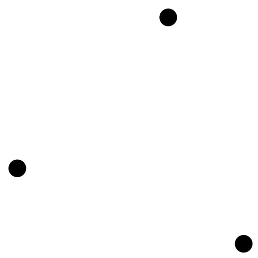
$H_0(Y) = ker(g_0)/Im(g_1) = C_0(Y)/\langle y - x, z - y \rangle \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}/\mathbb{Z} \oplus \mathbb{Z} \cong \mathbb{Z}$.

$H_1(Y) = ker(g_1)/Im(g_2) = \langle a + b + c \rangle / \langle a + b + c \rangle \cong \{0\}$.

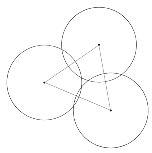
Notice that $ker(g_2) = \{0\}$ and $C_3(Y) = \{0\}$ so $H_2(Y) \cong \{0\}$. In fact, for $n \geq 2, H_n(Y) \cong \{0\}$.

So we have that $H_n(Y) = \begin{cases} \mathbb{Z} & n = 0 \\ 0 & n \geq 1 \end{cases}$.
Again this should be expected since the disc has no tunnels and is just one connected component. So $\beta_0 = 1$ and β_n for $n \geq 1 = 0$.

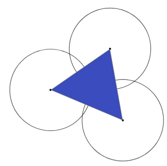
Now we build simplicial complexes to note the differences in homology between them. Consider the following set of points in $\mathbb{R}^2$ with the Euclidean metric:



We build a Čech complex and a Rips complex using these points as our vertex set.



(a) Čech complex



(b) Rips complex

It is quite easy to see that the Čech complex consists of three 1-simplices and three 0-simplices while the Rips complex consists of one 2-simplex in addition to the same three 1-simplices and 0-simplices the Čech complex has. Also each of these complexes is homeomorphic to the circle and the disc. From earlier calculations, we know that the homology groups H_n for the circle and the disc are different for $n \geq 1$ and so the choice of simplicial complex affects the resulting homology groups.

Chapter 4

Applying Persistent Homology to Topological Data Analysis

The concept of TDA employs the fact that data has shape. What this means is that the data points in the given data set were sampled from a topological space. In some sense, persistent homology seeks to "recreate" the topological space from which the data points were sampled. While it may not be possible to fully recreate the topological space with all of its geometric features, with persistent homology, we are able to gain relevant information about the shape the data was sampled from.
In this chapter, we look at some definitions in persistent homology. We also briefly discuss the methods involved in analyzing data using persistent homology before discussing an area of active research in TDA; statistical interpretation of topological summaries. Finally, we analyze sample data using the methods discussed in this chapter.

4.1 Persistent Homology

Persistent homology is the flagship tool of TDA [14]. It extends the concept of homology to filtered simplicial complexes, a sequence of subcomplexes that are ordered by inclusion. It also helps us identify the changes in homology that occur on an object with respect to a particular parameter. This allows us to see which features of the dataset remain over time. Persistence diagrams and persistence barcodes(also known as barcodes) are topological signatures that can be obtained from computing persistent homology. In this section, we define some terms that are commonly used in persistent homology.

Let K be a simplicial complex and $f : K \to \mathbb{R}$ be an increasing function. Then, for each $a \in \mathbb{R}$ the set $f^{-1}(-\infty, a]$ is a subcomplex of K. Suppose that $a_0 = -\infty$ and a_i is in the range of f for each $i \in \mathbb{N}$. Suppose that $K_i = f^{-1}(-\infty, a_i]$ for each i, then the sequence $\emptyset = K_0 \subset K_1 \subset \cdots \subset K_n = K$ is known as a *filtration* of K and the function f is known as *a filtration function* on K.

Let K be a simplicial complex. Suppose $\emptyset = K_0 \subset K_1 \subset \cdots \subset K_n = K$ is a filtration of K. Since for all $i < j, K_i \subset K_j$, then the inclusion map from each $K_i \to K_j$ induces a homomorphism $f_p^{i,j} : H_p(K_i) \to H_p(K_j)$ for each dimension of homology groups p. The filtration with the inclusion map between successive subsets give rise to a sequence of p-homology groups connected by the induced homomorphisms:

$$0 = H_p(K_0) \xrightarrow{f_p^{0,1}} H_p(K_1) \xrightarrow{f_p^{1,2}} \cdots \xrightarrow{f_p^{n-1,n}} H_p(K_n) = H_p(K)$$

The *p-th persistent homology groups* $H_p^{i,j}$ are defined as the images of the homomorphisms $f_p^{i,j}$ for $0 \leq i \leq j \leq n$.

The *p-th persistent Betti number* is the rank of the p-th persistent homology group. That is, $\beta_p^{i,j} = rank(H_p^{i,j})$.

For $i = j, H_p^{i,j} = H_p^{i,i} = H_p(K_i)$.

For $i < j$:
A class λ in the p-th homology group of K_i is said to be *born at* K_i if $\sigma \notin H_p^{i-1,i}$.

A class λ in the p-th homology group of K_i is said to *die entering* K_j if $f_p^{i,j-1}(\sigma) \notin H_p^{i-1,j-1}$ but $f_p^{i,j}(\sigma) \in H_p^{i-1,j}$.

Let f be a filtration function on a simplicial complex K and $K_0 \subset K_1 \subset \cdots \subset K_n = K$ be a filtration of K. Suppose $K_i = f^{-1}(-\infty, a_i]$ for $i \in \mathbb{N}$ and $a_i \in \mathbb{R}$ for all i and $i < j$:

Suppose λ is the homology class born at K_i which dies entering K_j. The difference $a_j - a_i$ is called *the persistence* of λ.

A persistence diagram is a multiset that is the union of a finite multiset of points in $\overline{\mathbb{R}}^2$ with the multiset of points on the diagonal $\Delta = \{(x, y) \in \mathbb{R}^2 \mid x = y\}$, where each point on the diagonal has infinite multiplicity.

4.1.1 Persistent Homology Pipeline

With the increase in the popularity of PH and TDA in general, different methods for the analysis of datasets in different domains have been put together. In this section, we cover one of such methods. The aim is to provide the reader with a clear path for data analysis using persistent homology.

1. Given the dataset, we decide at this point the kind of complex to use. We have only discussed simplicial complexes in this essay but other types of complexes, including cubical complexes (which are generalizations of the square to higher dimensions) exist. The kind of data in the dataset should ultimately determine the type of complex to use. For instance, given data points from an image, it is best to use the cubical complex since that would model the image more precisely than a simplicial complex.
   However, defining a suitable distance function on most datasets allows us to view them as finite metric spaces also known as point clouds. For data points from an image, since each pixel contains information on the color of the image at different points, the whole image can be thought of as belonging to $\mathbb{R}^{m \times n}$ with the Euclidean

metric, where n is the number of pixels in the image and m is the length of the vector each pixel contains. For this essay, we shall assume all of our datasets are point clouds.

2. After deciding on the kind of complex to use to model the dataset, the next step is to choose the type of that complex you would use. For instance, having chosen a simplicial complex we would need to decide on whether to use a Čech complex, a Rips complex or some other type of simplicial complex. Ultimately, what we are trying to do in this step is to build a filtration of the simplicial complex. It is noteworthy that it is computationally more expensive to use the Čech complex than to use the Rips complex as is illustrated in section 3.3.2. This, among other things, should affect the choice of simplicial complex.

3. The next thing to do after choosing the particular kind of complex to use is to extract topological information from the filtration and represent it visually. What this means is that for instance, choosing to use a Rips complex you compute the homology groups for each subcomplex in the filtration. Computing the homology groups give us insight into the topological features of the dataset. We then represent this information about the homology groups visually using the persistence diagram, the persistence barcode(or barcode for short) or some other visual representation. Depending on the choice of representation, different diagrams could be drawn for the different homology groups of the filtration or they could be represented on one diagram with the information for each homology group being color coded.

4. After summarizing topological information in a preferred representation (that is either a barcode or a persistence diagram), the next thing to do is to interpret the the results of the computations captured in the signature. Representing the results of the computation visually is not enough. As Otter et al. [27] put it, data sets may be from a model of random networks hence the need to average results over multiple realizations. This is an active area of research in the field and we discuss it in the following section.

4.2 Statistics and TDA

We discussed, in section 4.1.1, that after computing persistent homology it is important to represent the information in a topological signature, and in step 4 of the same section we see that it is important to interpret the information captured in the signature. As a field whose formation is largely attributed to works mainly by algebraic topologists, it is no surprise then, that the majority of theory developed in TDA focuses on topological signatures and the mathematical properties they capture. While this provides rich theory that allows significant information about the underlying shape of the data to be computed, a fundamental part of data analysis remains underdeveloped; the quantification of the uncertainty and noise in the data. Integrating statistical methodology with TDA is essential for using TDA for data analysis as it, for instance, allows us to separate the true features of the data from those caused by noise. Different

methods for integrating statistics and TDA exist.
In this section we focus on the study of the geometric properties of the space of persistence diagrams; a metric space with persistence diagram points.
Recall that in the definition of a persistence diagram it was said that each point on the diagonal has infinite multiplicity. In general, it is persistence diagrams with similar multiplicity that are compared. This happens because the comparison is done by studying bijections between the elements of the persistence diagrams. We define a distance function that is, not only popular in TDA literature but also, useful in applications:

Let X and Y be persistence diagrams and let $p \in [1, \infty]$. The *p-th Wasserstein distance* between X and Y is defined as

$$W_p[d](X, Y) = \inf_{\phi: X \to Y} \left[\sum_{x \in X} d[x, \phi(x)]^p \right]^{1/p}$$

for $p \in [1, \infty)$ and as

$$W_\infty[d](X, Y) = \inf_{\phi: X \to Y} \sup_{x \in X} d[x, \phi(x)]$$

for $p = \infty$, where d is a metric on $\mathbb{R}^2$ and ϕ ranges over all bijections from X to Y.
When $p = \infty$, the p-th Wasserstein distance is called the *Bottleneck metric.*

A property of the space of persistence diagrams with the Wasserstein metric is that it is complete and separable. This property makes it possible to define the Fréchet mean and Fréchet variance on the metric space. This makes it possible for statistical tools to be used in the context of TDA.
Analyzing the space of persistence diagrams with statistics is an area of active research in TDA. In summary, research in the area of statistical TDA has so far been able to deduce that points closest to the diagonal are features created by noise whereas those farthest from the diagonal are true topological features of the dataset.
Various libraries that implement the computation of the Wasserstein and bottleneck distances exist. For example the `DIONYSUS` library in C++, which is the oldest of these libraries, allows for the implementation of the computation of the Wasserstein distances. For more information on these libraries see Otter et al. [27].

4.3 Computing Persistent Homology With Python

We have so far talked about simplicial complexes, and the process to follow when computing persistent homology for a dataset. In this section we take a look at an example that illustrates the process of computing persistent homology using python. We make use of the Python library `Ripser` which was built using the C++ package Ripser as the core computational engine. The Python library is referred to as `Ripser.py` to differentiate it from the C++ package. According to the official documentation, [29] `Ripser.py` provides an intuitive interface for visualizing persistence diagrams. In addition to the `Ripser.py` library, python functions created by Elizabeth Munch, [24] were used in this example.

4.3.1 Python Code for Computing Persistent Homology

The data points generated for this example were sampled from an annulus. Python code was used to plot the data points in $\mathbb{R}^2$. The Rips complex was used as our choice of simplicial complex and the filtration was generated for different values of the radius. For this example, the homology groups are computed only up to the first dimension that is, only the 0-dimensional and 1-dimensional homology groups are computed.

We now look at what the different blocks of code do:

```python
#---Basic imports
import numpy as np
import matplotlib.pyplot as plt
import ripser
import teaspoon.MakeData.PointCloud as makePtCloud
import teaspoon.TDA.Draw as Draw
import math
from random import random
```

This code block imports the different libraries to be used later in our work. `Numpy` allows for numerical analysis in python while matplotlib allows us to plot different graphs and charts. `Ripser` is the TDA library that helps us generate our persistence diagrams. `Teaspoon` is a signal processing python library and we make use of the `MakeData` and `TDA.Draw` modules to help us generate the point cloud and persistence diagrams. The `math` and `random` libraries allow us to use math functions and generate random integers to be used.

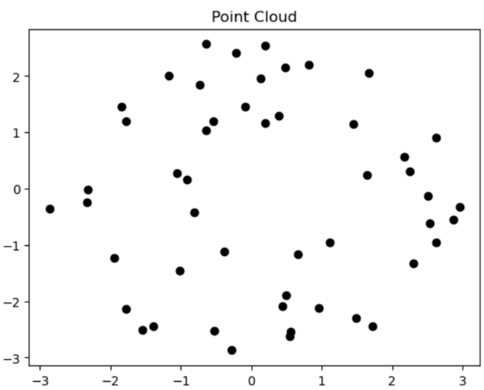
```python
def draw_Simplicial_Complex(matrix_np, distance):
    if distance>0:
        for i in range(len(matrix_np)):
            for j in range(i+1, len(matrix_np)):
                x=np.array([P[i],P[j]])
                if abs(math.dist(x[0],x[1]))<distance:
                    x_values = [x[0][0], x[1][0]]
                    y_values = [x[0][1], x[1][1]]
                    plt.plot(x_values, y_values, color="black",
markersize=0.01)
    plt.scatter(P[:,0],P[:,1], color="black")

    for i in range(len(matrix_np)-2):
        for j in range(i+1, len(matrix_np)-1):
            for k in range(j+1, len(matrix_np)):
                x=np.array([P[i],P[j],P[k]])
                for arr in range(1,len(x)-1):
                    statement = abs(math.dist(x[arr-1], x[arr]))<distance
 and abs(math.dist(x[arr],x[arr+1]))<distance and
abs(math.dist(x[arr-1],x[arr+1]))<distance
                    if statement:
                        plt.fill([x[arr-1][0],x[arr][0],x[arr+1][0]]
,[x[arr-1][1],x[arr][1],x[arr+1][1]],"b")
    plt.scatter(P[:,0],P[:,1], color="black")
    diagram_Complex = plt.show()
```

```python

def drawTDAtutorial(P,diagrams, R = 2):
    fig, axes = plt.subplots(nrows=1, ncols=3, figsize = (20,5))

    # Draw point cloud
    plt.sca(axes[0])
    plt.title('Point Cloud')
    plt.scatter(P[:,0],P[:,1])

    # Draw diagrams
    plt.sca(axes[1])
    plt.title('0-dim Diagram')
    Draw.drawDgm(diagrams[0])

    plt.sca(axes[2])
    plt.title('1-dim Diagram')
    Draw.drawDgm(diagrams[1])
    plt.axis([0,R,0,R])
```

In this block of code we define two functions; draw_Simplicial_Complex and drawTDAtutorial. The former takes in a point cloud and a nonnegative real number as arguments, and returns a simplicial complex. The latter takes in the point cloud, a list and an integer value and outputs a graph of the point cloud, and the 0-dimensional and 1-dimensional persistence diagrams. The *math.dist* method used in the draw_Simplicial_Complex funciton computes the Euclidean distance between its arguments.

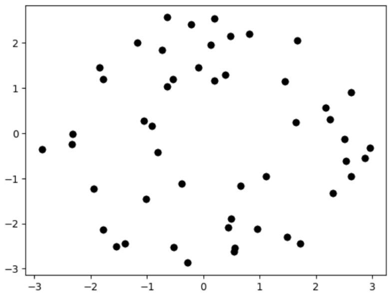
```python
r=0.5
R=3
N=50
P = makePtCloud.Annulus(N=N, r=r, R=R)
plt.title("Point Cloud")
plt.scatter(P[:,0],P[:,1], color = "black");
```

Over here we visualize our point cloud. The *.Annulus* method samples N data points from within an annulus whose inner radius is r and outer radius is R. Initializing our values for r and R as 0.5 and 3 respectively we receive the following output:

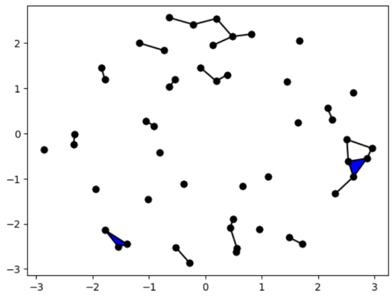



```python
distances = [0,0.5,1.0,1.5,2.0,2.5]
for i in distances:
    draw_Simplicial_Complex(P, i)
```

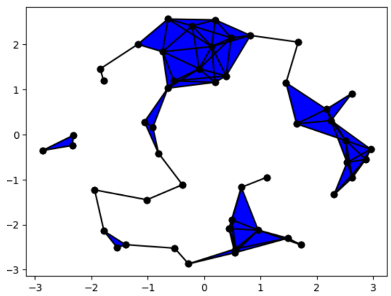
In this next bit of code we initialize a list for the distances and call the draw_Simplicial_Complex function for each of these distances to get the filtration of the Rips complex on the data points. This code outputs the following:



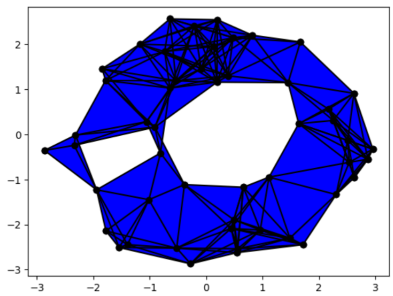
Rips complex for distance=0



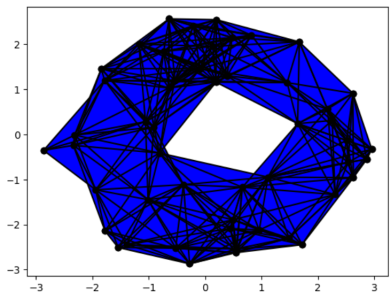
Rips complex for distance=0.5



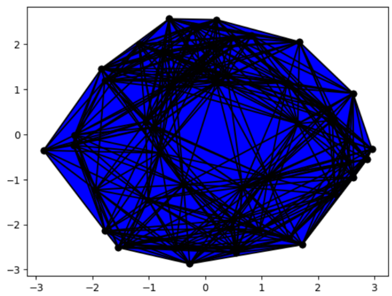
Rips complex for distance=1.0



Rips complex for distance=1.5



Rips complex for distance=2.0

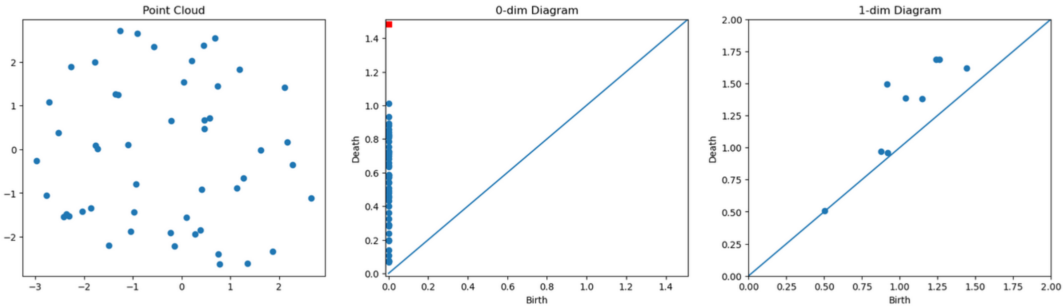


Rips complex for distance=2.5

Figure 4.1: Filtration of the rips complex on the data points sampled from the annulus

```
diagrams = ripser.ripser(P)['dgms']
drawTDAtutorial(P, diagrams)
```

Finally, this last bit of code generates the 0-dimensional and 1-dimensional persistence diagrams on the dataset. Running this code results in the following:



We have seen in this example how the whole process of computing persistent homology on a dataset can be done using python. Other libraries for the computation of persistent homology exist in python and in other programming languages. A table comparing the different libraries in terms of efficiency, among others, can be found in Otter et al.[27].

Chapter 5

Conclusion

Analyzing data is an important skill in our world today. There are different methods for data analysis to choose from and while some do well with data that is not complex, they may not be very efficient in the analysis of big data. Topological Data Analysis (TDA) is a field of mathematics that develops tools using concepts from algebraic topology and computational geometry, among other fields, to help in the efficient analysis of big data. It has proven to be very successful in this regard. Persistent homology is one of such tools developed by TDA. While there are different tools TDA has successfully developed, this essay focuses on the use of persistent homology to analyze big data.

Throughout this essay, we have talked about concepts from topology and algebraic topology that are useful in PH and TDA in general. We have also discussed persistent homology and a pipeline for the computation of persistent homology for a given dataset. Finally, we discussed the statistical interpretation of topological summaries, an area of research that seeks to perform statistical calculations on the space of persistence diagrams.

Bibliography

[1] Mehmet E Aktas, Esra Akbas, and Ahmed El Fatmaoui. "Persistence homology of networks: methods and applications". In: *Applied Network Science* 4.1 (2019), pp. 1–28.

[2] J D Bowen. *Fundamental Group is Independent of Base Point for Path-Connected Space*. Jan. 2009. URL: https://proofwiki.org/wiki/Fundamental_Group_is_Independent_of_Base_Point_for_Path-Connected_Space.

[3] Gunnar Carlsson. "Topology and data". In: *Bulletin of the American Mathematical Society* 46.2 (2009), pp. 255–308.

[4] Frédéric Chazal and Bertrand Michel. "An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists". In: *Frontiers in Artificial Intelligence* 4 (2021), p. 667963.

[5] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. "Stability of persistence diagrams". In: (2005), pp. 263–271.

[6] Brandon Doherty. "The Cech complex in Topological Data Analysis". In: (2018).

[7] Herbert Edelsbrunner, John Harer, et al. "Persistent homology-a survey". In: *Contemporary mathematics* 453 (2008), pp. 257–282.

[8] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.

[9] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. "Topological persistence and simplification". In: *Proceedings 41st annual symposium on foundations of computer science*. IEEE. 2000, pp. 454–463.

[10] Ulderico Fugacci et al. "Persistent Homology: a Step-by-step Introduction for Newcomers." In: (2016), pp. 1–10.

[11] Timothy Gowers, June Barrow-Green, and Imre Leader. *The Princeton companion to mathematics*. Princeton University Press, 2008.

[12] Jakob Hansen and Robert Ghrist. "Toward a spectral theory of cellular sheaves". In: *Journal of Applied and Computational Topology* 3.4 (2019), pp. 315–358.

[13] Allen Hatcher. *Algebraic topology*. Cambridge University Press, 2005.

[14] Felix Hensel, Michael Moor, and Bastian Rieck. "A survey of topological machine learning methods". In: *Frontiers in Artificial Intelligence* 4 (2021), p. 681108.

[15] j2kun. *The čech complex and the Vietoris-rips complex*. Apr. 2016. URL: https://jeremykun.com/2015/08/06/cech-vietoris-rips-complex/.

[16] Michael Kerber. "Persistent homology: state of the art and challenges". In: *Internationale Mathematische Nachrichten* 231.1 (2016), pp. 15–33.

[17] Erwin Kreyszig. *Introductory functional analysis with applications*. Vol. 17. John Wiley & Sons, 1991.

[18] Krishnavedala. *A torus with aspect ratio 3 as the product of a smaller (red) and a bigger (magenta) circle.* Apr. 2014. URL: `https://en.wikipedia.org/wiki/Torus#/media/File:Torus_cycles.svg`.

[19] mathtuition88. *Proof of Associativity of Operation * on Path-homotopy Classes.* June 2015. URL: `https://mathtuition88.com/2015/07/11/proof-of-associativity-of-operation-on-path-homotopy-classes/`.

[20] J Peter May. *A concise course in algebraic topology*. University of Chicago press, 1999.

[21] Yuriy Mileyko, Sayan Mukherjee, and John Harer. “Probability measures on the space of persistence diagrams”. In: *Inverse Problems* 27.12 (2011), p. 124007.

[22] Sidney A Morris. *Topology without tears*. University of New England, 1989.

[23] Elizabeth Munch. “A user's guide to topological data analysis”. In: *Journal of Learning Analytics* 4.2 (2017), pp. 47–61.

[24] Elizabeth Munch. *Python Tutorial on Topological Data Analysis*. Feb. 2021. URL: `https://mybinder.org/v2/gh/lizliz/TDA-Python-Workshop-JMM21/HEAD?filepath=PythonTDA_Intro.ipynb`.

[25] James R Munkres. *Elements of algebraic topology*. CRC press, 2018.

[26] James R Munkres. *Topology*. Vol. 2. Prentice Hall Upper Saddle River, 2000.

[27] Nina Otter et al. “A roadmap for the computation of persistent homology”. In: *EPJ Data Science* 6 (2017), pp. 1–38.

[28] Václav Snášel et al. “Geometrical and topological approaches to Big Data”. In: *Future Generation Computer Systems* 67 (2017), pp. 286–296.

[29] Christopher Tralie and Nathaniel Saul. *Ripser.py 0.6.4 documentation*. 2019. URL: `https://jeremykun.com/2015/08/06/cech-vietoris-rips-complex/`.

[30] Christopher Tralie, Nathaniel Saul, and Rann Bar-On. “Ripser.py: A Lean Persistent Homology Library for Python”. In: *The Journal of Open Source Software* 3.29 (Sept. 2018), p. 925. DOI: `10.21105/joss.00925`. URL: `https://doi.org/10.21105/joss.00925`.

[31] Shmuel Weinberger. “What is... persistent homology”. In: *Notices of the AMS* 58.1 (2011), pp. 36–39.