MASTER'S DEGREE THESIS

Master of Science in Advanced Mathematics and Mathematical Engineering

Title: General Persistence Theory: Towards Multiparameter Persistent Homology

Author: Manuel Lecha Sánchez

Advisor: Julian Pfeifle

Department: Mathematics

Academic year: 2021-2022

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

I wish to express my gratitude to my supervisor Julian Pfeifle for his dedication and guidance during the whole process. I would also like to thank my family and friends for their unconditional love and support.

Abstract

Topological data analysis (TDA) emerged as a field of research aiming to obtain topological descriptors of datasets. Persistent Homology was elaborated in the early 2000's as a first theoretical approach to obtain such descriptors. In particular, Persistent Homology seeks to estimate the homology of data's underlying topology. The general pipeline involves two main steps: Given a dataset $\mathcal{X}$, we

1. Build a commutative diagram $F(\mathcal{X})$ of topological spaces, also known as filtration, which we presume might contain meaningful information about the actual topology of $\mathcal{X}$,

2. Compute the n-th homology of the diagram $F(\mathcal{X})$ to algebraically analyse the resulting object $H_n \circ F(\mathcal{X})$, the persistence module.

This work aims to explore Persistence Theory in its full generality and comprehensively introduce its principal background. As a particular instance of persistence theory, we first examine one parameter persistence theory, which arises from uniparametric filtrations of spaces. Although enjoying many excellent properties, it often does not capture the structure of interest in the dataset. Therefore, we are motivated to consider multiparametric filtrations, which led us to Multiparamer Persistence Theory. Multiparameter persistence modules are more flexible and richer than the ones arising from one parameter persistence. However, as a result, they become more complex and opaque, which motivates us to seek alternative invariants.

Keywords

TDA, Topological Data Analysis, General Persistence Theory, Persistent Homology, Multiparameter Persistent Homology

Contents

Chapter 1

Introduction

Fueled by statistics and computer science advances, Machine Learning has taken off in the past few decades as a powerful tool for scientific and technological development. In this context, Topological data analysis (TDA) emerged as a field of research aiming to obtain topological descriptors of datasets. TDA uses Algebraic Topology, Statistics and Computation Theory techniques to estimate topological quantities of datasets.

The set of tools encompassed by TDA is relatively diverse; for instance, we could find the Mapper algorithm [SMC07], which generates a simplified topological graph for a given set of points. However, we will devote our entire work to Persistent Homology, considered the TDA's flagship tool, which has been applied in a large number of fields as diverse as medicine [NLC11], genomics [SRL+21], neuroscience [SEN+20], astrophysics [CkCH+22], finance [GK18] and deep learning [BGND+20].

A little motivation example to Persistent Homology

Assume we want to estimate the homology (cf. Section 3.2) of a regular submanifold $X \in \mathbb{R}^d$ from a discrete sample of it $\mathcal{X}$ (cf. Figure 1.1).

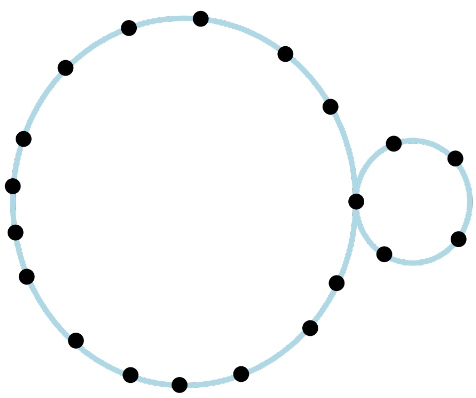


Figure 1.1: The topological space X and a discrete sample of it X.

To generate possible meaningful topological features from the point cloud, we need to provide it with some richer topology. A first approach could be to choose a parameter $\epsilon \geq 0$ and consider the ϵ-thickening of $\mathcal{X}$:

$$\mathcal{X}_\epsilon = \cup_{x\in\mathcal{X}} B(x,\epsilon),$$

where $B(x,\epsilon)$ is the ball of radius ϵ centered at x, as it is done in In Figure 1.2.

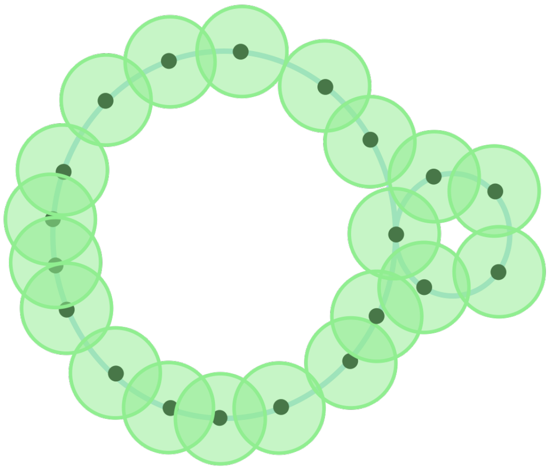


Figure 1.2: Well-chosen ϵ-thickening of $\mathcal{X}$.

In Figure 1.2 the ϵ-thickening $\mathcal{X}_\epsilon$ captures relevant information about the shape of X. In particular, it preserves the two one-dimensional holes of the space. However, this is a consequence of a well-chosen ϵ. For instance, by increasing not so much the value of ϵ, the topological space $\mathcal{X}_\epsilon$ might lose meaningful features from $\mathcal{X}$ (cf. Figure 1.3).

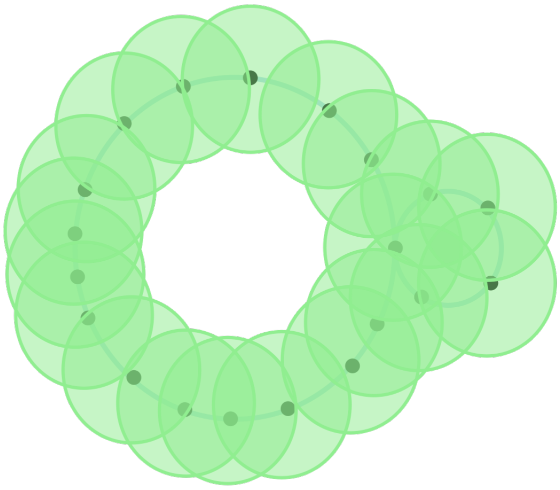


Figure 1.3: ϵ-thickening of $\mathcal{X}$ just capturing 1 of the two holes of X.

In Figure 1.4, we represent a sample of space which, at a large scale, looks like a circular shape, but at a small scale, it behaves as X. This example further illustrates the problem of tuning the ϵ parameter for recovering the topology of X, which is completely sensitive to scales.

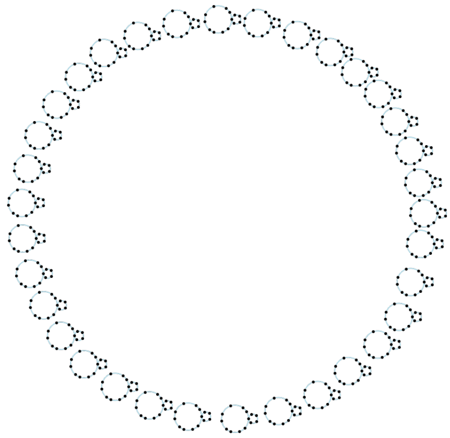


Figure 1.4: Space with different topological structures at distinct scales.

The idea of Persistent Homology arose from answering that problem with a simple quote: "Do not choose". In Persistent Homology, given a parameter dependant topological construction, as the ϵ-thickening, we encode all possible resulting topological spaces in a single mathematical object for which we can compute its homology. Informally, the pipeline of persistent homology involves two fundamental steps: Given a dataset $\mathcal{X}$, we

1. Build a commutative diagram $F(X)$ of topological spaces whose structure reveals information about $\mathcal{X}$ and the underlying space of $\mathcal{X}$. This is formalised with the notion of filtration of topological spaces, discussed in Section 3.1.
2. Compute the n-th homology of the diagram $H_n \circ F(X)$ and use topological and algebraic tools to analyse the resulting object, the persistence module. In Section 3.2 we explain why are we allowed to compute $H_n \circ F(X)$.

Structure and Key Themes

This work explores several major topics, which, as a whole, underly a significant portion of the Persistent Homology Theory:

Category Theory Category Theory is the general study of mathematical structures and their relations. This theory has progressively become the primary language of most of the advances in Persistent Homology. In Section 2.1, we introduce the fundamentals which are necessary to read the rest of our work.

Simplicial complexes In Persistent Homology, we need to represent metric spaces computationally. In Section 2.2, we introduce simplicial complexes, which are collections of higher dimensional analogues to points, lines and triangles that fit together nicely. Since they are provided with a combinatorial structure, they are suitable for being represented in computers. Finally, we explore two ways to obtain a simplicial complex from a point cloud: the Čech complex and the Rips complex.

Filtrations In Section 3.1, we introduce a crucial concept to Persistent Homology; filtrations of topological spaces. We formally present filtrations as functors from partially order sets to topological spaces, satisying some additional convenient properties. As a result, filtrations allow us to encode nested sequences of topological spaces following a partial order. Filtrations over a product of totally order sets are known as multifiltrations or multiparameter filtrations. Moreover, we introduce the most well-known type of filtration; the sub-level sets filtration. The sub-level sets filtration associated to a topological space $\mathbb{X}$, a partial order $P_{\leq}$, and a not necessarily continuous function $u : \mathbb{X} \to \mathbf{P}$ is the functor $\mathcal{S}_u : P \to \mathbf{Top}$ such that $\mathcal{S}_u(a) := \{x \in \mathbb{X} \mid u(x) \leq a\}$. As one might appreciate, this construction is fairly general. For instance, by choosing $d_{\mathcal{X}} : \mathbb{R}^n \to \mathbb{R}$ to be the distance function to $\mathcal{X}$, the sub-level set filtration $\mathcal{S}_{d_{\mathcal{X}}}$ corresponds to the ϵ-thickening filtration of our previous example.

Homology In Section 3.2 we define homology over simplicial complexes, and provide an example of a computation. Moreover, we introduce singular homology, which is defined over any topological space. More important, we prove that homology is a functor. This is the key result that allows us to compute the homology of a nested sequence of topological spaces as a whole. As a result, we obtain a persistence module.

The category of persistence modules A persistence module is a functor from a partially ordered set to the category of vector spaces. Persistence modules appear as a result of post-composing the n-th singular homology functor with a given filtration of a topological space. In Section 4.1, we delve into the category of persistence modules. It is fundamental for Persistent Homology to seek convenient and meaningful ways to represent persistence modules in computers. As a result, understanding how persistence modules decompose becomes a relevant task. Therefore, we dedicate a considerable part of the section to discussing the decomposition and classification results of persistence modules. Pointwise finite dimensional (p.f.d) persistence modules are those whose image is always a finite dimensional vector space. It turns out that every p.f.d persistence module decomposes uniquely into a direct sum of indecomposable persistence modules. We define a particular instance of persistence module, the interval persistence module, which is worth $\mathbb{k}$ ($\mathbb{k}$ is a field) only in a given interval of the partial order, and 0 otherwise. We prove that interval persistence modules are always indecomposables in the category. Finally, we introduce a correspondence between persistence modules and graded modules which Carlsson and Zoromodian first introduced in [ZC04].

Distances on the category of persistence modules Although it is crucial to understand how persistence modules decompose, it is much more critical for Machine Learning and Data Analysis tasks to have a suitable way to compare them. Therefore, metrics in persistence modules become essential to Persistent Homology. In Section 4.2 we introduce a generalised version of a distance which was first introduced by Chazal in [CCSG+09], the interleaving distance, which intends to measure how far two persistent modules are from being isomorphic. If two persistence modules are isomorphic, they are at zero interleaving distance. Later, we present a matching distance defined over families of indecomposables, the Bottleneck distance. Real data comes with noise, therefore, if we define a distance over persistence modules, we need small perturbations in our data not to significantly perturb the resulting persistence module with respect to it. Theoretical results around that concern are known as Stability results. The interleaving problem, which consists on determining whether the Bottleneck distance is equal to the Interleaving distance, is another important concern for Persistence Theory.

The theory of One Parameter Persistent Homology In Section 5.1, for the first time in our work, we delve into a particular instance of persistence theory: one parameter persistence theory. The theory of one parameter persistence is the theory associated with p.f.d persistence modules over total order categories. In particular, we associate $1D$-persistence to the theory of persistence modules over $\mathbb{R}_{\leq}$ or $\mathbb{N}_{\leq}$. First, this section presents some of the most popular $1D$-Filtrations arising from simplicial complexes: the Čech and the Rips filtrations. Later, we delve into the decomposition and parametrization of $1D$-persistence modules, which turn out to nicely decompose into interval persistence modules. As a result, we can represent every $\mathbb{R}_{\leq}$-persistence module as a unique multi-set of intervals of $\mathbb{R}_{\leq}$, called barcode. Therefore, obtaining a convenient and meaningful representation for $1D$ persistence modules in computers. After discussing the decomposition of $1D$-persistence modules, we adapt the distances introduced in Section 4.2 to the $1D$ theory. Therefore, we present the Interleaving and Bottleneck distance versions for $1D$-persistence. As a consequence to the decomposition of $1D$-persistence modules into interval persistence modules, we can algorithmically compute the Bottleneck distance via matchings of intervals of $\mathbb{R}_{\leq}$. As a significant contribution, and regarding the notion of metrics in $1D$ persistence, Lesnick proved in [Les11] the Isometry Theorem for $1D$-persistence, therefore establishing the equivalence between the Bottleneck distance and the Interleaving distance. Before this result, there was no explicit algorithm to compute the interleaving distance between two $1D$-persistence modules. Furthermore, the Isometry Theorem will help us prove our first Stability result for $1D$-persistence; perturbing data by an ϵ results in a barcode at a distance at most ϵ from the original one. Moreover, we will prove a similar result for sub-level sets filtrations. The $1D$-persistence theory enjoys many excellent properties, and it could seem that it has no drawbacks. However, this section introduces some concerns that motivate us to go beyond the one-parameter case. First, as we experimentally exemplify, it is susceptible to data points that differ significantly from the rest, namely, outliers, which commonly appear in real datasets. Second, $1D$-persistence is highly sensitive to density variations. Finally, we illustrate that $1D$-persistence might be unable to capture information that higher parameter filtrations can.

The theory of Multiparameter Persistent Homology Multiparameter Persistent Homology (cf. Section 5.2), first introduced in [CZ09], is the persistence theory associated with multifiltrations, i.e., with persistence modules indexed by the product of total order categories. In particular, we associate multiparameter persistence with persistence modules over $\mathbb{R}^d_{\leq}$. Although the $1D$-persistence theory enjoys many excellent properties, several reasons motivate us to consider multifiltrations, some already mentioned. As we will see, multiparameter persistence is a natural choice when working with:

- Datasets with outliers,
- Data with significant variations of density,
- Data that is naturally equipped with one or several real-valued functions, such as time-varying data,
- Data with tendrils.

However, these are not the only ones; many other natural examples may arise in a particular problem. We introduce models of multifiltrations that commonly appear in the multiparameter setting, some of them fairly suitable to work with the type of data mentioned above. Later we explore the decomposition of multiparameter persistence modules. MPH yields data invariants that are more flexible and richer than the ones arising from $1D$-persistence. However, as a result, in MPH, we also deal with more complex objects. Multiparameter persistence modules can be viewed as commutative diagrams of vector spaces, which have been deeply studied in quiver representation theory. This theory provides a very insightful perspective on the classification of multiparameter persistence modules; they have wild representation types. Therefore, any sort of classification or parametrization is hopeless. Later, we discuss how barcodes in the multiparameter setting do not share the same nice properties as in $1D$-persistence. Finally, we introduce three simple invariants that serve as a surrogate for the $1D$-barcode in applications and play a central role in MPH; the Hilbert function, the rank invariant, and the multigraded Betti numbers.

Chapter 2

Preliminaries

In this chapter, we first present basic concepts and definitions from category theory. Second, we introduce simplicial complexes, which provide a combinatorial way to represent topological spaces in computers. In particular, we explain different ways to obtain a simplicial complex from a given point cloud.

2.1. Basic Category Theory

2.1.1 Basic definitions on categories

Let $\mathbb{k}$ denote a field.

Definition 2.1.1. A **category** is a sextuple $\mathbf{C} = (\mathcal{O}, \mathcal{A}, \mathrm{dom}, \mathrm{cod}, \circ, \mathrm{Id})$ where

- $\mathcal{O}$ is the class of **objects** of $\mathbf{C}$,
- $\mathcal{A}$ is the class of **arrows** or **morphisms**,
- dom and cod are functions of the form $\mathcal{A} \overset{\mathrm{dom}}{\underset{\mathrm{cod}}{\rightrightarrows}} \mathcal{O}$, and for any arrow $A \xrightarrow{f} B \in \mathcal{A}$, the $\mathrm{dom}(A \xrightarrow{f} B) = A$ is the **domain**, and $\mathrm{cod}(A \xrightarrow{f} B) = B$ is the **codomain**,
- $\circ$ or **composition law** is a function $\circ(f, g)$ from the set of composable pairs
$$\mathcal{CP} = \{(f, g) \mid f, g \in \mathcal{A} \text{ and } \mathrm{dom}(f) = \mathrm{cod}(g)\}$$
into $\mathcal{A}$, denoted as $f \circ g$, such that $\mathrm{dom}(f \circ g) = \mathrm{dom}(g)$ and $\mathrm{cod}(f \circ g) = \mathrm{cod}(f)$,
- Id or **identity law** is a function $\mathcal{O} \xrightarrow{\mathrm{Id}} \mathcal{A}$ such that for any $A \in \mathcal{O}$, its identity $\mathrm{Id}(A) = \mathrm{Id}_A \in \mathcal{A}$ is the unique arrow with $\mathrm{dom}(\mathrm{Id}_A) = A = \mathrm{cod}(\mathrm{Id}_A)$ that satisfies that $(f, \mathrm{Id}_\star)$ and $(\mathrm{Id}_\star, g)$ are composable pairs, and $f \circ \mathrm{Id}_\star = f$ and $\mathrm{Id}_\star \circ g = g$;

such that the following properties hold:

1. **Associative Composition Law:** If (f, g) and (h, f) are composable pairs, then $h \circ (f \circ g) = (h \circ f) \circ g$.
2. **Smallness of morphism class:** For any pair of objects $a, b \in \mathcal{O}$, $Hom(a, b)$ is a set defined as:
$$Hom(A, B) = \{f \in \mathcal{A} \mid A \xrightarrow{f} B\}.$$

Definition 2.1.2. Let $\mathbf{C} = (\mathcal{O}, \mathcal{A}, \mathrm{dom}, \mathrm{cod}, \circ, \mathrm{Id})$ be a category,

- $\mathbf{C}$ is **small** if $\mathcal{O}$ and $\mathcal{A}$ are both sets (excluding classes).
- $\mathbf{C}$ is **finite** if $\mathcal{O}$ and $\mathcal{A}$ are *finite sets*.
- $\mathbf{C}$ is **discrete** if elements in $\mathcal{A}$ are just precisely the identity arrows.
- $\mathbf{C}$ is **connected** if for any pair of objects (A, B), $Hom(A, B) \neq \emptyset$.

Since for every object in a category $\mathbf{C}$ we have its associated identity arrow, we will omit them from diagrams.

Definition 2.1.3 (Quasi/Partial/Total order)**.** A set $\mathcal{S}$ together with a transitive and reflexive binary relation $\leq$ induces a small category (**preorder** or **quasi-order** category) with $\mathcal{O} = \mathcal{S}$ and $Hom(a, b) = a \to b$ if $a \leq b$ or $Hom(a, b) = \emptyset$ otherwise. A preorder category such that for every pair of objects the set $Hom(a, b) \cup Hom(b, a)$ contains at most one arrow, is said to be a **partial order** or **poset** category $\mathbf{P}_{\leq}$. If the set $Hom(a, b) \cup Hom(b, a)$ contains exactly one arrow, the category is a **total order** category $\mathbf{T}_{\leq}$.

Example 2.1.4 (Total order category)**.** Any ordinal number $n \in \mathbb{N}$ defines a total ordered category $\mathbf{n}$ whose object set is $\mathcal{O} = (\{0, \dots, n-1\}$ and whose morphisms are induced by the usual total order in $\mathbb{N}$.

$$\mathbf{0} : \quad \emptyset$$

$$\mathbf{1} : \quad 0$$

$$\mathbf{2} : \quad 0 \longrightarrow 1$$

$$\begin{tikzcd} \mathbf{3} : & 0 \arrow[r] \arrow[d] & 1 \arrow[dl] \\ & 2 & \end{tikzcd}$$

Figure 2.1: Ordinal number categories **0**,**1**,**2** and ,**3**.

Example 2.1.5 (Common categories)**.** Some well known examples of categories are:

- **Set:** $\mathcal{O} =$ Sets, $\mathcal{A} =$ Set functions.

- **Grp:** $\mathcal{O}$ = Groups, $\mathcal{A}$ = Group homomorphisms.
- **Top:** $\mathcal{O}$ = Topological spaces, $\mathcal{A}$ = Continuous maps.
- **R-Mod/Mod-R:** $\mathcal{O}$ = Left/Right-Modules over a ring R, $\mathcal{A}$ = Linear maps.
- **Vect(k):** $\mathcal{O}$ = Vector spaces over $\Bbbk$, $\mathcal{A}$ = Linear maps.

Special arrows

Let **C** be a category.

Definition 2.1.6. An arrow $A \xrightarrow{f} B$ is **invertible** in **C** if there exists an arrow $A \xleftarrow{f'} B$ such that $f \circ f' = \mathrm{Id}_B$ and $f' \circ f = \mathrm{Id}_A$. Such an arrow is called an **isomorphism**. Two objects A, B are **isomorphic** if there exists an isomorphism $A \xrightarrow{f} B$. In that case we say that $A \cong B$.

Definition 2.1.7. An arrow $A \xrightarrow{m} B$ is a **monic** if whenever there exist parallel arrows $D \overset{f_1}{\underset{f_2}{\rightrightarrows}} A$, then $m \circ f_1 = m \circ f_2$ implies $f_1 = f_2$.

Definition 2.1.8. An arrow $A \xrightarrow{ep} B$ is an **epi** if whenever there exist parallel arrows $B \overset{f_1}{\underset{f_2}{\rightrightarrows}} C$, then $f_1 \circ ep = f_2 \circ ep$ implies $f_1 = f_2$.

Special objects

Let **C** be a category.

Definition 2.1.9. An object A of **C** is **initial** if for every element B in **C**, there exists a *unique* arrow $A \to B$.

Definition 2.1.10. An object B of **C** is **terminal** if for every element A in **C**, there exists a *unique* arrow $A \to B$.

Definition 2.1.11. An object A of **C** is **zero** if it is initial and terminal at the same time.

Example 2.1.12 (Zero object in **Vect(k)**)**.** The zero vector space $\{0\}$ is the zero object of **Vect(k)**. First of all, it is terminal, since for every vector space V of the category, there exists a unique morphism

$$V \longrightarrow \{0\},$$

namely, the one that sends every element of V to the zero element. Moreover, it is initial since there exists a unique morphism from $\{0\}$ to every other vector space V,

$$\{0\} \longrightarrow V,$$

namely, the one that sends the unique element in $\{0\}$ to the zero vector $0 \in V$.

Definition 2.1.13. Given an object A in **C** a **subobject** of A is a pair (B, m) where $B \xrightarrow{m} A$ is a monomorphism.

Building categories from categories

Definition 2.1.14 (Subcategory)**.** Given two categories $\mathbf{C}$ and $\mathbf{C}^{'}$, we say that $\mathbf{C}^{'}$ is a subcategory of $\mathbf{C}$ if we have that

1. $\mathcal{O}^{'} \subseteq \mathcal{O}$,
2. $\mathcal{A}^{'} \subseteq \mathcal{A}$,
3. domains, codomains, composition law and identity law in $\mathbf{C}^{'}$ are just restrictions of the ones in $\mathbf{C}$ to $\mathcal{O}^{'}$ and $\mathcal{A}^{'}$.

If for every pair of objects A, B in $\mathbf{C}^{'}$ we have $Hom_{\mathbf{C}'}(A, B) = Hom_{\mathbf{C}}(A, B)$, then $\mathbf{C}^{'}$ is a **full subcategory** of $\mathbf{C}$.

Definition 2.1.15 (Opposite category)**.** Given a category $\mathbf{C} = (\mathcal{O}, \mathcal{A}, \mathrm{dom}, \mathrm{cod}, \circ, \mathrm{Id})$, its opposite category is $\mathbf{C}^{op} = (\mathcal{O}, \mathcal{A}, \mathrm{cod}, \mathrm{dom}, \star, \mathrm{Id})$, where arrow domains and codomains are switched and $f \star g = g \circ f$.

Definition 2.1.16 (Product category)**.** If $\mathbf{C}_1, \dots, \mathbf{C}_n$ are categories, then its **product category** $\mathbf{C}_1 \times \cdots \times \mathbf{C}_n$ is constructed by setting

- $\mathcal{O} = \mathcal{O}_1 \times \cdots \times \mathcal{O}_n$,
- $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$,
- $(f_1, \dots, f_n) \circ (g_1, \dots, g_n) = (f_1 \circ g_1, \dots, f_n \circ g_n)$ whenever $(f_1, g_1), \dots, (f_n, g_n)$ are composable pairs in $\mathbf{C}_1, \dots, \mathbf{C}_n$ respectively.

Observation 2.1.17. Let $\mathbf{T}^1_{\leq}, \dots, \mathbf{T}^n_{\leq}$ be total order categories, then its product category $\mathbf{T}^1_{\leq} \times \cdots \times \mathbf{T}^n_{\leq}$ is a partial order category such that there exists at most one arrow $(A_1, \dots, A_n) \xrightarrow{f} (B_1, \dots, B_n)$ whenever there exist arrows $A_1 \xrightarrow{f_1} B_1, \dots A_n \xrightarrow{f_n} B_n$ in $\mathbf{T}^1_{\leq}, \dots, \mathbf{T}^n_{\leq}$ respectively.

Example 2.1.18. Let $\mathbf{R}_{\leq}$ be total order category induced by the real numbers together with its usual total order. Then, the product category $\mathbf{R}_{\leq} \times \cdots \times \mathbf{R}_{\leq}$ is the poset category $\mathbf{R}^n_{\leq}$ induced by the usual partial order in $\mathbb{R}^n$.

2.1.2 Functors and natural transformations

In this section we are going to introduce functors and natural transformations, which respectively can be understood as morphisms between categories and arrows between them.

Definition 2.1.19. A functor $F : \mathbf{C} \to \mathbf{D}$ betweem two categories $\mathbf{C}$ and $\mathbf{D}$ consists of a pair of functions:

- An object function F that sends any $c \in \mathbf{C}$ to $F(c) \in \mathbf{D}$,
- An arrow function F that sends any morphism $C \xrightarrow{f} C'$ in $\mathbf{C}$ to $F(C \xrightarrow{f} C') \in \mathcal{A}_{\mathbf{D}}$ and satisfies two additional properties:

1. **Composition law preservation:** For any composable pair $(f, g) \in \mathbf{C}$ we have $F(f \circ g) = F(f) \circ F(g)$.
2. **Identity preservation:** For every object $C \in \mathbf{C}$ we have $F(\mathrm{Id}(C) \in \mathbf{C}) = \mathrm{Id}_{F(C)} \in \mathbf{D}$.

Example 2.1.20. For any category $\mathbf{C}$ its **identity functor** is

$$\mathrm{Id}_{\mathbf{C}} := \begin{cases} \mathrm{Id}_{\mathbf{C}}(A) = A \text{ for any object } A \in \mathbf{C}, \\ \mathrm{Id}_{\mathbf{C}}(f) = f \text{ for any morphism } f \in \mathbf{C}. \end{cases}$$

Remark 2.1.21. The composition of a composable pair of functors is a functor.

From identity preservation and composition law preservation any pair of objects $A, B \in \mathbf{C}$ satisfy that $F(Hom(A, B)) \subseteq Hom(F(A), F(B))$. We say that F is:

- **faithful** if $F|_{Hom(A,B)}$ is *injective* for any pair of objects $A, B \in \mathbf{C}$.
- **full** if $F|_{Hom(A,B)}$ is *surjective* for any pair of objects $A, B \in \mathbf{C}$.
- **fully faithful** if it is both faithful and full.

Definition 2.1.22. Let $F, G : \mathbf{C} \to \mathbf{D}$ be functors between categories $\mathbf{C}$ and $\mathbf{D}$. Then a **natural transformation** $\tau : F \to G$ assigns to each $C \in \mathbf{C}$ a morphism $\tau_C : F(C) \to G(C)$ such that for any arrow $C \xrightarrow{f} C'$ in $\mathbf{C}$, the following diagram commutes.

$$\begin{tikzcd} F(C) \arrow[r, "\tau_C"] \arrow[d, "F(f)"'] & G(C) \arrow[d, "G(f)"] \\ F(C') \arrow[r, "\tau_{C'}"'] & G(C') \end{tikzcd}$$

If for every object in $C \in \mathbf{C}$, τ_C is an invertible arrow, we say that τ is a **natural equivalence** or an **equivalence of categories**.

2.1.3 Limits and colimits

Definition 2.1.23 (I-shaped diagram in $\mathbf{C}$)**.** Let $\mathbf{I}$ be a small category usually known as the indexing category. An **I-shaped diagram** in $\mathbf{C}$is a functor $F : \mathbf{I} \to \mathbf{C}$.

$$\begin{tikzcd} \bullet \arrow[r] \arrow[d] & \bullet \arrow[d] \\ \bullet \arrow[r] & \bullet \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r] \arrow[d] & B \arrow[d] \arrow[r, shift left] \arrow[r, shift right] & E \arrow[d] \\ C \arrow[r] & D & F \end{tikzcd} \qquad \begin{tikzcd} {\color{red}A} \arrow[r] \arrow[d] & {\color{green}B} \arrow[d] \\ {\color{blue}C} \arrow[r] & {\color{brown}D} \end{tikzcd}$$

Figure 2.2: Indexing category $\mathbf{I}$, category $\mathbf{C}$, and $F(\mathbf{I})$ for a diagram functor $F : \mathbf{I} \to \mathbf{C}$.

Definition 2.1.24 (Category of $\mathbf{I}$-shaped diagrams in $\mathbf{C}$)**.** The **category of I-shaped diagrams in C** is the *functor category* whose objects are $\mathbf{I}$-shaped diagrams in $\mathbf{C}$, and whose morphisms are natural transformations between them. We denote it as $\mathbf{C}^{\mathbf{I}}$.

Definition 2.1.25 (Constant diagram functor)**.** Given some object $X \in \mathbf{C}$, the **constant diagram functor X** is $X \in \mathcal{O}(\mathbf{C}^{\mathbf{I}})$ such that

$$X := \begin{cases} X(A) = X \text{ for every } A \in \mathcal{O}(I), \\ X(f) = Id_X \text{ for every } f \in \mathcal{A}(I). \end{cases}$$

Definition 2.1.26 (Cone/Cocone)**.** A **cone** in $\mathbf{C}^{\mathbf{I}}$ is a morphism $c : X \to F$ where X is a constant diagram functor. Dually, a **cocone** is a morphism from $F \in \mathbf{C}^{\mathbf{I}}$ to a constant functor X.

$$\begin{tikzcd}
X \arrow[dd, "\tau_C"'] \arrow[dr, "\tau_A"] \arrow[drr, "\tau_B"] \arrow[ddr, "\tau_D"] & & \\
& A \arrow[r] \arrow[dl] & B \arrow[dl] \\
C \arrow[r] & D &
\end{tikzcd}
\qquad
\begin{tikzcd}
& A \arrow[r] \arrow[dl] \arrow[ddr, "\tau_A"] & B \arrow[dl] \arrow[dd, "\tau_B"] \\
C \arrow[r] \arrow[drr, "\tau_C"'] & D \arrow[dr, "\tau_D"] & \\
& & X
\end{tikzcd}$$

Figure 2.3: Cone (left) and cocone (right) commuting diagrams example.

Definition 2.1.27 (Limit/Colimit)**.** The **limit** of an **I**-shaped diagram $F \in \mathbf{C}^{I}$ (if exists) is the universal cone to F. That is, a cone $\eta : \lim F \to F$ such that

for any other cone $\tau : X \to F$, there exist a unique arrow $X \xrightarrow{f} \lim F$ such that $\tau = \eta \circ f$. Dually, the **colimit** of an **I**-shaped diagram (if exists) is its universal cocone.

$$\begin{tikzcd}
X \arrow[d, dashed, "f"] \arrow[dd, bend right=60, "\tau"'] \\
limF \arrow[d, "\eta"] \\
F
\end{tikzcd}
\qquad
\begin{tikzcd}
X \\
colimF \arrow[u, dashed, "f"'] \\
F \arrow[u, "\eta"'] \arrow[uu, bend right=60, "\tau"']
\end{tikzcd}$$

Figure 2.4: Limit/Colimit diagram visualization.

2.1.4 Limits and colimits: Convenient examples

Products and coproducts

Definition 2.1.28 (Product/Coproduct)**.** Given a category **C** and a discrete indexing category **I**, the limit/colimit of an **I**-shaped diagram functor $F : \mathbf{I} \to \mathbf{C}$ is called **product**/**coproduct**.

$$
\mathbf{I} : \bullet \quad \bullet
\qquad
\begin{tikzcd}
 & H \arrow[d, dashed, "h"] \arrow[ddl, bend right, "h_1"'] \arrow[ddr, bend left, "h_2"] & \\
 & A \times B \arrow[dl, "\pi_A"'] \arrow[dr, "\pi_B"] & \\
A & & B
\end{tikzcd}
\qquad
\begin{tikzcd}
 & H & \\
 & A \coprod B \arrow[u, dashed, "h"'] & \\
A \arrow[ur, "\iota_A"] \arrow[uur, bend left, "h_1"] & & B \arrow[ul, "\iota_B"'] \arrow[uul, bend right, "h_2"']
\end{tikzcd}
$$

Figure 2.5: Examples of product/coproduct diagrams from a two object discrete indexing category.

Example 2.1.29 (Poset product/coproduct)**.** Given a poset $\mathbf{P}_{\leq}$, the product of n of its elements, $p_1, \dots, p_n$, is precisely the **greatest lower bound** of them all. To show this, we focus on the case of the product of two elements a and b in $\mathbf{P}_{\leq}$. By definition, the product $a \times b$ must satisfy $a \times b \leq a$ and $a \times b \leq b$. Moreover, the universal property of the product must be satisfied, or in this case, if there exists a $c \leq a$ and $c \leq b$, then $c \leq a \times b$.

$$
\begin{tikzcd}
 & c \arrow[d, dashed, "\leq"] \arrow[ddl, bend right, "\leq"'] \arrow[ddr, bend left, "\leq"] & \\
 & a \times b \arrow[dl, "\leq"'] \arrow[dr, "\leq"] & \\
a & & b
\end{tikzcd}
$$

By the duality product-coproduct, one then can easily check that the coproduct of n elements in a poset is the **lowest upper bound** of that collection of elements.

$$
\begin{tikzcd}
 & c & \\
 & a \coprod b \arrow[u, dashed, "\leq"'] & \\
a \arrow[ur, "\leq"] \arrow[uur, bend left, "\leq"] & & b \arrow[ul, "\leq"'] \arrow[uul, bend right, "\leq"']
\end{tikzcd}
$$

Equalizers and Coequalizers

Again, here we present two mutually dual constructions which can be defined by means of the notion of limit and colimit.

In particular, when possible, those constructions will allow us to define the notion of kernel and co-kernels.

Definition 2.1.30 (Equalizer/Coequalizer)**.** Given a category $\mathbf{C}$ and the indexing category $\mathbf{I}$

$$\mathbf{I} : \bullet \rightrightarrows \bullet$$

the limit/colimit o an $\mathbf{I}$-shaped diagram functor $F : \mathbf{I} \to \mathbf{C}$ is called **equalizer**/**coequalizer**.

Let us unwrap this definition.

Definition 2.1.31 (Explicit version)**.** Given a category **C** and two objects A and B together with two parallel morphisms f and g such that

$$A \underset{g}{\overset{f}{\rightrightarrows}} B$$

an **equalizer** consists of an object $EQ(f,g)$ and a morphism $eq(f,g)$ satisfying that $f \circ eq(f,g) = g \circ eq(f,g)$, the latter being universal.

$$\begin{tikzcd} EQ(f,g) \arrow[r, "eq(f,g)"'] & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \\ C \arrow[u, dashed, "s"] \arrow[ur, bend right, "m"'] & & \end{tikzcd}$$

The **coequalizer** is the dual notion, therefore the object $COEQ(f,g)$ and the morphism $coeq(f,g)$ satisfying that $coeq(f,g) \circ f = coeq(f,g) \circ g$, the latter being universal.

$$\begin{tikzcd} A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \arrow[r, "coeq(f,g)"] \arrow[dr, bend right, "m"'] & COEQ(f,g) \arrow[d, dashed, "s"] \\ & & C \end{tikzcd}$$

Equalizers and coequalizers allow us to define the notion of kernel and cokernel in a categorical sense. To define kernels/cokernels we need zero morphisms in our category.

Definition 2.1.32. Given a category **C** and a morphism $f : A \to B$, the **kernel**/**cokernel** of f is the equalizer/coequalizer of f and the 0 morphism.

$$A \underset{0}{\overset{f}{\rightrightarrows}} B$$

In general we think of the kernel as a subobject of A and the cokernel as a subobject of B.

Example 2.1.33. Let *Forg* be the forgetful functor mapping **Vect(k)** to **Set**. In the category **Vect(k)**

- every morphism has a kernel. The kernel of an arrow $f : V_1 \to V_2$ is the set-theoretic preimage $Forg(f)^{-1}(0)$ with the vector space structure. In other words, the usual vector space kernel.
- every morphism has a cokernel. The cokernel of an arrow $f : V_1 \to V_2$ is the quotient of V_2 by the image $Im(f)$.

2.1.5 Abelian categories

Definition 2.1.34 (Ab-category)**.** An **Ab-category** has the property that for every pair of objects A, B, the hom-set $Hom(A,B)$ has the structure of an abelian group with bilinear composition:

$$
f \circ (g + h) = (f \circ g) + (f \circ h),
$$
$$
(f + g) \circ h = (f \circ h) + (g \circ h).
$$

Observe that in an Ab-category all equalizers are kernels.

Definition 2.1.35 (Biproduct)**.** Let **C** be a category with zero morphisms.

$$
\begin{tikzcd}
A_1 \arrow[r, shift right, "i_1"'] & A_1 \oplus \cdots \oplus A_n \arrow[l, shift right, "p_1"'] \arrow[r, shift left, "p_n"] \arrow[d, shift left] & A_n \arrow[l, shift left, "i_n"] \\
& \cdots \arrow[u, shift left] &
\end{tikzcd}
$$

such that,

$$
p_k i_k = id_{A_k}, \text{ and } p_l \circ i_k = 0 \text{ for } k \neq l,
$$

and such that

$$
(A_1 \oplus \cdots \oplus A_n, p_k) \text{ and } (A_1 \oplus \cdots \oplus A_n, i_k)
$$

are respectively product and coproduct for the $A_1, \dots, A_k$.

Definition 2.1.36 (Additive category)**.** An **additive category** is an Ab-category that has binary biproducts for every finite set of objects.

Definition 2.1.37 (Endomorphism ring)**.** Given an object A of an additive category, the **endomorphism ring** $\mathrm{End}(A)$ of A is the set $\mathrm{Hom}(A, A)$ together with the addition and composition of morphisms as addition and multiplication operations, respectively.

An endomorphism ring $\mathrm{End}(A)$ is **local** if for every morphism $f \in \mathrm{End}(A)$, either f or $1 - f$ is a unit.

Definition 2.1.38 (Abelian category)**.** An **abelian category C** is an additive category in which

1. A null object exists,
2. Binary biproducts for every finite set of objects exist,
3. Every morphism has a kernel and a cokernel,
4. Every monic arrow is a kernel, and every epi is a cokernel.

The category **Vect(k)** is a classical example of an abelian category.

Definition 2.1.39 (bicomplete)**.** A category is **bicomplete** if binary biproducts exist for every set of objects.

Definition 2.1.40 (AB2-category)**.** An **AB2-category** is a bicomplete abelian category .

Vect(k) and $\mathbb{R}$ are examples of AB2-categories.

2.2. Simplicial complexes

We introduce Simplicial complexes, which allow us to represent metric spaces combinatorially. Therefore, being suitable for computers.

2.2.1 Basic notions

Let $\mathcal{X} = \{x_0, ..., x_k\}$ be a set of points in $\mathbb{R}^d$, then the convex hull $\text{conv}(\mathcal{X})$ is the set of all its possible convex combinations

$$\text{conv}(\mathcal{X}) = \{\sum_{i=0}^{k} \lambda_i x_i \mid \sum_{i=0}^{k} \lambda_i = 1, \lambda_i \geq 0\,, \forall i\} \tag{2.1}$$

If $\mathcal{X}$ is a set of affinely independent points, then $\Delta = \text{conv}(\mathcal{X})$ is a **geometric k-simplex**. We introduce the following terminology related to a geometric simplex Δ:

- **Dimension:** $\dim(\Delta) = k$.
- **Vertices of Δ:** $x_0, \ldots, x_k$.
- **Edges of Δ:** $\text{conv}(\{x_i, x_j\})$ for $i \neq j$.
- If $\tau = \text{conv}(\mathcal{Y})$ for $\mathcal{Y} \subset \mathcal{X}$, we say that τ is a **face** of Δ. If the dimension of τ is $\dim(\tau) = k - 1$, then τ is a **facet**.

Definition 2.2.1. A **geometric simplicial complex** $K \subset \mathbb{R}^d$ is a set of simplices that satisfies:

- If $\sigma \in K$ and τ is a face of σ, then $\tau \in K$.
- If $\sigma, \tau \in K$, then $\sigma \cap \tau$ is either empty or a face of both.

For computational reasons in later sections it will be very convenient to deal with objects provided with a certain combinatorial structure. Therefore, we will be dealing with the purely combinatorial counterpart to the geometric simplicial complex, known as an abstract simplicial complex.

Definition 2.2.2. An **abstract simplicial complex** is a pair $\mathcal{K} = (V, \Sigma)$, where V is a finite set of vertices, and Σ is a collection of subsets of Σ such that for all element $\sigma \in \Sigma$, $\tau \subseteq \sigma$ implies $\tau \in \Sigma$.

A few accompanying definitions to Definition 2.2.2:

- An element σ of Σ is an **abstract simplex** of $\mathcal{K}$. It is a **k-simplex** if $|\sigma| = k + 1$.
- If $\tau \subseteq \sigma \in \mathcal{K}$, then τ is a face of σ. If the dimension τ is $\dim(\tau) = \dim(\sigma) - 1$, then it is a **facet** of σ.
- The **dimension** $\dim(\mathcal{K})$ of $\mathcal{K}$ is the maximal dimension of a simplex in $\mathcal{K}$.

- An abstract simplicial complex $\mathcal{L}$ is a **subcomplex** of $\mathcal{K}$ if $\mathcal{L} \subseteq \mathcal{K}$.
- For $0 \leq n \leq \dim(\mathcal{K})$ the n-**skeleton** Σ^n of $\mathcal{K}$ is the abstract simplicial subcomplex of $\mathcal{K}$ consisting of all abstract simplices of $\mathcal{K}$ with dimension at most n.

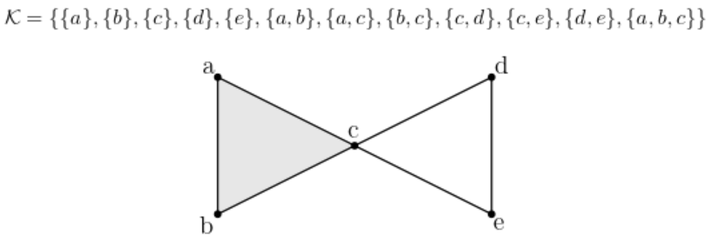


Figure 2.6: Example of an abstract simplicial complex.

It is easy to turn a geometric simplicial complex into an abstract simplicial complex: replace coordinates by unique labels. To turn an abstract simplicial complex into a geometric complex it is necessary to choose coordinates of vectors that satisfy the requirements for a geometric simplicial complex cf. Definition 2.2.1. This, can be always done:

Theorem 2.2.3. *Every abstract simplicial complex $\mathcal{K}$ with n vertices admits a* **geometric realization** *in $\mathbb{R}^{n-1}$.*

Proof. Let $\mathcal{L}$ be the full abstract simplicial complex with n vertices. Then, $\mathcal{K}$ is an abstract simplicial subcomplex of L. Since L admits a realization in $\mathbb{R}^{n-1}$, then $\mathcal{K}$ does to. $\square$

Note that we have proven the existence of geometric realizations. Realizations in lower-dimensional spaces may exist.

Abstract simplicial complexes provide a combinatorial description for topological spaces. Therefore, simplicial maps are the combinatorial analogue to continuous functions.

Definition 2.2.4 (Simplicial map)**.** Let $\mathcal{K} = (V_{\mathcal{K}}, \Sigma_{\mathcal{K}})$ and $\mathcal{L} = (V_{\mathcal{L}}, \Sigma_{\mathcal{L}})$ two abstract simplicial complexes. A **simplicial map** is an assignment $f : V_{\mathcal{K}} \to V_{\mathcal{L}}$ on the vertices of $\mathcal{K}$ and $\mathcal{L}$ such that for every $\sigma = \{v_0, \ldots, v_k\} \in \Sigma_{\mathcal{K}}$, its image $f(\sigma) = \{f(v_0), \ldots, f(v_k)\} \in \Sigma_{\mathcal{L}}$.

Example 2.2.5. Let $\mathcal{K}$ be an abstract simplicial complex as the one in Figure **??**. The map $f : \mathcal{K} \to \mathcal{K}$ such that $a \mapsto a$; $b \mapsto b$; $c \mapsto b$ is a simplicial map.

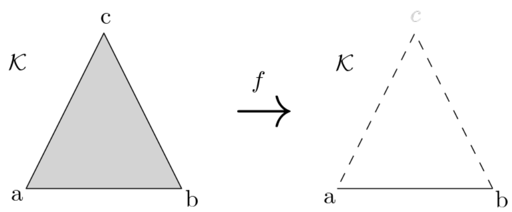


Figure 2.7: Simplicial map.

In Section 3.2.2 the following notion is going to be strictly necessary:

Definition 2.2.6 (Oriented simplex)**.** An **oriented** k-simplex $\sigma = [v_0, \dots, v_k]$ is a k-simplex with a fixed order in its vertices such that for any permutation $\pi : \{0, \dots, k\} \to \{0, \dots, k\}$ we have the following identification:

$$\sigma = (-1)^{\text{sgn}(\pi)}[v_0, \dots, v_k],$$

where sgn is the signature of the permutation, i.e., its value is 0 if the permutation is obtainable from an even number of two-element swaps, and 1 otherwise.

Oriented simplices induce orientations in their facets.

Definition 2.2.7 (Induced orientation on facets)**.** Let $\sigma = [v_0, \dots, v_k]$ be an oriented simplex. The **induced orientation** of the facet of σ obtained by suppressing the vertex v_i is

$$(-1)^i[v_0, ..., \hat{v}_i, ..., v_k].$$

Definition 2.2.8 (Oriented simplicial complex)**.** An abstract simplicial complex is **oriented** if each of its abstract simplices is oriented.

Giving an orientation to an abstract simplicial complex $\mathcal{K}$ is picking up an orientation to each of its simplices. An usual way for doing so, is to define a total order on the vertices of $\mathcal{K}$ and use its induced orientations.

2.2.2 Simplicial complexes from data

Let $\mathcal{X}$ be a finite set of points lying in a metric space $(\mathbb{M}, d)$, i.e., a set together with a notion of distance. In this section, we briefly introduce different ways to obtain a simplicial complex from $\mathcal{X}$.

The Čech complex

Definition 2.2.9 (Covering of X)**.** A **covering** $\mathcal{U} = \{U_i\}_{i \in I}$ of $\mathcal{X}$ is a family of subsets of $(\mathbb{M}, d)$ indexed by some I such that $\mathcal{X} \subseteq \bigcup_{i \in I} U_i$.

Definition 2.2.10. The **nerve** $\mathcal{N}(\mathcal{U})$ of the covering $\mathcal{U}$ is the abstract simplicial complex on I in which $\{i_0, ... i_n\}$ spans an n-simplex if and only if $\bigcap_{i \in \{i_0, ... i_n\}} U_i \neq \emptyset$.

In general, the nerve complex $\mathcal{N}(\mathcal{U})$ of some covering $\mathcal{U} = \{U_i\}_{i \in I}$ does not necessarily reflect any information about the topology. However, it does under certain conditions.

Definition 2.2.11. Let X, Y be topological spaces, and $I = [0, 1]$. A map

$$F : X \times I \to Y$$

is called a **homotopy**. If $f_t(x) = F(x, t)$, then F is called a homotopy from f_0 to f_1. Two maps f, g are **homotopic**, written $f \simeq g$ if there exists a homotopy F such that $f = f_0$ and $g = f_1$.

Definition 2.2.12. Let X, Y be topological spaces. We say that X is **homotopy equivalent** to Y (written $X \simeq_h Y$), if there are maps

$$f : X \to Y, \;\; g : Y \to X,$$

such that

$$g \circ f \simeq \text{Id}_X, f \circ g \simeq \text{Id}_Y.$$

There exist many versions of the following theorem, that provide different criteria for guaranteeing that the nerve preserves some topological quantity of the covering. However, we consider it convenient to present the following one:

Theorem 2.2.13 (Nerve Theorem)**.** *If $\mathcal{U}$ is a finite, closed, convex cover of $\mathcal{X}$, then $\mathcal{X} \simeq \mathcal{N}(\mathcal{U})$ ($\simeq$ as homotopy equivalence).*

One of the most common ways to obtain a simplicial complex from a point cloud relies on the notion of the nerve.

Definition 2.2.14 (The Čech complex)**.** Let $\mathcal{U}_\epsilon = B_\epsilon(x)_{x\in\mathcal{X}}$ be the collection of balls of radius $\epsilon \geq 0$ centered at the points of $\mathcal{X}$. Then, $\mathcal{U}_\epsilon$ is a covering of $\mathcal{X}$. The **Čech complex** subject to $\mathcal{X}$ and ϵ is $\check{\mathrm{C}}\mathrm{ech}(\mathcal{X}, \epsilon) := \mathcal{N}(\mathcal{U}_\epsilon)$.

The algorithm to compute a Čech complex from a given $\mathcal{X}$ has complexity $\mathcal{O}(nd)$ (see [DI12]). Therefore, sometimes it might be convenient to rely in the following simpler construction.

Vietoris-Rips complex

Definition 2.2.15 (Diameter of $\mathcal{X}$)**.** The diameter of $\mathcal{X}$ is

$$\mathrm{diam}(\mathcal{X}) = \sup_{x,y\in\mathcal{X}} \mathrm{d}(x, y).$$

Definition 2.2.16. The **Vietoris-Rips** complex $\mathrm{Rips}(\mathcal{X}, \epsilon)$ is the abstract simplicial complex such that:

1. The vertex set is $\mathcal{X}$.
2. A subset $\sigma \subseteq \mathcal{X}$ is a simplex if and only if $\mathrm{diam}(\sigma) \leq \epsilon$

Observation 2.2.17. A subset $\mathcal{Y} \subseteq \mathcal{X}$ has diameter $\mathrm{diam}(Y) \leq \epsilon$ if and only if for every pair of vertices $x_i, x_j \in \mathcal{Y}$, the distance $d(x_i, x_j) \leq \epsilon$. Therefore, the Rips complex is completely determined by its 1-skeleton, as opposite from the Čech complex, which is not.

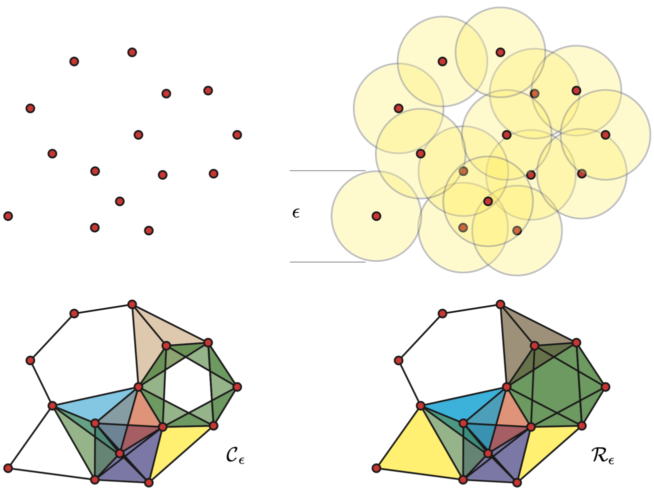


Figure 2.8: Čech complex reconstruction (lower -left) and Vietoris Rips reconstruction (lower-right) over a set of points for a certain ϵ parameter. Extracted from [Ghr07].

Chapter 3

Filtrations and Homology: The two main ingredients

We introduce filtrations as the functorial way to understand a nested sequence of topological spaces indexed by an arbitrary poset category. Filtrations are among the philosophical roots of TDA, which aims to estimate topological quantities from data. Next, we introduce homology of topological spaces, a classical invariant which informally quantifies the number of n-dimensional holes of our space. First, we will present simplicial homology, the most naive but practical version. Later, we will introduce singular homology, defined over any topological space. Moreover, we will briefly explain why homology is, in fact, a functor. Postcomposing a filtration functor with the homology functor will take us directly to persistent homology theory.

3.1. Filtrations

To get an intuition, observe that we can think of $\mathbb{N}$ as a directed graph or diagram where the nodes are the natural numbers, and the edges are directed arrows linking each number with its successor. Thus, allowing us to take a path from one number to any of its successors. The latter corresponds to the usual total order category of the natural numbers, denoted as $\mathbb{N}_{\leq}$.

$$0 \xrightarrow{a_{01}} 1 \xrightarrow{a_{12}} 2 \xrightarrow{a_{23}} 3 \xrightarrow{a_{34}} \dots$$

Given a simplicial complex K, a $\mathbb{N}_{\leq}$-simplicial complex filtration of K, is a nested sequence of simplicial complexes indexed by the natural numbers $\mathbb{N}$ such that $K_i \subseteq K_j$ whenever $i \leq j$, together with inclusion maps $\iota_{i(i+1)}$.

$$K_0 \xhookrightarrow{\iota_{01}} K_1 \xhookrightarrow{\iota_{12}} K_2 \xhookrightarrow{\iota_{23}} K_3 \xhookrightarrow{\iota_{34}} \dots$$

If instead of dealing with just one copy of $\mathbb{N}_{\leq}$, we consider the product $\mathbb{N}_{\leq} \times \mathbb{N}_{\leq}$ of two copies of $\mathbb{N}_{\leq}$,

$$\begin{tikzcd}
\vdots & \vdots & \vdots & \\
(2,0) \arrow[r] \arrow[u] & (2,1) \arrow[r] \arrow[u] & (2,2) \arrow[r] \arrow[u] & \dots \\
(1,0) \arrow[r] \arrow[u] & (2,1) \arrow[r] \arrow[u] & (2,2) \arrow[r] \arrow[u] & \dots \\
(0,0) \arrow[r] \arrow[u] & (0,1) \arrow[r] \arrow[u] & (0,2) \arrow[r] \arrow[u] & \dots
\end{tikzcd}$$

we will be dealing with a $\mathbb{N}_{\leq} \times \mathbb{N}_{\leq}$-simplicial complex **bifiltration** of K. Recall from 2.1.17 that $\mathbb{N}_{\leq} \times \mathbb{N}_{\leq}$ is a partial order.

In this particular case, our bifiltration is a grid of simplicial complexes indexed by the biproduct of the natural numbers $\mathbb{N} \times \mathbb{N}$ satisfying the same principle as our filtration above, $K_i \subseteq K_j$ whenever $i \leq j$ in the induced partial order.

$$\begin{tikzcd}
\vdots & \vdots & \vdots & \\
K_{(2,0)} \arrow[r] \arrow[u] & K_{(2,1)} \arrow[r] \arrow[u] & K_{(2,2)} \arrow[r] \arrow[u] & \dots \\
K_{(1,0)} \arrow[r] \arrow[u] & K_{(2,1)} \arrow[r] \arrow[u] & K_{(2,2)} \arrow[r] \arrow[u] & \dots \\
K_{(0,0)} \arrow[r] \arrow[u] & K_{(0,1)} \arrow[r] \arrow[u] & K_{(0,2)} \arrow[r] \arrow[u] & \dots
\end{tikzcd}$$

More generally, we introduce the following definition of a filtration:

Definition 3.1.1 (Filtration)**.** Let $\mathbf{P}_{\leq}$ be some poset category and let $\mathbf{C}$ be a category. Moreover, let X be an object of $\mathbf{C}$. A **filtration** is a functor $F_X : \mathbf{P}_{\leq} \to \mathbf{C}$ such that:

- $F_X(a) := X_a$ where X_a is a subobject of X,
- X_a is a subobject of X_b whenever $a \to b$ in $\mathbf{P}_{\leq}$,
- $F_X(a \to b)$ is the morphism $X_a \hookrightarrow X_b$ where $\hookrightarrow$ is a monomorphism between subobjects X_a and X_b.

From now, we consider topological filtrations $F : \mathbf{P}_{\leq} \to \mathbf{Top}$, sometimes restricting to simplicial complex filtrations.

From 2.1.17 we know that given total order categories $\mathbf{T}^1_{\leq}, \dots, \mathbf{T}^n_{\leq}$ the product category $\mathbf{T}^1_{\leq} \times \dots \times \mathbf{T}^n_{\leq}$ is a partial order category.

Definition 3.1.2 (n-parameter filtration)**.** An $\mathbf{T}^1_{\leq} \times \dots \times \mathbf{T}^n_{\leq}$-filtration is an n-**parameter filtration**. For $n \geq 2$ they are usually referred to as **multifiltrations**.

Filtrations over $\mathbb{N}_{\leq}, \mathbb{Z}_{\leq}$ or $\mathbb{R}_{\leq}$ are examples of 1-parameter filtrations. Filtrations over $\mathbb{N}^n_{\leq}, \mathbb{Z}^n_{\leq}, \mathbb{R}^n_{\leq}$ for $n \geq 2$ are examples of multifiltrations.

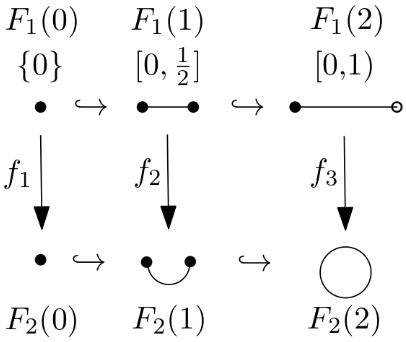
Observation 3.1.3. $\mathbf{P}_{\leq}$-filtrations on **Top** conform the functor category $\mathbf{Filt}_{\mathbf{P}_{\leq}}$. Morphisms in $\mathbf{Filt}_{\mathbf{P}_{\leq}}$, i.e., natural transformations, are of the following form: Given two elements F_1 and F_2 in $\mathbf{Filt}_{\mathbf{P}_{\leq}}$, a morphism $f : F_1 \to F_2$ is induced by a collection of arrows f_a in **Top** for every $a \in \mathbf{P}_{\leq}$ making the following diagram commute whenever $a \leq b$ in $\mathbf{P}_{\leq}$:

$$\begin{tikzcd} F_1(a) \arrow[r, hook, "F_1(a\to b)"] \arrow[d, "f_a"'] & F_1(b) \arrow[d, "f_b"] \\ F_2(a) \arrow[r, hook, "F_2(a\to b)"] & F_2(b) \end{tikzcd}$$

Example 3.1.4. Given the following poset

$$\begin{tikzcd} P_{\leq} \ := \ 0\bullet \arrow[r] \arrow[rr, bend right] & 1\bullet \arrow[r] & 2\bullet \end{tikzcd}$$

setting $f_1(t) = f_2(t) = f_3(t) := e^{2\pi i t}$ the collection $f(a) = f_a$ for $a \in P_{\leq}$ induces a morphism between the following F_1 and F_2 filtrations.



3.1.1 Sub-level set filtration

One of the main or canonical constructions for filtrations of topological spaces is the following:

Definition 3.1.5 (Sub-level set filtration)**.** The **sub-level sets filtration** associated to a topological space $\mathbb{X}$, a poset $\mathbf{P}$, and a not necessarily continuous function $u : \mathbb{X} \to \mathbf{P}$ is the functor $\mathcal{S}_u : P \to \mathbf{Top}$ such that

$$S_u := \begin{cases} \mathcal{S}_u(a) := \{x \in \mathbb{X} \mid u(x) \leq a\}, \\ \mathcal{S}_u(a \leq b) := \mathcal{S}_u(a) \hookrightarrow \mathcal{S}_u(b), \text{ where } \hookrightarrow \text{ represents the inclusion morphism.} \end{cases}$$

Example 3.1.6. If $\mathbf{P} = \mathbb{R}_{\leq}$, the sub-level set filtration for a given function $u : \mathbb{X} \to \mathbb{R}$ is

$$S_u := \begin{cases} \mathcal{S}_u(a) = u^{-1}((-\infty, a)), \\ \mathcal{S}_u(a \leq b) := u^{-1}((-\infty, a)) \hookrightarrow u^{-1}((-\infty, b)). \end{cases}$$

Sub-level set filtrations are functors from a poset to the category **Top**. From a practical perspective, we will be mostly interested in filtrations over simplicial complexes.

3.2. Simplicial and Singular Homology

This section briefly recalls some of the basics of simplicial and singular homology. First, we introduce the generic setting for defining homology. Second, we provide the reader with a gentle, informal, intuitive introduction to simplicial homology. Moreover, we briefly introduce singular homology and discuss its functoriality. This fact allows us to compute the homology of a topological space filtration, which leads to persistent homology theory.

3.2.1 Chain complexes and Homology

Let **C** be an abelian category.

Definition 3.2.1 ($\mathbb{Z}$-graded chain complex)**.** A $\mathbb{Z}$**-graded chain complex** in **C** is

- A collection of objects $\{C_n\}_{n\in\mathbb{Z}}$,
- and of morphisms $\partial_n : C_n \to C_{n-1}$

$$\dots \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \dots C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} \dots$$

  such that

$$\partial_n \circ \partial_{n+1} = 0$$

  for all $n \in \mathbb{Z}$.

In a chain complex, the morphisms ∂_n are called **boundary maps** or **differentials**, and the elements of C_n **n-chains**.

Definition 3.2.2 (chain map)**.** Given $A_\bullet, B_\bullet$ chain complexes in **C**, a **chain map** is a collection of morphisms in $\{f_n : A_n \to B_n\}_{n\in\mathbb{Z}}$ in **C** such that $f_n \circ \partial_n^A = \partial_n^B \circ f_{n+1}$.

Chain complexes with maps between them form the category of chain complexes $\mathbf{Ch}(\mathbf{C})_\bullet$.

Definition 3.2.3 (Cycles/Boundaries/Homology)**.** Given $C_\bullet \in \mathbf{Ch}(\mathbf{C})_\bullet$

- elements in

$$Z_n := \mathrm{Ker}(\partial_n : C_n \to C_{n-1})$$

  are called **n-cycles**,
- elements in

$$B_n := \mathrm{Img}(\partial_{n+1} : C_{n+1} \to C_n)$$

  are called **n-boundaries**. Since $\partial_n \circ \partial_{n+1} = 0$ we have

$$0 \subseteq B_n \subseteq Z_n \subseteq C_n$$

- the **cokernel**

$$H_n := Z_n / B_n$$

  by the inclusion is called the n-th chain homology of $C_\bullet$.

Definition 3.2.4 (Exact/Short exact sequence)**.** An **exact sequence** in **C** is a chain complex $C_\bullet \in \mathbf{Ch}(\mathbf{C})_\bullet$ such that for each $n \in \mathbb{N}$, $H_n(C_\bullet) = 0$. We say that a it is **short exact** if is an exact sequence of the following form:

$$\dots 0 \to A \to B \to C \to 0 \to \dots$$

A functor between abelian categories is **exact** if it sends short exact sequences to short exact sequences.

3.2.2 A gentle introduction to simplicial homology

We introduce in a very gentle manner the basic notions regarding simplicial homology, a topological invariant defined over **oriented** simplicial complexes cf. Section 2.2.8. For doing so, we will rely on an example, allowing us to introduce the necessary concepts. As we will see in the next section 3.2.3, more refined versions of homology exist. Nevertheless, for us it will be sufficient to grasp the basic ideas of simplicial homology.

Example 3.2.5. Suppose we want to compute the simplicial homology of the boundary T of a triangle, which is topologically equivalent to S^1.

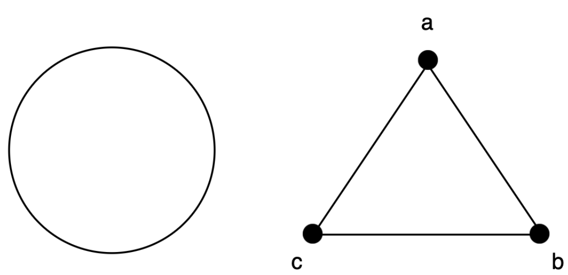


Figure 3.1: S^1 and an equivalent simplicial complex.

The triangle is a 1-dimensional oriented abstract simplicial complex

$$\mathcal{T} := (V = \{a, b, c\}, \Sigma = \{[a], [b], [c], [a, b], [b, c], [a, c]\}),$$

which has

$$\begin{aligned}\Sigma_0 &= \{[a], [b], [c]\},\\ \Sigma_1 &= \{[a, b], [b, c], [a, c]\},\\ \Sigma_k &= \emptyset \text{ for } k \geq 2\end{aligned}$$

as its n-dimensional skeleta.

Recall from Definition 3.2.3 that homology is defined over chain complexes of an abelian category. Therefore, we have to define chain complexes from oriented simplicial complexes. We consider the n distinct $\mathbb{k}$-vector spaces with bases, the elements of the n-th skeleton of our simplicial complex. We denote them as $C_n(\mathcal{K}, \mathbb{k})$.

In example 3.2.5 we have

- $C_2(\mathcal{T}, \Bbbk) = \{0\}$,
- $C_1(\mathcal{T}, \Bbbk) = \Bbbk[a, b] \oplus \Bbbk k[b, c] \oplus \Bbbk[a, c] \cong \Bbbk^3$,
- $C_0(\mathcal{T}, \Bbbk) = \Bbbk[a] \oplus \Bbbk[b] \oplus \Bbbk[c] \cong \Bbbk^3$.

Definition 3.2.6 (Boundary of a simplex)**.** The **boundary** of a simplex $\sigma \in \mathcal{K}$ is the collection of all of its facets .

Definition 3.2.7 (n-boundary map)**.** The **n-boundary map** $\partial_n : C_n(\mathcal{K}, \Bbbk) \to C_{n-1}(\mathcal{K}, \Bbbk)$ is the linear map defined over the elements σ of the basis of $C_n(\mathcal{K}, \Bbbk)$ such that its image $\partial_n\sigma$ is the sum of the facets of σ equipped with the induced order (cf. Definition 2.2.7) from σ, i.e.,

$$\partial_n([v_0, .., v_n]) := \sum_{i=0}^{n} (-1)^i [v_0, ..., \hat{v}_i, ..., v_n].$$

The boundary map satisfies the following property, which intuitively says that a simplex cannot be boundary of any other simplex living two dimensions above.

Proposition 3.2.8. $\partial_{n-1} \circ \partial_n = 0$

Proof.

$$\begin{aligned}
&\partial_{n-1} \circ \partial_n([v_0, .., v_n]) = \\
&\sum_{i=0}^{n} (-1)^i \partial_{n-1}([v_0, ..., \hat{v}_i, ..., v_n]) = \\
&\sum_{j<i} (-1)^i (-1)^j [v_0, ..., \hat{v}_j, ..., \hat{v}_i, ..., v_n] + \\
&\sum_{j>i} (-1)^i (-1)^{j-1} [v_0, ..., \hat{v}_i, ..., \hat{v}_j, ..., v_n] = 0.
\end{aligned}$$

□

Therefore, $(C_\bullet(\mathcal{K}, \Bbbk), \partial_\bullet)$ is a chain complex cf. Definition 3.2.1 called a **simplicial chain complex**.

$$\dots\ C_{n+2} \xrightarrow{\partial_{n+2}} C_{n+1} \xrightarrow{\partial_{n+1}} C_n \xrightarrow{\partial_n} \dots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_0} C_0 \xrightarrow{0} 0$$

In Example 3.2.5, we obtain

$$0 \to C_2(\mathcal{T}, \Bbbk) \xrightarrow{\partial_2} C_1(\mathcal{T}, \Bbbk) \xrightarrow{\partial_1} C_0(\mathcal{T}, \Bbbk) \xrightarrow{\partial_0 = 0} 0$$

The **n-chains** $c \in C_n(\mathcal{K}, \Bbbk)$ can be written as a formal linear combination of the form

$$c = \sum_{\sigma \in \Sigma_n} n_\sigma \sigma, \quad n_\sigma \in \Bbbk,$$

where every σ is an oriented n-simplex of our simplicial complex.

From Definition 3.2.3 we define simplicial homology over a **simplicial chain complex** as follows:

Definition 3.2.9. Given an ordered abstract simplicial complex K with its corresponding simplicial chain complex $(C_\bullet(\mathcal{K}, \Bbbk), \partial_\bullet)$, the $\mathbf{n^{th}}$ **simplicial homology vector space** of $\mathcal{K}$ over $\Bbbk$ is the quotient $H_n(\mathcal{K}, \Bbbk) = \mathrm{Ker}(\partial_n)/\mathrm{Img}(\partial_{n+1}) = Z_n(\mathcal{K}, \Bbbk)/B_n(\mathcal{K}, \Bbbk)$.

Example 3.2.10. (Example 3.2.5, continued) To finish with our discussion on simplicial homology we compute

$$H_1(\mathcal{T}, \Bbbk) = H_1((C_\bullet(\mathcal{T}, \Bbbk), \partial_\bullet)) = \frac{Z_1(\mathcal{T}, \Bbbk)}{B_1(\mathcal{T}, \Bbbk)} = \frac{Ker(\partial_k)}{Img(\partial_{k+1})}.$$

Observe that since we do not have any 2-simplex, $\mathrm{Im}(\partial_2) = \{0\}$ and therefore, $H_1(\mathcal{T}, \Bbbk) = \ker(\partial_1)$. Therefore, first we apply ∂_1 to C_1, which yields

$$\partial_1(C_1(\mathcal{T}, \Bbbk)) = \alpha_1([b] - [a]) + \alpha_2([c] - [b]) + \alpha_3([c] - [a]),$$

for $\alpha_1, \alpha_1, \alpha_3 \in \Bbbk$. We rearrange the expression to

$$\partial_1(C_1(\mathcal{T}, \Bbbk)) = -(\alpha_1 + \alpha_3)[a] + (\alpha_1 - \alpha_2)[b] + (\alpha_2 + \alpha_3)[c],$$

and can see that it vanishes if and only if $\alpha_1 = \alpha_2 = -\alpha_3$. Thus, $\ker(\partial_1) \cong \Bbbk$, and $H_1(\mathcal{T}, \Bbbk) \cong \Bbbk$, which informally reflects that our simplicial complex has a 1-dimensional hole.

3.2.3 Singular homology and the functoriality of homology

We first present singular homology. Therefore, in the same way as in Section 3.2.2 we need to define a chain complex from any topological space. Let R be a commutative ring with unity.

A **topological n-simplex** Δ^n is a geometric simplex together with the subspace topology of $\mathbb{R}^{n+1}$. Given X in **Top**, a **singular n-simplex** is just an arrow $\sigma \in \mathcal{A}(\mathbf{Top})$ of the form $\sigma : \Delta^n \mapsto X$, i.e., an element of $Hom(\Delta^n, X)$. For convenience, we denote $Hom(\Delta^n, X)$ as $\Sigma_n(X)$.

Definition 3.2.11. An injective continuous map between topological spaces $f : X \to Y$ is an embedding if f yields an homeomorphism between f and $f(X)$.

The word 'singular' is used to express that σ is not a nice embedding but can have 'singularities' where its image does not look at all like a simplex. For instance, $\sigma : \Delta^n \to X$ can be the constant map $\sigma(\Delta^n) = x$ for some $x \in X$.

Definition 3.2.12 (Singular n-chain)**.** For $n \in \mathbb{N}$, a **singular n-chain** is an element of the free R-module $R[\Sigma_n(X)]$.

The n-boundary map definition is defined by the same formula as in Section 3.2.2, but instead of suppressing the vertex, $[v_0, ..., \hat{v}_i, ..., v_n]$, we compose σ with the following canonical injection of its facets preserving the ordering of the vertices of Δ^n. The **ith-simplicial facet map** is the monomorphism $\delta_i : \Delta_{n-1} \to \Delta_n$ that injects a $(n-1)$-topological simplex to the n-simplex that has a 0 in its i-th component, for $n \geq 0$ and $0 \leq i \leq n$.

Definition 3.2.13 (Boundary map)**.** The n-boundary $\partial_n(\sigma)$ of $\sigma \in \Sigma_n(X)$ is

$$\partial_n(\sigma) := \sum_{k=0}^{n} (-1)^k (\sigma \circ \delta_k).$$

As in Section 3.2.2, the very foundation of the entire theory relies in the following proposition:

Proposition 3.2.14. *The boundary map satisfies*

$$\partial_n \circ \partial_{n+1} = 0.$$

Proof. Analogous to the proof of 3.2.8. □

Therefore, given $X \in \mathbf{Top}$, the sequence of modules $R[\Sigma_\bullet(X)]$ together with the collection of morphisms $\partial_\bullet$ forms a chain complex

$$C_\bullet(X) := C_\bullet(R[\Sigma_\bullet(X)], \partial_\bullet) := C_\bullet(R[Hom(\Delta^\bullet, X)], \partial_\bullet)$$

called the **singular complex** or **alternating face complex** of X.

Therefore, from Definition 3.2.3, given a topological space X with its corresponding singular complex, the **nth-singular homology module** over R is the quotient

$$H_n^{sing}(X) := \frac{\text{Ker}(\partial_n)}{\text{Img}(\partial_{n+1})} = \frac{Z_k(X)}{B_k(X)}.$$

We briefly discuss why singular homology is a functor, therefore,

Proposition 3.2.15. *Given $X, Y \in$ **Top** and $f \in Hom(X, Y)$, the map f can be extended to $f_\# \in Hom_{Ch_\bullet}(C_\bullet(X), C_\bullet(Y))$.*

Proof. Given f, for any $\sigma \in \Sigma_n(X)$ we can define the map $f \circ \sigma : \Sigma_\bullet(X) \to \Sigma_\bullet(Y)$, which linearly extends in the obvious way to elements of $R[\Sigma_n(X)]$. Therefore, since σ are the basis elements of $R[\Sigma_n(X)]$, our map extends to $f_\# : C_\bullet(X) \to C_\bullet(Y)$. Finally, proving that $f_\#$ is a chain map consists in just noticing that for $u \in R[\Sigma_k(X)]$, $\partial(f_\# \circ u) = f_\#(\partial \circ u)$. □

Therefore, the assignment $F : \mathbf{Top} \mapsto \mathbf{Ch_\bullet}(\mathbf{Mod(R)})$ such that

$$F := \begin{cases} F(X) = C_\bullet(X), \\ F(X \xrightarrow{f} Y) = f_\# \end{cases}$$

is a functor.

Proposition 3.2.16. *Z_n^{sing}, B_n^{sing} and H_n^{sing} are functors to* **Mod(R)**.

Proof. First, let's prove that Z_n is a functor. Let $A_\bullet, D_\bullet \in \mathbf{Ch_\bullet}$ and $f : A_\bullet \to D_\bullet$ be a chain map. Suppose $z \in Z_n(A_\bullet)$, then $z \in A_n$ and $\partial_n(z) = 0$. Thus, $f(z) \in D_n$, and by the chain map commutativity, $\partial_n(f(z))) = f(\partial_n(z)) = f(0) = 0$. Therefore, $f(z) \in Z_n(B_\bullet)$, and $Z_n(f) := f|_{Z_n(A)}$ making Z_n a functor. By analogous reasons, $B_n(f) := f|_{B_n(A)}$ induces a functor. Hence, for each coset in $\frac{Z_n(A)}{B_n(A)}$ one can define $f_\#(z + B_n(A)) = f(z) + B_n(B)$ as a well-defined quotient map. Moreover, it is straightforward that $id_\bullet = id$ and $(g \circ f)_\bullet = g_\bullet \circ f_\bullet$, thus, H_n^{sing} is a functor from $Ch_\bullet(Mod(R))$ to **Mod(R)**. □

Therefore, from Proposition 3.2.15 and Proposition 3.2.16 we have that $H_n^{sing} \circ F : \mathbf{Top} \to \mathbf{Mod(R)}$ is a functor. We ignore F and simply say that homology is a functor $H_n^{sing} : \mathbf{Top} \to \mathbf{Mod(R)}$. Finally, we want to add that Whenever simplicial homology can be defined over a space, it is isomorphic singular homology.

Chapter 4

General Persistence Theory

Historically, the notion of persistence first arose early in the 2000s in Afra's Zoromodian PhD Thesis [Zom01]. He developed an algorithm to compute the appearance and disappearance of a cycle in a nested sequence of simplicial complexes.

Later, in 2004, Gunnar Carlsson and Afra Zoromodian introduced the notion of persistence module in the article [ZC04]. Here, they define a persistence module as a sequence of **k**-vector spaces together with linear maps. However, it was not until Frédéric Chazal introduced the category of persistence modules in [CCSG+09] that they began to be studied as objects on their own. A persistence module is the result of post-composing a filtration of a topological space with singular or simplicial homology and, therefore, a functor.

This chapter is devoted to introducing the category of persistence modules in its full generality. First, we present some of the main results regarding the decomposition of persistence modules. Those results are essential to developing convenient ways for representing them on computers. Additionally, since the primary goal of persistent homology is to be practical and helpful for machine learning and statistics purposes, we introduce two notions of distance in the category; the interleaving distance and the bottleneck distance. Finally, we establish the algebraic relation between persistent modules and modules.

4.1. The category of persistence modules $\mathbf{Pers}(\mathbb{k}^{\mathbf{P}})$

4.1.1 Introduction to $\mathbf{Pers}(\mathbb{k}^{\mathbf{P}})$

Recall from Definition 2.1.3 that a set P endowed with a partial order binary relation $\leq$ determines a partial order category, i.e. a preorder such that whenever there exists a morphism $a \to b$ and a morphism $b \to a$, this implies that $a = b$. We denote the latter category as $\mathbf{P}_{\leq}$. We recall the formal definition of a filtration.

Definition 4.1.1 (Definition 3.1.1)**.** Let $\mathbf{P}_{\leq}$ be some poset category and let $\mathbf{C}$ be a category. Moreover, let X be an object of $\mathbf{C}$. A **filtration** is a functor $F_X : \mathbf{P}_{\leq} \to \mathbf{C}$ such that:

- $F_X(a) := X_a$ where X_a is a subobject of X,
- X_a is a subobject of X_b whenever $a \to b$ in $\mathbf{P}_{\leq}$,

- $F_X(a \to b)$ is the morphism $X_a \hookrightarrow X_b$ where $\hookrightarrow$ is a monomorphism between subobjects X_a and X_b.

Suppose that we are working with a certain filtration F, a **persistence module** is the object resulting from post-composing $H_k^{sing} \circ F$.

Definition 4.1.2. The **category of** $\Bbbk$**-persistence modules** $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$, or simply **Pers** if the context allows it, is the functor category **Funct(P$_\leq$,Vect(k))** whose objects are functors from the partial order category $\mathbf{P}_\leq$ to the category of **k**-vector spaces **Vect(k)**, and whose morphisms are natural transformations. Therefore, for a given M and N in $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$, an morphism $\phi : N \to M$ is a collection of arrows $\phi_a \in \mathcal{A}(\mathbf{Vect(k)})$ such that for every $a \in \mathbf{P}_\leq$ whenever $a \leq b$ in $\mathbf{P}_\leq$ the following diagram commutes:

$$\begin{array}{ccc} N(a) & \xrightarrow{N(a\to b)} & N(b) \\ {\scriptstyle\phi_a}\downarrow & & \downarrow{\scriptstyle\phi_b} \\ M(a) & \xrightarrow{M(a\to b)} & M(b) \end{array}$$

Remark 4.1.3. It is a classical result that functor categories inherit the properties of their target category. In particular, a functor category whose target is abelian is abelian too. Therefore, since **Vect(k)** is an abelian category, in fact, an Ab2-category, $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$ is so. Thus, $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$ has a zero object, is bicomplete, has kernels and cokernels, etc

We show how we can inherit several notions and constructions in $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$ from working pointwise in **Vect(k)**.

Given M and N in $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$, an arrow $\phi : N \to M$ is an **isomorphism** of persistence modules if it is pointwise an isomorphism of vector spaces. In other words, ϕ is an isomorphism if $\phi_a : N(a) \to M(a)$ is an isomorphism for every a in $\mathbf{P}_\leq$. Following the same philosophy, a morphism of persistence modules is injective/surjective if it is injective/surjective pointwisely in **Vect(k)**.

Likewise, we define the **zero object** of $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$ as the functor $0_P : \mathbf{P}_\leq \to \mathbf{Vect(k)}$ that sends every $a \in \mathbf{P}_\leq$ to the zero of **Vect(k)**. It is easily checked that O_P it is both initial and terminal: First, for any other persistence module M, there exists a unique arrow to O_P that sends every $M(a)$ to the 0 vector space. Conversely, there exists a unique arrow from O_P to any other M in $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$, the one that sends each $O_P(a)$ to $0 \in M(a)$.

Definition 4.1.4 (Persistent sub-module)**.** A **persistent sub-module** L of $M \in \mathbf{Pers}(\Bbbk^{\mathbf{P}})$ is a functor $L : \mathbf{P}_\leq \to \mathbf{Vect(k)}$ satisfying that

1. its associated **object function** assigns to each $a \in P_\leq$ a vector space $L(a) \subseteq M(a)$ and,
2. its **arrow function** is the arrow function of M restricted for every $a \to b$ to the space in $L(a)$ such that has value in $L(b)$.

Definition 4.1.5 (Quotient persistence module)**.** Let L be a sub-module of M. The **quotient persistence module** is the functor $M/L : \mathbf{P}_\leq \to \mathbf{Vect(k)}$

- whose **object function** assigns to each $a \in \mathbf{P}_\leq$ the value $(M/L)(a) = M(a)/L(a)$.

- whose **arrow function** is the unique linear map $(M/L)(a \to b)$ defined as follows:

  Let π_a be the projection of $M(a)$ onto the quotient $M(a)/L(a)$ and analogously for π_b. Moreover, let us define f as $f := \pi_b \circ M(a \to b)$. Then $(M/L)(a \to b)$ is the unique function satisfying the following diagram resulting from the **universal property of quotients of vector spaces**

$$\begin{tikzcd} M(a) \arrow[r, "f"] \arrow[d, "\pi_a"] & M(b)/L(b) \\ M(a)/L(a) \arrow[ur, dashed, "(M/L)(a\to b)"'] & \end{tikzcd}$$

We continue working pointwise in **Vect(k)** to introduce the following constructions on **Pers**$(\mathbb{k}^{\mathbf{P}})$.

- **Direct sums**: Given M and N in **Pers**$(\mathbb{k}^{\mathbf{P}})$, the direct sum $M \oplus N$ is the functor that sends every $a \in P_{\leq}$ to $(M \oplus N)(a) := M(a) \oplus N(a)$.

  Since **Pers**$(\mathbb{k}^{\mathbf{P}})$ is additive, an object M of **Pers**$(\mathbb{k}^{\mathbf{P}})$ distinct from 0_P is **indecomposable** if for any isomorphism $M \simeq M_1 \oplus M_2$ either M_1 or M_2 is equal to 0_P.

- **Images**: Letting $\phi : N \to M$ a morphism, we pointwisely take into account the image of ϕ, which is $Im(\phi(a)) \subseteq M(a)$. Since the diagram

$$\begin{tikzcd} N(a) \arrow[r, "N(a\to b)"] \arrow[d, "\phi_a"'] & N(b) \arrow[d, "\phi_b"] \\ M(a) \arrow[r, "M(a\to b)"] & M(b) \end{tikzcd}$$

  commutes, we have that $M(a \to b)(Im(\phi_a)) \subseteq Im(\phi_b)$ and so we get $Im(\phi)$ as a submodule of M.

- **Kernels**: Similarly, considering pointwise the kernels of ϕ, we define $Ker(\phi)$ as a submodule of N.

- **Cokernels**: The cokernel of a morphism $f : N \to M$ can be built up as the quotient module of M by the image submodule associated to f.

- ...

We have briefly introduced the category **Pers**$(\mathbb{k}^{\mathbf{P}})$. The ultimate goal of utilising persistence theory for machine learning and statistics forces us to look for convenient ways to represent persistence modules in computers. As mathematicians, we try to do so by studying how persistence modules decompose and how they can be classified. We hope to find nice families of indecomposable elements that are easy to parameterise and allow us to express any persistence module in a given **Pers**$(\mathbb{k}^{\mathbf{P}})$.

Definition 4.1.6 (Pointwise finite dimensional persistence modules)**.** A persistence module $M \in$ **Pers**$(\mathbb{k}^{\mathbf{P}})$is **pointwise finite dimensional**, or **p.f.d**, if for every $a \in \mathbf{P}_{\leq}$, its vector space $M(a)$ is finite dimensional. We denote the subcategory of p.f.d persistence modules as $\mathbf{Pers_f}(\mathbb{k}^{\mathbf{P}})$.

The following theorem is due to Bakke Botnan and Crawley-Boevey.

Theorem 4.1.7 ([BCB18, Theorem 1.1])**.** *For any poset $\mathbf{P}_{\leq}$, any persistence module $M \in \mathbf{Pers_f}(\Bbbk^{\mathbf{P}})$ is isomorphic to a direct sum of indecomposable persistence modules with local endomorphism ring (Definition 2.1.37).*

The Krull–Remak–Schmidt–Azumaya theorem [Azu50, Theorem 1] asserts that persistence modules that decompose into a direct sum of indecomposables with local endomorphism ring do so in a unique way, up to reordering of isomorphisms. That is, given $M \in \mathbf{Pers_f}(\Bbbk^{\mathbf{P}})$, if there exist families $\mathcal{B}(M)$ and $\mathcal{B}(M)'$ of indecomposables such that $M \cong \bigoplus_{I \in \mathcal{B}(M)} I$ and $M \cong \bigoplus_{I \in \mathcal{B}(M)'} I$, then there exists a bijection $\sigma : \mathcal{B}(M) \to \mathcal{B}(M)'$ such that $I \cong \sigma(I)$ for all I in $\mathcal{B}(M)$. We say that any category satisfying that property satisfies [KS1].

Proposition 4.1.8. *Let $\bigoplus_{i\in I} M_i \in \mathbf{Pers_f}(\Bbbk^{\mathbf{P}})$. Then $\bigoplus_{i\in I} M_i \cong \prod_{i\in I} M_i$.*

Proof. We work pointwise in $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$. Since $\mathbf{Vect(k)}$ is bicomplete, for every $a \in \mathbf{P}_{\leq}$ we have $(\bigoplus_{i\in I} M_i)(a) \cong \bigoplus_{i\in I} M_i(a) \cong \prod_{i\in I} M_i(a)$, which again is finite dimensional. Therefore, $\prod_{i\in I} M_i(a) \cong (\prod_{i\in I} M_i)(a)$ which again is p.f.d. □

We say that a category that satisfies Proposition 4.1.8 for every object decomposing into a collection of indecomposables, satisfies [KS2]. An additive category satisfying [KS1] and [KS2] is a **Krull-Schmidt category**.

Proposition 4.1.9. *$\mathbf{Pers_f}(\Bbbk^{\mathbf{P}})$ is a Krull-Schmidt category.*

Proof. The proof follows from [BCB18, Theorem 1.1], [Azu50, Theorem 1] and Proposition 4.1.8. □

Therefore, every p.f.d persistence module decomposes uniquely into a direct sum of indecomposable persistence modules. However, we still do not know what these indecomposables look like or if they have any feasible representation for computers.

4.1.2 Interval persistence modules

In this subsection we take a look at a special kind of persistence modules, namely interval persistence modules. First, it is convenient to introduce the notion of interval in a generic partial order category.

Definition 4.1.10 (Interval on a poset)**.** Given a partial order set $\mathbf{P}_{\leq}$, we say that an **interval** of $\mathbf{P}_{\leq}$ is a non-empty convex subset of $\mathbf{P}_{\leq}$ satisfying the connectivity axiom. That is, it is a subset $I \subset \mathbf{P}_{\leq}$ satisfying the following properties:

(I1) $I \neq \emptyset$.

(I2) For every pair of elements a, b in I, if there exists an element p in $\mathbf{P}_{\leq}$ such that $a \leq p \leq b$, then p is in I.

(I3) [Connectivity axiom] For every pair of elements a, b in I there exists a sequence $s = \{s_i\}_{i=0,\dots,n}$ with $s_0 = a, s_n = b$, such that each element is comparable with its consecutive one, if it exists.

For example, an interval in $\mathbb{N}_{\leq}$ looks like this:

$$\dots \longrightarrow s-1 \longrightarrow s \longrightarrow \dots \longrightarrow t \longrightarrow t+1 \longrightarrow \dots$$

for $s, t \in \mathbb{N}$, $s \leq t$. An example of an interval in $\mathbb{N}^2$ is

$$\begin{tikzcd}
(0,2) \arrow[r] & (1,2) \arrow[r] & (2,2) \\
(0,1) \arrow[r] \arrow[u] & (1,1) \arrow[r] \arrow[u] & (2,1) \arrow[u] \\
(0,0) \arrow[r] \arrow[u] & (1,0) \arrow[r] \arrow[u] & (2,0) \arrow[u]
\end{tikzcd}$$

The connectivity axiom ensures that every two elements in an interval are connected by an arrow. Therefore, we avoid situations like the following one:

$$\begin{tikzcd}
\bullet \arrow[r] & \bullet & & & \bullet \arrow[r] & \bullet \\
\bullet \arrow[r] \arrow[u] & \bullet \arrow[u] & & & \bullet \arrow[r] \arrow[u] & \bullet \arrow[u] \\
& & & \bullet \arrow[ull] \arrow[ur] & &
\end{tikzcd}$$

Definition 4.1.11. Given a poset $\mathbf{P}_{\leq}$, we refer to a multiset of intervals in $\mathbf{P}_{\leq}$ as a **barcode**.

Let $\mathbf{Pers}(\mathbb{k}^{\mathbf{P}})$ be associated to $\mathbf{P}_{\leq}$. For each interval in $\mathbf{P}_{\leq}$, we define the following persistence module:

Definition 4.1.12 (Interval persistence module)**.** Let $I \subset \mathbf{P}_{\leq}$ be an interval. The **interval persistence module $\mathbf{k}^I$** is the persistence module in $\mathbf{Pers}(\mathbb{k}^{\mathbf{P}})$ such that:

$$\mathbf{k}^I(a) := \begin{cases} k \text{ if } a \in I, \\ 0 \text{ otherwise.} \end{cases}, \quad \mathbf{k}^I(a \to b) := \begin{cases} id_k \text{ if } a, b \in I, \\ 0 \text{ otherwise.} \end{cases}$$

An interval persistence module just differs from 0_P in the support of the interval. For example, an interval persistence module over $\mathbb{N}$ looks like this:

$$\dots \longrightarrow 0 \longrightarrow k \xrightarrow{\text{id}} \dots \xrightarrow{\text{id}} \mathbb{k} \longrightarrow 0 \longrightarrow \dots$$

Interval persistence modules satisfy the following nice property.

Proposition 4.1.13. $\mathbf{k}^I$ *is an indecomposable in* $\mathbf{Pers}(\mathbb{k}^{\mathbf{P}})$.

Proof. We take a look at $End(\mathbf{k}^I)$ (Definition 2.1.37). The endomorphism ring $End(\mathbf{k}^I) = Hom(\mathbf{k}^I, \mathbf{k}^I)$ is given by all natural transformations $\varphi : \mathbf{k}^I \to \mathbf{k}^I$. Therefore, they must satisfy the commutativity diagram

$$\begin{tikzcd}
\mathbf{k}^I(a) \arrow[r, "\mathbf{k}^I(a\to b)"] \arrow[d, "\varphi_a"'] & \mathbf{k}^I(b) \arrow[d, "\varphi_b"'] \\
\mathbf{k}^I(a) \arrow[r, "\mathbf{k}^I(a\to b)"] & \mathbf{k}^I(b)
\end{tikzcd}$$

The connectivity axiom of intervals ensures that any two elements in an interval are connected by an arrow. Therefore, by connectivity and the commutativity of the diagram we get that any φ must be a multiplication by a constant $c \in \mathbf{k}$. Thus, $End(\mathbf{k}^I) \cong \mathbf{k}$. Suppose that $\mathbf{k}^I \cong M \oplus N$ for persistence modules M and N. Then, $End(\mathbf{k}^I) \cong End(M \oplus N) \cong \mathbf{k}$. Since $End(M) \oplus End(N)$ is a subspace of $End(M \oplus N)$ either $End(M) = 0$ or $End(N) = 0$ implying $M = 0$ or $N = 0$. $\square$

Definition 4.1.14 (Interval decomposable persistence module)**.** A $\mathbf{P}_{\leq}$-persistence module $M \in \mathbf{Pers}(\mathbb{k}^{\mathbf{P}})$ is **interval decomposable** if there exists a (possibly infinite) multiset of intervals $\mathbb{B}(M)$ such that

$$M \cong \bigoplus_{I \in \mathbb{B}(M)} \mathbf{k}^I.$$

Remark 4.1.15. In Proposition 4.1.13 we proved that $End(\mathbf{k}^I) \cong \mathbf{k}$. Therefore, interval persistence modules have local endomorphism rings. It follows from the Azumaya–Krull–Remak–Schmidt theorem [Azu50, Theorem 1] that the multiset $\mathbb{B}(M)$ is unique.

Definition 4.1.16 (Persistence module barcode)**.** Given a interval decomposable persistence module M, its associated **barcode** $\mathbb{B}(M)$ is the unique multiset of intervals that characterises its decomposition.

4.1.3 Where does the word module come from? A nice equivalence of categories

This section introduces an equivalence of categories that provided the first decomposition and classification results of persistent homology. The main reason for the name persistence modules is that Carlsson and Zoromodian proved in [ZC04] that elements in $\mathbf{Pers}(\mathbb{k}^{\mathbb{N}})$ are really modules in the sense of abstract algebra. A refined version of this statement first appeared in [CZ09] for $\mathbf{Pers}(\mathbb{k}^{\mathbb{Z}^{\mathbf{d}}})$ and later for $\mathbf{Pers}(\mathbb{k}^{\mathbb{R}^{\mathbf{d}}})$ in [Les12]. Modules have been intensely studied in the field of Mathematics. Therefore, a proven correspondence to persistence modules was an opportunity for a more refined understanding and the first classification results. In this section, we aim to give a sketch of the proof for $\mathbf{Pers}(\mathbb{k}^{\mathbb{Z}^{\mathbf{d}}})$ following [CZ09].

Let $\mathbb{k}[t_1, \dots, t_d]$ be the commutative polynomial ring in d variables. A **monomial** in the indeterminates $t_1, \dots, t_d$ is a product of the form

$$t_1^{e_1} t_2^{e_2} \dots t_d^{e_d}$$

where $e_i \in \mathbb{N}$. We write $e = (e_1, \dots, e_d) \in \mathbb{N}^d$ and t^e to $t_1^{e_1} t_2^{e_2} \dots t_d^{e_d}$. A **polynomial** $p \in \mathbf{k}[t_1, \dots, t_d]$ is a formal linear combination of monomials with coefficients in $\Bbbk$:

$$p = \sum_e c_e t^e,$$

where $c_e \in \mathbf{k}$.

An $\mathbb{N}^d$-graded ring is a ring R equipped with a decomposition of abelian groups $R \cong \bigoplus_e R_e$, $e \in \mathbb{N}^d$, so that the multiplication has the property $R_u \cdot R_v \subseteq R_{u+v}$.

The prototypical case of a graded ring is the polynomial ring $\Bbbk[t_1, \dots, t_d]$. The grading of $\Bbbk[t_1, \dots, t_d]$, which we will also denote as A^d is

$$A^d \cong \bigoplus_e \Bbbk t^e.$$

For example, we visualize the monomials in A^2 as the $\mathbb{N}^2$ grid

$$\begin{tikzcd}
\dots \arrow[r] & \dots \arrow[r] & \dots \arrow[r] & \dots \\
c_{(0,2)} \arrow[r] \arrow[u] & c_{(1,2)} \arrow[r] \arrow[u] & c_{(2,2)} \arrow[r] \arrow[u] & \dots \arrow[u] \\
c_{(0,1)} \arrow[r] \arrow[u] & c_{(1,1)} \arrow[r] \arrow[u] & c_{(2,1)} \arrow[r] \arrow[u] & \dots \arrow[u] \\
c_{(0,0)} \arrow[r] \arrow[u] & c_{(1,0)} \arrow[r] \arrow[u] & c_{(2,0)} \arrow[r] \arrow[u] & \dots \arrow[u]
\end{tikzcd}$$

We obtain a polynomial by giving values to the nodes of the grid. For example, the polynomial $3t_2^2 + 7t_1t_2^2 + 4t_1^2t_2 + 2$ in A^2 can be represented as

$$\begin{tikzcd}
\dots \arrow[r] & \dots \arrow[r] & \dots \arrow[r] & \dots \\
3 \arrow[r] \arrow[u] & 7 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & \dots \arrow[u] \\
0 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & 4 \arrow[r] \arrow[u] & \dots \arrow[u] \\
2 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & \dots \arrow[u]
\end{tikzcd}$$

A graded module is an abelian group M together with a decomposition $M \cong \bigoplus_e M_e$ and an R-module structure $R_u \cdot M_e \subseteq M_{u+e}$.

Let us denote the i-th standard basis vector of $\mathbb{Z}^d$ by e_i.

Definition 4.1.17 (Graded module over A^d)**.** An graded module M over A^d is an abelian group M equipped with a decomposition of vector spaces

$$M \cong \bigoplus_{a \in \mathbb{N}^d} M_a$$

such that

$$t_i M_a \subset M_{a+e_i}$$

for every $a \in \mathbb{N}^d$ and every indeterminate t_i of A^d.

Graded modules over A^d form a category **GradMod(d,k)** whose morphisms $f : M \to N$ satisfy $f(M_a) \subset N_a$ for all $a \in \mathbb{N}^d$.

Definition 4.1.18 (Homogeneous element)**.** Given a graded A^d-module, an element $m_a \in M$ is **homogeneous** if $m_a \in M_a$ for some $a \in \mathbb{N}^d$. Given the decomposition as a direct sum of vector spaces of M, every element $m \in M$ decomposes as a sum $\sum_{i=0}^{k} m_i$ of k homogeneous elements.

In [ZC04] Carlsson and Zoromodian first proved the correspondence between $R[t]$-graded modules, that is, graded modules over a ring R with indeterminate t, and $\mathbf{Pers}(\Bbbk^{\mathbb{N}})$. Later on, they generalised this correspondence to $\mathbf{Pers}(\Bbbk^{\mathbb{N}^\mathbf{d}})$, which Lesnick refined to $\mathbf{Pers}(\Bbbk^{\mathbb{R}^\mathbf{d}})$ in his thesis [Les12].

We sketch the proof given by Carlsson and Zoromodian.

Theorem 4.1.19. $\mathbf{Pers}(\Bbbk^{\mathbb{N}^\mathbf{d}})$ *is isomorphic to* **GradMod(d)**.

Proof. First, we define a functor $F : \mathbf{Pers}(\Bbbk^{\mathbb{N}^\mathbf{d}}) \to \mathbf{GradMod(d)}$ such that for a given $M \in \mathbf{Pers}(\Bbbk^{\mathbb{N}^\mathbf{d}})$ it satisfies

$$F(M) = \bigoplus_{a \in \mathbb{N}^d} M(a).$$

We define the action of $\Bbbk[t_1, \dots, t_n]$ on $F(M)$ as follows:

1. **Indeterminates acting on homogeneous elements of** $F(M)$: Let $m \in F(M)$ be an homogeneous element, so that $m \in M(a)$ for some $a \in \mathbb{N}^d$. We define the action of t_i on m by
$$t_i(m) := M(a \to a + e_i)(m).$$
This is a morphism $M(a) \to M(a + e_i)$ sending the homogeneous element m of $M(a)$ into an homogeneous element of $M(a + e_i)$.

2. **Indeterminates acting on** $F(M)$**:** The action of t_i on $F(M)$ follows by linearity. Recall that every element $m \in F(M)$ decomposes as a sum $\sum_{j=0}^{k} m_j$ of k homogeneous elements. Therefore, we define the action of an indeterminate t_i on an element m as
$$t_i(m) = \sum_{j=0}^{k} t_i(m_j).$$

3. $\Bbbk[t_1, \dots, t_d]$ **acting on** $F(M)$**:** The action of indeterminates on $F(M)$ directly extends to an action of $\Bbbk[t_1, \dots, t_d]$ on $F(M)$. For example,
$$(2t_3^2 t_2 + t_1)(m) = 2(t_3 \circ t_3 \circ t_2)(m) + t_1(m).$$

Therefore, $F(M)$ is a well-defined graded module. Moreover, natural transformations extend to morphisms in the natural way.

Conversely, we define a functor $G : \mathbf{GradMod(d)} \to \mathbf{Pers}(\Bbbk^{\mathbb{N}^{\mathbf{d}}})$ such that given a graded module $M \cong \bigoplus_{a \in \mathbb{N}^d} M_a$, its assignment $G(M)$ is the persistence module such that for every $a, b \in \mathbb{N}^d$ with $a \leq b$ satisfies:

1. $G(M)(a) := M(a)$,
2. $G(M)(a \to b) := t^{b-a}|_{M(a)}$.

Thus, F and G define an equivalence of categories. $\square$

4.2. Distances on $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$: Interleaving and Bottleneck distance

Previously, we have highlighted the importance of understanding decomposition and classification of persistence modules as a first step toward finding convenient ways to represent them on computers. However, we must not just focus our discussion around that purpose. Machine learning and statistics need tools to compare the objects they are dealing with in a specific task. Therefore, metrics in $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$ become essential for TDA.

With that purpose in mind, first, we are going to present an extended pseudo-distance in the isomorphism classes of persistence modules, called the interleaving distance. Informally, it measures how far are two persistence modules from being isomorphic. The interleaving distance was first presented by Frédérich Chazal in [CCSG+09].

Later, we will present the Bottleneck distance, an extended pseudo-distance defined over pairs of collections of indecomposable objects in a category.

4.2.1 Interleaving distance

Definition 4.2.1. A **flow** on $\mathbf{P}_{\leq}$ is a functor $T : ([0, +\infty)) \to End(\mathbf{P}_{\leq})$ such that

1. $T(0) = id_{\mathbb{P}_{\leq}}$,
2. $T(\epsilon_1) \circ T(\epsilon_2) = T(\epsilon_1 + \epsilon_2)$ for all $\epsilon_1, \epsilon_2 \in [0, +\infty)$.

Example 4.2.2. For $P = \mathbb{R}^n_{\leq}$, the functor $T : ([0, +\infty)) \to End(\mathbb{R}^n_{\leq})$ given by

$$T := \begin{cases} T(\epsilon)(a) := a + \epsilon \text{ for any } a, \epsilon \in \mathbb{R}^n_{\leq}, \\ T(\epsilon_1 \leq \epsilon_2)(a + \epsilon_1) := a + \epsilon_2 \text{ for any } a, \epsilon_1, \epsilon_2 \in \mathbb{R}^n_{\leq} \text{ with } \epsilon_1 \leq \epsilon_2, \end{cases}$$

is a flow on $\mathbb{R}^n_{\leq}$. First, $T(0)(a) = a$. Moreover, $(T(\epsilon_1) \circ T(\epsilon_2))(a) = T(\epsilon_1)(a + \epsilon_2) = a + \epsilon_1 + \epsilon_2$.

Definition 4.2.3 (ϵ-shift module)**.** Let T be a flow on $\mathbf{P}_{\leq}$ and M a persistence module in $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$. Given $\epsilon \geq 0$, the ϵ-**shift** $M[\epsilon]_T$ of M along the flow T is

$$M[\epsilon]_T := M \circ T(\epsilon).$$

For a given $a \in \mathbf{P}_{\leq}$, we have that $M[\epsilon]_T(a) = (M \circ T(\epsilon))(a) = M(T(\epsilon)(a))$ is the element of the persistence module shifted by the endomorphism $T(\epsilon)$ of $\mathbf{P}_{\leq}$.

For $0 \leq \epsilon_1 \leq \epsilon_2$, the flow evaluated on an arrow $T(\epsilon_1 \leq \epsilon_2)$ induces a morphism of persistence modules. First, $T(\epsilon_1 \to \epsilon_2)$ is the morphism of endomorphisms $T(\epsilon_1) \to T(\epsilon_2)$. Therefore, $M \circ (T(\epsilon_1) \to T(\epsilon_2))(a) = M \circ (T(\epsilon_1)(a) \to T(\epsilon_2)(a)))$ is a morphism of persistence modules subject to $T(\epsilon_1 \leq \epsilon_2)$. Thus, for $0 \leq \epsilon_1 \leq \epsilon_2$, $T(\epsilon_1 \leq \epsilon_2)$ induces a morphism of persistence modules

$$\tau^M_{\epsilon_1,\epsilon_2}(M) : M[\epsilon_1]_T \to M[\epsilon_2]_T.$$

Definition 4.2.4 (ϵ-smoothing morphism)**.** Given a flow T on $\mathbf{P}_{\leq}$, a persistence module M over the same $\mathbf{P}_{\leq}$, and $\epsilon \geq 0$, the induced morphism by $T(0 \leq \epsilon)$

$$\tau^M_\epsilon(M) : M \to M[\epsilon]_T,$$

is the ϵ-**smoothing morphism** of M over T.

Definition 4.2.5 (ϵ-interleaving)**.** Let T be a flow on a poset $\mathbf{P}_{\leq}$ and $\epsilon \geq 0$. Let M and N be two persistence modules over $\mathbf{P}_{\leq}$. An (ϵ, T)-interleaving is a pair of morphisms

$$(f : M \to N[\epsilon]_T, g : N \to M[\epsilon]_T)$$

that satisfy the following commutative diagram

$$\begin{tikzcd}
M \arrow[r, "f"] \arrow[dr] \arrow[rr, bend left, "\tau^M_{2\epsilon}"] & N[\epsilon]_T \arrow[r, "{g[\epsilon]}"] \arrow[dr] & M[2\epsilon]_T \\
N \arrow[r, "g"] \arrow[ur] \arrow[rr, bend right, "\tau^N_{2\epsilon}"'] & M[\epsilon]_T \arrow[r, "{f[\epsilon]}"] \arrow[ur] & N[2\epsilon]_T
\end{tikzcd}$$

In other words, $g[\epsilon] \circ f = \tau^M_{2\epsilon}$ and $f[\epsilon] \circ g = \tau^N_{2\epsilon}$. In this case, we say that M and N are ϵ-interleaved, and write $M \sim^T_\epsilon N$, or $M \sim_\epsilon N$ if the context allows it.

Observation 4.2.6. If two persistent modules M and N are 0-interleaved they have a pair of morphisms f, g such that $(g \circ f)_a = id_{M_a}$ and $(f \circ g)_a = id_{N_a}$ for every $a \in \mathbf{P}_{\leq}$. Therefore, they are isomorphic.

Observation 4.2.7. Being ϵ-interleaved is not a transitive relation since $M \sim^T_\epsilon N \sim^T_\epsilon L$ does not imply $M \sim^T_\epsilon N$. Therefore, it is not an equivalence relation.

Remark 4.2.8. Given a category $\mathbf{C}$ and a poset $\mathbf{P}_{\leq}$ with a flow T, the definition of interleaving naturally extends to the category **Funct($\mathbf{P}_{\leq}$,C)**. Therefore, we can talk of ϵ-interleaved filtrations. Hence, the functoriality of homology makes the following statement hold: *"ϵ-interleaved filtrations induce ϵ-interleaved persistent modules."*

We define the interleaving distance as follows:

Definition 4.2.9 (Interleaving distance)**.** Let $\mathbf{P}_{\leq}$ be a poset and T a flow defined over it. Moreover, let M and N be two persistent modules in $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$. The **interleaving distance** between M and N with respect to T is:

$$d_I^T(M,N) = \inf\{\epsilon \geq 0 | M \sim_\epsilon^T N\}.$$

We sometimes omit the superindex T which refers to the flow.

Observation 4.2.10. The interleaving distance between two persistence modules can be infinite. For example, let $\mathbf{P}_{\leq} = \mathbf{2}$ be the ordinary number category with the usual order. We define a flow in $\mathbf{2}$ by $T(\epsilon) := \epsilon \, mod(2)$. Then, any pair of persistence modules M and N in $\mathbf{Pers}(\mathbf{k}^{\mathbf{2}})$ are either at distance 0 (isomorphic), 1, or ∞.

Theorem 4.2.11 ([dSMS17, Theorem 2.5])**.** *The interleaving distance with respect to a flow T of $\mathbf{P}_{\leq}$ is an extended pseudo-distance on $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$. In other words, for any persistence modules $M, N, L \in \mathbf{Pers}(\Bbbk^{\mathbf{P}})$, it satisfies:*

1. $d_I(M,N) \in [0,+\infty)$,
2. $d_I(M,N) = d_I(N,M)$,
3. $d_I(M,N) \leq d_I(M,L) + d_I(N,L)$.

Given two persistent modules M and N, if $M \cong N$, they are 0-interleaved, and therefore, $d_I(M,N) = 0$. However, the converse is not always true.

Example 4.2.12. Take $M \in \mathbf{Pers}(\mathbf{k}^{\mathbb{R}})$ such that

$$M := \begin{cases} M(0) = \mathbf{k}, \\ M(a) = 0 \text{ if } a \neq 0. \end{cases}$$

We have that M is not isomorphic to $0 \in \mathbf{Pers}(\mathbf{k}^{\mathbb{R}})$. However, if we consider the flow T over $\mathbb{R}_{\leq}$ as in Example 4.2.2

$$T(\epsilon)(a) = a + \epsilon,$$

we get $M \sim_\epsilon^T 0$ for all $\epsilon \geq 0$. Thus, $d_I^T(M,0) = 0$.

4.2.2 Bottleneck distance

Let $\mathbf{K}$ be a Krull-Schmidt category with an extended pseudo-distance in its object set

$$d : \mathcal{O}(\mathbf{K}) \times \mathcal{O}(\mathbf{K}) \to [0,+\infty),$$

such that if $A, B \in \mathbf{K}$ are isomorphic, then $d(A,B) = 0$. For $A \in \mathbf{K}$, we denote by $\mathcal{B}(A)$ to the set of indecomposables such that $A \cong \bigoplus_{I \in \mathcal{B}(A)} I$.

Definition 4.2.13. (ϵ-matching) Let $A, B \in \mathcal{O}(\mathbf{K})$, and $\epsilon \geq 0$. An ϵ-**matching** between $\mathcal{B}(A)$ and $\mathcal{B}(B)$ is a bijection $\sigma : \mathcal{I}_A \subseteq \mathcal{B}(A) \to \mathcal{I}_B \subseteq \mathcal{B}(B)$ such that

1. for every $I_a \in \mathcal{I}_A$ the quantity $d(I_a, \sigma(I_a)) \leq \epsilon$,
2. for every $I \in \mathcal{B}(A)/\mathcal{I}_A$ or $I \in \mathcal{B}(B)/\mathcal{I}_B$ we have $d(I,0) \leq \epsilon$.

Remark 4.2.14. Since for $A, B \in \mathbf{K}$ such that $A \cong B$ the distance $d(A, B) = 0$, the ϵ-matching does not depend on the set of representatives.

Two objects in $\mathbf{K}$ are ϵ-matched if there exist an ϵ-matching between their sets of indecomposables.

Definition 4.2.15 (Bottleneck distance)**.** Let A and B be objects in $\mathbf{K}$. The **bottleneck distance** associated to d between A and B is

$$d_B(A, B) = \inf\{\epsilon \geq 0 | A \text{ and } B \text{ are } \epsilon\text{-matched}\}.$$

4.2.3 Relevant challenges related to distances in persistence theory

Stability Theory In TDA, we aim to estimate topological quantities from datasets to learn from them. Therefore, if we provide $\mathbf{Pers}(\mathbb{k}^{\mathbf{P}})$ with a notion of distance, as we did, we need *small perturbations* in our data not to significantly perturb the resulting persistence module with respect that distance. Determining whether the latter holds or not is a relevant challenge for persistence homology theory. In the field, results concerning that fact are usually called Stability Theorems.

The Isometry Problem Given a Krull-Schmidt category $\mathbf{K}$ associated to d, the isometry problem is to determine whether $d = d_B$, i.e., when does the isometry hold. Assume that the isometry theorem holds for d_I, and so $d_I = d_B$, then it is possible to compute d_I from d_B, or to translate stability results from d_I to d_B.

Chapter 5

Persistent Homology: From one to multiple parameters

An **invariant** is a function ϕ from a collection of persistence modules into a set S such that if $M \cong N$, then $\phi(M) = \phi(N)$. If the converse holds, the invariant is **complete**. In general, the persistence homology pipeline looks as follows:

$$\xrightarrow{1.} \boxed{\text{Data}} \xrightarrow{2.} \boxed{\text{Filtration}} \xrightarrow{3.} \boxed{\text{Persistence Module Invariant}} \xrightarrow{4.}$$

We distinguish four different steps:

1. **Obtaining data**: Obtaining the data from which we want to estimate the homology groups of its underlying space and perform any learning task.

2. **Filtration/Complexification**: This step consists in building up a meaningful filtration from data. That means a filtration that we presume might reveal relevant aspects of its topology. Usually, it is convenient to work with filtrations provided with a combinatorial structure. Therefore, it is common to turn data into simplicial complexes as in Section 2.2.2 and to encode the full reconstruction in a single simplicial complex filtration. The filtration step is among the philosophical roots of TDA since it aims to encode the entire incremental space reconstruction to forget about parameter tuning.

3. **Compression:** The compression process aims to obtain a compressed representation of the filtration containing information on its homology. It is comprised of two different steps. The first one is computing the n-th singular/simplicial homology of the filtration, from which we obtain a persistence module. The second one is computing an invariant that represents the persistence module.

4. **Learning:** Learn from the representation of the persistent module that has been obtained in the compression step.

In this chapter, we first introduce one parameter persistent homology. We present some of the most popular $1D$-Filtrations. Second, we delve into the decomposition of $1D$-persistence modules, which turn out to be interval decomposable. Later, we introduce the interleaving and bottleneck distances associated to the $1D$-persistence theory, and discuss the most relevant stability results of $1D$-persistence. Finally, we show that $1D$-persistent homology can be extremely sensitive to outliers and that considering multiparameter filtrations we capture more information about the space. Therefore, we motivate the study of multiparameter persistent homology or MPH.

After presenting $1D$-persistent homology, we delve into MPH. We introduce some of the most common multifiltrations. Later, we delve into quiver representation theory. This, allows us to adapt some results from the quiver theory to the classification of multiparameter persistence modules, which turns out to be hopeless. Finally, we present invariants which intend to play a similar role to the barcodes of the $1D$ case.

5.1. One parameter persistent homology

The theory of one parameter persistence is the theory associated with total order categories $\mathbf{T}_{\leq}$. Therefore, a one-parameter or $1D$-persistence module is an object from $\mathbf{Pers}(\Bbbk^{\mathbf{T}})$. In general, the main example of one-parameter persistence module is given by

$$M_n^u := H_n^{sing} \circ \mathcal{S}(u),$$

where H_n^{sing} is the n-th homology functor and $\mathcal{S}(u)$ is the sub-level set filtration (Def. 3.1.5) subject to $u : \mathbb{X} \to \mathbf{T}_{\leq}$, where $\mathbb{X}$ is the topological space whose homology groups we want to compute. Therefore, $1D$-persistence encompasses the theory associated with $\mathbb{N}_{\leq}, \mathbb{Z}_{\leq}, \mathbb{R}_{\leq}$. However, people in the field usually refer to one-persistent homology as the theory around $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$, which is built up from sub-level set filtrations from functions $u : \mathbb{X} \to \mathbf{R}_{\leq}$ (see Example 3.1.6).

5.1.1 1D-Filtrations

Let $\mathcal{X}$ be a point cloud in $\mathbb{R}^n$ endowed with the euclidean distance.

There is no generic recipe for obtaining the best $1D$-filtration. Depending on the problem, and in the information we want to obtain, it might be convenient to follow different paths. In Section 2.2.2 we provided two distinct ways to obtain simplicial complexes from $\mathcal{X}$, namely, Čech (Definition 2.2.14) and Vietoris Rips (Definition 5.1.7) complexes. We introduce their filtered versions, which in fact are particular instances of sub-level set filtrations.

Čech complex filtration

Definition 5.1.1 (Čech complex filtration)**.** For $\mathcal{X} \subset \mathbb{R}^n$, the **Čech complex filtration** is the filtration $\text{Čech}(\mathcal{X})$ such that $\text{Čech}(\mathcal{X})(\epsilon) = \text{Čech}(\mathcal{X}, \epsilon)$ for $\epsilon \geq 0$.

We introduce analogues of Theorem 2.2.13 for filtrations. Let F be a $\mathbf{P}_{\leq}$-filtration of topological spaces.

Definition 5.1.2 (Covering filtration)**.** A **covering filtration** of F is a set of filtrations U such that $U(a)$ is a covering of $F(a)$ for every $a \in \mathbf{P}_{\leq}$.

Theorem 5.1.3 (Persistence Nerve Theorem (Weak version)). *If for every $a \in \mathbf{P}_{\leq}$, we have that $U(a)$ satisfies the hypotheses of Theorem 2.2.13 for $F(a)$, namely that $U(a)$ is a finite, closed, convex covering of $F(a)$, then $N(U)$ is homotopy equivalent to F.*

For the strong and general version of Theorem 5.1.3 we refer the reader to [CS18].

Definition 5.1.4 (Growing balls sub-level set filtration). Let $d_{\mathcal{X}}$ be the minimum distance from a point $x \in \mathbb{R}^n$ to $\mathcal{X}$:

$$\mathrm{d}_{\mathcal{X}}(x) := \min_{y \in \mathcal{X}} \| x - y \| .$$

The **growing balls $\mathbb{R}^+$-filtration** is the sub-level set filtration $S_{d_{\mathcal{X}}}$.

Remark 5.1.5. For any $\epsilon \geq 0$, we have that $S_{d_{\mathcal{X}}}(\epsilon)$ is the union of balls $B_\epsilon(x)$ for each $x \in \mathcal{X}$.

Lemma 5.1.6. $\check{\mathrm{C}}\mathrm{ech}(\mathcal{X}) \simeq S_{d_{\mathcal{X}}}$

Proof. For each $\epsilon \geq 0$, $S_{d_{\mathcal{X}}}(\epsilon)$ is a finite, closed, convex cover of $\mathcal{X}$. Therefore, the lemma follows from Theoren 5.1.3. □

Vietoris-Rips filtrations

Definition 5.1.7. The **Vietoris-Rips filtration** is the filtration $\mathrm{Rips}(\mathcal{X})$ such that for each $\epsilon \geq 0$, $\mathrm{Rips}(\mathcal{X})(\epsilon) = \mathrm{Rips}(\mathcal{X}, \epsilon)$.

Definition 5.1.8 (Diameter of $\mathcal{X}$). The diameter of $\mathcal{X}$ is

$$\mathrm{diam}(\mathcal{X}) = \sup_{x,y \in \mathcal{X}} \mathrm{d}(x, y).$$

Observation 5.1.9. By definition, the sub-level set filtration S_{diam} is equivalent to $\mathrm{Rips}(\mathcal{X})$, since for each $d \in \mathbb{R}^+$, some $\sigma \subseteq \mathcal{X}$ is contained in both if and only if $\mathrm{diam}(\sigma) \leq d$.

The interleaving property

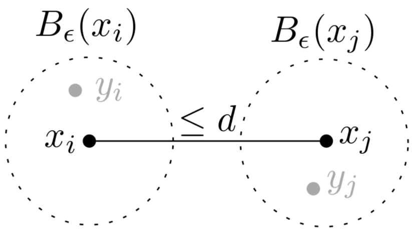
Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be a finite point cloud in $\mathbb{R}^n$.

Definition 5.1.10 (ϵ-perturbation). Given $\mathcal{X}$, a finite point cloud $\mathcal{Y} = \{y_1, \ldots, y_n\}$ is an ϵ-perturbation of X if $d(x_i, y_i) \leq \epsilon$ for all $1 \leq i \leq \epsilon$.

Proposition 5.1.11. *Choose $\epsilon \geq 0$ and let $\mathcal{Y}$ be an ϵ-perturbation of $\mathcal{X}$. Then:*

1. *The filtrations $\mathrm{Rips}(\mathcal{X})$ and $\mathrm{Rips}(\mathcal{Y})$ are 2ϵ-interleaved.*
2. *The filtrations $\check{\mathrm{C}}\mathrm{ech}(\mathcal{X})$ and $\check{\mathrm{C}}\mathrm{ech}(\mathcal{X})$ are ϵ-interleaved.*

Proof. From the triangle inequality, it is apparent in the following figure that, given any subset $\sigma \subseteq \mathcal{X}$ with diameter d, its corresponding subset in $\tau \subseteq \mathcal{Y}$ has diameter at most $d + 2\epsilon$.



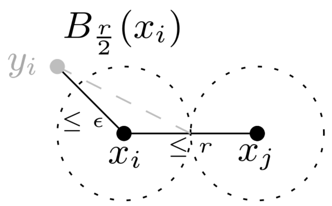
Therefore, if $\sigma \in \text{Rips}(\mathcal{X})(d)$, then $\tau \in \text{Rips}(\mathcal{X})(d + 2\epsilon)$. Thus, we define the following simplicial maps:

- The map $f : \text{Rips}(\mathcal{X}) \to \text{Rips}(\mathcal{Y})[2\epsilon]$ as $f(x_i) = y_i$,
- The map $g : \text{Rips}(\mathcal{Y}) \to \text{Rips}(\mathcal{X})[2\epsilon]$ as $g(y_i) = x_i$.

Note that the ϵ-smoothing morphisms for $\text{Rips}(\mathcal{X})$ and $\text{Rips}(\mathcal{Y})$ are simply the inclusion morphisms. Therefore, since f and g commute with inclusions, the following diagram commutes, and so *Rips*$(\mathcal{X})$ and *Rips*$(\mathcal{Y})$ are 2ϵ-interleaved.

$$\begin{tikzcd}
\text{Rips}(\mathcal{X}) \arrow[r, "f"] \arrow[dr, hook] & \text{Rips}(\mathcal{Y})[2\epsilon] \arrow[r, "g[\epsilon]"] \arrow[dr, hook] & \text{Rips}(\mathcal{X})[4\epsilon] \\
\text{Rips}(\mathcal{Y}) \arrow[r, "g"] \arrow[ur, hook] & \text{Rips}(\mathcal{X})[2\epsilon] \arrow[r, "f[\epsilon]"] \arrow[ur, hook] & \text{Rips}(\mathcal{Y})[4\epsilon]
\end{tikzcd}$$

The following figure shows that for analogous reasons $\check{\text{C}}\text{ech}(\mathcal{X})$ and $\check{\text{C}}\text{ech}(\mathcal{X})$ are ϵ-interleaved.



□

Proposition 5.1.11 will be crucial for establishing the first stability results in Section 5.1.3.

5.1.2 Decomposition and parametrization

This section introduces the decomposition and parametrization of $1D$-persistence modules. To begin with, we introduce the first decomposition result ever for persistence modules, presented by Carlsson and Zoromodian in their joint article [ZC04]. The proof is very simple and relies on the categorical correspondence proven in Theorem 4.1.19.

Definition 5.1.12 (Finitely generated persistence module)**.** Given a $P_{\leq}$-persistence module M and an homogeneous element $m \in M(a)$, we write $gr(m) = a$. We say that $S \subset \cup_{a \in P_{\leq}} M(a)$ is a **set of generators** for M if for every $m \in \cup_{a \in P_{\leq}} M(a)$, we have that

$$m = \sum_{i=0}^{n} c_i M(gr(m_i) \to gr(m))(m_i),$$

with $m_i \in S$ and $c_i \in \Bbbk$. We say that M is finetely generated if there exists a finite set of generators for M.

Theorem 5.1.13 (Structure Theorem for PIDs)**.** *Let D be a principal ideal domain. Then every finitely generated module M over D is isomorphic to a direct sum of cyclic D-modules. That is, there is a unique decreasing sequence of proper ideals $(d_1) \supseteq (d_2) \supseteq \cdots \supseteq (d_m)$ such that*

$$M \cong D^{\beta} \oplus \left(\bigoplus_{i=1}^{m} D/(d_i) \right), \tag{5.1}$$

where $d_i \in D$, and $\beta \in \mathbb{Z}$. Moreover, every graded module over a graded principal ideal domain D, decomposes uniquely into the form:

$$M \cong \left(\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} D \right) \oplus \left(\bigoplus_{j=1}^{m} \Sigma^{\gamma_j} D/(d_j) \right), \tag{5.2}$$

where $d_j \in D$ are homogenous elements such that $(d_1) \supseteq (d_2) \supseteq \cdots \supseteq (d_m), \alpha_i, \gamma_j \in \mathbb{Z}$, and Σ^{α} denotes an α-shift upward in grading.

Definition 5.1.14 (Finite type)**.** Let M be a persistence module in $\mathbf{Pers}(\Bbbk^{\mathbf{P}})$. We say that M is of **finite type** if $M(a)$ is finitely generated for all $a \in \mathbb{N}$ and there exists some $b \in \mathbb{N}$ such that $M(b \leq c)$ is an isomorphism for all $c \geq b$. We denote the category of $\mathbb{N}$-persistence modules of finite type as $\mathbf{Pers_{ft}}(\Bbbk^{\mathbb{N}})$.

Theorem 5.1.15. *Every persistent module in $\mathbf{Pers_{ft}}(\Bbbk^{\mathbb{N}})$ is interval decomposable (Definition 4.1.14).*

Proof. Let $\Bbbk[t]$ be the polynomial ring with indeterminate t. By Theorem 4.1.19 we know that $\mathbf{Pers}(\Bbbk^{\mathbb{N}^{\mathbf{d}}})$ is equivalent to **GradMod(d,k)**. In particular, for $d = 1$ the functor F of Theorem 4.1.19 is

$$F(M) = \bigoplus_{a \in \mathbb{N}} M(a)$$

The action of $\Bbbk[t]$ on $F(M)$ is defined as in the Theorem.

Restricting to $\mathbf{Pers_{ft}}(\Bbbk^{\mathbb{N}})$ implies that $F(M)$ is finitely generated (cf. Definition 5.1.14. From a classical algebraic result, if $\Bbbk[t]$ is a principal ideal domain. Therefore, from the Structure Theorem 5.1.13 we get the following isomorphism

$$F(M) \cong \left(\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} \Bbbk[t] \right) \oplus \left(\bigoplus_{j=1}^{m} \Sigma^{\gamma_j} \Bbbk[t]/(t^{n_j}) \right). \tag{5.3}$$

Using the inverse functor G in Theorem 4.1.19, equation 5.3 proves the existence of an interval decomposition for M given by intervals of the form $(\alpha_i, +\infty)$ and $(\gamma_j, \gamma_j + n_j)$. $\square$

Recall that the decomposition of every interval persistence module is unique (up to isomorphism) by Remark 4.1.15. Therefore, every persistence module in $\mathbf{Pers_{ft}}(\Bbbk^{\mathbb{N}})$ is determined by its barcode (Definition 4.1.16). Hence, the parametrization of persistence modules in $\mathbf{Pers_{ft}}(\Bbbk^{\mathbb{N}})$ is given by the intervals shown in Theorem 5.1.15. This fact provided a simple way to represent persistent modules in computers. Therefore, it was crucial to practical applications of persistence homology.

Later, Crawley-Boevey generalised this result to $\mathbf{Pers_{f}}(\Bbbk^{\mathbb{R}})$ in [CB12, Theorem 1.1]. Moreover, in 2018, Bakke-Botnan and Crawle-Boevey proved in their joint article [BCB18] the following theorem:

Theorem 5.1.16 ([BCB18, Theorem 1.2])**.** *Pointwise finite-dimensional persistence modules over a totally ordered set decompose into interval modules.*

Therefore, Theorem 5.1.16 was the ultimate refinement of the $1D$-persistence decomposition theorem.

Remark 5.1.17. Barcodes in $\mathbf{Pers_f k^T}$ are a complete invariant.

Barcodes on $\mathbf{Pers_{f}}(\Bbbk^{\mathbb{R}})$ have nice visual representations, which definitely is another advantage for data analysis tasks.

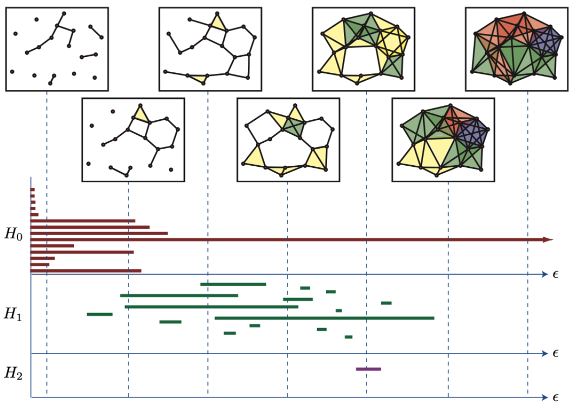


Figure 5.1: Example of a barcode visual snapshot of the homology groups H_0, H_1, H_2 of a given simplicial complex filtration. Extracted from [Ghr07].

Sometimes it is convenient to deal with an equivalent representation of the barcode, the persistence diagram.

Definition 5.1.18 (Persistence diagram)**.** Let M be a persistent module in $\mathbf{Pers_{f}}(\Bbbk^{\mathbb{R}})$ with barcode $\mathbb{B}(M)$. We denote by B the finite multiset of points in the plane $\{(inf(I), sup(I))\}_{I\in\mathbb{B}(M)}$. The persistence diagram of M is $Dgm_M = B \cup \Delta$ where $\Delta = \{(x,x) | x \in \mathbb{R}\}$.

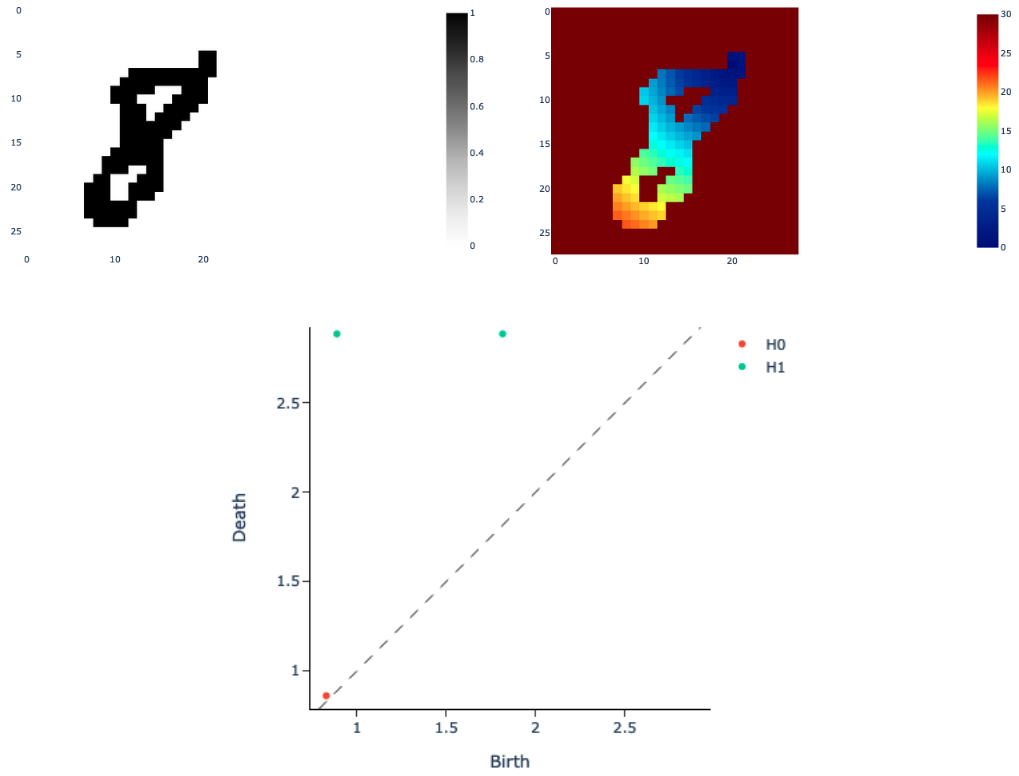


Figure 5.2: Example of a persistence diagram obtained from an image of a figure eight in the database MNIST [Den12] following the pipeline in article [GT19]. First, we pick a grayscale image of a figure eight. Second, we binarize the image and apply a radial filter. Third, we compute persistent homology from a density filtration. The persistence diagram reflects the two 1-dimensional holes of the figure in H_1. Moreover, the only generator of H_0 shows that the figure consists of only one connected component.

5.1.3 Distances and stability results in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$

Interleaving distance in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ Recall from Example 4.2.2 that the functor $T : ([0, +\infty)) \to End(\mathbb{R}_{\leq})$ given by

$$T := \begin{cases} T(\epsilon)(a) := a + \epsilon \text{ for any } a, \epsilon \in \mathbb{R}_{\leq}, \\ T(\epsilon_1 \leq \epsilon_2)(a + \epsilon_1) := a + \epsilon_2 \text{ for any } a, \epsilon_1, \epsilon_2 \in \mathbb{R}_{\leq} \text{ with } \epsilon_1 \leq \epsilon_2, \end{cases}$$

defines a flow in $\mathbb{R}$. We define the interleaving distance (Definition 4.2.9) for $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ to be the what induced by the flow T.

Bottleneck distance in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ Recall from Proposition 4.1.9 that $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ is Krull-Schmidt. Moreover, recall from Section 5.1.2 that every element in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ is interval decomposable [CB12, Theorem 1.1]. We define the Bottleneck distance (Definition 4.2.15) in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ as the what induced by ϵ-matchings subject to d_I^T.

Observation 5.1.19. For sake of simplicity we restrict to finite intervals. Let $\langle a, b\rangle$ be any interval (left closed/open and right closed/open) of $\mathbb{R}$ with $a, b \in \mathbb{R}$. Given two interval modules $\Bbbk^{\langle a,b\rangle}$ and $\Bbbk^{\langle c,d\rangle}$ in $\mathbf{Pers}(\Bbbk^{\mathbb{R}})$, their interleaving distance (Definition 4.2.15) is:

$$d_I^T(\Bbbk^{\langle a,b\rangle}, \Bbbk^{\langle c,d\rangle}) = \max(|c-a|, |d-b|).$$

Observation 5.1.20. Let $\Bbbk^{\langle a,b\rangle}$ be an interval module in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ and 0_P the zero of $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$. Their interleaving distance is:

$$d_I^T(\Bbbk^{\langle a,b\rangle}, 0) = \frac{b-a}{2}.$$

We defined ϵ-matchings in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ (Definition 4.2.13) to be induced by d_I^T. Therefore, considering Observations 5.1.19 and 4.2.8, we give the following explicit definition.

Definition 5.1.21. (ϵ-matching in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$) Let $M, N \in \mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$, and $\epsilon \geq 0$. An ϵ-**matching** between the barcodes $\mathbb{B}(M)$ and $\mathbb{B}(N)$ is a bijection $\sigma : \mathcal{I}_M \subseteq \mathbb{B}(M) \to \mathcal{I}_N \subseteq \mathbb{B}(N)$ such that

1. for every $I_M \in \mathcal{I}_M$ their interleaving distance is

$$\mathrm{d}(I_M, \sigma(I_M)) = \max(|\sup(I_M) - \sup(\sigma(I_M))|, |\inf(I_M) - \inf(\sigma(I_M))|) \leq \epsilon,$$

2. for every $I \in \mathcal{B}(M)/\mathcal{I}_M$ or $I \in \mathbb{B}(N)/\mathcal{I}_N$, we have $\dfrac{sup(I) - inf(I)}{2} \leq \epsilon$.

The **Bottleneck distance in** $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ is induced by the ϵ-matchings in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$associated to d_I^T.

Remark 5.1.22. Given two barcodes $\mathbb{B}(M)$ and $\mathbb{B}(\mathbb{N})$, the Bottleneck distance in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ can be computed in $\mathcal{O}(n^{1.5} log(n))$ where $n = |\mathbb{B}(M)| + |\mathbb{B}(N)|$. See the Hera software article [KMN17].

Isometry Theorem As a major contribution, Lesnick proved in [Les11] the **Isometry Theorem** (cf. Section 4.2.3) for $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$endowed with d_I^T, which establishes the equivalence between the bottleneck distance and the interleaving distance.

Theorem 5.1.23 (The Isometry Theorem, [Les11, Theorem 3.4])**.** *For a pair of persistence modules $M, N \in \mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$the interleaving distance between M and N equals the bottleneck distance between its barcodes $\mathbb{B}(M)$ and $\mathbb{B}(N)$*

$$d_I^T(M, N) = d_B(\mathbb{B}(M), \mathbb{B}(N)).$$

Before this result, there was no clear algorithm to compute the interleaving distance between two elements in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$. But now this can be achieved by just computing the bottleneck distance between their barcodes. Therefore, the Isometry Theorem is crucial for making persistent homology into a practical and helpful tool for the analysis of data.

Stability results for $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ Real data always comes with a certain level of noise, therefore, stability results are essential for TDA. For instance, suppose that we want to learn how to classify two certain individuals from some given samples. If we intend to use barcodes as descriptors of those samples, we need that noise perturbs the barcodes in at most a controlled way. In fact, we would like relevant features to persist and still be reflected in the barcodes. The following results we introduce formalise this notion of stability by upper-bounding the bottleneck-distance. In fact, they assert that the map from data to barcodes is 1-Lipschitz continuous with respect to suitable choices of metrics.

Let $\mathcal{X} = \{x_1, \dots, x_n\}$ be a finite point cloud in $\mathbb{R}^n$.

Proposition 5.1.24. *Choose $\epsilon \geq 0$ and let $\mathcal{Y}$ be an ϵ-perturbation of $\mathcal{X}$. Then the following bounds for the bottleneck distance hold:*

$$
\begin{aligned}
d_B(H_n^{simp} \circ \mathrm{Rips}(\mathcal{X}), H_n^{simp} \circ \mathrm{Rips}(\mathcal{Y})) &\leq 2\epsilon, \\
d_B(H_n^{simp} \circ \check{\mathrm{C}}\mathrm{ech}(\mathcal{X}), H_n^{simp} \circ \check{\mathrm{C}}\mathrm{ech}(\mathcal{Y})) &\leq \epsilon.
\end{aligned}
$$

Proof. The proof follows from the interleaving property of Čech and Vietoris Rips complexes (Proposition 5.1.11), the functoriality of homology and the Isometry Theorem for $1D$-persistence (Theorem 5.1.23). □

We generalise the result of Proposition 5.1.24 to any simplicial sub-level set filtration.

Proposition 5.1.25. *Let $\mathcal{K}$ be a simplicial complex, and let $u, v : \mathcal{K} \to [0, \infty)$. Then*

$$
d_B(H_n^{simp} \circ S_u, H_n^{simp} \circ S_v) \leq \| u - v \|_\infty = \max_{\sigma \in \mathcal{K}} |u(\sigma) - v(\sigma)| = \epsilon.
$$

Proof. The proof is fundamentally similar to Proposition 5.1.11.

Let f, g be the identity morphism on the vertices. Because $\epsilon = \| u - v \|_\infty$, the fact that $v \in S_u(r)$ implies $v \in S_v(r + \epsilon)$, and $v \in S_v(r)$ implies $v \in S_u(r + \epsilon)$. Note that the same argument works for simplices. Therefore, the following diagram commutes:

$$
\begin{tikzcd}
S_u \arrow[r, "f"] \arrow[dr] & S_v[\epsilon] \arrow[r, "g[\epsilon]"] \arrow[dr] & S_u[2\epsilon] \\
S_v \arrow[r, "g"] \arrow[ur] & S_u[\epsilon] \arrow[r, "f[\epsilon]"] \arrow[ur] & S_v[2\epsilon]
\end{tikzcd}
$$

The rest follows from he functoriality of homology and the Isometry Theorem for $1D$-persistent homology (Theorem 5.1.23). □

In fact, the following holds:

Theorem 5.1.26 (Bottleneck stability in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ [CCSG+09, CdSGO12])**.** *For any topological space $\mathbb{X}$ and any functions $u, v : \mathbb{X} \to \mathbb{R}_{\leq}$ such that $H_n^{sing} \circ S_u$ and $H_n^{sing} \circ S_v$ are p.f.d, we have*

$$
d_B(H_n^{sing} \circ S_u, H_n^{sing} \circ S_v) \leq \| u - v \|_\infty = \sup_{p \in \mathbb{X}} |u(p) - v(p)| = \epsilon.
$$

5.1.4 Why go beyond 1D-persistent homology?

As we have seen, $1D$-persistence enjoys many excellent properties. Therefore, at this point, it could seem that it has no drawbacks. However, this section introduces some examples of concerns that motivate us to go beyond the one-parameter case.

Motivation 1: Sensitivity to outliers

Although we can use the results of Section 5.1.3 to endow $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}})$ with metrics that ensure certain stability under noise, this is not sufficient for many applications. Real-world data does not usually come with just noise, but with data points that differ significantly from the rest of the sample, namely, outliers. The Rips and Čech filtrations introduced in Section 5.1.1 are highly sensitive to outliers. To illustrate, consider the three data sets shown in Figure 5.3.

The first data set corresponds to a uniform high-density sample of a circular shape in $\mathbb{R}^3$. The second is a sample of the same circular shape but with added noise, while the third differs from the first by just a few outliers. From the first and the second persistence diagrams of H_1, we observe that the most significant feature (the $1d$-hole of our circular shape) remains almost invariant, as expected from Proposition 5.1.24. Likewise, as an expected consequence, a few features with a short persistence appear, namely, noise. However, between the persistence barcodes of the first and the third data sets, in which the third differs from the first only by the addition of a few points, we obtain a completely different barcode. As it can be appreciated, the significant feature loses relevance, and some new considerable ones appear. Therefore, illustrating the sensitivity of $1D$-persistence to outliers. Moreover, Rips and Čech complex filtrations are insensitive to the geometrical features of high-density regions, as one can easily figure out. This motivates us to consider multifiltrations which may be richer and more stable. In particular, a natural solution could be to consider density as a parameter of our filtration.

Motivation 2: Example of lack of information captured

Consider the sets

$$X' = \{(x,x) \in \mathbb{R}^2 \mid x \in [-1,0)\} \cup \{(x,-x) \in \mathbb{R}^2 \mid x \in [-1,1]\},$$
$$X'' = \{(x,x) \in \mathbb{R}^2 | x \in [0,1]\},$$

and let $X = X' \cup X'' \subset \mathbb{R}^2$, endowed with the usual topology as shown in Figure 5.4.

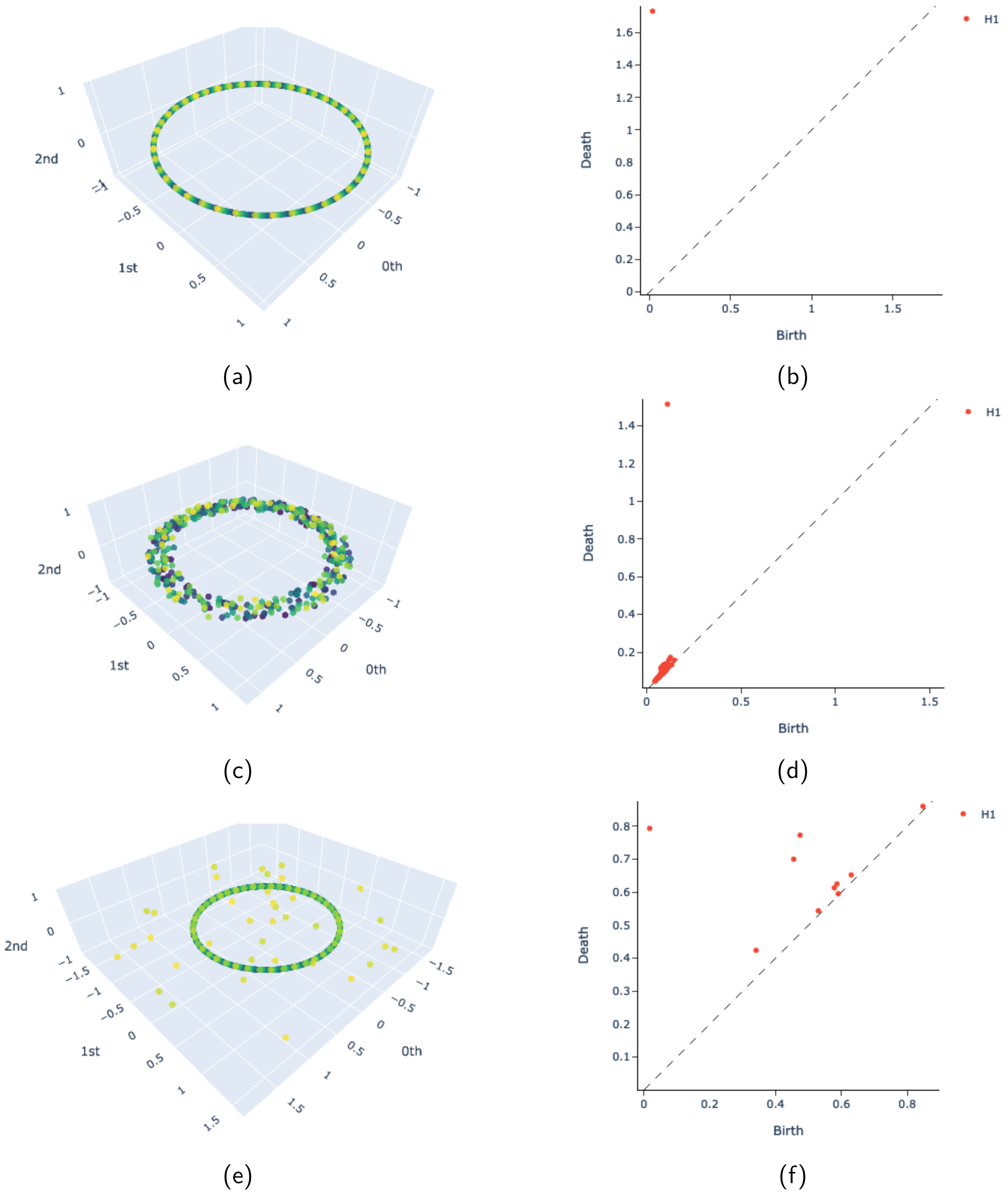


Figure 5.3: **a)** Point cloud of circular shape without noise. **b)** The persistence diagram of H_1 of a). **c)** Point cloud of a circle sampled with noise. **d)** Persistence diagram of H_1 of c). **e)** Point cloud of a circle without noise with an small amount of outliers. **f)** Persistence diagram of H_1 of e).

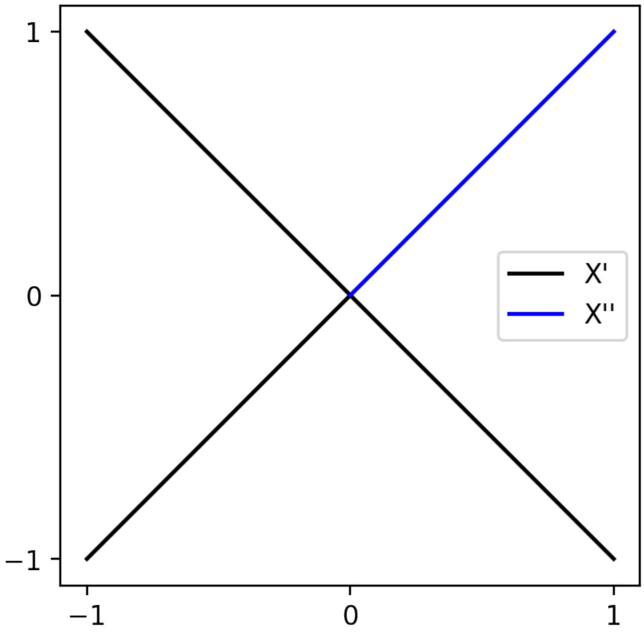


Figure 5.4: Topological space $X \in \mathbb{R}^2$

Let π_y be the second coordinate projection consider the following functions $u, v : X \to \mathbb{R}_{\leq}$ that send a general point $p = (x, y) \in X$ to

$$u(p) := \pi_y(p) = y,$$
$$v(p) := \begin{cases} \pi_y(p) = y \text{ if } p \in X', \\ 0 \text{ if } \in X'' \end{cases}$$

Let $\bullet^1$ be the 1-point discrete space, and $\bullet^2$ the 2-point discrete space.

We introduce homotopy equivalences $(\simeq_h)$ of the preimages of open intervals (y_a, y_b) with $y_a, y_b \in \mathbb{R}_{\leq} \cup \infty$ under u and v. This will allow us to easily compute their 0th-homology.

$$u^{-1}((y_a, y_b)) \simeq_h \begin{cases} \bullet^1 \text{ if } 0 \in (y_a, y_b), \\ \bullet^2 \text{ if } 0 \notin (y_a, y_b) \text{ and } (y_a, y_b) \cap [-1, 1] \neq \emptyset, \\ \emptyset \text{ otherwise.} \end{cases}$$
$$v((y_a, y_b)) \simeq_h \begin{cases} \bullet^1 \text{ if } (y_a, y_b) \cap [0, 1] \neq \emptyset, \\ \bullet^2 \text{ if } y_b \in [-1, 0], \\ \emptyset otherwise. \end{cases}$$

Therefore, we get

$$H_0 \circ S_u \cong H_0 \circ S_v,$$

whence

$$d_I(H_0 \circ S_u, H_0 \circ S_v) = 0.$$

However, it is straightforward that

$$\parallel u - v \parallel_\infty = \max_{p \in X} |u(p) - v(p)| = 1.$$

Therefore, we have illustrated how in even a very simple situation, one-parameter persistent homology is not able to capture sufficient topological information to distinguish two filtrations coming from relatively distinct functions.

Let $\mathbb{U}$ denote the sub-poset of $\mathbb{R}^{op} \times \mathbb{R}$ consisting of intervals (a, b) where $a \leq b$. Given $w : \mathbb{X} \to \mathbb{R}$, we define a refinement of the sub-level set filtration as a functor $\mathcal{L}_w : \mathbb{U} \to \mathbf{Top}$ that satisfies

$$\mathcal{L}_w := \begin{cases} \mathcal{L}_w((a,b)) = w^{-1}([a,b]), \\ \mathcal{L}_w((a,b) \to (c,d)) = w^{-1}([a,b]) \hookrightarrow w^{-1}([c,d]). \end{cases}$$

The filtration $\mathcal{L}_w$ is known as the **inter-level set filtration**. See [CdSGO12] for a discussion of its decomposition theory, and [BL18] for its algebraic stability results.

By its nature, one can easily see that in our example, $\mathcal{L}_u$ is capable to distinguish that $u^{-1}([0,1]) \simeq_h \bullet^2$, something that $S_u(a)$ does not do, since $S_u(a) \simeq_h \bullet^1$ for any $a \in [0,1]$. Therefore, we have informally illustrated that considering new posets may help us to capture more information that might be relevant for our purposes.

5.2. Multiparameter persistent homology

Although the $1D$-persistence theory enjoys many excellent properties, several data archetypes motivate us to consider multifiltrations, some of them already mentioned in Section 5.1.4. The theory of multiparameter persistence homology, or MPH, is the persistence theory associated to multifiltrations. Therefore, multiparameter persistence is a natural choice when working with:

- Datasets with outliers,
- Data with significant variations of density,
- Data endowed with one or several real-valued functions, such as time-varying data.
- Data with tendrils.

A multiparameter persistence module is an object from $\mathbf{Pers}(\Bbbk^{\mathbf{P^n}})$, where $\mathbf{P}^n_{\leq} = \mathbf{T}^{\mathbf{1}}_{\leq} \times \ldots \mathbf{T}^{\mathbf{n}}_{\leq}$ is a product of total order categories. The main example of a multiparameter persistence module is

$$M^u_n := H^{sing}_n \circ \mathcal{S}_u,$$

where H^{sing}_n is the n-th homology functor and $\mathcal{S}_u$ is the sub-level set filtration (Definition 3.1.5) associated to a function of the form $u : \mathbb{X} \to \mathbf{P}^n_{\leq}$. However, we usually refer to multiparameter persistent homology as the theory related with elements in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}^{\mathbf{d}}})$, therefore, associated with sub-level set filtrations $\mathcal{S}_u$ with $u : \mathbb{X} \to \mathbf{R}^{\mathbf{d}}_{\leq}$.

In this section, we first introduce examples of multifiltrations that commonly appear in the multiparameter setting, most of them will naturally arise as a solution to deal with datasets with some of above mentioned properties. For instance, multifiltrations which intend to diminish the impact that outliers have in $1D$-persistence (cf. Section 5.1.4).

Section 5.2.2, it is not all roses for MPH. It turns out that in contrast to $1D$-persistence, the decompostion of multiparameter persistence modules is a hopeless task. Multiparameter persistence modules do not decompose in $\mathbb{R}^d$-interval persistence modules or even into persistence modules defined over nice regions of $\mathbb{R}^d$. Moreover, the barcode for interval decomposable persistence modules is not a complete invariant as in the $1D$ case. A first approach to solve both problems is to define new invariants intending to play a similar role to barcodes.

5.2.1 Multifiltrations

We introduce a few common examples of multi-parameter filtrations that naturally arise in the MPH setting.

As illustrated in Section 5.1.4, $1D$-persistence is very sensitive to density; low-density regions, namely outliers, significantly perturb barcodes, while on the other hand, geometrical features in high-density regions may remain uncaptured. We present one construction that may be helpful to solve that problem.

Definition 5.2.1 (Degree Rips Bifiltration). Let $\mathcal{X}$ be a point cloud in $\mathbb{R}^n$, and $\epsilon \geq 0 \in R$, and $d \geq 0 \in \mathbb{N}$. Let $\mathrm{DRips}(\mathcal{X}, \epsilon, d)$ be the maximal subcomplex of $\mathrm{Rips}(\mathcal{X}, \epsilon, d)$, whose vertices are at least linked by an edge with other $d-1$ vertices. The **degree Rips Bifiltration** is the filtration $\mathrm{DRips}(\mathcal{X})$ such that $\mathrm{DRips}(\mathcal{X})((\epsilon, d)) = \mathrm{DRips}(\mathcal{X}, \epsilon, d)$.

Note that the Degree Rips Bifiltration captures density information, therefore, possibly diminishing the impact of outliers in the filtration. We introduce the sub-level Rips Bifiltration, which is a fairly general type of filtration. In fact, the degree Rips Bifiltration is a particular instance of it.

Sub-level Rips Bifiltration

Definition 5.2.2. Given a topological space $\mathbb{X}$ and a not necessarily continuous function $u : \mathbb{X} \to \mathbb{R}$, the sub-level Rips bifiltration is the functor $Rips_u : \mathbb{R} \times [0, \infty) \to \mathbf{Top}$ that satisfies

$$\mathrm{Rips}_u := \begin{cases} \mathrm{Rips}_u((\rho, \epsilon)) = \mathrm{Rips}(u^{-1}((-\infty, \rho]))_\epsilon, \\ \mathrm{Rips}_u((\rho_a, \epsilon_a) \to (\rho_b, \epsilon_b)) = \mathrm{Rips}(u^{-1}((-\infty, \rho_a]))_{\epsilon_a} \hookrightarrow \mathrm{Rips}(u^{-1}((-\infty, \rho_b]))_{\epsilon_b}. \end{cases}$$

Similarly, we can define the sub-level Čech Bifiltration by using Čech instead of Rips. We introduce examples of sub-level Rips bifiltrations that are helpful for the types of datasets before mentioned.

Example 5.2.3 (Density and outliers). Let us consider a probability density function δ on $X \in \mathbb{R}^n$ and a function $u : X \to \mathbb{R}$ defined by

$$u(x) := \frac{1}{\delta(x)}. \tag{5.4}$$

Note that u induces a sub-level set filtration $S_u : \mathbb{R} \to X$ from high-density regions to lower ones. This motivates us to introduce consider Rips_u, which is a **density-Rips bifiltration**. Other examples come from using Gaussian density functions or nearest neighbour density functions. We present the **density-Rips bifiltration** a possible solution to the problem Section 5.1.4.

Example 5.2.4 (Data with tendrils)**.** Another bifiltration that commonly arises in MPH is the **eccentricity-Rips bifiltration**. The eccentricity Rips bifiltration is the sub-level set Rips bifiltration associated to the **eccentricity function**

$$u : \mathcal{X} \to \mathbb{R}, \text{ such that } u(x) \mapsto \frac{1}{\frac{1}{|X|}\sum_{y \in X} d(x, y)}.$$

The eccentricity Rips bifiltration might be convenient when working with dataset with tendrils emanating from a central core. The bifiltration is able to separate the central core from the spikes, therefore, separating them into clusters.

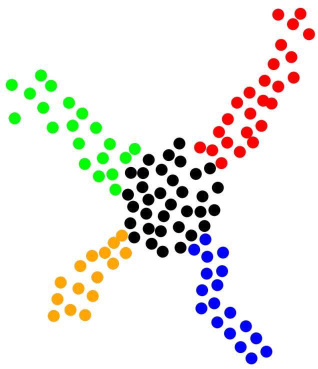


Figure 5.5: Dataset with tendrils emanating from a central core

Example 5.2.5 (Data with intrinsic functions)**.** Another natural motivation to multiparameter persistent homology are data sets $\mathcal{X}$ which come equipped with an intrinsic function $u : \mathcal{X} \to \mathbb{R}$, such as time-dependent data. In this case, the filtering function u is given by $u(x) = t$ where t is the time associated with x.

We refer the reader to [BL22, Section 5] for more common examples of multifiltrations.

5.2.2 Quiver Representations and the Decomposition of Multiparameter persistence modules

In this section we explore the decomposition of multiparameter persistence modules. As we have seen, MPH yields data invariants that are more flexible and richer than the ones arising from $1D$-persistence. As a result, in MPH we also deal with more complex objects. We investigate quiver representation theory and relate it with MPH. From doing so, we discover that any sort of classification of multiparameter persitence modules is completely hopeless.

Introduction to Quiver Representation Theory

Definition 5.2.6 (Quiver)**.** A quiver **Q** is a quadruple $Q = (Q_0, Q_1, h, t)$, where Q_0 is the set of **vertices**, Q_1 the set of **arrows** and h and t are maps assigning to each arrow $a \in Q_1$ its head $h(a) \in Q_0$ and its tail $t(a) \in Q_0$.

Quivers can be thought of as directed graphs, or multigraphs with possibly infinitely many nodes and arrows. In fact, in some cases it is useful to represent them as directed graphs with one vertex per element in Q_0 and one edge (t_a, h_a) per element in Q_1. Whenever both Q_0 and Q_1 are finite sets, we say that the quiver is **finite**.

Here is an example of a graphical representation of a quiver:

$$\bullet 1 \xrightarrow{a} \bullet 2 \xleftarrow{b} \bullet 3 \xrightarrow{c} \bullet 4 \xleftarrow{d} \bullet 5$$

Definition 5.2.7 (Quiver representation)**.** A **representation** of a quiver $Q = (Q_0, Q_1, h, t)$ over a field $\Bbbk$ is a tuple $\mathbb{V} = (V, f)$ where $V = \{V_i \mid i \in Q_0\}$ is a set of $\Bbbk$-vector spaces associated to each of the vertices in V, and $f = \{f_a : V_{t(a)} \to V_{h(a)} \mid a \in Q_1\}$ is a set of $\Bbbk$-linear maps associated to the arrows in Q_1.

We say that a representation $\mathbb{V}$ is **finite-dimensional** if all the vector spaces V_i are finite dimensional. :

Definition 5.2.8 (Quiver subrepresentation)**.** We say that $\mathbb{W} = (W, g)$ is a **subrepresentation** of $\mathbb{V} = (V, f)$ if every $W_i \subseteq V_i$ and $g_a = f_a|_{W_{t(a)}}$.

Now we introduce morphisms between Quivers:

Definition 5.2.9 (Morphism of quiver representations)**.** A **morphism** of representations of Q, $\phi : \mathbb{V} \to \mathbb{W}$ is a collection of linear maps $\phi_i : V_i \to W_i$, such that for every $a \in Q_1$ the following diagram commutes:

$$\begin{tikzcd}
V_{t(a)} \arrow[r, "f_a"] \arrow[d, "\phi(t(a))"'] & V_{h(a)} \arrow[d, "\phi(h(a))"] \\
W_{t(a)} \arrow[r, "g_a"] & W_{h(a)}
\end{tikzcd}$$

For any two morphisms $\phi : \mathbb{V} \to \mathbb{W}$ and $\rho : \mathbb{W} \to \mathbb{X}$ of distinct representations of a quiver Q, the **composition** is $(\rho \circ \phi)_i = \rho_i \circ \phi_i$ for every $i \in V$. Following the same recipe, the **identity morphism** $Id_{\mathbb{V}} : \mathbb{V} \to \mathbb{V}$ of a representation $\mathbb{V}$ can be defined pointwise as $(Id_{\mathbb{V}})_i := Id_{V_i}$.

The category of representations of a quiver Q over a field $\Bbbk$ together with morphisms between them is denoted by $\mathbf{Rep_k(Q)}$. The subcategory of finite dimensional representations is denoted as $\mathbf{rep_k(Q)}$.

Following the same recipe as above and working pointwise one can deduce the following constructions over the category:

- There is a **zero object**, mainly the one assigning to each vertex of the quiver the zero space and to each of the arrows the zero linear map between vector spaces.
- Given any two objects of the category $\mathbb{V}$ and $\mathbb{W}$ we can define a biproduct (**direct sum**) pointwise as follows:
$$(\mathbb{V} \oplus \mathbb{W})_i = V_i \oplus W_i,$$
If in any representation $\mathbb{X} \cong \mathbb{V} \oplus \mathbb{W}$ both $\mathbb{V}$ and $\mathbb{W}$ are distinct from the zero object, we say that $\mathbb{X}$ is a **decomposable** representation. Otherwise, we say it is **indecomposable**.

- The **quotient** of $\mathbb{V}$ by a subrepresentation $\mathbb{W}$ of $\mathbb{V}$ can be defined pointwise as $(V/W)_i = V_i/W_i$. The new induced maps will be the ones induced by the quotient.
- Every morphism $\phi : \mathbb{V} \to \mathbb{W}$ has **kernels**, **images**, and **cokernels**, again easily defined pointwise.

In fact, $\mathbf{Rep_k(Q)}$ is an abelian category. Therefore, morphisms ϕ in the category are **monomorphims** whenever $ker(\phi) = 0$, **epimorphism** whenever $coker(\phi) = 0$ and **isomorphism** only if monomorphism and epimorphism.

We introduce some basic concepts regarding the classification of finite-dimensional representations of finite quivers up to isomorphism.

Theorem 5.2.10 (Krull, Remak, Schmidt)**.** *Given a finite quiver Q, a representation* $\mathbb{V} \in \mathbf{rep_k(Q)}$ *decomposes uniquely (up to permutations) into a finite sum of indecomposable representations*

$$\mathbb{V} = \bigoplus_{j=1}^{k} \mathbb{V}_j.$$

This theorem allows us to focus our attention on the classification of indecomposables up to isomorphism.

Example 5.2.11. A finite dimensional representation of the quiver

$$\bullet 1 \xrightarrow{a} \bullet 2$$

is a pair of vector spaces V_1, V_2 together with a linear map f_a between them:

$$V_1 \xrightarrow{f_a} V_2$$

Up to isomorphism, the only possible indecomposable representations of Q are

$$\begin{aligned} \mathbb{U} &= (\mathbb{k}, 0, 0) \\ \mathbb{V} &= (0, \mathbb{k}, 0) \\ \mathbb{W} &= (\mathbb{k}, \mathbb{k}, \mathrm{id}) \end{aligned}$$

First of all, it is automatic that $\mathbb{U}$ and $\mathbb{V}$ are indecomposable. The only way one could expect to decompose $\mathbb{W}$ would be as the direct sum $\mathbb{U} \oplus \mathbb{V} \cong (\mathbb{k}, \mathbb{k}, 0)$, which definitely

Now, we claim the following:

Lemma 5.2.12. $\mathbb{U}$, $\mathbb{V}$ *and* $\mathbb{W}$ *are the unique indecomposables in* $rep_k(Q)$.

Proof. Assume there exists an indecomposable representation $\mathbb{I} = (\mathbb{k}^n, \mathbb{k}^m, f_a)$ distinct from $\mathbb{U}$, $\mathbb{V}$ and $\mathbb{W}$.

But then

$$\mathbb{I} = (\mathbb{k}^n, \mathbb{k}^m, f_a) \cong (ker(f_a) \oplus V, \mathbb{k}^m, f_a)$$

for some $V \subset \mathbb{k}^m$, and so

$$\mathbb{I} \cong (ker(f_a), 0, 0) \oplus (V, \mathbb{k}^m, \bar{f}_a).$$

However, since $\mathbb{I}$ is indecomposable, either $ker(f_a) = 0$ or $m = 0$.

If $m = 0$, then

$$\mathbb{I} = (\mathbb{k}^n, 0, 0) \cong \bigoplus_{i=1}^{n} \mathbb{U}.$$

Therefore, f_a must be injective.

If, on the other hand, $m \neq 0$ but f_a is injective, we get that

$$\mathbb{I} \cong (\mathbb{k}^n, f, Im(f)) \oplus (0, W, 0),$$

for some vector space W. Again, since we have assumed $\mathbb{I}$ to be indecomposable, either $W = 0$, meaning that f is an epimorphism, or $n = 0$.

In the second case,

$$\mathbb{I} = (0, \mathbb{k}^m, 0) \cong \bigoplus_{i=1}^{m} \mathbb{V}.$$

Therefore, f must be an isomorphism with $n = m$ which automatically makes

$$\mathbb{I} \cong \bigoplus_{i=1}^{n} \mathbb{W},$$

which finally proves that any finite dimensional representation of Q, will be isomorphic to a representation as follows:

$$\mathbb{U}^n \oplus \mathbb{V}^m \oplus \mathbb{W}^r,$$

for all $n, m, r \in \mathbb{N}$. □

In fact, given V_1 and V_2, we can choose bases for V_1 and V_2 in such a way that the matrix representation of f_a is

$$M_{f_a} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \qquad V_1 \xrightarrow{M_{f_a}} V_2,$$

where $r_a = rank(f_a)$ and I_r is the $r \times r$ identity matrix. Therefore, two representations $\mathbb{X} = (X_1, X_2, f_a)$, $\mathbb{Y} = (Y_1, Y_2, g_a)$ are isomorphic if and only if

$$\begin{aligned} d_1 &= dim(X_1) = dim(Y_1), \\ d_2 &= dim(X_2) = dim(Y_2) \qquad \text{and} \\ r_a &= rank(f_a) = rank(g_a), \end{aligned}$$

making every representation $\mathbb{X}$ isomorphic to

$$X \cong \mathbb{U}^{r-d_1} \oplus \mathbb{V}^{r-d_2} \oplus \mathbb{W}^r.$$

Whenever we obtain a finite number of possible indecomposables in $\mathbf{rep_k(Q)}$ for a certain finite quiver as in 5.2.11, we say that our quiver is of **finite type**.

Example 5.2.13. Working over an algebraically closed field $\Bbbk$, consider the quiver

$$\begin{tikzcd} \bullet 1 \arrow[loop above, "a"] \end{tikzcd}$$

and a representation $\mathbb{V} = (V_1, f_a)$ of the quiver "diagram", where f_a is an endomorphism of V_1.

Therefore, given any V_1, we choose a basis such that the matricial representation of f_a is in Jordan normal form.

$$J = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_p \end{bmatrix}, \text{ where each } J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}.$$

Thus, two representations $\mathbb{V} = (V_1, f_a)$ and $\mathbb{W} = (W_1, g_a)$ are isomorphic only if f_a and g_a have the same Jordan normal form. Therefore, the indecomposable representations coincide with the possible distinct Jordan blocks, which are parametrized by the continuous parameter λ and the discrete parameter n that measures the dimension of the block.

Whenever we obtain countably many one parameter families of indecomposables in $\mathbf{rep_k(Q)}$, we say that our quiver is **tame**. Intuitively this means that we can list all indecomposables by parametrizing them, and thus, still have a certain control or knowledge about them.

Example 5.2.14. The following quiver is known as the extended Kronecker quiver K_3:

$$\begin{tikzcd} \bullet 1 \arrow[r, bend left, "a"] \arrow[r, "b"] \arrow[r, bend right, "c"'] & \bullet 2 \end{tikzcd}$$

Surprisingly it turns out that the following proposition regarding K_3 holds.

Proposition 5.2.15. *Solving the classification problem for K_3 would imply solving the classification problem for any quiver Q.*

For a complete proof of this, see [Bar][Proposition 2.1]. Here, we will only outline the tree main steps of the proof.

Let L_t be the quiver having one vertex and t loops. In particular, L_1 is the quiver of Example 5.2.13. However, we will consider the cases with $t \geq 2$.

- **Step 1:** *Solving the classification problem for K_3 implies solving the classification problem for L_2 (cf Figure 5.6).*

A representation of L_2 is given by

$$\mathbb{V} = (V_1, f_a : V_1 \to V_1, f_b : V_2 \to V_2).$$

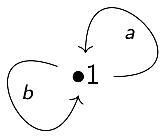


Figure 5.6: The quiver L_2

Therefore, two representations

$$\begin{aligned} \mathbb{V} &= (V_1, f_a : V_1 \to V_1, f_b : V_1 \to V_1), \\ \mathbb{W} &= (W_1, g_a : W_1 \to W_1, g_b : W_1 \to W_1) \end{aligned}$$

are isomorphic if and only if there exists an invertible map $\phi : V_1 \to W_1$ such that

$$\begin{aligned} \phi \circ f_a \circ \phi^{-1} &= g_a, \\ \phi \circ f_b \circ \phi^{-1} &= g_b. \end{aligned}$$

Expressed in terms of the matricial representations of each of the linear maps, say A_a, A_b, B_a and B_b for f_a, f_b, g_a and g_b respectively, and the matricial representation P for ϕ, the latter conditions translate to

$$F \cdot A_a \cdot F^{-1} = A_b \text{ and } F \cdot B_a \cdot F^{-1} = B_b.$$

Therefore, classifying representations of L_2 is contained in the problem of classifying matrices up to simultaneous conjugation relation. Gelfand and Ponomarev in [GP70] proved that this problem in turn contains the problem of classifying n-tuples of matrices up to simultaneous conjugation. Therefore, the problem for the case $n = 2$ contains the problem for an arbitrary n.

As in [Bar][Proposition 2.1, Step 1], given a representation $\mathbb{V}$ of L_2 we can construct a representation of K_3, say $K_3(\mathbb{V})$ as

$$\begin{tikzcd} V_1 \oplus V_1 \arrow[r, bend left=40, "{[1_{V_1},0]}"] \arrow[r, "{[0,1_{V_1}]}"] \arrow[r, bend right=40, "{[f_a,f_b]}"'] & V_1 \end{tikzcd} .$$

At this point it can be proven that $End_{L_2}(\mathbb{V})$ is isomorphic to $End_{K_3}(K_3(\mathbb{V}))$, and that moreover, an element is indecomposable in $\mathbb{V}$ if and only if it is indecomposable in $K_3(\mathbb{V})$, and that the same holds for isomorphic representations.

- **Step 2:** Solving the classification for K_3 implies solving the classification for L_t with $t \geq 2$.

In [Bar][Proposition 2.1, Step 2] they define a representation $L_2(\mathbb{V})$ of L_2 from a representation $\mathbb{V} = (V_1, l_1, \dots, l_t)$ of L_t. First of all, this is done by setting $L_2(V_1) = V_1^{t+1}$. Secondly, the maps $L_2(\mathbb{V})_a$ and $L_2(\mathbb{V})_b$ are defined as matrices with t blocks $(1_{V_1}, \dots, 1_{V_1})$ and $(f_1, \dots, f_t)$ in the superdiagonal, respectively, and zeros elsewhere. From this, it can be proven that the classification for L_t, $t \geq 2$ is contained in the classification of L_2.

- **Step 3:** Solving for K_3 implies solving for any quiver.

Consider a quiver Q with vertex set $Q_0 = \{1, \dots, n\}$ and arrows $Q_1 = \{a_1, \dots, a_r\}$, and suppose we have a representation $\mathbb{V} = (V_1, \dots, V_n, f_1, \dots, f_r)$ of Q. Then we define a representation $L_t(\mathbb{V}) = (L_1, b_1, \dots, b_t)$ of L_t, with $t = n + r$, such that

$$L_1 = V_1 \oplus V_2 \cdots \oplus V_n,$$

and that b_i is represented as the block matrix with only non-zero block 1_{V_i} at position (i, i) for $1 \leq i \leq n$, and b_{n+j} is the block matrix with only non-zero block f_j at position $(t(f_j), h(f_j))$ for $1 \leq j \leq r$. Then, $\mathbb{V}$ is indecomposable if and only if $L_t(\mathbb{V})$ is indecomposable and two representations are isomorphic if they are under $L_t(\bullet)$. Therefore, this shows that the classification of Q is included in the classification of L_t.

In Example 5.2.14 we have seen a quiver whose classification is hopeless. As stated, solving the classification for K_3 would imply solving the classification problem any arbitrary quiver Q. Therefore, having any kind of control over isomorphism classes or indecomposables turns out to be an impossible task. Whenever the latter phenomenon occurs for a given quiver Q, we say that it of **wild type**.

Definition 5.2.16. More formally, a given finite quiver Q is of **wild type** if for any other finite quiver Q', there is a functor $F : \mathbf{rep_k(Q)} \to \mathbf{rep_k(Q')}$ such that

1. F is fully faithful,
2. F is exact, i.e., it preserves exactness (cf. Def. 3.2.4).

Example 5.2.17. The n-subspace quiver is given by a set of $n+1$ vertices and n arrows $a_i : i \to n+1$. For $n = 1$, it is the quiver of example 5.2.11.

For $n = 5$, this corresponds to the quiver with diagram

$$\begin{tikzcd}
\bullet 1 \arrow[rr, "a_1"] & & \bullet 6 & & \bullet 5 \arrow[ll, "a_5"'] \\
& \bullet 2 \arrow[ur, "a_2"] & & \bullet 4 \arrow[ul, "a_4"'] & \\
& & \bullet 3 \arrow[uu, "a_3"] & &
\end{tikzcd}$$

Let $K(\vec{\alpha})$ be the $(n+1) \times (n+1)$ matrix having the n-dimensional vector $\vec{\alpha}$ as superdiagonal and zeros elsewhere. By setting $\lambda = 0$, it turns out that any representation of the following form describes an n-parameter family of non-isomorphic indecomposables.

$$\begin{tikzcd}
\mathbb{K}^{n+1} \arrow[r, "\begin{bmatrix} 1 \\ 0 \end{bmatrix}"] & \mathbb{K}^{2(n+1)} & \mathbb{K}^{n+1} \arrow[l, "\begin{bmatrix} 1 \\ K(\vec{\alpha}) \end{bmatrix}"'] \\
\mathbb{K}^{n+1} \arrow[ur, "\begin{bmatrix} 0 \\ 1 \end{bmatrix}"] & \mathbb{K}^{n+1} \arrow[u, "\begin{bmatrix} 1 \\ 1 \end{bmatrix}"] & \mathbb{K}^{n+1} \arrow[ul, "\begin{bmatrix} 1 \\ J(\lambda) \end{bmatrix}"']
\end{tikzcd}$$

Therefore, the number of non-isomorphic indecomposables increases indefinitely with n. Thus, as in Example 5.2.14, any sort of classification for the representations of the 5-subspace quiver is hopeless. Likewise, we say that the 5-subspace quiver is of **wild type**.

Regarding the classification of quiver representations there are 3 theorems, which altogether are very disappointing.

Theorem 5.2.18 (Drozd,[Dro80])**.** *Every finite quiver is of* **finite**, **tame** *or* **wild** *type.*

Theorem 5.2.19 (Gabriel's Theorem)**.** *A connected quiver is of* **finite** *type if and only if its associated graph is a Dynkin diagram, cf. Figure 5.7.*

Theorem 5.2.20 (Gabriel, Donovan-Freislich, Nazarova)**.** *A connected quiver is* **tame** *if and only if its associated graph is an Euclidean diagram, cf. Figure 5.8.*

These theorems leave us with a very particular kind of trichotomy in the classification of quiver representations. In particular, **finite** and **tame** quivers are very few and concrete. Hence, most of quivers will be of **wild type**, making it impossible to classify their representations. This will have disastrous consequences in the first steps towards multiparameter persistence homology.

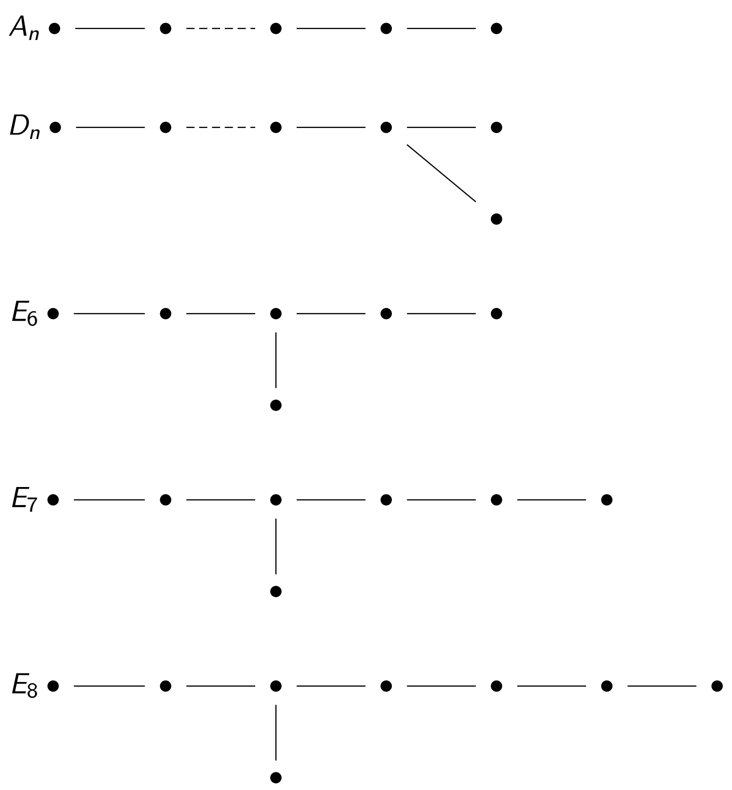


Figure 5.7: Dynkin diagrams.

For further details in the topic of quiver representation theory we suggest [Bar15] as a generic introduction to the field. In addition, we recommend [Oud15] for a brief overview of the topic with persistent homology.

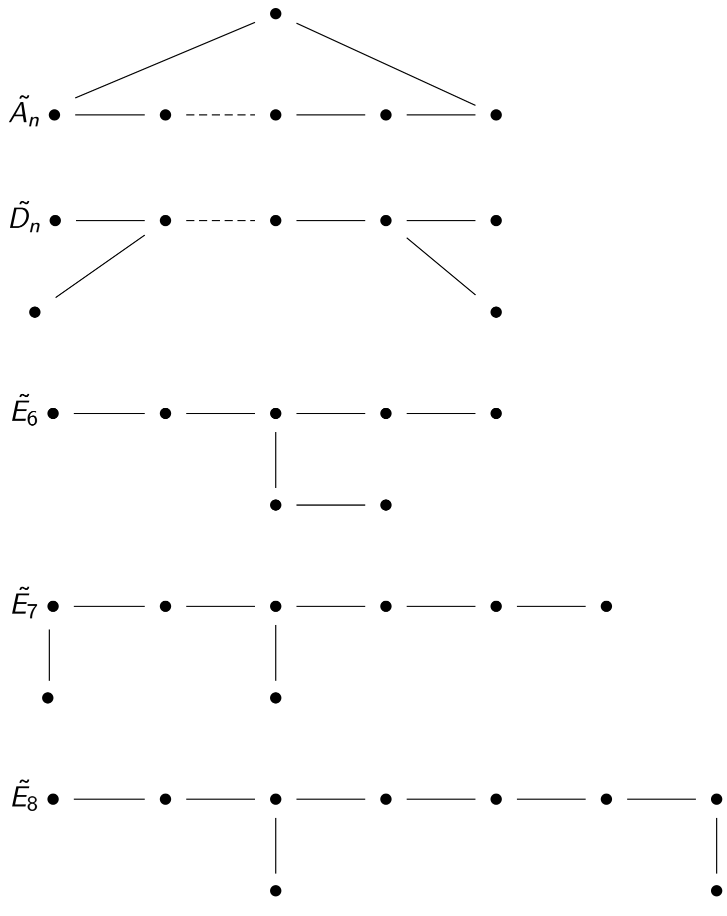


Figure 5.8: Euclidean diagrams.

Multiparameter persistence is going wild

Recall from Example 5.2.17 that the n-subspace quiver is given by a set of $n+1$ vertices and n arrows $a_i : i \to n+1$. Therefore, for the case $n = 5$, the 5-subspace or 5-star quiver Q_5 is the one whose underlying diagram is the following.

$$\begin{tikzcd}
\bullet 1 \arrow[rr, "a_1"] & & \bullet 6 & & \bullet 5 \arrow[ll, "a_5"'] \\
& \bullet 2 \arrow[ur, "a_2"] & & \bullet 4 \arrow[ul, "a_4"'] & \\
& & \bullet 3 \arrow[uu, "a_3"] & &
\end{tikzcd}$$

From this quiver we can infer the wildness of the classification and parametrization of multi-parameter persistent homology already from the base case $n = 2$.

Any sort of representation of Q_5, is going to be given by $\mathbb{V} = (V_1, V_2, \dots, V_6, f_1, f_2, \dots, f_6)$ and can be graphically represented with the diagram

$$\begin{tikzcd}
V_1 \arrow[rr, "f_1"] & & V_6 & & V_5 \arrow[ll, "f_5"'] \\
& V_2 \arrow[ur, "f_2"] & & V_4 \arrow[ul, "f_4"'] & \\
& & V_3 \arrow[uu, "f_3"] & &
\end{tikzcd}$$

We know from Example 5.2.17 that Q_5 is a wild quiver. As it is shown, increasing the dimension of the vector spaces involved in its representation of Fig. 5.9 keeps arbitrarily increasing the number of indecomposable classes appearing in $\mathbf{rep(Q_5)}$.

$$\begin{tikzcd}
\mathbb{K}^{n+1} \arrow[rr, "f_1=\begin{bmatrix}1\\0\end{bmatrix}"] & & \mathbb{K}^{2(n+1)} & & \mathbb{K}^{n+1} \arrow[ll, "f_5(\vec{\alpha})=\begin{bmatrix}1\\K(\vec{\alpha})\end{bmatrix}"'] \\
\mathbb{K}^{n+1} \arrow[urr, "f_2=\begin{bmatrix}0\\1\end{bmatrix}"] & & \mathbb{K}^{n+1} \arrow[u, "f_3=\begin{bmatrix}1\\1\end{bmatrix}"] & & \mathbb{K}^{n+1} \arrow[ull, "f_4(\lambda)=\begin{bmatrix}1\\J(\lambda)\end{bmatrix}"']
\end{tikzcd}$$

Figure 5.9: n-parameter indecomposable representation of $\mathbf{rep(Q_5)}$.

Notice that we can directly relate Q_5 with a poset P_5 containing 5 elements and whose order is determined by the direction of the arrows in Q_5. In this particular case, this relation is bijective. In fact, the left direction is always true; any poset can be seen as a quiver. Therefore, a representation of Q_5 can be seen as a representation of P_5, i.e., a P_5-persistent module over some field $\mathbb{K}$.

When we talk about multi-parameter persistence modules, we are referring to those whose underlying posets are either $\mathbb{R}^d, \mathbb{Z}^d$ or $\mathbb{N}^d$. Therefore, for instance, it will be sufficient to embed a wise choice of a P_5-persistent module into a $\mathbb{N}^2_{\leq}$-persistent module to show that the classification for bipersistent modules is already wild.

Obviously, that P_5-persistent module is the above representation Q_5 Fig 5.9. For convenience, we will just show the restriction to the $[5] \times [5]$ grid representation assuming zeros elsewhere.

$$
\begin{tikzcd}
\mathbb{K}^{n+1} \arrow[r, "f_1"] & \mathbb{K}^{2(n+1)} \arrow[r] & \mathbb{K}^{2(n+1)} \arrow[r] & \mathbb{K}^{2(n+1)} \arrow[r] & \mathbb{K}^{2(n+1)} \\
0 \arrow[r] \arrow[u] & \mathbb{K}^{n+1} \arrow[r, "f_2"] \arrow[u, "f_2"] & \mathbb{K}^{2(n+1)} \arrow[r] \arrow[u] & \mathbb{K}^{2(n+1)} \arrow[r] \arrow[u] & \mathbb{K}^{2(n+1)} \arrow[u] \\
0 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & \mathbb{K}^{n+1} \arrow[r, "f_3"] \arrow[u, "f_3"] & \mathbb{K}^{2(n+1)} \arrow[r] \arrow[u] & \mathbb{K}^{2(n+1)} \arrow[u] \\
0 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & \mathbb{K}^{n+1} \arrow[r, "f_4(\lambda)"'] \arrow[u, "f_4(\lambda)"] & \mathbb{K}^{2(n+1)} \arrow[u] \\
0 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & 0 \arrow[r] \arrow[u] & \mathbb{K}^{n+1} \arrow[u, "f_5(\vec{\alpha})"']
\end{tikzcd}
$$

Figure 5.10: Restriction to the $[5] \times [5]$ grid of our wild persistent P_5-persistent module $\mathbb{N}^2_{\leq}$-embedding.

We have not formalised the discussion into categorical language by building up an exact and fully faithful functor $F : \mathbf{rep}(\mathbf{Q_5}) \to \mathbf{Pers}(\mathbb{k}^{\mathbb{N}^2})$ since we think the construction is pretty obvious and formalising would not contribute to our work.

Regarding the representation of $[n] \times [m]_{\leq}$-persistence modules, the following theorem has been proven:

Theorem 5.2.21 ([Les94]Theorem 2.5)**.** *The posets*

1. $\{[1] \times [m], [n] \times [1], [2] \times [2], [2] \times [3], [2] \times [4], [3] \times [2], [4] \times [2]\}$ *are of* **finite type***,*
2. $\{[2] \times [5], [3] \times [3], [5] \times [2]\}$ *are* **tame***,*
3. **wild** *otherwise.*

Barcodes aren't good

In Section 5.1.2 we prove that in $1D$-persistence, indecomposables are exactly interval persistence modules. In Section 5.2.2, we show that in the multiparameter persistence case, this is far from being true. Moreover, barcodes in $\mathbf{Pers_f}(\mathbb{k}^{\mathbb{R}})$ satisfy an additional nice property that they do not in the multiparameter case. Given a peristence module $M \in \mathbf{Pers_f}(\mathbb{k}^{\mathbb{R}})$ we have

$$\mathrm{rank}(M(a \leq b)) = |\{I \in \mathbb{B}(M) | a, b \in I\}|,$$

namely the features that are alive from a to b. Therefore, the rank being a complete invariant of the decomposition of M (see [CZ09, Theorem 5]). However, we can easily contradict this fact for multiparameter barcodes.

Let $\mathbf{3}$ denote the ordinal category as in Example 2.1.4. Moreover, let P_3 be the poset $P_3 = \mathbf{3} \times \mathbf{3}$. Assume we have the following P_3-persistence module.

$$
\begin{tikzcd}
\Bbbk \arrow[r, "\mathrm{id}"] & \Bbbk \arrow[r] & 0 \\
\Bbbk \arrow[u, "\mathrm{id}"] \arrow[r, "{[1,0]^T}"] & \Bbbk^2 \arrow[u, "{[1,0]}"'] \arrow[r, "{[1,1]}"'] & \Bbbk \arrow[u] \\
0 \arrow[u] \arrow[r] & \Bbbk \arrow[u, "{[0,1]^T}"'] \arrow[r, "\mathrm{id}"'] & \Bbbk \arrow[u, "\mathrm{id}"']
\end{tikzcd}
$$

Suppose that we want to find an interval decomposition. Therefore, because of the identity maps, each of the following two subgrids must belong to an interval.

$$
\begin{tikzcd}
\Bbbk \arrow[r, "\mathrm{id}"] & \Bbbk \arrow[r] & 0 \\
\Bbbk \arrow[u, "\mathrm{id}"] \arrow[r, "{[1,0]^T}"] & \Bbbk^2 \arrow[u, "{[1,0]}"'] \arrow[r, "{[1,1]}"'] & \Bbbk \arrow[u] \\
0 \arrow[u] \arrow[r] & \Bbbk \arrow[u, "{[0,1]^T}"'] \arrow[r, "\mathrm{id}"'] & \Bbbk \arrow[u, "\mathrm{id}"']
\end{tikzcd}
\quad
\begin{tikzcd}
\Bbbk \arrow[r, "\mathrm{id}"] & \Bbbk \arrow[r] & 0 \\
\Bbbk \arrow[u, "\mathrm{id}"] \arrow[r, "{[1,0]^T}"] & \Bbbk^2 \arrow[u, "{[1,0]}"'] \arrow[r, "{[1,1]}"'] & \Bbbk \arrow[u] \\
0 \arrow[u] \arrow[r] & \Bbbk \arrow[u, "{[0,1]^T}"'] \arrow[r, "\mathrm{id}"'] & \Bbbk \arrow[u, "\mathrm{id}"']
\end{tikzcd}
$$

At this point, either the two are distinct intervals, or they belong to the same interval, as in the following figure:

$$
\begin{tikzcd}
\Bbbk \arrow[r, "\mathrm{id}"] & \Bbbk \arrow[r] & 0 \\
\Bbbk \arrow[u, "\mathrm{id}"] \arrow[r, "{[1,0]^T}"] & \Bbbk^2 \arrow[u, "{[1,0]}"'] \arrow[r, "{[1,1]}"'] & \Bbbk \arrow[u] \\
0 \arrow[u] \arrow[r] & \Bbbk \arrow[u, "{[0,1]^T}"'] \arrow[r, "\mathrm{id}"'] & \Bbbk \arrow[u, "\mathrm{id}"']
\end{tikzcd}
$$

However, by definition, if they belong to the same interval, $\mathrm{rank}(M_a \to M_b)$ must be at least 1 for every $a \leq b$. Therefore, since $\mathrm{rank}(M((1,0)) \to M((1,2))) = 0$, they must belong to distinct intervals I and J as in the figure:

$$
I :=
\begin{tikzcd}
\Bbbk \arrow[r, "\mathrm{id}"] & \Bbbk \arrow[r] & 0 \\
\Bbbk \arrow[u, "\mathrm{id}"] \arrow[r, "\mathrm{id}"] & \Bbbk \arrow[u, "\mathrm{id}"'] \arrow[r] & 0 \arrow[u] \\
0 \arrow[u] \arrow[r] & 0 \arrow[u] \arrow[r] & 0 \arrow[u]
\end{tikzcd}
\quad
J :=
\begin{tikzcd}
0 \arrow[r] & 0 \arrow[r] & 0 \\
0 \arrow[u] \arrow[r] & \Bbbk \arrow[u] \arrow[r, "\mathrm{id}"'] & \Bbbk \arrow[u] \\
0 \arrow[u] \arrow[r] & \Bbbk \arrow[u, "\mathrm{id}"'] \arrow[r, "\mathrm{id}"'] & \Bbbk \arrow[u, "\mathrm{id}"']
\end{tikzcd}
$$

However, if they do, $\text{rank}((I \oplus J)((1,0)) \to (I \oplus J)((1,2))) = 0$ as it apparent in the following figure:

$$\begin{array}{ccccc}
\Bbbk & \xrightarrow{\text{id}} & \Bbbk & \xrightarrow{0} & 0 \\
\oplus & & \oplus & & \oplus \\
0 & \xrightarrow{0} & \Bbbk & \xrightarrow{\text{id}} & \Bbbk
\end{array}$$

Since the dimension of $M(a)$ has been fulfilled by $(I \oplus J)(a)$ for every $a \in P_3$, no more intervals can take part on the decomposition. Therefore, proving that the number of persitent features between $a \leq b$ is not equal to the sum of the ranks of the interval persistence modules ranks in which they simultaneously live.

5.2.3 Invariants

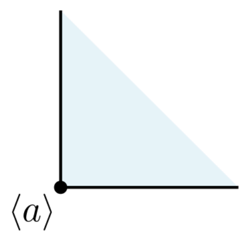
In Section 5.2.2, we have seen that the decomposition of persistence modules in $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}^\mathbf{d}})$ is hopeless. Moreover, we have provided an example illustrating that multiparameter persistence barcodes do not reflect the rank as they do in the one parameter case.

We introduce a few invariants that intend to play a similar role to barcodes in $1D$-persistence. We are at the early stages of developing sufficient theoretical background to understand which of those might be a better choice for topological data analysis techniques. There are many reasonable options; the best depends on the application and the user's purpose.

Free modules and resolutions

Definition 5.2.22 (Free module)**.** Let $\mathbf{P}^n_{\leq} = \mathbf{T}^\mathbf{1}_{\leq} \times \ldots \mathbf{T}^\mathbf{n}_{\leq}$, where each $\mathbf{T}^\mathbf{i}_{\leq}$ is a total order category, and let denote $\langle a \rangle$ to the interval $\{b \in \mathbf{P}^n_{\leq} \mid b \geq a\}$. A persistence module $M \in \mathbf{Pers}(\Bbbk^{\mathbf{P^n}})$ is **free** if it is interval decomposable and every interval is of the form $\langle a \rangle$.

Therefore, a multiparameter persistence module that decomposes in intervals of the form:



Definition 5.2.23 (Free resolution)**.** A **free resolution** $\mathcal{F}$ of a persistence module M is an exact sequence of free persistence modules

$$\ldots \xrightarrow{\partial_3} F_2 \xrightarrow{\partial_2} F_1 \xrightarrow{\partial_1} F_0 \xrightarrow{\partial_0} M \to 0.$$

Therefore, a free resolution is a chain complex of free persistence modules with vanishing homology. Thus, it satisfies that for each $n \in \mathbb{N}$, the persistence submodule $ker(\partial_n)$ of $\mathcal{F}_n$ can be covered by the free persistence module $\mathcal{F}_{n+1}$, in particular, we have $ker(\partial_n) = Img(\partial_{n+1})$.

Definition 5.2.24 (Trivial resolutions)**.** A **trivial resolution** is a direct sum of free resolutions of the form

$$\cdots \to 0 \to 0 \to F \xrightarrow{\mathrm{id}_F} F \to 0 \to \cdots \to 0.$$

Definition 5.2.25 (Minimal resolution)**.** A free resolution $\mathcal{F}$ of a persistence module M **is a minimal free resolution** if every resolution $\mathcal{F}'$ is isomorphic to the sum of $\mathcal{F}$ and a trivial resolution.

Definition 5.2.26 (Free presentation)**.** A **free presentation** $\mathcal{F}$ of a persistence module M is an exact sequence of free persistence modules

$$F_1 \xrightarrow{\partial_1} F_0 \xrightarrow{\partial_1} M \to 0.$$

Therefore, every free presentation is a free resolution.

Definition 5.2.27 (Finitely presented persistence module)**.** A peristence module M is finitely presented if it exist a free presentation $\mathcal{F}$ of M where each of the persistence modules F_0 and F_1 are finitely generated.

The Hilbert function

Definition 5.2.28 (Hilbert function)**.** The **Hilbert function** associated to M is the function $H_M : \mathbb{R}^d \to \mathbb{N}$ that satisfies:

$$H_M(a) := \dim(M(a)).$$

Example 5.2.29.

$$\begin{tikzcd}
\Bbbk \arrow[r, "\mathrm{id}"] & \Bbbk \arrow[r] & 0 & & 1 \arrow[r] & 1 \arrow[r] & 0 \\
\Bbbk \arrow[u, "\mathrm{id}"] \arrow[r, "{[1,0]^T}"] & \Bbbk^2 \arrow[u, "{[1,0]}"'] \arrow[r, "{[1,1]}"'] & \Bbbk \arrow[u] \arrow[r, "H_M"] & & 1 \arrow[u] \arrow[r] & 2 \arrow[u] \arrow[r] & 1 \arrow[u] \\
0 \arrow[u] \arrow[r] & \Bbbk \arrow[u, "{[0,1]^T}"'] \arrow[r, "\mathrm{id}"'] & \Bbbk \arrow[u, "\mathrm{id}"'] & & 0 \arrow[u] \arrow[r] & 1 \arrow[u] \arrow[r] & 1 \arrow[u]
\end{tikzcd}$$

Slightly perturbing our space might result in a completely different Hilbert function. However, it has been proved in a recent paper by Steve Oudot and Luis Sccocola [OS21] that for a particular choice of metric, the Hilber functions satisfy a certain notion of stability for finitely presentable 2-parameter persistence modules.

The rank invariant and the fibered barcode

The rank invariant was introduced by Carlsson and Zoromodian in his joint article [CZ09] and it is an obvious refinement of the Hilbert function.

Definition 5.2.30 (The rank invariant)**.** Given a persistence module $M \in$ its **rank invariant** of M is the function $\mathrm{rank}_M(a \leq b) := \mathrm{rank}(M(a) \to M(b))$.

From Section 5.2.2, the rank invariant is an incomplete invariant for multiparameter persistence modules, but complete for $1D$-persistence modules. The rank invariant is equivalent to the **fibered barcode**.

Let $\mathcal{L}$ be the set of parametrized affine lines in $\mathbb{R}^d$ with non-negative slope. For a given $M \in$ and $L \in \mathcal{L}$, the restriction of M to L is a functor $M|_L : L \to \mathbf{Vect(k)}$. Since M is p.f.d, then $M|_L : L \to \mathbf{Vect(k)}$is a one parameter p.f.d. persistence module. Therefore, $M|_L$ is interval decomposable, where the intervals are intervals of L. We denote its barcode by $\mathbb{B}(M|_L)$.

Definition 5.2.31 (The fibered barcode)**.** The **fibered barcode** of a $M \in \mathbf{Pers_f}(\Bbbk^{\mathbb{R}^{\mathbf{d}}})$ is the function F with domain $\mathcal{L}$ that satisfies

$$F(L) = \mathbb{B}(M|_L).$$

Proposition 5.2.32. *The rank invariant is equivalent to the fibered barcode.*

Proof. For any $a \leq b \in \mathbb{R}^d$ consider the unique line L passing through a and b. Since, the rank invariant is a complete invariant for $1D$-persistence modules, we have that $rank_M(a \leq b) = \{I \in \mathbb{B}(M|_L) | a, b \in I\}$ and that the collection of ranks $\{rank_M(a \leq b)\}_{a \leq b}$ determines $\mathbb{B}(M|_L)$. $\square$

Therefore, for $n \geq 2$ is an incomplete invariant.

Theorem 5.2.33. *[Lan14, Theorem 1] Given two persistence modules $M, L \in$ $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}^{\mathbf{d}}})$ that are ϵ-interleaved, and any line L of finite slope $\vec{\Delta} = (\delta_1, \dots, \delta_d)$, we have:*

$$d_B(\mathbb{B}(M|_L), \mathbb{B}(N|_L)) \leq \frac{\epsilon}{\delta *} d_I(M, N),$$

where $\delta * = \min_i(\delta_i) > 0$.

Multigraded Betti Numbers

Theorem 5.2.34 (Hilbert Theorem)**.** *[Pee10] Let $M \in$ $\mathbf{Pers_f}(\Bbbk^{\mathbb{R}^{\mathbf{d}}})$ be a finitely presented multiparameter persistence module, then M admits a minimal resolution by free modules of length at most n, which is unique up to isomorphism*

$$F_n \xrightarrow{\partial_n} \dots \xrightarrow{\partial_3} F_2 \xrightarrow{\partial_2} F_1 \xrightarrow{\partial_1} F_0 \xrightarrow{\partial_0} M \to 0.$$

Therefore, the generators of each of the free modules F_n of the resolution are characterised by the isomorphism class of M.

Definition 5.2.35 (Multigraded Betti numbers)**.** Let $M \in \mathbf{Pers_f}(\Bbbk^{\mathbb{R}^{\mathbf{d}}})$ be a persistence module which has a minimal resolution

$$F_n \xrightarrow{\partial_n} \dots \xrightarrow{\partial_3} F_2 \xrightarrow{\partial_2} F_1 \xrightarrow{\partial_1} F_0 \xrightarrow{\partial_0} M \to 0,$$

then, the i-th graded Betti number of M is

$$\beta_i(M) = \mathbb{B}(F_i).$$

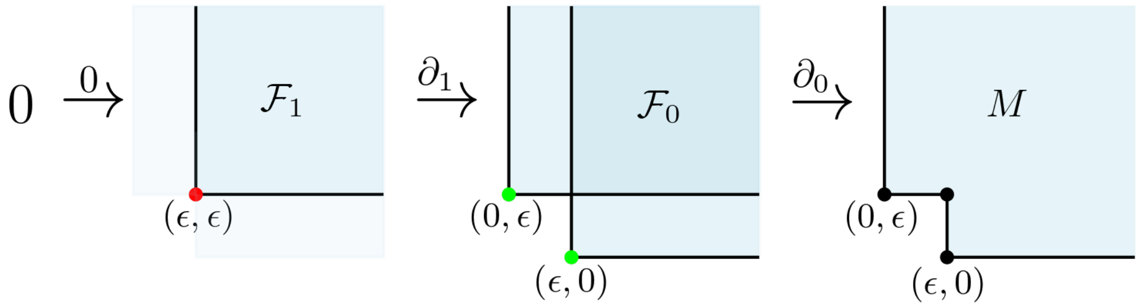


Figure 5.11: Minimal resolution of a bipersistence module

Example 5.2.36 (Computing bigraded Betti numbers)**.** Let M be the bipersistence module

$$M(a) := \begin{cases} \mathbb{k} \text{ if } a \in \mathbb{R}^+/[0,\epsilon)^2, \\ 0 \text{ otherwise.} \end{cases}, \quad M(a \to b) := \begin{cases} id_{\mathbb{k}} \text{ if } a, b \in \mathbb{R}^+/[0,\epsilon)^2, \\ 0 \text{ otherwise.} \end{cases}.$$

represented in Figure 5.11. To compute the bigraded Betti numbers of M we need to generate a minimal free resolution of M.

We start covering M with the free module $\mathcal{F}_0$ generated by the sum of the two interval modules $\mathbf{k}^{\langle(\epsilon,0)\rangle}$ and $\mathbf{k}^{\langle(0,\epsilon)\rangle}$ (cf. Figure 5.11). Observe that to cover M with free indecomposables we need $\mathbf{k}^{\langle(\epsilon,0)\rangle}$ and $\mathbf{k}^{\langle(0,\epsilon)\rangle}$, if suppressing any of the two, M can not be covered. Therefore, a presentation from $\mathcal{F}_0$ is minimal. We define the morphism of persistence modules

$$\partial_0 : \mathbf{k}^{\langle(\epsilon,0)\rangle} \oplus \mathbf{k}^{\langle(0,\epsilon)\rangle} \to M,$$

such that for every $a \in \mathbb{R}^2$ is the linear map

$$\begin{aligned} \partial_0(a) : \mathbf{k}^{\langle(0,\epsilon)\rangle}(a) \oplus \mathbf{k}^{\langle(\epsilon,0)\rangle}(a) &\longrightarrow M(a) \\ (x,y) &\longmapsto x+y \end{aligned}$$

Therefore, we have the following situation

$$\partial_0(a) := \begin{cases} 1.\ Id_{\mathbb{k}} \text{ if } a \in (\langle(\epsilon,0)\rangle \cup \langle(0,\epsilon)\rangle)/\langle(\epsilon,\epsilon)\rangle, \\ 2.\ x+y \text{ if } a \in \langle(\epsilon,\epsilon)\rangle, \\ 3.\ 0 \text{ otherwise.} \end{cases},$$

For 1. and 3., it is straightforward that $\ker(\partial_0(a)) = \{0\}$. However, if $a \in \langle(\epsilon,\epsilon)\rangle$, we have $\partial_0(a)((x,0)) = \partial_0(a)((0,y))$, and so $\ker(\partial_0(a)) \cong \mathbb{k}$. Therefore, the persistence submodule $ker(\partial_0)$ of $\mathcal{F}_0$ is isomorphic to the interval persistence module $ker(\partial_0) \cong \mathbf{k}^{\langle(\epsilon,\epsilon)\rangle}$. Moreover, we have that the image persistence module of ∂_0 is $Img(\partial_0) \cong M$.

Finally, we define $\mathcal{F}_1$ as $\mathcal{F}_1 = \mathbf{k}^{\langle(\epsilon,\epsilon)\rangle}$ (cf. Figure 5.11). Then, we conclude by setting

$$\delta_1 := Id_{\mathcal{F}_1}.$$

Hence, the bigraded Betti numbers of M are

$$\beta_2(M) = \emptyset,\ \ \beta_1(M) = \{(\epsilon,0)\},\ \ \beta_0(M) = \{(\epsilon,0),(0,\epsilon)\}.$$

We refer the reader to [OS21] for a discussion on the stability of multigraded Betti numbers.

To conclude with the section we use the RIVET software [The20] to provide the reader with an example of a computation and a visualization of the three discussed invariants resulting from the degree Rips Bifiltration of the dataset (a) of Example 5.1.4, the circular shape dataset.

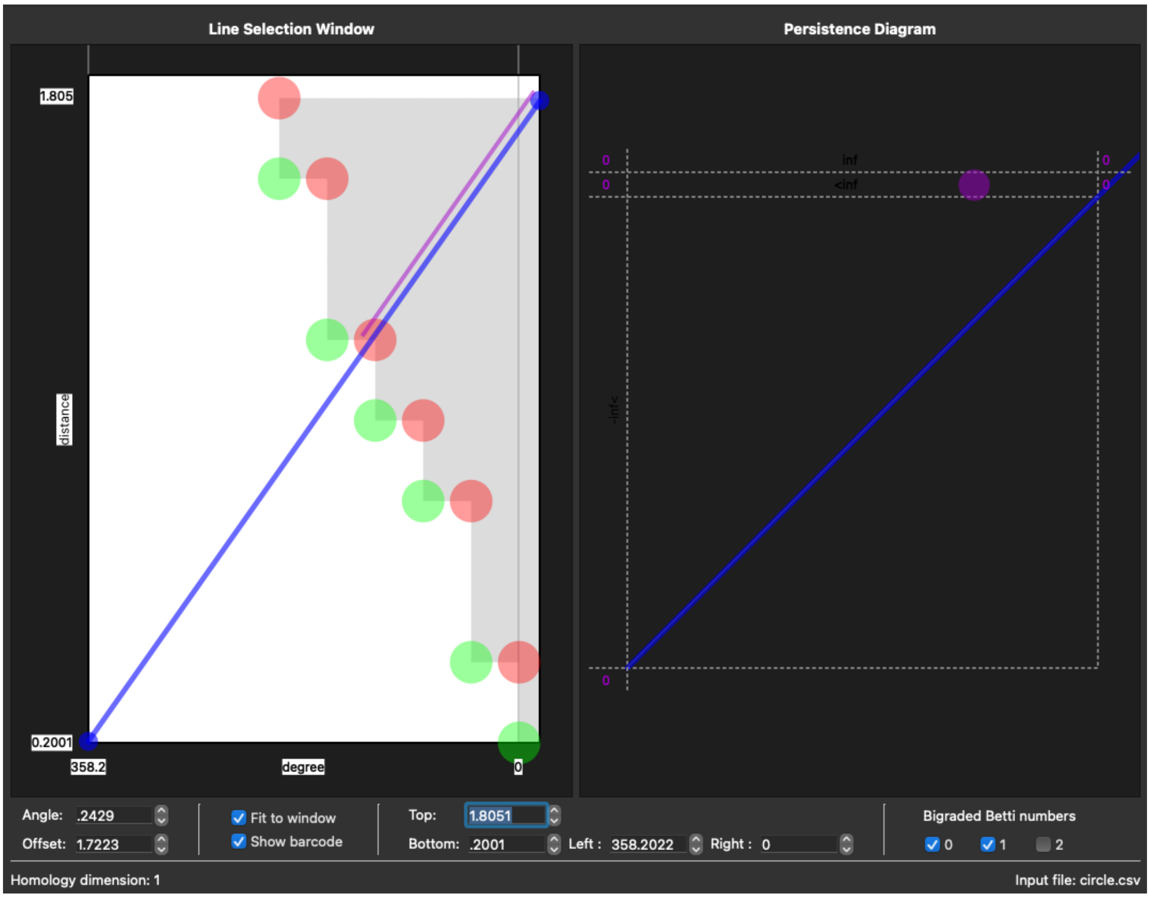


Figure 5.12: The RIVET visualization of the Hilbert function (greyscale shading) , an instance of the fibered barcode (the query line L shown in blue, and the barcode shown in purple) and the 0^{th} (green) and 1^{st} (red) Betti numbers of the degree Rips Bifiltration of dataset (a) of Example 5.1.4

Chapter 6

Final comments and future work

In this work, we have explored Persistence Theory in its full generality. As a particular instance, we first examine $1D$-persistence, which enjoys many excellent and convenient properties for data analysis. However, as demonstrated in Section 5.1.4, it might often be insufficient to capture the structure of interest in the dataset. In Section 5.2, we illustrate how multiparameter persistence is a natural choice when working with:

- Datasets with outliers,
- Data with significant variations of density,
- Data that is naturally equipped with one or several real-valued functions, such as time-varying data.
- Data with tendrils.

As a result, we genuinely believe that Multiparameter Persistence will become as practical and as widely used as $1D$-persistence, although its realm is far more complex (cf. Section 5.2.2).

We want to express our intention to focus our future studies on the following two key themes:

- The development of a general theoretical framework of bottleneck stable invariants for multiparameter persistence modules [OS21, BOOS22, BOO21].
- Investigating the connections between Sheaf Theory and Multiparameter Persistence Theory for seeking richer invariants [Ber20, KS17].

Bibliography

[Azu50] Gorô Azumaya. Corrections and supplementaries to my paper concerning krull-remak-schmidt's theorem. *Nagoya Mathematical Journal*, 1:117–124, 1950.

[Bar] M. Barot. Representation of quivers.

[Bar15] M. Barot. *Introduction to the Representation Theory of Algebras*. Springer International Publishing, 2015.

[BCB18] Magnus Bakke Botnan and William Crawley-Boevey. Decomposition of persistence modules, 2018.

[Ber20] Nicolas Berkouk. *Persistence and Sheaves: from Theory to Applications*. PhD thesis, 09 2020.

[BGND$^+$20] Rickard Brüel-Gabrielsson, Bradley J. Nelson, Anjan Dwaraknath, Primoz Skraba, Leonidas J. Guibas, and Gunnar Carlsson. A topology layer for machine learning, 2020.

[BL18] Magnus Botnan and Michael Lesnick. Algebraic stability of zigzag persistence modules. *Algebraic &amp Geometric Topology*, 18(6):3133–3204, oct 2018.

[BL22] Magnus Bakke Botnan and Michael Lesnick. An introduction to multiparameter persistence, 2022.

[BOO21] Magnus Bakke Botnan, Steffen Oppermann, and Steve Oudot. Signed barcodes for multi-parameter persistence via rank decompositions and rank-exact resolutions, 2021.

[BOOS22] Magnus Bakke Botnan, Steffen Oppermann, Steve Oudot, and Luis Scoccola. On the bottleneck stability of rank decompositions of multi-parameter persistence modules, 2022.

[CB12] William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules, 2012.

[CCSG$^+$09] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry*, SCG '09, page 237–246, New York, NY, USA, 2009. Association for Computing Machinery.

[CdSGO12] Frederic Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules, 2012.

[CkCH+22] Pierre Christian, Chi kwan Chan, Anthony Hsu, Feryal Ozel, Dimitrios Psaltis, and Iniyan Natarajan. Topological data analysis of black hole images, 2022.

[CS18] Nicholas J. Cavanna and Donald R. Sheehy. The generalized persistent nerve theorem, 2018.

[CZ09] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Discrete & Computational Geometry*, 42(1):71–93, 2009.

[Den12] Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

[DI12] Stefan Dantchev and Ioannis Ivrissimtzis. Efficient construction of the čech complex. *Computers & graphics.*, 36(6):708–713, March 2012.

[Dro80] Ju A Drozd. Tame and wild matrix problems. 1980.

[dSMS17] Vin de Silva, Elizabeth Munch, and Anastasios Stefanou. Theory of interleavings on categories with a flow, 2017.

[Ghr07] Robert Ghrist. Barcodes: The persistent topology of data. Technical report, 2007.

[GK18] Marian Gidea and Yuri Katz. Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications*, 491:820–834, Feb 2018.

[GP70] I.M. Gelfand and V.A. Ponomarev. Problems of linear algebra and classification of quadruples of subspaces in a finite-dimensional vector space. *Colloq. Math. Soc. Janos Bolyai*, 5:163–237, 01 1970.

[GT19] Adélie Garin and Guillaume Tauzin. A topological "reading" lesson: Classification of mnist using tda, 2019.

[KMN17] Michael Kerber, Dmitriy Morozov, and Arnur Nigmetov. Geometry helps to compare persistence diagrams. *Journal of Experimental Algorithmics (JEA)*, 22:1–20, 2017.

[KS17] Masaki Kashiwara and Pierre Schapira. Persistent homology and microlocal sheaf theory, 2017.

[Lan14] Claudia Landi. The rank invariant stability via interleavings. *CoRR*, abs/1412.3374, 2014.

[Les94] Zbigniew Leszczyński. On the representation type of tensor product algebras. *Fundamenta Mathematicae*, 144(2):143–161, 1994.

[Les11] Michael Lesnick. The optimality of the interleaving distance on multidimensional persistence modules. *CoRR*, abs/1106.5305, 2011.

[Les12] Michael Lesnick. *Multidimensional interleavings and applications to topological inference*. PhD thesis, 2012.

[NLC11] Monica Nicolau, Arnold J. Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011.

[OS21] Steve Oudot and Luis Scoccola. On the stability of multigraded betti numbers and hilbert functions, 2021.

[Oud15] Steve Y. Oudot. *Persistence Theory: From Quiver Representations to Data Analysis*. Number 209 in Mathematical Surveys and Monographs. American Mathematical Society, 2015.

[Pee10] I. Peeva. *Graded Syzygies*. Algebra and Applications. Springer London, 2010.

[SEN+20] Bernadette J. Stolz, Tegan Emerson, Satu Nahkuri, Mason A. Porter, and Heather A. Harrington. Topological data analysis of task-based fmri data from experiments on schizophrenia, 2020.

[SMC07] Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007.

[SRL+21] Amelia Shoemark, Bruna Rubbo, Marie Legendre, Mahmoud R. Fassad, Eric G. Haarman, Sunayna Best, Irma C.M. Bon, Joost Brandsma, Pierre-Regis Burgel, Gunnar Carlsson, Siobhan B. Carr, Mary Carroll, Matt Edwards, Estelle Escudier, Isabelle Honoré, David Hunt, Gregory Jouvion, Michel R. Loebinger, Bernard Maitre, Deborah Morris-Rosendahl, Jean-Francois Papon, Camille M. Parsons, Mitali P. Patel, N. Simon Thomas, Guillaume Thouvenin, Woolf T. Walker, Robert Wilson, Claire Hogg, Hannah M. Mitchison, and Jane S. Lucas. Topological data analysis reveals genotype–phenotype relationships in primary ciliary dyskinesia. *European Respiratory Journal*, 58(2), 2021.

[The20] The RIVET Developers. Rivet, 2020.

[ZC04] Afra Zomorodian and Gunnar E. Carlsson. Computing persistent homology. In Jack Snoeyink and Jean-Daniel Boissonnat, editors, *Proceedings of the 20th ACM Symposium on Computational Geometry, Brooklyn, New York, USA, June 8-11, 2004*, pages 347–356. ACM, 2004.

[Zom01] Afra Joze Zomorodian. *Computing and Comprehending Topology: Persistence and Hierarchical Morse Complexes*. PhD thesis, 2001. AAI3023247.