A Study of Persistent Homology In Topological Data Analysis

A thesis submitted to
Indian Institute of Science Education and Research Pune
in partial fulfilment of the requirements for the
Mathematics M.Sc Degree Program
under the supervision of
Prof. Mainak Poddar

by
Trishartadeb Mistri
March, 2024

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road, Pashan, Pune India 411008

Certificate

This is to certify that this thesis entitled "*A Study of Persistent Homology In Topological Data Analysis*" submitted towards the partial fulfilment of the Mathematics M.Sc Degree Program at the Indian Institute of Science Education and Research Pune represents work carried out by *Trishartadeb Mistri* under the supervision of *Prof. Mainak Poddar*.

Prof. Mainak Poddar
Master's Thesis Supervisor

Declaration

I hereby declare that the matter embodied in the report entitled ***A Study of Persistent Homology In Topological Data Analysis*** are the results of the study carried out by me at the Department of Mathematics, Indian Institute of Science Education and Research, Pune, under the supervision of Mainak Poddar and the same has not been submitted elsewhere for any other degree.

Trishartadeb Mistri

This thesis is dedicated to my parents

Acknowledgements

I am profoundly grateful to my supervisor, Prof. Mainak Poddar, for his patient guidance, insightful feedback, and unwavering support throughout every stage of this thesis. I also thank Prof. Rama Mishra for her invaluable advice on future research directions and for helping me navigate life's challenges with wisdom and composure.
My heartfelt appreciation goes to my friend Kwanit Gangopadhyay, whose expertise in coding and constructive discussions were crucial for building the Čech and Delaunay complexes at the heart of this work. I am blessed to share both joys and struggles with friends Drishti, Shinjini, Nilay, Manjima, Aniket, Haraprasad, Veronica (my badminton partner), Savita, Nipurn, Gyan, and Shantanu whose camaraderie and laughter lightened long days of calculation and proof checking. I am equally indebted to my seniors Dhrubajyoti, Subhash, Abir, Suthirtha, Sakshi, and Sandip for their mentorship and encouragement whenever I faced obstacles.
Finally, I extend my thanks to all the faculty members of the Mathematics Department at IISER Pune whose courses deepened my understanding and inspired my research, and to the administrative and technical staff for their prompt assistance. To my family, thank you for your boundless love and belief in me, which sustained my spirit through every setback. Your support has made my journey at IISER Pune both memorable and meaningful.

Abstract

This thesis studies how persistent homology can recover stable topological features from high-dimensional noisy data. It describes a mathematical setup with homology theory and simplicial complexes (like Rips and Cech complexes) where multi-scale data representations can be built. By building persistence modules and applying algebraic techniques over polynomial rings, the thesis highlights efficient algorithms for the computation of the lifetime of features like connected components and loops, with both theoretical arguments and the usage of computational software for topological data analysis.

Contents

1 Introduction

Modern science and engineering are now characterized by an unprecedented influx of data. This surge is driven both by innovative experimental techniques and by the widespread availability of high-powered computing resources. Moreover, the very nature of the data has evolved. In many cases, data is now represented as exceedingly long vectors, in which only a handful of coordinates are truly relevant to the issues at hand—and these key coordinates are often not identified in advance. Compounding this challenge is the high-dimensionality of the data, which greatly hampers visualization efforts.

Additionally, today's data is frequently noisier and more incomplete than in the past, a trend particularly evident in biological datasets produced by high-throughput methods such as microarrays. Our current analytical capabilities, both in terms of volume and in understanding the data's structure, are struggling to keep pace with this rapid production.

In these notes, we study how the disciplines of geometry and topology can contribute valuable insights to data analysis following the works of Gunnar Carlsson and Afra Zomorodian . We do not make any claims to originality.

Geometry, fundamentally the study of distance functions, is especially relevant when dealing with large finite sets of data, where distances between points play a critical role. The formalism formulated for integrating topological and geometric approaches focuses on point clouds, i.e., finite collections of points equipped with a distance function. These point clouds are interpreted as noisy samples drawn from an underlying geometric object.

This geometric approach to the analysis of complex datasets is summarized by the following key points.

2 Persistence and Homology

2.1 Introduction

When analyzing the intrinsic properties of a space X, one of the first considerations is its breakdown into path-connected components. This natural partition not only segments X but also yields an invariant: the number of distinct components, which encapsulates a fundamental qualitative feature of the space. Moreover, there exist several additional qualitative characteristics worth examining.

Homology, for instance, is a tool used to detect and measure holes within a complex of simplics. Since a collection of simplicies build a simplicial complex, it becomes essential to establish a computational framework grounded in these building blocks in order to derive meaningful insights. A formal exposition of this framework will be presented alongside a simple illustrative example.

Let take an abstract simplicial complex W of dimension n, and let $\mathcal{F}$ be a chosen field of coefficients.

Example.[1] Consider $\mathcal{C}, \mathcal{D}$ in the image below.

Figure 1: C and D

Both spaces under consideration are path-connected; however, they differ in a fundamental way. The space resembling the letter "C" on the right contains no distinct loops, whereas the space resembling the letter "D" on the left exhibits only a single loop. This discrepancy in loop structure is maintained under continuous deformations. Consequently, by formalizing this

property with rigorous mathematical language, one can clearly distinguish between these two spaces. Here, we suggest to the study of loops and their higher-dimensional analogues as "connectivity information." In this context, partitioning a space into its path-connected components is viewed as 0th-level connectivity information, 1st-level connectivity information considers the presence of loops, and this hierarchy extends to higher dimensions.

2.2 Homology Theory

For more details, we can see this reference [6]

2.2.1 Simplicial Homology

A a topological space X be Δ-***complex*** constructed by attaching standard simplices together via continuous maps, called *characteristic maps*, satisfying:

1. The map $\sigma\colon \Delta^n \to X$ is injective on the interior of Δ^n.
2. For each face F of Δ^n, the restriction $\sigma|_F$ factors through some $\sigma'\colon \Delta^{n-1} \to X$.

This structure allows more flexibility than simplicial complexes, enabling the modeling of spaces that lack a strict triangulation.

Definition 2.1 (Affine Subspaces of $\mathbb{R}^n$ and Linear Maps)**.** *Let* $\mathbb{R}^n = \{(y_1, \ldots, y_n) : y_i \in \mathbb{R},\, 1 \leq i \leq n\}$ *denote the standard* n*-dimensional Euclidean space. We define:*

$$\mathbb{R}^n_+ = \{(y_1, \ldots, y_n) \in \mathbb{R}^n : y_n \geq 0\}, \quad J = [-1, 1] \subset \mathbb{R}, \quad \mathbb{N} = \{0, 1, 2, \ldots\} \subset \mathbb{Z}.$$

Alternatively, A translation of an n*-dimensional vector subspace define an affine subspace* $W \subset \mathbb{R}^m$ *of dimension* n *. That is,* V *is affine if for any* $b_1, \ldots, b_r \in V$ *and coefficients* $\mu_1, \ldots, \mu_k \in [0, 1]$ *satisfying* $\sum_{j=1}^{k} \mu_i = 1$*, we have:*

$$\sum_{j=1}^{k} \mu_i a_i \in V.$$

A function $f : W \to \mathbb{R}^d$, where V is an affine subspace of $\mathbb{R}^m$, is said to be affine linear *if it preserves affine combinations:*

$$f\left(\sum_{j=1}^{k} \lambda_j a_j\right) = \sum_{j=1}^{k} \mu_j f(a_j).$$

If W is an affine subspace of $\mathcal{R}^m$ with dimension $m-1$, then its complement $\mathbb{R}^m \setminus W$ consists of exactly two connected components, denoted H_1 and H_2. The unions $W \cup H_1$ and $W \cup H_2$ are called the (closed) half-spaces determined by W.

Furthermore, a collection of points $w_1, \ldots, w_n \in \mathcal{R}^m$ is said to be affinely independent if the smallest affine subspace containing them is $(n-1)$-dimensional. Equivalently, the vectors $w_2 - w_1, \ldots, w_n - w_1$ are linearly independent.

Definition 2.2 (Simplices And Polytope)**.** *[3]*

A subset $C \subseteq \mathbb{R}^m$ is called convex *when any two points $a, b \in C$, the line segment connecting them, denoted ab, is totally contained inside C.*

Given any subset $A \subseteq \mathbb{R}^m$, the convex hull *of A, written $\langle A \rangle$, is defined as the smallest convex set containing A.*

A polytope *is the convex hull containing a finite set of points in $\mathbb{R}^m$. A polytope* C *is called an p-dimensional polytope (or simply, an p-polytope) if the smallest affine subspace that contains* C *has dimension p. By convention, a polytope of dimension -1 is the empty set.*

A point w is called vertex in a polytope C *if, whenever w lies on a line segment $ab \subset C$, it must be that $w = a$ or $w = b$. It follows that any p-polytope has at least $p+1$ vertices.*

*If an p-polytope has exactly $p+1$ vertices, then it is called an p-*simplex*. In this case, the convex hull of the set $\{w_0, w_1, \ldots, w_n\}$, denoted by $\langle w_0, w_1, \ldots, w_n \rangle$, forms an p-simplex if and only if the points $w_0, w_1, \ldots, w_n$ are* affinely independent*.*

We denote an p-simplex with vertices $w_0, w_1, \ldots, w_n$ by $[w_0, w_1, \ldots, w_n]$.

For a r-simplex $A = [w_0, w_1, \ldots, w_r]$, the barycenter *of $\mathcal{D}$ is given by:*

$$\hat{\mathbb{D}} := \frac{1}{r+1} \sum_{i=0}^{r} w_i.$$

Let $[w_0, \ldots, w_p]$ be an p-simplex. If we remove any one of its $p+1$ vertices, the remaining p vertices span an $(p-1)$-simplex. Such a simplex is called a *face* of $[w_0, \ldots, w_n]$.
The *boundary* of an p-simplex Δ^n, denoted $\partial\Delta^n$, is the union of all its $(p-1)$-dimensional faces.

$\overset{\circ}{\Delta}{}^p$ denoted the *interior* (or *open simplex*) of Δ^p , is defined as the complement of its boundary within itself:

$$\overset{\circ}{\Delta}{}^p = \Delta^p \setminus \partial\Delta^p.$$

1. A *1-simplex* is defined as the convex hull of two distinct points, say v_0 and v_1, in a given space. When these points are embedded in a metric space (e.g., $\mathbb{R}^n$) with a finite distance between them, the 1-simplex is a closed line segment whose length is the distance $d(v_0, v_1)$.

2. ***1-Skeleton* of a triangulation is the union of 1- of that triangulation**

Let us describe the simplicial complex groups of a Δ-complex X and let $\Delta_p(X)$ be the free abelian group with basis the open p-simplicies e^p_α of X.

A p-chain is an element of $\Delta_p(X)$. Take $C_p = \Delta_p(X)$

As illustrated by the structure of simplices, the boundary of an p-simplex $[w_0, \ldots, w_n]$ is composed of its $(p-1)$-dimensional faces. Each face is obtained by omitting one vertex from the full simplex, and is denoted $[w_0, \ldots, \widehat{w_i}, \ldots, w_p]$, where the hat symbol $\widehat{w_i}$ indicates that the vertex w_i is excluded. In terms of chains, the boundary of the simplex is given by the alternating sum of these faces:

$$\partial_p[w_0, \ldots, w_p] = \sum_{j=0}^{p} (-1)^j [w_0, \ldots, \widehat{w_i}, \ldots, w_p].$$

Now, the boundary homomorphism on $\Delta - complex$ X is

$$\partial_n : \Delta_p(\mathcal{X}) \to \Delta_{p-1}(\mathcal{X})$$

by the values on basis elements:

$$\partial_p(\sigma_\alpha) \ = \ \sum_i (-1)^i \sigma_\alpha | [w_0, \ldots, \widehat{w}_i, \ldots, w_n]$$

Theorem 2.1. *The composition of boundary operators is always zero. That is, for any $n \geq 2$, the following composition vanishes:*

$$\Delta_p(\mathcal{X}) \xrightarrow{\partial_p} \Delta_{p-1}(\mathcal{X}) \xrightarrow{\partial_{p-1}} \Delta_{p-2}(\mathcal{X}), \quad so\ \partial_{p-1} \circ \partial_p = 0.$$

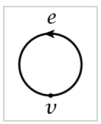
Such a sequence is called a **chain complex.** This implies the inclusion $Im\partial_{n+1} \subset Ker\partial_n$.
So, the n-th **homology group** of the chain complex to be the quotient $H_n = Ker\partial_n / Im\partial_{n+1}$, where the elements of $Ker\partial_n$ are called cycles and elements of $Im\partial_{n+1}$ are the boundary.

Example 2.1. *Let $X = S^1$, the circle, represented as a simplicial complex having one vertex v and one edge e. Then the chain groups $\Delta_\prime(S^1)$ and $\Delta_1(S^1)$ are both isomorphic to $\mathbb{Z}$, and the boundary map ∂_1 is trivial since $\partial e = v - v = 0$.*

For $p \geq 2$, there are no p-simplices in this complex, so $\Delta_p(S^1) = 0$. Therefore, the homology groups are:

$$H_p^\Delta(S^1) \cong \begin{cases} \mathbb{Z} & for\ p = 0, 1, \\ 0 & for\ p \geq 2. \end{cases}$$

This example illustrates a general principle: since 0 is the boundary map



of chain complex, then the homology groups are naturally isomorphic to the

chain groups themselves.

We see that to define simplicial homology, the topological space must have a triangulation or be expressed as a simplicial complex. However, all topological spaces do have that property. For that reason, we need to introduce Singular homology by considering all continuous maps from standard simplices into a space, which applies to any topological space. This universality makes it a truly general homology theory.

2.2.2 Singular Homology

[6]

A *singular p-simplex* in a topological space Y is defined as a continuous map

$$\sigma : \Delta^p \to Y.$$

The term "singular" indicates that the map σ need not be an embedding; it may exhibit singularities where its image fails to resemble a geometric simplex. The only requirement is that σ is continuous.

Let $C_n(Y)$ denote the free abelian group generated by the set of all singular p-simplices in Y. Elements of $C_n(Y)$, called *singular p-chains*, are finite formal sums

$$\sum_i p_i \sigma_i,$$

where $p_i \in \mathbb{Z}$ and $\sigma_i : \Delta^p \to Y$ are singular p-simplices.

The *boundary map*

$$\partial_p : C_n(Y) \to C_{p-1}(Y)$$

is defined using the same alternating sum formula as for simplicial chains:

$$\partial_p(\sigma) = \sum_{i=0}^{p} (-1)^i \sigma \circ [w_0, \ldots, \widehat{w_i}, \ldots, w_p],$$

where $[w_0, \ldots, \widehat{v_i}, \ldots, w_p]$ denotes the i-th face of the standard p-simplex (with the i-th vertex omitted), and is identified canonically with Δ^{p-1}. Thus, each composition $\sigma \circ [w_0, \ldots, \widehat{w_i}, \ldots, v_p]$ is a singular $(p-1)$-simplex.

We often write the boundary operator simply as ∂ when the context makes the dimension clear. The boundary map satisfies the key identity:

$$\partial_p \circ \partial_{p+1} = 0,$$

or concisely, $\partial^2 = 0$. This allows us to define the *singular homology groups* of X as

$$H_p(X) = Ker\partial_p / \operatorname{im} \partial_{p+1}.$$

Theorem 2.2 (Decomposition over Path-Components)**.** *Let Y be a topological space with path-components Y_β. Then, for each $p \geq 0$, the homology group $H_p(X)$ decomposes as a direct sum:*

$$H_p(X) \cong \bigoplus_\alpha H_p(Y_\beta).$$

Theorem 2.3 (Homology in Degree Zero)**.** *If Y is nonempty and path-connected, then the zeroth homology group satisfies*

$$H_0(Y) \cong \mathbb{Z}.$$

In general, for any space X, the group $H_0(X)$ is isomorphic to a direct sum of copies of $\mathbb{Z}$, one for each path-component of X.

Remark 2.1. *If X consists of a single point, then its homology groups are trivial in all positive degrees:*

$$H_p(Y) = 0 \quad \textit{for all } p > 0,$$

while the $\prime$ th homology group is isomorphic to $\mathbb{Z}$, that is,

$$H_0(Y) \cong \mathbb{Z}.$$

Relation between Simplicial and Singular Homology

- **Definition Domains:** Simplicial homology is defined for spaces with a simplicial complex structure, while singular homology is defined for

all topological spaces.

- **Construction:** Simplicial homology uses combinatorial simplices and their face relations, whereas singular homology uses continuous maps from standard simplices into the space.
- **Generality:** Singular homology is more general; it applies to any topological space, even those without a triangulation.
- **Equivalence:** For any space that is the geometric realisation of a simplicial complex, then the simplicial and singular homology groups are the same.
- **Use:** Simplicial homology is often easier to compute when a simplicial structure is present, but singular homology is more flexible and theoretically complete. One can say that simplicial homology is the special case of singular homology.

2.3 Homotopy and Invariance of Homology

[6]

As mentioned at the beginning of this chapter, the power of homology lies in its invariance under continuous deformation. This property allows us to classify spaces up to homotopy and forms the foundation of topological data analysis. In this section, we formalize this idea.

Definition 2.3 (Homotopic Maps)**.** *Two continuous maps $f, g : X \to Y$ are said to be* homotopic *if there exists a continuous function*

$$H : Y \times [0, 1] \to Z$$

such that $H(y, 0) = f(z)$ and $H(y, 1) = g(y)$ for all $y \in Y$. In this case, we write $f \simeq g$ and say that f and g are homotopic.

Definition 2.4 (Homotopy Equivalence)**.** *Two topological spaces Z and Y*

are said to be homotopy equivalent *if there exist continuous maps*

$$f : Y \to Z \quad \textit{and} \quad g : Z \to Y$$

such that the compositions satisfy $g \circ f \simeq \mathrm{id}_Y$ *and* $f \circ g \simeq \mathrm{id}_Z$*. In this case, we say* Z *and* Y *are homotopy equivalent, and denote this by*

$$Z \overset{h}{\sim} Y$$

One of the most important results in singular homology is that **homotopy equivalent spaces have isomorphic homology groups**. To prove this, we start with a continuous map $f : Y \to Z$, which induces a group homomorphism $f_* : H_n(Y) \to H_n(Z)$ on homology for each p. If f is a homotopy equivalence, then f_* will be an isomorphism.

To build this homomorphism, we begin at the chain level. Given a singular n-simplex $\sigma : \Delta^p \to X$, we define $f_\sharp(\sigma) = f \circ \sigma : \Delta^p \to Y$. Then, we apply f to the image of σ. Then, we extend this linearly to all p-chains by:

$$f_\sharp\left(\sum_j n_j \sigma_j\right) = \sum_j n_j f_\sharp(\sigma_j)$$

Now we have a map between chain complexes $f_\sharp : C_p(X) \to C_p(Y)$. We can verify that this commutes with the boundary map:

$$f_\sharp(\partial\sigma) = f_\sharp\left(\sum_{j=0}^{n}(-1)^j \sigma|[w_0, \ldots, \hat{w}_i, \ldots, w_p]\right) = \sum_{j=0}^{p}(-1)^j f\circ\sigma|[w_0, \ldots, \hat{v}_j, \ldots, v_p] = \partial f_\sharp(\sigma)$$

This confirms that $f_\sharp \circ \partial = \partial \circ f_\sharp$, so the following diagram commutes:

$$\begin{array}{ccccccccc}
\cdots & \xrightarrow{\partial} & C_{p+1}(Y) & \xrightarrow{\partial} & C_p(Y) & \xrightarrow{\partial} & C_{p-1}(Y) & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle f_\sharp} & & \downarrow{\scriptstyle f_\sharp} & & \downarrow{\scriptstyle f_\sharp} & & \\
\cdots & \xrightarrow{\partial} & C_{p+1}(Z) & \xrightarrow{\partial} & C_p(Z) & \xrightarrow{\partial} & C_{p-1}(Z) & \longrightarrow & \cdots
\end{array}$$

This kind of diagram, where corresponding squares commute, is called a **commutative diagram**. It means that any path from one object to another (following arrows) gives the same result, no matter which route is taken.

Because $f_\sharp$ preserves boundaries and cycles (i.e., it maps cycles to cycles and boundaries to boundaries), it induces a map on homology:

$$f_* : H_p(Y) \to H_p(Z)$$

A map f_* between homology groups has been constructed starting from a continuous map f between spaces. If f is a homotopy equivalence, then f_* is shown to be an isomorphism, demonstrating that singular homology is a **homotopy invariant**.

Theorem 2.4. *[6]*

Homomorphisms between the homology groups of two chain complexes are induced by a chain map between them.

Theorem 2.5. *If two continuous maps $f, g : Y \to Z$ are homotopic, then the same homomorphism on homology is induced by both maps, i.e.,*

$$f_* = g_* : H_n(Y) \to H_n(Z)$$

for all n.

Corollary 1. *If a map $f : Y \to Z$ is a homotopy equivalence, then it induces isomorphisms on all homology groups:*

$$f_* : H_P(Y) \to H_p(Z)$$

for every p.

- **Functoriality:** Ley Y be any topological space, abelian group B, and integer $k \geq 0$, and for any continuous map $f : Y \to Z$ and $g : Z \to X$, there exists an induced homomorphism

  $$H_k(f, B) : H_k(Y, B) \to H_k(Z, B).$$

  Additionally, the following properties hold:

  $$H_k(f \circ g, B) = H_k(f, B) \circ H_k(g, B), \quad H_k(\mathrm{Id}_Y, B) = \mathrm{Id}_{H_k(Y,B)}.$$

  These conditions are referred to as *functoriality.*

- **Normalization:** The homology group $H_0(*, B)$ of a space $*$ containing one point is isomorphic to the group B, i.e.,

  $$H_0(*, B) \cong B.$$

- **Betti Numbers:** For any field F, the homology group $H_k(X, F)$ is a vector space over F. The dimension of this vector space, if finite, is

denoted by

$$\beta_k(X, F),$$

and is denoted as the k-th Betti number of X with coefficients in F. Intuitively, this Betti number $\beta_k(X, F)$ represents the number of independent surfaces of k-dimension in the space. Whenever two spaces are homotopy equivalent, their corresponding Betti numbers are identical.

Example 2.2. *For any topological space X with finite components, $\beta_0(X)$ is the number of path components.*

Example 2.3. *The first Betti number β_1 of the letter D above is one, and for the letter C it is zero, whatever field F is. In this case, it gives a formalisation of the number of loops present in the space.*

Remark 2.2. *The Betti numbers depend on the choice of the coefficient field F. In most of the calculations in this thesis, we will use coefficients in the field F_2 of 2 elements.*

The definition of homology applicable to all topological spaces is singular homology. Singular homology depends on the linear algebra of infinitely generated modules over the ring $\mathbb{Z}$ to describe homology groups. However, for computational purposes, this is not a good choice.

While computations can be done using different methods, such as the long exact sequence of a pair, the long exact Mayer-Vietoris sequence, the excision theorem, or spectral sequences, direct computation from the definition is not feasible for general spaces.

Nevertheless, when a space is provided with certain structures, it often leads to finite linear algebra problems that yield correct answers, i.e., answers that agree with the singular homology technique. A particularly nice example occurs when the space is described as a simplicial complex.

Definition 2.5. *Let W be a finite set. An **abstract simplicial complex** L on W is a family of non-empty subsets of W, such that if $w \in L$ and $t \subseteq w$ is non-empty, then $t \in L$.*

In some book or paper, the non-empty condition in the definition of an abstract simplicial complex is omitted, and the empty set is an abstract simplex of dimension -1.

- An abstract simplex a is an element of L. The dimension of a, denoted $\dim(a)$, is $|a| - 1$.
- If $t \subseteq a \in L$, then:
  - t is a face of a.
  - a is a coface of t.
  - t is a facet of a if $\dim(t) + 1 = \dim(a)$.
- The dimension $\dim(L)$ of L is the maximal dimension of a simplex in L.
- The (closed) star of a vertex $v \in K$ is defined as $\mathrm{St}_K(w) = \mathrm{St}(w) = \{w \in K \mid a \cup \{v\} \in K\} \subseteq K$.
- The link of a vertex $v \in K$ is defined as $\mathrm{Lk}_K(w) = \mathrm{Lk}(w) = \{t \in \mathrm{St}(w) \mid w \notin t\} \subseteq \mathrm{St}(w)$.

Intuitively, a simplicial complex provides a way to describe a space as a collection of fundamental building blocks: points, line segments, triangles, and their higher-dimensional counterparts. In some topological spaces, simplicial complexes give a simple combinatorial framework. Due to this, they are often used to approximate topological spaces in different ways.

A important aspect for this section is that for simplicial complexes, homology computations reduce to the linear algebra of finitely generated $\mathbb{Z}$-modules. We elaborate on this in detail. Given a simplicial complex $Y = (W, \Sigma)$, we denote by Σ_p the subset of Σ containing all simplices $\sigma \in \Sigma$ with $\#(\sigma) = p + 1$. The elements of Σ_p are called p-simplices. The group of p-chains in Y is the formal free abelian group generated by elements of Σ_p, written as $C_p(X)$.

If we impose a total ordering on the vertex set W, we can define operators $d_j : \Sigma_p \to \Sigma_{p-1}$ for $0 \leq j \leq p$, given by

$$d_j(\sigma) = \sigma - \{s_j\},$$

where s_j is the j-th element of σ under the chosen ordering. This leads to the definition of the boundary operators

$$\partial_p : C_p(X) \to C_{p-1}(X),$$

given by

$$\partial_p = \sum_{i=0}^{p} (-1)^i d_i.$$

Since the groups $C_p(X)$ have bases given by Σ_p, these boundary operators can be represented by matrices $D^{(p)}$, where columns correspond to elements of Σ_p and rows to elements of Σ_{p-1}. For $\sigma \in \Sigma_p$ and $\tau \in \Sigma_{p-1}$, the matrix entry $D^{(p)}_{\tau\sigma}$ satisfies: - $D^{(p)}_{\tau\sigma} = 0$ if $\tau \not\subseteq \sigma$, - $D^{(p)}_{\tau\sigma} = (-1)^i$ if $\tau \subseteq \sigma$ and obtained by removing the i-th vertex of σ.

A fundamental observation is that the composition of two consecutive boundary operators is always zero:

$$\partial_p \circ \partial_{p+1} = 0.$$

As a result, the image of ∂_{p+1} is contained in the Kernel of ∂_p, allowing the definition of simplicial homology as

$$H_p^{\text{simp}}(Y, \mathbb{Z}) = Ker(\partial_p)/\text{im}(\partial_{p+1}).$$

This basis-independent definition can be translated into computations involving matrices $\{D^{(p)}\}_{p\geq 0}$. The final homology computations typically rely on reducing these matrices to their *Smith normal form*. It follows that $H_p^{\text{simp}}(Y, \mathbb{Z})$ is always canonically isomorphic to the singular homology of the geometric realization $|Y|$. Thus, for simplicial complexes, homology can be computed algorithmically, although finding the Smith normal form of large

matrices may be computationally expensive.

2.4 Constructing coverings and complexes.

[7] Topological methods often require a simplicial complex as input. However, in many cases, the main interest are not naturally given in this form. Consequently, a fundamental step in topological analysis involves constructing simplicial complexes.

We introduce various methods for generating complexes from a finite collection of points in a metric space (Point cloud). These points may be considered as a sample of a geometric shape, a subset of numerical data in $\mathbb{R}^n$, or as other types of structured datasets.

Our discussion will focus on two key properties that such constructions should satisfy:

- **Relationship with the underlying shape:** This is often ensured by the *nerve theorem*, which establishes a connection between the constructed complex and the topological features of the original space.
- **Stability to perturbations:** A desirable property of these constructions is robustness to small variations in input data. This is formalized by the *interleaving property*, which provides a measure of similarity between different complexes.

Rips complexes provide a straightforward method for constructing a simplicial complex from a finite set of points.

Definition 2.6 (Filtration of a Simplicial Complex)**.** *Let K be the simplicial complex. the filtration of k is a collection of subcomplexes $\{K_r\}_r$, where r is parameter. it satisfying the condition that $K_r \subseteq K_{r'}$ whenever $r < r'$*

A notable example is the Rips filtration on a finite set S, which defines a filtration of the $(|S|-1)$-simplex.

2.4.1 Rips complexes

Definition 2.7 (**Rips complexes**)**.** *[7]*

Let Y be a metric space, and let $S \subset Y$ be a subset contain finite elements. Given a scale $r \geq 0$, the Rips complex or Rips(S, r) is an abstract simplicial complex with the following properties:

1. *S is the vertex set.*
2. *A subset $s \subseteq S$ construct a simplex if and only if* Diam(s) *has length $\leq r$.*

Rips complexes are a special case of clique complexes. Given a graph G with vertex set V and edge set E, the *clique complex* of G is the abstract simplicial complex whose vertex set is V, and a subset $s \subseteq V$ is said to form a simplex if and only if every pair of vertices in s is connected by an edge in G. A Rips complex is exactly the clique complex of its 1-skeleton.

The *diameter* of a finite subset $A \subseteq X$ in a metric space Y is defined as: Diam(A) = maximum distance between two points x,y in A.

The diameter of X itself is given by:
Diam$(X) = \sup_{x,y \in X} d(x, y)$.

A few observations:

- Rips complexes are referred to as *Vietoris-Rips complexes.*
- The complex Rips(S, r) provides a combinatorial representation of S at scale r.

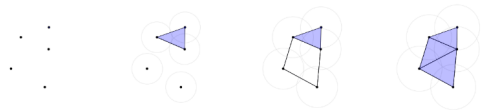


Figure 2: Five points in the plane and three of them are Rips complexes Rips(S,r). Visualization is assisted by circles as radius r increases around each point.

- Diam(s) represents the diameter of s. The condition Diam$(s) \leq r$ ensures that the distance between any two vertices of s is at most r.

- It is straightforward to verify that Rips complexes are indeed abstract simplicial complexes: if s is a simplex, then every subset of s is also a simplex.

Definition 2.8. *Let Y be a metric space and let $S \subset Y$ be a finite subset. The Rips filtration on S is the collection of abstract simplicial complexes*

$$\{\mathrm{Rips}(S, r)\}_{r \geq 0},$$

together with inclusion maps

$$i_{r_1, r_2} : \mathrm{Rips}(S, r_1) \hookrightarrow \mathrm{Rips}(S, r_2) \quad \textit{whenever } r_1 \leq r_2.$$

More generally, a *filtration* of a simplicial complex K is a family of subcomplexes $\{K_r\}_{r \geq 0}$ indexed by a parameter r such that

$$K_r \subseteq K_{r'} \quad \text{for all } r < r'.$$

• In particular, the Rips filtration on a finite set S is a filtration of the $(|S| - 1)$-simplex.

Thus, a Rips filtration gives the whole collection of Rips complexes on S. While a single Rips complex depends on the choice of the scale r. In the definition of persistent homology ,the filtration as a whole is independent of any particular scale. Filtrations play a fundamental role .

2.4.2 Čech complexes

Definition 2.9 (Čech complex). *[7]*

Let X be a metric space and let $S \subset X$ be a finite subset. Choose a scale $r \geq 0$. The Čech complex $\check{C}(S, r)$ is an abstract simplicial complex defined by the following rules:

1. *The vertex set is S.*

2. *A subset $s \subseteq S$ is a simplex if and only if*

$$\bigcap_{x \in s} B(x, r) \neq \varnothing.$$

Recall: $B(x, r) = \{y \in X \mid d(x, y) \leq r\}$ is a closed ball.

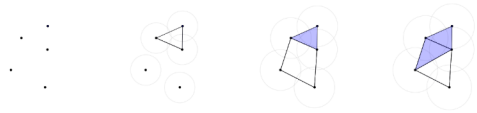


Figure 3: Five points in the plane and three corresponding Čech complexes Cech(S,r).

Remark 2.3. *A few comments about the definition:*

- *Čech complexes are a classical topological construction used in many contexts throughout topology.*
- *$\check{C}(S, r)$ represents a combinatorial snapshot of S at scale r.*
- *It is easy to verify that Čech complexes are also abstract simplicial complexes.*

Remark 2.4. *Some properties of the Čech complexes:*

1. *While harder to compute, Čech complexes are attractive due to a well-understood geometric interpretation, which will be explained in the next section within the context of nerve complexes.*
2. *$\check{C}(S, r)$ is an abstract simplicial complex, typically not embeddable in X, although it is often homotopy equivalent to a subset of X.*
3. *For r is smaller than one half of the smallest distance between points in S, $\check{C}(S, r)$ is a discrete set.*
4. *For r at least as large as twice the largest pairwise distance between points in S, the Čech complex $\check{C}(S, r)$ is regarded as the $(|S| - 1)$-simplex.*

5. *If $r_1 \leq r_2$, then $\check{C}(S, r_1) \subseteq \check{C}(S, r_2)$.*

6. *It is straightforward to show that:*

$$\check{C}(S, r) \subseteq \mathit{Rips}(S, 2r).$$

This follows from the fact that if two r radius balls overlap, then the distance between their centers must be at most $2r$.

7. *Similarly, one can observe that:*

$$\mathit{Rips}(S, r) \subseteq \check{C}(S, r).$$

A more refined inclusion holds in Euclidean spaces, where Jung's theorem ensures that:

$$\mathit{Rips}(S, r\sqrt{2}) \subseteq \check{C}(S, r).$$

Theorem 2.6. *Let M be a compact Riemannian manifold. Then there exists a positive number ϵ such that $\check{C}(M, r)$ is homotopy equivalent to M whenever $r \leq \epsilon$. Moreover, for every $r \leq \epsilon$, there exists a finite subset $W \subseteq M$ such that the subcomplex*

$$\check{C}(W, r) \subseteq \check{C}(M, r)$$

is also homotopy equivalent to M.

Definition 2.10. *Let Y be a metric space and let $S \subset Y$ be a finite subset (a sample from Y). The **Čech filtration** $\check{\mathcal{C}}(S)$ is the collection of abstract simplicial complexes*

$$\{\check{C}(S, r)\}_{r \geq 0}$$

along with the inclusion maps

$$i_{r_1, r_2} : \check{C}(S, r_1) \hookrightarrow \check{C}(S, r_2) \quad \textit{for all } r_1 \leq r_2.$$

As with the Rips filtration, the Čech filtration provides a scale-free approximation of the underlying space using simplicial complexes.

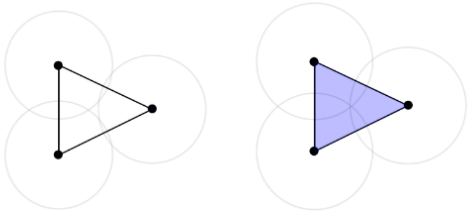


Figure 4: Let there be three points in the plane which span a triangle in the Čech complex if and only if the three closed balls of radius r centered at the points have a common intersection. On the other hand, the configuration on the left, consisting of three edges forming a cycle without a filled-in triangle, cannot arise as a Rips complex of any triple of points.

Smallest Example Demonstrating a Difference

The smallest example that highlights the difference between the Čech and Vietoris-Rips complexes occurs when considering three points forming an equilateral triangle in a metric space:

- *In the* ***Čech complex*** $\check{C}(S, r)$*, a 2-simplex (a filled triangle) appears when the three pairwise intersecting balls of radius r have a common intersection.*
- *In the* ***Vietoris-Rips complex*** $VR(S, r)$*, a 2-simplex forms as soon as all three edges exist (i.e., when all three pairwise distances are within r), without requiring a common intersection.*

This difference makes the Čech complex a more precise representation of the topology of the underlying space but computationally more expensive, whereas the Rips complex serves as an efficient approximation.

Theorem 2.7. *We have the following inclusions between the Čech and Vietoris-Rips complexes:*

$$\check{C}(X, r) \subseteq VR(X, 2r) \subseteq \check{C}(X, 2r).$$

Theorem 2.8 (Jung's Theorem)**.** *Let $F \subset \mathbb{R}^n$ be a finite subset with diameter D. Then F is contained in a ball of radius at most*

$$D\sqrt{\frac{n}{2(n+1)}}.$$

For $Y = (\mathbb{R}^n, d_2)$, we obtain the inclusion:

$$Rips(S, r\sqrt{\frac{2(n+1)}{n}}) \subseteq \check{C}(S, r).$$

Remarks

The factor $r\sqrt{2}$ appearing in Remark 7 plays a crucial role in relating the Vietoris-Rips and Čech complexes. Specifically, it provides an upper bound on the scaling relationship between the two filtrations, ensuring that the Rips complex remains a good approximation of the Čech complex while maintaining computational efficiency.

The special case of Nerve Complexes are Čech Complexes

2.4.3 Nerve Complexes

Definition 2.11 (**Nerve complexes**). *Let*

$$\mathcal{U} = \{U_1, U_2, \ldots, U_m\}, \text{ where } m \in \mathbb{N}$$

*The **nerve** of $\mathcal{U}$, denoted by $N(\mathcal{U})$, is the abstract simplicial complex defined as follows:*

1. *The vertex set of $N(\mathcal{U})$ is $\mathcal{U}$, i.e., the elements $U_1, U_2, \ldots, U_m$.*
2. *A finite subset $s \subseteq \mathcal{U}$ forms a simplex in $N(\mathcal{U})$ if and only if*

$$\bigcap_{U \in s} U \neq \varnothing.$$

Remark 2.5. *A Čech complex is the nerve of the corresponding collection of r-balls, i.e.,*

$$\check{C}(S, r) = N\big(\{B(s, r)\}_{s \in S}\big).$$

Another example is the Delaunay triangulation, which is the nerve of the Voronoi diagram.

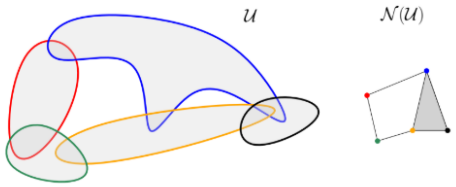


Figure 5: A example of Nerve Complex

Theorem 2.9 (Nerve Theorem)**.** *Let it be supposed that X is a topological space and that a numerable open covering $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ has been given for X. Furthermore, it is assumed that for every non-empty finite subset $S \subseteq A$, the intersection*

$$\bigcap_{s \in S} U_s$$

is either contractible or is empty. Under these conditions, the nerve $N(\mathcal{U})$ of the covering is shown to be homotopy equivalent to X.

Even the Vietoris–Rips complex is considered computationally costly, as the entire metric space is included in its vertex set. To address this issue in the study of subspaces of Euclidean space, an effective solution is provided by the use of the Voronoi decomposition.

Remark 2.6. *1. Cech complex is defined for finite subsets. So in theorem 2.6, what is Cech complex of M?*

2. Can one always find a covering that satisfies the conditions in the nerve theorem?

Answer :

By definition, every point x in a topological space lies in at least one set from an open cover. In particular, for any point $x \in M$, there exists a collection of open sets from the cover that all contain x. The intersection of such a subcollection is guaranteed to be nonempty because x lies in each of them.

In the setting of a compact Riemannian manifold, the open cover is often chosen such that each set (typically a geodesic ball) is contractible, and every finite intersection is either contractible or empty. If several of these sets all contain the same point x, then their intersection is clearly nonempty.

The Čech complex $\check{C}(M, \epsilon)$ is defined as the *nerve* of this open cover. That is, the vertices correspond to the balls $B(x, \epsilon)$, and a finite subset of these balls spans a simplex if and only if their intersection is nonempty.

Since M is compact, we can extract a finite subcover of such balls. The Nerve Theorem ensures that the nerve (i.e., the Čech complex) of this finite cover is homotopy equivalent to M. Therefore, $\check{C}(M, \epsilon)$ captures the correct topological information of the manifold for sufficiently small ϵ.

2.4.4 Voronoi Diagram and Delaunay Triangulation

Throughout this section, let $S \subset \mathbb{R}^2$ be a finite subset satisfying a general position property: no four points of S lie on the same circle.

We begin by presenting the Voronoi diagram of S, which decomposes the plane into specific regions. For each $s \in S$, define the *Voronoi region* of s as

$$V_u = \{x \in \mathbb{R}^2 \mid \forall\, s \in S \setminus \{u\},\ d(x, u) \leq d(x, s)\}.$$

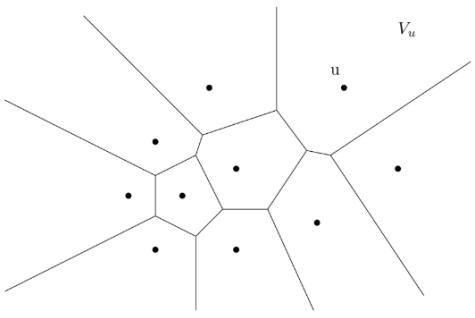


Figure 6: An example of a Voronoi decomposition.

If a pair of Voronoi regions, say V_{s_1} and V_{s_2}, have a non-empty intersection, then—due to the general position condition—this intersection is a (bounded or unbounded) line segment called a *Voronoi edge* and lies on the bisector between s_1 and s_2.

If a triple of Voronoi regions V_{s_1}, V_{s_2}, and V_{s_3} have a non-empty intersection, then this intersection is a point called a *Voronoi vertex*. Since this point lies on all three pairwise bisectors, it is the center of the circle passing through s_1, s_2, and s_3.

The **Voronoi diagram** (or decomposition) of a set S is defined as the collection of its Voronoi regions, edges, and vertices.

Voronoi regions and Voronoi diagrams can be similarly defined for subsets of $\mathbb{R}^n$. In this higher-dimensional setting, the general position property is stated as follows: no collection of $n + 2$ points from S lies on the same sphere.

If, for some point w, there exist two distinct points in S that are both closest to w, then w lies on the corresponding **Voronoi edge**. Likewise, if there are three such closest points, then w is a **Voronoi vertex**. The aforementioned general position criterion (in the plane, for example) ensures that no point has four closest points in S.

Definition 2.12. *The **Delaunay triangulation** on a finite set $S \subset \mathbb{R}^2$, denoted by $D(S)$, is the triangulation of S satisfying the following properties:*

- *Its vertices are all points of S.*
- *A pair of points $x, y \in S$ form an edge if and only if their corresponding Voronoi regions satisfy:*
$$V_x \cap V_y \neq \varnothing.$$
- *Three points $x, y, z \in S$ form a triangle if and only if their Voronoi regions satisfy:*
$$V_x \cap V_y \cap V_z \neq \varnothing.$$

This definition ensures that $D(S)$ is indeed recognized as a planar triangulation of S. Interestingly, it can be observed that an edge xy in a Delaunay triangulation may be partially extended outside the union $V_x \cup V_y$.

Furthermore, an edge xy is referred to as a **boundary edge** (i.e., one that is contained in precisely one triangle) if and only if the intersection $V_x \cap V_y$ is found to be unbounded. Similarly, a vertex x is said to be a **boundary vertex** (i.e., an endpoint of a boundary edge) if and only if V_x is identified as unbounded.

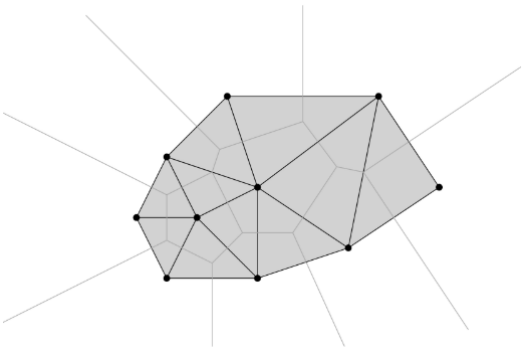


Figure 7: An example of a Delaunay triangulation with its underlying Voronoi decomposition.

2.5 Persistent Homology

Let $Y \subset \mathbb{R}^n$ be a subspace and suppose we have a method for sampling points from Y (possibly with noise), with the resulting finite sample denoted by $\mathcal{Y}$. In general, one cannot compute the Betti numbers of Y directly from the sample $\mathcal{Y}$; however, if Y is a Riemannian manifold, then under certain conditions its Betti numbers can be recovered. To tackle this challenge, we use the Čech complex $\check{C}(Y, \epsilon)$ associated with a covering of Y by balls of radius ϵ. If the centers of these balls lie on a submanifold $M \subseteq \mathbb{R}^n$ and $\mathcal{Y}$ is sufficiently dense in M, then $\check{C}(Y, \epsilon)$ coincides with the Čech complex of a covering of M; furthermore, if ϵ is chosen small enough, every ball is geodesically convex and the complex accurately computes the homology of M. Note that for any two parameters $\epsilon_1 \leq \epsilon_2$, there is a natural inclusion $\check{C}(\mathcal{Y}, \epsilon_1) \subseteq \check{C}(\mathcal{Y}, \epsilon_2)$. For example, consider a Čech complex constructed on a finite set of points in the Euclidean plane that are arranged roughly in

a circular pattern (see Figure 8). Although the main shape appears circular, computing its homology yields a first Betti number of three because the complex captures not only the large loop but also two smaller loops due to sampling errors or noise. Increasing ϵ may fill these spurious holes, but it might also introduce a new hole elsewhere, resulting in an incorrect Betti number. Importantly, the complex corresponding to a smaller ϵ is naturally included in the one defined for a larger ϵ; by examining the induced homology maps, it is observed that the two small cycles are eliminated in the larger complex, while the large cycle is preserved, thereby producing the correct topological feature. This observation forms the basis of a systematic computational method that summarizes the behavior of homology over all scales, providing a robust framework for persistent homology.[1]

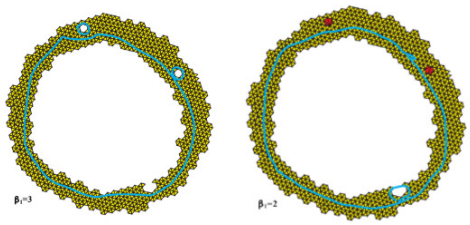


Figure 8: An example of a Čech complex constructed on a finite set of points in $\mathbb{R}^2$.

2.6 Persistence

Consider filtration of simplicial complex K,

$$K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K_i \subseteq \cdots \subseteq K_{i+p}$$

In persistent homology we quantify the persistence of homological features. The p-persistent kth homology group of a space (or filtered complex) at level K_i is defined as

$$H_k^{i,p} = \frac{Z_k^i}{B_k^{i+p} \cap Z_k^i}, \tag{4}$$

where Z_k^i is the group of k-cycles in K_i and B_k^{i+p} is the group of k-boundaries

in K_{i+p}. This quotient is considered well-defined because both B_k^{i+p} and Z_k^i are viewed as subgroups of C_{i+p}^k, and thus their intersection is regarded as a subgroup of Z_k^i. The p-persistent kth Betti number $\beta_{i,p}^k$ is defined to be the rank of the free part of $H_k^{i,p}$.

A homology class is said to be p-persistent if it appears (or is "born") at some level Y_i of the filtration and continues to exist (remains nontrivial) at the level Y_{i+p}. In other words, it survives for at least p steps of the filtration without being "killed" (i.e., without becoming trivial by turning into a boundary).

Alternatively, one may consider the induced map

$$\eta_k^{i,p} : H_k^i \to H_{i+p}^k,$$

which sends each homology class at level K_i to its corresponding class in K_{i+p}. The image of this map is isomorphic to $H_k^{i,p}$.

Here the author want to mean that we can capture the persistent homology groups by looking at the images of the inclusion-induced maps between the homology groups of successive complexes in the filtration. This approach naturally selects the cycles that survive across multiple levels of the filtration (i.e., remain nontrivial) and discards those that vanish; hence, it defines the persistent homology groups even though the original maps might not be injective on the entire homology groups.

2.6.1 The Persistence Module

In this section, we present an alternative framework for understanding persistent homology by unifying the homology of all levels of a filtered complex into a single algebraic object, known as the *persistence module*. Instead of analyzing each filtration level in isolation, we aggregate the homology groups across the entire filtration, which allows us to study the evolution of topological features in a coherent manner.

Definition 2.13 (persistence complex)**.** *Let $\{C_i^*\}_{i\geq 0}$ be a family of chain*

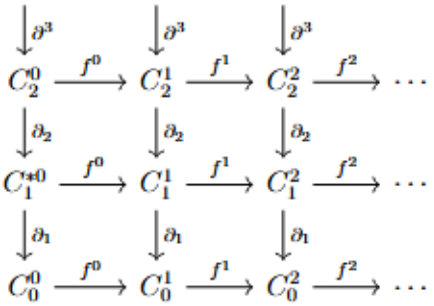
complexes over a ring R together with chain maps

$$f_i \colon C_i^* \to C_{i+1}^*.$$

These maps satisfy the natural compatibility conditions, forming a commutative diagram:

$$C_0^* \xrightarrow{f_0} C_1^* \xrightarrow{f_1} C_2^* \xrightarrow{f_2} \cdots .$$

A filtered complex K with its inclusion maps naturally gives rise to such a persistence complex.



Next,

Definition 2.14 (persistence module)**.** *A persistence module over a commutative ring R is a collection of R-modules $\{M_i\}_{i\geq 0}$, together with homomorphisms*

$$\varphi_i \colon M_i \to M_{i+1}.$$

For example, if we let $M_i = H^k(C_i^*)$, the homology groups of the chain complexes in the persistence complex, then the maps φ_i are induced by the chain maps f_i, and the resulting collection $\{H^k(C_i^*), \varphi_i\}$ forms a persistence module.

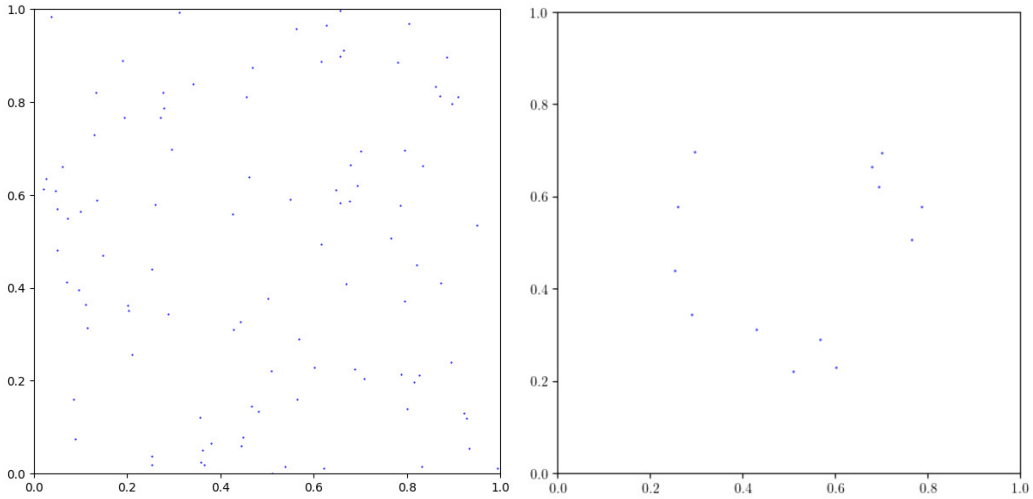
We say that a **persistence complex (or persistence module) is of *finite type*** if each component C_i^* (or M_i) is a finitely generated R-module, and there exists an integer m such that f_i (or φ_i) is an isomorphism for all

$i \geq m$. In practice, when the filtered complex K is finite (This means that the simplicial complex K has finitely many simplices in total), it induces a persistence complex of finite type, and its homology forms a persistence module of finite type. This unified perspective not only streamlines the computation of persistent homology but also lays the groundwork for establishing a correspondence between persistence modules and graded modules over the polynomial ring $R[t]$. In subsequent sections, we will exploit this correspondence to derive efficient algorithms for computing persistent homology over various coefficient rings

2.6.2 Delaunay Homology on Point Cloud Data

Let us consider a point cloud $\mathbb{X}$ containing 600 points. We restrict our attention to the points lying within an annular region with an inner radius of 0.5 and an outer radius of 2. The Delaunay complex is constructed from this restricted set of points, where only the vertices corresponding to these points are included in the simplicial complex. This ensures that the resulting complex captures the geometric and topological features of the underlying data distribution.

By computing the homology of this Delaunay complex, we can analyze its topological characteristics, such as the number of connected components (H_0), loops (H_1).



Second picture: This is the restriction of these points in the annulus.

Second picture: This is the restriction of these points in the annulus.

Figure 9:
First picture: This is 100 points data.
Second picture: This is the restriction of these points in the annulus.

2.7 Algebra

A polynomial $f(t)$ with coefficients in R (in this thesis, Rshould be a commutative ring) is defined as a formal sum of the form:

$$f(t) = \sum_{i=0}^{\infty} a_i t^i, \quad a_i \in \mathbb{R},$$

where t is an indeterminate and $a_i = 0$ for all but finitely many i. The commutative ring $\mathbb{R}[t]$ with unity is formed by the collection of all such polynomials. If R is a ring that has no zero divisors and every ideal in R is principal, then R is a *principal ideal domain* (PID). Examples of PIDs include $\mathbb{R}, \mathbb{Q}, \mathbb{Z}, \mathbb{Z}_p$ (for prime p), and $F[t]$ for any field F.

An algebraic structure $(R, +, \cdot)$ is called a *graded ring* if it is decomposed into a direct sum of Abelian groups, expressed as:

$$R = \bigoplus_{j \in \mathbb{Z}} R_j.$$

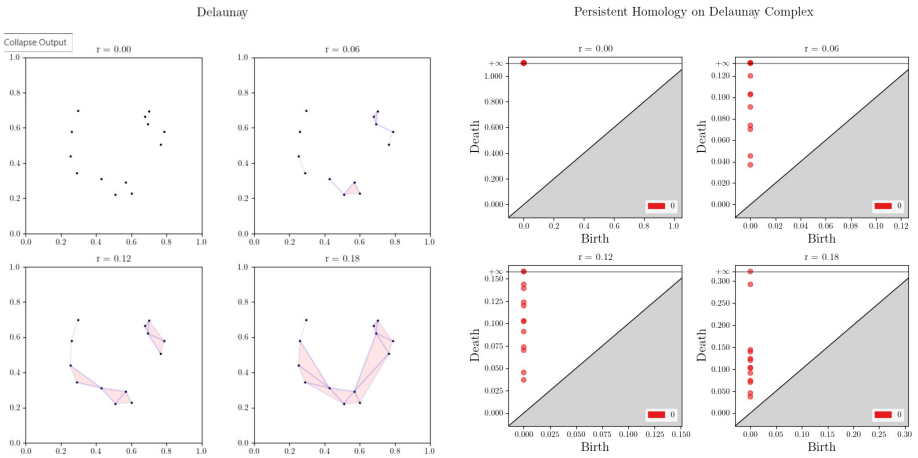


Figure 10:
First image: Use Delaunay triangulation on this new points data.
Second image: Use persistent homology on this.

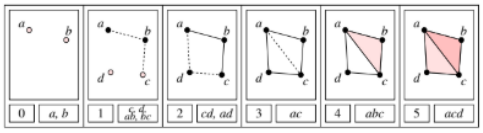


Figure 11: A filtered complex with newly added simplices highlighted.

The inclusion $R_g \cdot R_h \subseteq R_{g+h}$ is satisfied by the multiplication for all $g, h \in G$, ensuring that the product of homogeneous elements of degrees g and h is rendered homogeneous of degree $g + h$.

The multiplication in such a ring is defined via bilinear pairings:

$$R_g \otimes R_h \to R_{g+h}.$$

Elements within a single R_j are called *homogeneous elements* of degree i, and for any $e \in R_j$, we write $\deg(e) = j$.

Similarly, a *graded module* M over a graded ring R is equipped with a direct sum decomposition:

$$M = \bigoplus_{j \in \mathbb{Z}} M_j,$$

where the module action is given by bilinear pairings:

$$R_g \otimes M_h \to M_{g+h}.$$

Example 2.4. *Let $k[x] = \bigoplus_{n\in\mathbb{Z}} kx^n$ be the polynomial ring, where $n \in \mathbb{N}\cup 0$. Then $k[x]$ can be expressed as*

$$\cdots \oplus 0 \oplus 0 \oplus k \oplus kx \oplus kx^2 \oplus \cdots .$$

Here, the ring $k[x]$ is viewed as a non-negatively graded ring, with the terms corresponding to powers of x, and all negative-degree components being assigned the zero module.

A graded ring (or module) is *non-negatively graded* if $R_i = 0$ (or $M_i = 0$) for all $i < 0$. The polynomial ring $R[t]$ can be given a non-negative grading where each monomial t^n satisfies:

$$(t^n) = R \cdot t^n, \quad n \geq 0.$$

Theorem 2.10 (Structure Theorem)**.** *[8] [4]*

If D is a principal ideal domain (PID), then every finitely generated D-module is isomorphic to a direct sum of cyclic D-modules. That is, it is uniquely decomposed into the form

$$D^\beta \oplus \bigoplus_{i=1}^{m} \frac{D}{d_i D},$$

for $d_i \in D$, $\beta \in \mathbb{Z}$, such that $d_i \mid d_{i+1}$.

Similarly, every graded module M over a graded PID D decomposes uniquely into the form

$$\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} D \oplus \bigoplus_{j=1}^{m} \Sigma^{\gamma_j} \frac{D}{d_j D},$$

where $d_j \in D$ are homogeneous elements so that $d_j \mid d_{j+1}$, $\alpha_i, \gamma_j \in \mathbb{Z}$, and Σ^α denotes an α-shift upward in grading.

In both cases, the theorem decomposes the structures into free (left) and torsion (right) parts. In the latter case, the torsional elements are also homogeneous.

2.7.1 Reduction

The structure theorem for finitely generated modules over principal ideal domains (PIDs) is extensively utilised in persistent homology to classify topological features across scales. This theorem ensures that such modules are decomposed into direct sums of cyclic submodules, enabling systematic analysis of persistence modules. In subsequent sections, we will explore its implications in computational topology.
The standard method for computing homology is based on a reduction algorithm, which we illustrate using integer coefficients, since the integers form a familiar ring, though the method naturally extends to modules over any principal ideal domain (PID).

Since the chain group C_k is free and its oriented k-simplices serve as the standard basis. In this setting, the boundary operator

$$\partial_k : C_k \to C_{k-1}$$

is represented, concerning the standard bases of the chain groups, by an integer matrix M_k whose entries are drawn from $\{-1, 0, 1\}$. This matrix, called the *standard matrix representation* of ∂_k, has m_k columns and m_{k-1} rows, corresponding to the number of k-simplices and $(k-1)$-simplices, respectively. The null space of M_k represents the cycle group Z_k, while its column space corresponds to the boundary group B_{k-1} (as illustrated in Figure 3).

The reduction algorithm systematically transforms the matrix representing the boundary operator ∂_k into its Smith normal form (SNF) [1][2] by finding alternative bases for the chain groups. This diagonalisation is achieved through the following elementary operations:

Elementary Row Operations on M_k:

$$\widetilde{M_k} = \left[\begin{array}{ccc|c} b_1 & 0 & \cdots & 0 \\ 0 & b_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & b_{l_k} \\ \hline \multicolumn{3}{c|}{\mathbf{0}} & \mathbf{0} \end{array}\right],$$

1. Row exchange: Rows i and j are exchanged.
2. Row scaling: Row i is multiplied by -1.
3. Row combination: Row i is replaced by row $i + q \cdot$ row j, where $q \in \mathbb{Z}$ and $j \neq i$.

Analogous **elementary column operations** are applied to M_k. Each column (or row) operation corresponds to a change in the basis for C_k (or C_{k-1}). For example: - A type (3) column operation replaces the basis element $e_i \in C_k$ with $e_i + qe_j$. - A corresponding row operation replaces the basis element $\hat{e}_j \in C_{k-1}$ with $\hat{e}_j - q\hat{e}_i$.

These operations preserve the algebraic structure while diagonalising ∂_k, enabling computation of homology invariants.

The matrix M_k is systematically transformed into its Smith normal form through a sequence of elementary operations applied to the bases of C_k and C_{k-1}.

Suppose $l_k = \text{rank}(M_k) = \text{rank}(\widetilde{M_k})$, and let $b_i \geq 1$ for $1 \leq i \leq l_k$ with the divisibility property $b_i \mid b_{i+1}$. Although the algorithm can also compute the corresponding bases $\{e_j\}$ for C_k and $\{\hat{e}_i\}$ for C_{k-1}, this step is not strictly necessary if one is only interested in the decomposition. Once the normal form is determined in every dimension, a complete description of H_k is obtained:

(i) The torsion coefficients of H_{k-1} (denoted d_i in (1)) are precisely those diagonal entries b_i greater than 1.

(ii) The set $\{\, e_i \mid l_k + 1 \leq i \leq m_k\}$ forms a basis for Z_k. Consequently, $\text{rank}(Z_k) = m_k - l_k$.

(iii) The set $\{\, b_i\hat{e}_i \mid 1 \leq i \leq l_k\}$ forms a basis for B_{k-1}. Equivalently, $\text{rank}(B_k) = \text{rank}(M_{k+1}) = l_{k+1}$.

By combining (ii) and (iii), we obtain

$$\beta_k \;=\; \text{rank}(Z_k) \;-\; \text{rank}(B_k) \;=\; m_k - l_k - l_{k+1}. \tag{3}$$

Decomposition of H_{k-1} Torsion Submodule

The diagonal entries $b_i > 1$ in the Smith Normal Form (SNF) of ∂_k are the **torsion coefficients** of H_{k-1}. These correspond to cyclic summands of the form $\mathbb{Z}/b_i\mathbb{Z}$ in the decomposition:

$$H_{k-1} \cong \mathbb{Z}^{\beta} \oplus \left(\bigoplus_{i=1}^{l_k} \mathbb{Z}/b_i\mathbb{Z} \right),$$

where $\beta = m_{k-1} - l_k - l_{k-1}$ is the Betti number (rank of the free submodule).

Free Submodule

The rank β is determined by the dimensions of the cycle and boundary groups:

$$\beta = \text{rank}(Z_{k-1}) - \text{rank}(B_{k-1}) = (m_{k-1} - l_{k-1}) - l_k.$$

Homology Group: The quotient $H_{k-1} = Z_{k-1}/B_{k-1}$ inherits its decomposition from the SNF of ∂_k, where b_i (torsion coefficients) and β (Betti number) fully describe its structure.

Another example, consider the complex 11. The matrix representation of ∂_1 with respect to the standard bases is: where the rows and columns are

$$M_1 \;=\; \left[\begin{array}{c|ccccc} & ab & bc & cd & ad & ac \\ \hline a & -1 & 0 & 0 & -1 & -1 \\ b & 1 & -1 & 0 & 0 & 0 \\ c & 0 & 1 & -1 & 0 & 1 \\ d & 0 & 0 & 1 & 1 & 0 \end{array}\right],$$

labeled by their respective bases. Reducing this matrix to its Smith normal form yields: where $z_1 = ad - bc - cd - ab$ and $z_2 = ac - bc - ab$ form a basis for Z_1, while $\{\, d - c,\, c - b,\, b - a\}$ is a basis for B_0.

$$\tilde{M}_1 = \left[\begin{array}{c|ccccc} & cd & bc & ab & z_1 & z_2 \\ \hline d-c & 1 & 0 & 0 & 0 & 0 \\ c-b & 0 & 1 & 0 & 0 & 0 \\ b-a & 0 & 0 & 1 & 0 & 0 \\ a & 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

One can apply an analogous procedure to compute homology over graded PIDs. In such a scenario, one first represents ∂_k using the standard basis of C_k (which is itself homogeneous) and a homogeneous basis for Z_{k-1}. After reducing to the normal form, the decomposition described by direct sum (2) follows from the new basis $\{\hat{e}_j\}$ for Z_{k-1}:

(i) A row i with all zero entries contributes a free term shifted by $\alpha_i = \deg(\hat{e}_i)$.

(ii) A row with diagonal entry b_i contributes a torsional term with homogeneous element $d_j = b_j$ shifted by $\gamma_j = \deg(\hat{e}_j)$.

The reduction algorithm is executed in $O(m^3)$ time, where m denotes the total number of simplices in the complex K. A key practical concern is presented by the use of exact integer arithmetic in these transformations, as very large intermediate values may be produced, rendering the algorithm computationally intensive for large complexes.

.

2.7.2 Correspondence

Given a persistence module $M = \{M_i, \varphi_i\}_{i\geq 0}$ over a ring R, we establish a connection between persistence modules and graded modules over the polynomial ring $R[t]$. To achieve this, we equip $R[t]$ with its standard grading and define the associated graded module:

$$\alpha(M) = \bigoplus_{i=0}^{\infty} M_i.$$

Here, the module structure on $\alpha(M)$ is simply the sum of the structures on each individual M_i, and multiplication by t is defined as:

$$t \cdot (m_0, m_1, m_2, \dots) = (0, \varphi_0(m_0), \varphi_1(m_1), \varphi_2(m_2), \dots).$$

In this construction, multiplication by t shifts elements upward in the grading.

Artin-Rees Lemma and Persistent Homology

Theorem 2.11 (Artin-Rees Lemma)**.** *Let R be a Noetherian ring, $I \subseteq R$ an ideal, and $N \subseteq M$ finitely generated R-modules. There exists an integer $k \geq 1$ such that for all $n \geq k$:*

$$I^n M \cap N = I^{n-k}(I^k M \cap N).$$

Interpretation. The lemma guarantees stability in the I-adic topology: the intersection of N with powers of I in M eventually stabilizes. This property is critical for ensuring finite-type conditions in persistence modules. □

Connection to Persistent Homology

Persistent homology encodes topological features across scales as *persistence modules*, which are graded modules over the polynomial ring $R[t]$. The Artin-Rees lemma underpins the following key result:

Theorem 2.12 (Correspondence Theorem)**.** *[8] The category of persistence modules of finite type over R is equivalent to the category of finitely generated, non-negatively graded modules over $R[t]$. This equivalence is governed by the Artin-Rees lemma.*

Definition 2.15 (Stability of Persistence Modules)**.** *A persistence module $\{M_i, \phi_i\}_{i \in \mathbb{N}}$ is stable if the induced maps $\phi_i : M_i \to M_{i+1}$ satisfy Artin-Rees-like intersection conditions. Over a field F, this stability ensures a*

decomposition into barcodes:

$$M \cong \bigoplus_{i=1}^{n} \Sigma^{\alpha_i} F[t] \oplus \bigoplus_{j=1}^{m} \Sigma^{\gamma_j} \frac{F[t]}{(t^{n_j})},$$

where Σ^{α_i} denotes a grading shift.

Example: Barcode Decomposition Over Fields

When $R = F$ is a field, $F[t]$ is a principal ideal domain (PID). The Correspondence Theorem simplifies to:

Example 2.5. *Every finitely generated $F[t]$-module decomposes uniquely as:*

$$M \cong \underbrace{\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} F[t]}_{\textit{Infinite bars}} \oplus \underbrace{\bigoplus_{j=1}^{m} \Sigma^{\gamma_j} \frac{F[t]}{(t^{n_j})}}_{\textit{Finite bars}},$$

where t^{n_j} encodes the death time of the j-th feature.

Challenges in Non-Field Cases

For non-PIDs like $\mathbb{Z}[t]$, the classification of modules becomes highly nontrivial due to torsion and the absence of unique factorisation:

Remark 2.7. *Over $\mathbb{Z}[t]$, persistence modules may exhibit torsion (e.g., $\mathbb{Z}/2\mathbb{Z}$-summands), preventing a complete barcode decomposition. Invariants like persistence diagrams are instead used to track features.*

For References, see below
[8]. [5]. [2].
This result follows from the Artin-Rees theory in commutative algebra . A direct consequence of Theorem 2.12 is that there is no straightforward classification of persistence modules over a general ground ring, such as $\mathbb{Z}$. The classification of modules over $\mathbb{Z}[t]$ is highly complex, making it difficult to obtain a simple, complete description. While it is possible to define meaningful invariants for $\mathbb{Z}[t]$-modules, a full classification remains elusive.

However, when the base ring is a field F, the situation simplifies significantly. In this case, $F[t]$ is a principal ideal domain (PID), and its graded ideals are exclusively of the form (t^n). Consequently, the structure of any finitely generated graded $F[t]$-module is described by the direct sum decomposition from Theorem 5:

$$\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} F[t] \quad \oplus \quad \bigoplus_{j=1}^{m} \Sigma^{\gamma_j} F[t]/(t^{n_j}). \tag{5}$$

2.7.3 Classification via P-Intervals

To classify isomorphism classes of $F[t]$-modules, we introduce a combinatorial encoding based on *P-intervals*.

Definition 2.16 (P-interval)**.** *A* P-interval *is an ordered pair* (i, j) *with* $0 \leq i < j$ *and* $j \in \mathbb{Z}_\infty := \mathbb{Z} \cup \{+\infty\}$.

Each P-interval corresponds to a graded $F[t]$-module through a bijection Q. Specifically, for a P-interval (i, j), we define

$$Q(i, j) = \Sigma^i F[t]/(t^{j-i}).$$

For the special case where $j = +\infty$, we set

$$Q(i, +\infty) = \Sigma^i F[t].$$

Given a finite set of P-intervals $S = \{(i_1, j_1), (i_2, j_2), \ldots, (i_n, j_n)\}$, we define

$$Q(S) = \bigoplus_{l=1}^{n} Q(i_l, j_l).$$

Corollary 2. *The mapping* $S \mapsto Q(S)$ *defines a bijection between the set of finite P-intervals and the class of finitely generated graded modules over* $F[t]$. *Consequently, the isomorphism classes of persistence modules of finite type over* F *correspond bijectively to finite sets of P-intervals.*

Example: Persistent Homology of a Polygon with Interior

Suppose a polygon (its boundary) is generated at step 2 of a filtration, and its interior (the 2-cell) is generated at step 5. The persistent homology groups over a field F can be described as graded $F[t]$-modules using the structure theorem for finitely generated modules over a PID.

- **Connected component (H_0):** The component is born at step 2 and persists forever, corresponding to a free summand:

$$H_0 \cong \Sigma^2 F[t]$$

  where Σ^2 denotes a grading shift by 2 (birth at step 2).

- **1-dimensional hole (H_1):** The hole (the polygon's boundary) is born at step 2 and is filled (dies) at step 5, corresponding to a torsion summand:

$$H_1 \cong \Sigma^2 \frac{F[t]}{(t^3)}$$

  where the exponent 3 reflects the lifespan $5 - 2 = 3$.

Summary of the decomposition:

$$PH \cong \Sigma^2 F[t] \oplus \Sigma^2 \frac{F[t]}{(t^3)}$$

where PH denotes the persistent homology module, the first term represents the persistent connected component, and the second term represents the finite interval (the hole).

Corresponding barcodes:

- H_0: one infinite interval $[2, \infty)$
- H_1: one finite interval $[2, 5)$

2.8 Derivation

We use the small filtration in Figure 11 as a running example and compute over $\mathbb{Z}_2$ (although any field will do). By Theorem 2.10, the persistence module

corresponds to a $\mathbb{Z}_2[t]$–module. Table 12 reviews the degrees of the simplices in the filtration, viewed as homogeneous elements of this module.

a	b	c	d	ab	bc	cd	ad	ac	abc	acd
0	0	1	1	1	1	2	2	3	4	5

Figure 12: Degree of simplices of filtration in Figure

Throughout this section we use $\{e_j\}$ and $\{\hat{e}_i\}$ to denote homogeneous bases for C_k and C_{k-1}, respectively. Then, any matrix representation M_k of the boundary operator ∂_k (with respect to these bases) satisfies the basic property

$$\deg \hat{e}_i + \deg M_k(i,j) = \deg e_j, \tag{1}$$

where $M_k(i,j)$ is the entry in row i, column j. For example, one obtains

$$M_1 = \begin{pmatrix} 0 & 0 & t & t & 0 \\ 0 & 1 & t & 0 & t^2 \\ t & t & 0 & 0 & 0 \\ t & 0 & 0 & t^2 & t^3 \end{pmatrix}, \tag{7}$$

for ∂_1 in our example. (For instance, one can check that $M_1(4,4) = t^2$, since $\deg(\texttt{ad}) - \deg(a) = 2 - 0 = 2$, in agreement with Table 11.)

Since the standard bases for the chain groups are homogeneous, we wish to represent

$$\partial_k \colon C_k \to C_{k-1}$$

relative to the standard basis for C_k and a homogeneous basis for the cycle group Z_{k-1}. We then reduce the matrix and read off the description of H_k as discussed in Section 2.5. The process is done inductively in the dimension. The base case is trivial: since $\partial_0 \equiv 0$, we have $Z_0 = C_0$, and the standard basis works for representing ∂_1.

Now, assume we have a matrix representation M_k of ∂_k relative to the standard basis $\{e_j\}$ for C_k and a homogeneous basis $\{\hat{e}_i\}$ for Z_{k-1}. To proceed

inductively, we must compute a homogeneous basis for Z_k and represent ∂_{k+1} relative to C_{k+1} and this new basis. We begin by sorting $\{\hat{e}_i\}$ in *reverse* order of degree (as already reflected in Equation (7)), then transform M_k into its column-echelon form $\tilde{M}_k$, which is a lower staircase form (see Figure **??**). In this form, nonzero elements occur only below the "staircase" and the boxed entries (the pivots) mark the leading positions. The number of pivots (which equals the rank of M_k) indicates how many columns are *not* pivot columns; these nonpivot columns then form the desired basis for Z_k. For our example, one obtains

$$\tilde{M}_1 = \begin{pmatrix} \boxed{t} & 0 & 0 & 0 & 0 \\ t & \boxed{1} & 0 & 0 & 0 \\ 0 & t & \boxed{t} & 0 & 0 \\ 0 & 0 & t & 0 & 0 \end{pmatrix}, \tag{8}$$

where, for example,

$$z_1 = \mathtt{ad} - \mathtt{cd} - t\,\mathtt{bc} - t\,\mathtt{ab} \quad \text{and} \quad z_2 = \mathtt{ac} - t^2\,\mathtt{bc} - t^2\,\mathtt{ab}$$

form a homogeneous basis for Z_1.

The procedure to obtain the echelon form is essentially Gaussian elimination on the columns, using only elementary column operations of types (1) (column exchange) and (3) (adding a multiple of one column to another) to preserve homogeneity. Starting with the leftmost column, we eliminate nonzero entries in pivot rows in order of increasing row index. If necessary, we reorder columns so that the new pivots appear in the correct order.

Corollary 3 (Echelon Form)**.** *The pivots in the column-echelon form are identical to the diagonal elements in the Smith normal form. Moreover, the degree of the basis elements corresponding to the pivot rows remains the same in both forms.*

Proof. Since the basis elements $\{\hat{e}_i\}$ are sorted in decreasing order of degree, the degrees $\deg \hat{e}_i$ decrease from top to bottom. In any fixed column j, the degree $\deg e_j$ is constant. By Equation (1), we have

$$\deg M_k(i,j) = \deg e_j - \deg \hat{e}_i,$$

so the degree of each entry in a column increases as one moves downward. Eliminating the nonzero entries below each pivot using row operations (which do not alter the pivot values or the degrees of the row basis elements) results in a diagonal form (after further row and column swaps). Thus, the pivot entries (and the degrees of their corresponding basis elements) remain unchanged when passing to the Smith normal form. □

Corollary 4. *Let $\tilde{M}_k$ be the column-echelon form for ∂_k with respect to the bases $\{e_j\}$ for C_k and $\{\hat{e}_i\}$ for Z_{k-1}. If row i has a pivot $\tilde{M}_k(i,j) = t^n$, then it contributes $\Sigma^{\deg \hat{e}_i} F[t]/(t^n)$ to the description of H_{k-1}. Otherwise, it contributes $\Sigma^{\deg \hat{e}_i} F[t]$. Equivalently, one obtains the P-intervals $(\deg \hat{e}_i,\, \deg \hat{e}_i + n)$ and $(\deg \hat{e}_i, \infty)$, respectively, for H_{k-1}.*

For instance, in our example, $\tilde{M}_1(1,1) = t$, and since $\deg d = 1$, this corresponds to a contribution of $\Sigma^1 \mathbb{Z}_2[t]/(t)$ or, equivalently, a P-interval $(1,2)$ in the description of H_0.

Therefore, we have no need for explicit row operations. By simply deleting the rows corresponding to the pivot columns (one dimension lower), we obtain the desired representation of ∂_{k+1} in terms of the basis for Z_k. This completes the induction step.

For instance, in our example the standard matrix representation for ∂_2 is

$$M_2 = \begin{pmatrix} & abc & acd \\ ac & t & t^2 \\ ad & 0 & t^3 \\ cd & 0 & t^3 \\ bc & t^3 & 0 \\ ab & t^3 & 0 \end{pmatrix}.$$

To obtain a representation in terms of C_2 and the basis $\{z_1, z_2\}$ for Z_1 computed earlier, we simply eliminate the bottom three rows of M_2 that correspond to the pivot columns in $\tilde{M}_1$ (as seen in Equation (8)). This yields

$$\check{M}_2 = \begin{array}{c} \\ z_2 \\ z_1 \end{array} \begin{array}{c} \begin{array}{cc} abc & acd \end{array} \\ \begin{pmatrix} t & t^2 \\ 0 & t^3 \end{pmatrix} \end{array}$$

where the basis elements are given by

$$z_1 = \mathtt{ad} - \mathtt{bc} - \mathtt{cd} - \mathtt{ab}, \qquad z_2 = \mathtt{ac} - t^2\,\mathtt{bc} - t^2\,\mathtt{ab}.$$

This completes the derivation of the representation for ∂_{k+1} with respect to the computed basis for Z_k, and thus, the inductive process for obtaining homogeneous bases for the cycle groups is fully established.

Example 2.6. *Consider the following filtration:*

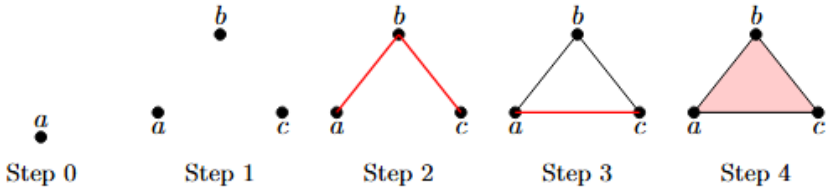


Figure 13: Filtration of a 2-simplex

Homology Computation

We aim to compute the persistent homology groups H_0 and H_1 as $\mathbb{Z}_2[t]$-modules.

Simplex	a	b	c	ab	bc	ca	abc
Degree	0	1	1	2	2	3	4

Table 1: Degrees of simplices in the filtration

Computing H_0

Let $\partial_1 : C_1 \to C_0$ be the boundary map , where $\{e_i\}$ and $\{f_i\}$ represent homogeneous bases for C_0 and C_1 respectively. The matrix representation M_1(using the (Equation 1) of ∂_1 is:

$$M_1 = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} ab & bc & ac \end{array} \\ \begin{pmatrix} t^2 & 0 & t^3 \\ t & t & 0 \\ 0 & t & t^2 \end{pmatrix} \end{array} \tag{10}$$

Through column echelon form reduction:

$$\check{M}_1 = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} ab & ac & z_1 \end{array} \\ \begin{pmatrix} t^2 & 0 & 0 \\ t & t & 0 \\ 0 & t & 0 \end{pmatrix} \end{array} \tag{11}$$

where $z_1 = ac + tab + tbc$ forms a homogeneous basis for Z_0.

Basis Transformation

Let $v \in C_1$ and $w \in C_0$ with bases $B_0 = [e_1, e_2, e_3]$ and $B_1 = [f_1, f_2, f_3]$. For the basis change $B_0 \to \widehat{B}_0$, there exists an invertible matrix P such that $\widehat{B}_0 = B_0 P$:

$$
\begin{aligned}
\partial_1 v &= w \\
w &= B_0[w]_{B_0} \\
&= \widehat{B}_0[w]_{\widehat{B}_0} \\
&= B_0 P[w]_{\widehat{B}_0} \\
\Rightarrow P^{-1}M[v]_{B_1} &= [\partial_1 v]_{\widehat{B}_0}
\end{aligned}
$$

The basis transformation corresponds to:

$$B_0 \longrightarrow B_0 P \longleftrightarrow M \longrightarrow P^{-1}M$$

$$B_0 \longrightarrow B_0 P^{-1} \longleftrightarrow M \longrightarrow PM$$

Matrix Transformation in $\mathbb{Z}_2$

The transformation matrix P satisfies $P = P^{-1}$ in $\mathbb{Z}_2$:

$$
\begin{aligned}
\begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \longrightarrow \begin{pmatrix} r_1 + r_2 t \\ r_2 \\ r_3 \end{pmatrix} &= \begin{bmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix} \\
&= P \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}, \quad P = \begin{bmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
&\text{where } P^2 \equiv I \pmod 2
\end{aligned}
$$

Applying row operations $r_2 \to r_2 + r_3$ on 11 equivalent to $c \to c + b$):

$$
\begin{array}{cc} & \begin{matrix} ab & bc & z_1 \end{matrix} \\
\check{M}_1 = \begin{matrix} a \\ b \\ c+b \end{matrix} & \begin{pmatrix} t^2 & 0 & 0 \\ t & 0 & 0 \\ 0 & t & 0 \end{pmatrix}
\end{array}
$$

Further operation $r_1 \to r_1 + tr_2$ (equivalent to $b \to at + b$):

$$\check{M}_1 = \begin{array}{c} \\ a \\ b+at \\ c+b \end{array} \begin{array}{c} \begin{array}{ccc} ab & bc & z_1 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 \\ \boxed{t} & 0 & 0 \\ 0 & \boxed{t} & 0 \end{pmatrix} \end{array} \tag{12}$$

Homology Results

The resulting homology groups are:

$$\begin{aligned} H_0 &= \frac{\langle a, at+b, c+b \rangle}{\langle t(at+b), t(b+c) \rangle} \\ &= \Sigma^0 \mathbb{Z}_2[t] \oplus \Sigma^1 \frac{\mathbb{Z}_2[t]}{\langle t \rangle} \oplus \Sigma^1 \frac{\mathbb{Z}_2[t]}{\langle t \rangle} \end{aligned}$$

Computing H_1

For the boundary map M_2:

$$M_2 = \begin{array}{c} \\ ab \\ bc \\ ac \end{array} \begin{array}{c} \begin{array}{c} abc \end{array} \\ \begin{pmatrix} t^2 \\ t^2 \\ t \end{pmatrix} \end{array}$$

After eliminating the top 2 rows from M_2 that are associated with pivots in $\check{M}_1$ 12. We get

$$\check{M}_2 = \begin{array}{c} \\ z_1 \end{array} \begin{array}{c} \begin{array}{c} abc \end{array} \\ \begin{pmatrix} t \end{pmatrix} \end{array}$$

where $z_1 = ca + t(ab) + t(cb)$. This gives:

$$H_1 = \frac{\langle bc + t(ab) + t(ca) \rangle}{\langle t(bc + t(ab) + t(ca)) \rangle} = \Sigma^3 \frac{\mathbb{Z}_2[t]}{\langle t \rangle}$$

2.9 Algorithm

Our algorithm for computing persistent homology over a field is derived from the standard reduction scheme and relies solely on elementary column operations. This approach allows us to avoid explicit matrix representations and row operations. Instead, we represent the boundary operators by maintaining an array T with a slot for each simplex in the filtration.

We first impose a total ordering on the simplices (for example, by degree and then by dimension) and mark those simplices that form the basis for the cycle groups. As each simplex σ_j is processed, the procedure RemovePivotRows is called to eliminate contributions from previously established pivot rows, updating its boundary chain d. If d becomes empty, σ_j is marked as a basis element; otherwise, the simplex corresponding to the maximum index in d is identified as the pivot. A persistence interval $(\deg(\sigma_i), \deg(\sigma_j))$ is then recorded (where σ_i is the pivot) and stored appropriately in the data structure T.

The main algorithm is outlined in Algorithm 1.

Algorithm 1 COMPUTEINTERVALS

for $k = 0$ *to* $\dim(K)$ **do**-_
$L_k \leftarrow \varnothing$ **for** $j = 0$ *to* $m - 1$ **do**
_ $d \leftarrow$ RemovePivotRows(σ_j) **if** $d = \varnothing$ **then**
_ Mark σ_j **else**
_ $i \leftarrow \max\{$ index in $d\}$ $k \leftarrow \dim(\sigma_i)$ Store j and d in $T[i]$ $L_k \leftarrow L_k \cup \{(\deg(\sigma_i), \deg(\sigma_j))\}$ **for** $j = 0$ *to* $m - 1$ **do**
_ σ_j is marked and $T[j]$ is empty $k \leftarrow \dim(\sigma_j)$ $L_k \leftarrow L_k \cup \{(\deg(\sigma_j), \infty)\}$

Algorithm 2 RemovePivotRows

$k \leftarrow \dim(\sigma)$ $d \leftarrow \partial_k \sigma$ Remove from d all terms corresponding to unmarked simplices **while** $d \neq \varnothing$ **do**-_
Let $i \leftarrow \max\{$ index in $d\}$ **if** $T[i]$ *is empty* **then**
_ **break else**
_ Let q be the coefficient of σ_i in $T[i]$ Update $d \leftarrow d - q^{-1}T[i]$ **return** d

Explanation:
The algorithm processes a filtered simplicial complex containing m simplices. For each simplex σ_j, it computes an updated boundary chain d by calling the RemovePivotRows procedure, which eliminates contributions from previously determined pivot columns. If the resulting chain d is empty, then σ_j is marked as a new basis element for the cycle group Z_k. Otherwise, the algorithm identifies the maximal index i in d as the pivot and records a persistence interval $(\deg(\sigma_i), \deg(\sigma_j))$ corresponding to the lifetime of the homological feature. In a subsequent pass, any marked simplex that does not appear as a pivot (i.e., its corresponding slot in T is empty) is assigned an infinite persistence interval. The data structure T keeps track of the pivots and the associated information, allowing the algorithm to operate efficiently by focusing on column operations alone.

References

[1] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[2] Frédéric Chazal, Vin De Silva, Marc Glisse, and Steve Oudot. *The structure and stability of persistence modules*, volume 10. Springer, 2016.

[3] Basudeb Datta. Minimal triangulations of manifolds. *arXiv preprint math/0701735*, 2007.

[4] D.S. Dummit and R.M. Foote. *Abstract Algebra*. Wiley, 2003.

[5] David Eisenbud. *Commutative algebra: with a view toward algebraic geometry*, volume 150. Springer Science & Business Media, 2013.

[6] Allen Hatcher. *Algebraic topology*. , 2005.

[7] Žiga Virk. *Introduction to Persistent Homology*. Založba UL FRI, 2022.

[8] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 347–356, 2004.