Introduction to Topological Data Analysis

Ismail Bouhaj under the supervision of prof. Kathryn Hess Bellwald

October 8, 2023

Content

1 Introduction

Inspired from Prof. Vidit Nanda's Lecture notes: "Computational Algebraic Topology".

This document provides a comprehensive exploration of the field of one-dimensional persistence modules, a profoundly influential domain within algebraic topology, which carries profound implications for the analysis of topological data. One-dimensional persistence modules offer a robust mathematical framework for examining the dynamic evolution of connected components within a topological space, across a spectrum of scales. In this context, we will delve into a deep analysis of the most seminal results to emerge from this discipline, elucidating both the theoretical advances and the practical applications that have ensued.

These one-dimensional persistence modules stand as indispensable tools, facilitating the revelation of concealed topological structures across an extensive spectrum of data types and domains. Whether applied to geometric data, network analysis, biological systems, or numerous other fields, their capacity to discern hidden topological patterns is unparalleled. By delving into the contents of this document, readers will gain profound insights into the transformative impact of one-dimensional persistence modules on our understanding of the intricate topological phenomena underlying complex datasets, illuminating a path toward innovative solutions and discoveries in various scientific and mathematical endeavors.

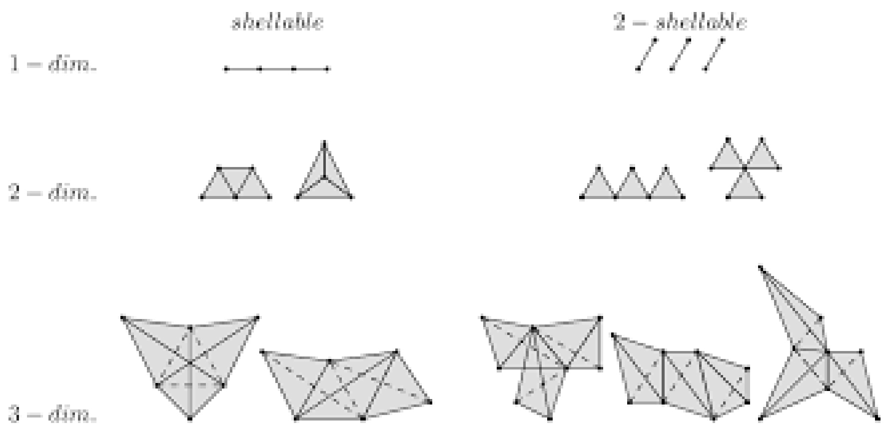
Furthermore, it's worth noting that our examination of one-dimensional persistence modules will be enriched by a category theory perspective. Category theory, a branch of abstract algebraic mathematics, provides a unifying framework for understanding mathematical structures and relationships. By adopting this perspective, we can explore the connections and transformations between different mathematical objects and their relationships within the realm of one-dimensional persistence modules. This categorical lens will allow us to uncover deeper insights into the algebraic and topological interplay inherent to these modules, ultimately contributing to a more profound understanding of their mathematical foundations and applications.

2 The category of simplicial complexes

2.1 Definitions

Definition 2.1 *Let V a finite non-empty set. We call $K \subset 2^V$ a simplicial complex if:*
-$\forall v \in V$ we have that $\{v\} \in K$.
-$\forall \sigma \in K$ we have $\tau \subset \sigma \Rightarrow \tau \in K$.

Remark 2.2 *The elements of V are called vertices.*
The non-empty elements of K are called simplices.
For each simplex σ we define the dimension of σ as $\dim \sigma = \#\sigma - 1$.
The dimension of K is defined as $\dim K = \max\{\dim \sigma | \sigma \in K\}$.
We call K_i the set of all i-dimensional simplices.



Definition 2.3 *Let K, K' two simplicial complexes with V, V' respectively their sets of vertices. A morphism from K to K' is an application $f : V \to V'$ such that for each $\sigma \in K$ we have that $f(\sigma) \in K'$.*

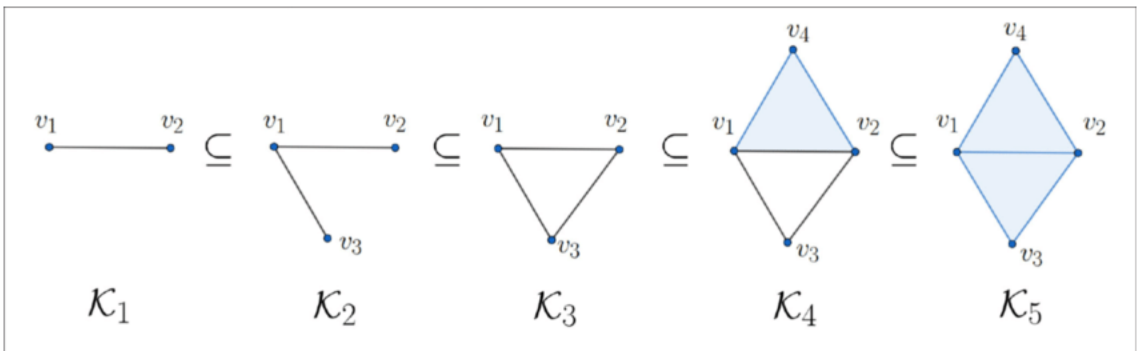
Remark 2.4 *One can easily see that these definitions satisfie the axiomes of a category. We call it the category of simplicial complexes* **SC**.

2.2 Filtration

Definition 2.5 *We say that σ is a face of τ, $\sigma \leq \tau$ if and only if $\sigma \subset \tau$. $L \subset K$ is a sub-complexe if the following holds: for each simplex $\sigma \in L$ if τ is a face of σ in K then $\tau \in L$.*

Definition 2.6 *A filtration of a simplicial complex F is a sequence of sub-complexes of the form:*

$$K_1 \subset K_2 \subset K_1 \subset \cdots \subset K_n = F$$

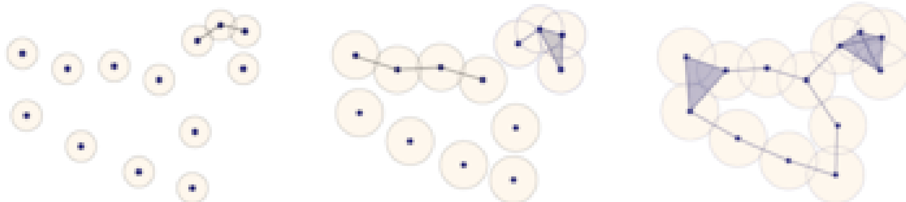


Usually, our data takes the form of a finite metric space (M, d), which can be represented as a list of pairwise distances between points or a weighted graph on the elements of M. In the next section, we will present an algorithm that constructs a filtered simplicial complex from a finite metric space.

2.3 An algorithm for filtration

Definition 2.7 *Let M a finite metric space. The Vietoris-Rips filtration of M is a family of simplicial complexes $(M, \mathbf{VR}_\epsilon)$ defined as follows:*

$$\{x_0, x_1, \ldots, x_k\} \in \mathbf{VR}_\epsilon \iff \forall i < j \leq k \ : d(x_i, x_j) \leq \epsilon$$



(a) Rips complex at some scale (b) ..at a larger scale. (c) ...and at an even larger scale.

It is easy to see that this definition is consistent, as the inclusion property and the singleton membership are verified.
Now observe that if $t_1 < t_2$, we immediately have that $\mathbf{VR}_{t_1}$ is a sub-complex of $\mathbf{VR}_{t_2}$.
The metric space and the distances between its points are finite. This gives us a finite number of t's such that $\mathbf{VR}_t \neq \mathbf{VR}_{t+\epsilon}$ for all $\epsilon > 0$. We call such t's critical points.
Thus, after ordering these t's increasingly, we revert to the definition of a filtration that can be indexed by natural numbers.
We will explore the motivation behind this after presenting persistence modules and persistence homology.

3 Homology

In order to construct homology groups, we need to introduce some preliminary tools. Specifically, we need to assign a notion of orientation to the vertices of a given simplicial complex, thereby creating a chain complex of vector spaces.

3.1 Orientation

Definition 3.1 *An orientation of a simplicial complexe K consist of ordering its vertecies i.e an injective function $o : K_0 \to \mathbb{N}$*

Remark 3.2 *the specific numbers assigned to each vertex are less important than the order in which the vertices are arranged. So, to simplify, we represent a simplex σ as a list like*

$(x_0, x_1, \ldots, x_k)$ where $(o(x_i))_{0\leq i\leq k}$ is an increasing sequence (which determines uniquely the order of x_i's).

3.2 The algebraic boundary functor

In this section we fix a field $\mathbb{K}$

Definition 3.3 *Let K a simplicial complexe. for $k \in \mathbb{N}$, $C_k(K)$ is the $\mathbb{K}$-vector space generate by spanning the k-simplicies (simplicies of dimension k) as a free family.*
We define $\partial_k^K : C_k(K) \to C_{k-1}(K)$ on the basis of $C_k(K)$ as follows:

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

where $\sigma_{-i} = (x_0, \cdots, x_{i-1}, x_{i+1}, \cdots, x_k)$

Definition 3.4 *Let $f \in \mathbf{SC}(K, L)$ be a morphism of simlicial complexes. the boundary map can be extented to f by $C_k f : C_k(K) \to C_k(L)$ defined on the basis of k-simplicies as follows:*

$$C_k f(\sigma) = \begin{cases} f(\sigma) & \textit{if } \dim f(\sigma) = k \\ 0 & \textit{otherwise} \end{cases}$$

Now we present some lemmas to conclude on the functoriality of the boundary map from the category of simplicial complexes **SC** to another cool category.

Lemma 3.5 *Let $f \in \mathbf{SC}(K, L)$. We have that for all $k > 0$ the following diagram commutes*

$$\begin{array}{ccc} C_k(K) & \xrightarrow{\partial_k^K} & C_{k-1}(K) \\ \downarrow{\scriptstyle C_k f} & & \downarrow{\scriptstyle C_{k-1} f} \\ C_k(L) & \xrightarrow[\partial_k^L]{} & C_{k-1}(L) \end{array}$$

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

In this step, we need to verify that the extension satisfies the most crucial relation in the action of functors on morphisms.

Lemma 3.6 *Let $g \in \mathbf{SC}(K, L)$ and $f \in \mathbf{SC}(L, G)$ we have that for all $k \geq 0$:*

$$C_k(f \circ g) = C_k f \circ C_k g$$

We can now confirm that the boundary operator is indeed a functor.

Proposition 3.7 *Let $K \in \mathbf{SC}$ for all $k > 0$ we have that:*

$$\partial_{k-1}^K \circ \partial_k^K = 0$$

This result is very important because it makes important diagrams commute allowing us to construct fundamental objects from a simplicial complex, like homology groups.

3.3 The category of chain complexes

Let $\mathbb{K}$ a field.

Definition 3.8 *A chain complexe consists of a functor F from the poset $\mathbb{N}^{op}$ to $\mathbf{Vect}_{\mathbb{K}}$ such that for all $k > 0$:*

$$F(k + 1 \geq k - 1) = 0$$

We define the category of chain complexes $\mathbf{Chain}_{\mathbb{K}}$ to be sub-category of $\mathbf{FUN}(\mathbb{N}^{op}, \mathbf{Vect}_{\mathbb{K}})$ with the property $F(k + 1 \geq k - 1) = 0$

Remark 3.9 *Note that this definition is consitant as the category of natural numbers is small.*
Concretely this means that we have the following objects $(C_k, d_k)_{k \in \mathbb{N}}$ such that for all $k > 0$.
$d_k \circ d_{k+1} = 0$ *this conditions implies that* $Img\ d_{k+1} \subset \ker d_k$

The morphisms from $(C_k, d_k)_{k\in\mathbb{N}}$ *to* $(C'_k, d'_k)_{k\in\mathbb{N}}$ *are* $(\Phi_k : C_k \to C'_k)_{k\in\mathbb{N}}$ *such that the following diagrams commute:*

$$\begin{array}{ccccccccccccc}
\cdots & \xrightarrow{d_{k+1}} & C_k & \xrightarrow{d_k} & C_{k-1} & \xrightarrow{d_{k-1}} & \cdots & \xrightarrow{d_3} & C_2 & \xrightarrow{d_2} & C_1 & \xrightarrow{d_1} & C_0 \\
 & & \downarrow{\scriptstyle \Phi_k} & & \downarrow{\scriptstyle \Phi_{k-1}} & & \downarrow{\scriptstyle \Phi_\bullet} & & \downarrow{\scriptstyle \Phi_2} & & \downarrow{\scriptstyle \Phi_1} & & \downarrow{\scriptstyle \Phi_0} \\
\cdots & \xrightarrow[d'_{k+1}]{} & C'_k & \xrightarrow[d'_k]{} & C'_{k-1} & \xrightarrow[d'_{k-1}]{} & \cdots & \xrightarrow[d'_3]{} & C'_2 & \xrightarrow[d'_2]{} & C'_1 & \xrightarrow[d'_1]{} & C'_0
\end{array}$$

in other words for all $k > 0$:

$$\Phi_{k-1} \circ d_k = d'_k \circ \Phi_k$$

3.4 The homology group functor

Definition 3.10 *Let* $(C_\bullet, d_\bullet) \in \mathbf{Chain}_{\mathbb{K}}$ *for each* $k \geq 0$ *we define the* k*-th homology group functor on the objects to be the quotient vector space:*

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

Proposition 3.11 *Let* $\Phi_\bullet \in \mathbf{Chain}_{\mathbb{K}}((C_\bullet, d_\bullet), (C'_\bullet, d'_\bullet))$. *For all* $k \geq 0$ *we have that* Φ_k *maps* $\ker d_k$ *to* $\ker d'_k$ *and* $Img\ d_{k+1}$ *to* $Img\ d'_{k+1}$

Remark 3.12 *That means that we have induced maps from morphisms of chain complexes on homology groups. Moreover one can show easily that these induced maps satisfy the functor property.*

Lemma 3.13 *The homology group* $H_\bullet$ *can be extended to morphisms of chain complexes to define a functor from* $\mathbf{Chain}_{\mathbb{K}}$ *to* $\mathbf{FUN}(\mathbb{N}, \mathbf{Vect}_{\mathbb{K}})$. *Where* $\mathbb{N}$ *is the category of natural numbers with only identity morphisms.*

Theorem 3.14 *The composition of the boundary functor and the homology functor defines a functor from the category of oriented simplicial complexes to the following category:* $\mathbf{FUN}(\mathbb{N}, \mathbf{Vect}_{\mathbb{K}})$.

Homology is a fundamental concept in algebraic topology. Intuitively it is cycles modulo boundaries. It helps us study the shape and structure of topological spaces by classifying cycles, which are like closed paths or loops within a space. These cycles may encircle regions or take on various forms.
The ”modulo boundaries” aspect involves recognizing that cycles can sometimes be boundaries of higher-dimensional objects. Homology allows us to understand these relationships and determine when cycles are equivalent if they differ only by boundaries. In essence, it's a way to categorize holes and voids within topological spaces by accounting for cycles and their relations to boundaries.

4 Persistence

Persistence in topological data analysis (TDA) refers to the study of how topological features, such as clusters, loops, or voids, evolve and persist across different scales or levels of resolution within a dataset. TDA employs a filtration process that gradually builds up the dataset's structure, and persistence diagrams are used to visualize and quantify the lifetime or persistence of these topological features throughout the filtration. These diagrams provide valuable insights into the dataset's underlying shape and are central to the analysis and interpretation of TDA results.

4.1 Persistence modules

Definition 4.1 *We define the category of persistence modules* $\mathbf{PM} := \mathbf{FUN}((\mathbb{N}, \leq), \mathbf{Vect}_{\mathbb{K}})$*. Let* $(V_\bullet, a_\bullet) \in \mathbf{FUN}((\mathbb{N}, \leq), \mathbf{Vect}_{\mathbb{K}})$ *for* $i \leq j$ *we define the associated homology group*

$$\mathbf{PH}_{i \to j}(V_\bullet, a_\bullet) = Img\ a_{i \to j}$$

The following diagram describes the nature of the objects and the morphisms between them:

$$\begin{array}{ccccccccccccc}
V_0 & \xrightarrow{a_0} & V_1 & \xrightarrow{a_1} & V_2 & \xrightarrow{a_2} & \cdots & \xrightarrow{a_{k-1}} & V_k & \xrightarrow{a_k} & V_{k+1} & \xrightarrow{a_{k+1}} & \cdots \\
{\scriptstyle f_0}\downarrow & & {\scriptstyle f_1}\downarrow & & {\scriptstyle f_2}\downarrow & & {\scriptstyle f_\bullet}\downarrow & & {\scriptstyle f_k}\downarrow & & {\scriptstyle f_{k+1}}\downarrow & & \\
V'_0 & \xrightarrow[a'_0]{} & V'_1 & \xrightarrow[a'_1]{} & V'_2 & \xrightarrow[a'_2]{} & \cdots & \xrightarrow[a'_{k-1}]{} & V_k & \xrightarrow[a'_k]{} & V_{k+1} & \xrightarrow[a'_{k+1}]{} & \cdots
\end{array}$$

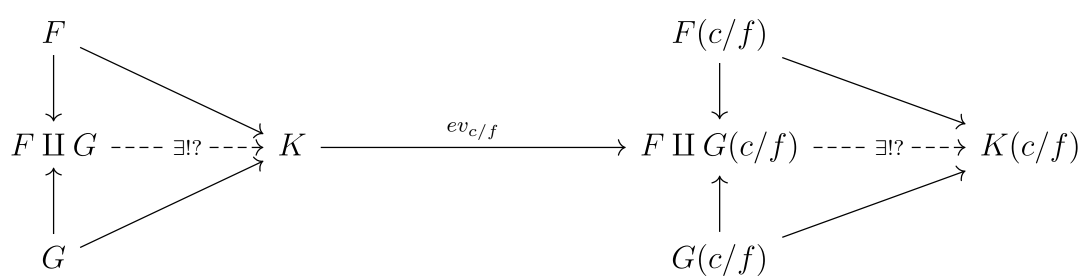
A persistence module is of finite type if for all $i \geq 0$ we have that $\dim V_i < \infty$ and $\exists j > 0$ such that for all $i \geq j$ we have that a_i is an isomorphism. One easily sees that theses conditions are satisfied by k-th homology groups taken from a filtred simplicial complex. But still it's not exactly the same thing as filtration is usually indexed by real numbers ϵ

4.2 Barcode

Topological barcodes allow for the extraction of crucial information about the structure of data. For instance, they can reveal how many connected components exist at each scale, what the dimensions of these components are, and how they evolve with scale. This information can be valuable for tasks such as image segmentation, anomaly detection, pattern recognition, and various other applications in data analysis.

Lemma 4.2 *Let* $\mathbf{C}, \mathbf{D} \in \mathbf{Cat}$ *and* $F, G \in \mathbf{FUN}(C, D)$ *we have that for all* $c \in C$*,*

$$F \amalg G(c) \cong F(c) \amalg G(c)$$



The category of functors inherits more generally limits and co-limits from its co-domain category.

Corollary 4.3 *Let* $(V_\bullet, a_\bullet), (V'_\bullet, a'_\bullet) \in \mathbf{PM}$ *we have that* $(V_\bullet, a_\bullet) \amalg (V'_\bullet, a'_\bullet) \cong (V_\bullet \bigoplus V'_\bullet, a_\bullet \bigoplus a'_\bullet)$.

Definition 4.4 *Let $c \in \mathbf{C}$ an object of a category. c is indecomposable iff the following holds:*

$$c \cong \coprod_{i \in I} U_i \iff \exists! i \in I,\ U_i \text{ not initial}$$

"The" initial object (up to isomorphism) is the following:

$$0 \xrightarrow{0} 0 \xrightarrow{0} 0 \xrightarrow{0} 0 \xrightarrow{0} \cdots$$

Lemma 4.5 *Let $0 \leq i \leq j \leq \infty$ we define $(I_\bullet^{i,j}, c_\bullet^{i,j})$ as follows: for all $i \leq k \leq j$ $I_k = \mathbb{K}$ and 0 otherwise furthermore for all $i \leq k < j$ c_k of rank 1.*

$$\cdots \longrightarrow 0 \xrightarrow{c_{i-1}} \mathbb{K} \xrightarrow{c_i} \mathbb{K} \xrightarrow{c_{i+1}} \cdots \xrightarrow{c_{j-2}} \mathbb{K} \xrightarrow{c_{j-1}} \mathbb{K} \xrightarrow{c_j} 0 \longrightarrow \cdots$$

These are indecomposable objects up to isomorphisms.

Theorem 4.6 *Let $(V_\bullet, a_\bullet) \in \mathbf{PM}$ a persistence module of finite type. There exists a unique set $\mathbf{Bar}(V_\bullet, a_\bullet)$ of indexes (i,j) $0 \leq i \leq j \leq \infty$ and a function $\mu(i,j) \in \mathbb{N}$ such that:*

$$(V_\bullet, a_\bullet) \cong \bigoplus_{(i,j)} (I_\bullet^{i,j}, c_\bullet^{i,j})^{\mu(i,j)}$$

There are several approaches and algorithms to extract the barcode of persistent homology from various data representations. The choice of method will depend on the specific needs of your application, the complexity of the data, and the available resources. It is essential to select the most suitable algorithm for your problem to obtain meaningful results in topological data analysis.

4.3 An equivalence of category

In this section, we aim to establish a notion of distance within the framework of persistence modules, as well as on the barcode representation. This will allow us to quantify the error introduced by various sources of noise, ensuring the consistency of our theoretical foundation. To achieve this, we first need to represent our persistence module using real numbers while maintaining a finite number of points where the space undergoes changes (adding the reference to the Vietoris-Rips complex). This represents an equivalent reformulation but one that can lead to a notion of distance. We will explain all of this in detail in the following section.

Definition 4.7 *An $\mathbb{R}^+$ indexed persistence module is a functor from $(\mathbb{R}^+, \leq)$ to $\mathbf{Vect}_{\mathbb{K}}$. We define the category of $\mathbb{R}^+$ indexed persistence modules by*

$$\mathbf{RPM} := \mathbf{FUN}((\mathbb{R}^+, \leq), \mathbf{Vect}_{\mathbb{K}})$$

This type of persistence module represents a more general scenario than persistence modules indexed by integers. However, in our case, we will consider a specific case referred to as "tame," which brings us back to our previous modeling approach.

Definition 4.8 *We call an $\mathbb{R}^+$ indexed persistence module $(V_\bullet, a_\bullet)$ tame iff for all $t \geq 0$ we have that $\dim V_t < \infty$ and there exists only a finite number t's such that $V_{t-\epsilon \leq t+\epsilon}$ is not an isomorphism for arbitary small $\epsilon > 0$*

We will present a rigourous proof that this particular category of tame $\mathbb{R}^+$ indexed persistence modules is equivalent to the category of finite type persistence modules.

Lemma 4.9 *The following functor F from the category of finite type persistence modules to the category of tame $\mathbb{R}^+$ indexed persistence modules is fully faithful and essentially surjective: $F(V_\bullet, a_\bullet)_t = (V_{\lfloor t \rfloor}, a_{\lfloor t \rfloor})$ and $F(f_\bullet : (V_\bullet, a_\bullet) \to (V'_\bullet, a'_\bullet))_t = f_{\lfloor t \rfloor}$*

This funtor defines an equivalence of category.

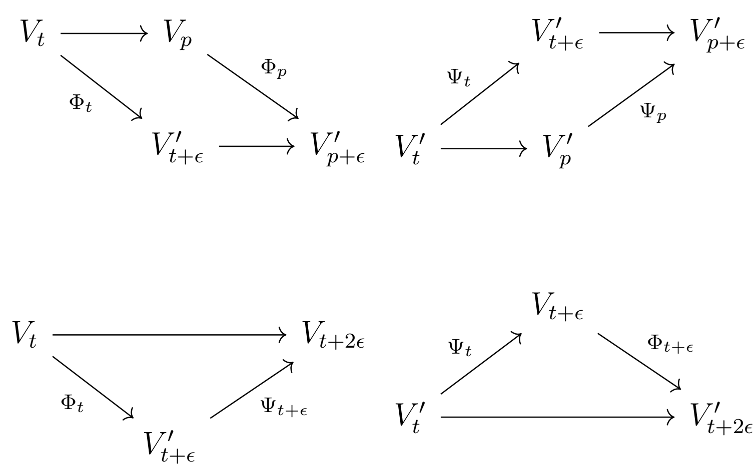
Remark 4.10 *One approach to convert a tame persistence module indexed over the positive real numbers, into a persistence module indexed over the natural numbers, is to establish a total ordering of the critical points according to their increasing values and then associate vector spaces to these points in the specified order. This allows us to systematically transform the original structure into a discrete or countable representation. In this context, the barcode will be represented as a collection of real-valued intervals.*

4.4 Interleaving Distance

Definition 4.11 *An ϵ-interleaving between persistence modules $(V_\bullet, a_\bullet)$ and $(V'_\bullet, a'_\bullet)$ consists of two families of mophismes:*

$$\{\Phi_t : V_t \to V'_{t+\epsilon} |\ t \geq 0\} \ \textit{and} \ \{\Psi_t : V'_t \to V_{t+\epsilon} |\ t \geq 0\}$$

such that for all $0 \leq t \leq p$ real numbers the following diagrams commute:



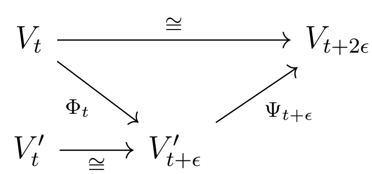
Definition 4.12 *We define the interleaving distance between two class of isomorphims of persistence modules by:*

$$d_{Int}((V_\bullet, a_\bullet), (V'_\bullet, a'_\bullet)) = inf\{\epsilon | \exists \ \textit{an } \epsilon\textit{-interleaving}\}$$

Let's verify that it is indeed a distance:

- by symmetry of $\Psi_\bullet$ and $\Phi_\bullet$ we have that $d_{Int}((V_\bullet, a_\bullet), (V'_\bullet, a'_\bullet)) = d_{Int}((V'_\bullet, a'_\bullet), (V_\bullet, a_\bullet))$
- $d_{Int}((V_\bullet, a_\bullet), (V'_\bullet, a'_\bullet)) = 0 \iff (V_\bullet, a_\bullet) \cong (V'_\bullet, a'_\bullet)$:
  let $(V_\bullet, a_\bullet) \cong_f (V'_\bullet, a'_\bullet)$ we have that f and f^{-1} define a 0-matching between $(V_\bullet, a_\bullet)$ and $(V'_\bullet, a'_\bullet)$. Thus $d_{Int}((V_\bullet, a_\bullet), (V'_\bullet, a'_\bullet)) = 0$.

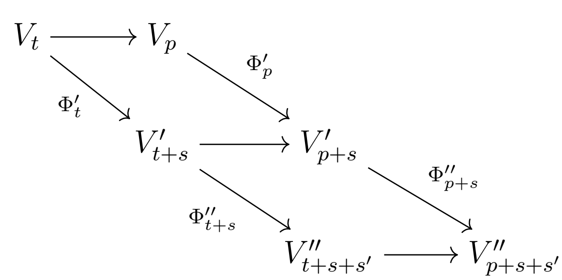
Now let $(V_\bullet, a_\bullet)$ and $(V'_\bullet, a'_\bullet)$ such that $d_{Int}((V_\bullet, a_\bullet), (V'_\bullet, a'_\bullet)) = 0$ let m be the minimal distance between critical points in both persistence modules. Consider an ϵ-matching of $(V_\bullet, a_\bullet)$ and $(V'_\bullet, a'_\bullet)$ such that $2\epsilon < m$. We have the folowwing diagram for all $t > 0$:

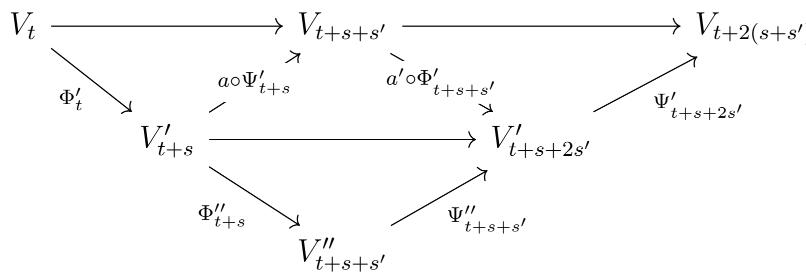


We have that: $a^{-1}_{t\leq t+2\epsilon} \circ \Psi_{t+\epsilon} \circ \Phi_t = id_{V_t}$ and $\Phi_t \circ a^{-1}_{t\leq t+2\epsilon} \circ \Psi_{t+\epsilon} = id_{V'_{t+\epsilon}}$

Thus Φ_t is an isomorphim, therefor $a'^{-1}_{t\leq t+\epsilon} \circ \Phi_t$ is an isomorphim from V_t to V'_t. In conclusion we have that $(V_\bullet, a_\bullet) \cong (V'_\bullet, a'_\bullet)$.

- $d_{Int}((V_\bullet, a_\bullet), (V''_\bullet, a''_\bullet)) \leq d_{Int}((V_\bullet, a_\bullet), (V'_\bullet, a'_\bullet)) + d_{Int}((V'_\bullet, a'_\bullet), (V''_\bullet, a''_\bullet))$:
  Let $a := d_{Int}((V_\bullet, a_\bullet), (V'_\bullet, a'_\bullet))$ and $b := d_{Int}((V'_\bullet, a'_\bullet), (V''_\bullet, a''_\bullet))$ and $\epsilon > 0$
  WLOG suppose that $a \leq b$ by the definitions of a and b we have s, s'-matchings between $(V_\bullet, a_\bullet), (V'_\bullet, a'_\bullet)$ and $(V'_\bullet, a'_\bullet), (V''_\bullet, a''_\bullet))$ respectively such that $a \leq s < a + \epsilon/2$ and $b \leq s' < b + \epsilon/2$ therefore we have the following families of maps:
  $\{\Phi'_t : V_t \to V'_{t+s} |\ t \geq 0\}$, $\{\Psi'_t : V'_t \to V_{t+s} |\ t \geq 0\}$, $\{\Phi''_t : V'_t \to V''_{t+s'} |\ t \geq 0\}$, $\{\Psi''_t : V''_t \to V'_{t+s'} |\ t \geq 0\}$
  Consider the following two families $\{\Phi_t = \Phi''_{t+s} \circ \Phi'_t : V_t \to V''_{t+s+s'} |\ t \geq 0\}$ and $\{\Psi_t = \Psi'_{t+s'} \circ \Psi''_t : V''_t \to V_{t+s+s'} |\ t \geq 0\}$. The following diagrams show that they define an $(s + s')$-matchings between $(V_\bullet, a_\bullet)$ and $(V''_\bullet, a''_\bullet)$ (we can do similarly to prove the commutativity of the other 2 diagrams)





In the second diagram we consider $a_{t+2s \leq t+s+s'} \circ \Psi'_{t+s}$ and $a'_{t+2s+s' \leq t+s+2s'} \circ \Phi'_{t+s+s'}$ we know that such a and a' exist because of the supposition $a \leq b$. In conclusion for all $\epsilon > 0$ we have a $(s+s') =: \delta$-matching between $(V_\bullet, a_\bullet)$ and $(V''_\bullet, a''_\bullet)$ such that $a+b \leq \delta < a+b+\epsilon$ which means that $d_{Int}((V_\bullet, a_\bullet),(V''_\bullet, a''_\bullet))$ is smaller than or equal to $a+b$.

4.5 Bottleneck Distance

A multiset is a collection in which elements are allowed to occur with multiplicities.

Definition 4.13 *For $\epsilon \geq 0$ an ϵ-matching, between two multisets of real valued intervals B and B', is a bijection $\rho : B_0 \to B'_0$ with $B_0 \subset B$, $B'_0 \subset B'$ satisfying:*
(1) Every $[s,t] \in (B - B_0) \cup (B' - B'_0)$ has length $t - s \leq 2\epsilon$,
(2) If $\rho[s,t] = [s_0, t_0]$ for some $[s,t] \in B_0$, then $|s - s_0| \leq \epsilon \geq |t - t_0|$.

Definition 4.14 *we define the bottleneck distance between two multisets of intervals:*

$$d_{Bot}(B, B') := \inf\{\epsilon | \ \exists \epsilon\text{-matching}\}$$

4.6 The stability Theorem, Dmitriy Morozov 2005

Theorem 4.15 *Let $(V_\bullet, a_\bullet)$, $(V'_\bullet, a'_\bullet)$ two tame persistence modules. We have that:*

$$d_{Int}((V_\bullet, a_\bullet),(V'_\bullet, a'_\bullet)) = d_{Bot}(Bar(V_\bullet, a_\bullet), Bar(V'_\bullet, a'_\bullet))$$

Let P a subset of a metric space (M,d) we define $P^{+\epsilon}$ to be $\bigcup_{x \in P} B(x, \epsilon)$

Theorem 4.16 *Let P, Q finite subsets of $\mathbb{R}^n$ such that $\exists\epsilon$ $P \subset Q^{+\epsilon}$ and $Q \subset P^{+\epsilon}$ Then for each dimension $k \geq 0$, the k-th persistent homology modules of the Vietoris-Rips filtrations $\mathbf{VR}_\bullet(P)$ and $\mathbf{VR}_\bullet(Q)$ are 2ϵ-interleaved.*

The Stability Theorems assures that the information obtained from the barcode and homology groups remains reliable and consistent when subjected to small perturbations or variations in the data. This robustness is crucial because real-world data often contains imperfections, and the Stability Theorem provides a mathematical foundation for trustworthiness in TDA results.

5 Summary

5.1 Finite Metric Space (Data):

The analysis begins with a finite metric space, which represents a set of data points where the distances between these points are defined. This data often comes from real-world sources and serves as the basis for topological analysis.

5.2 A simplicial complexes

To uncover the underlying topological features of the data, a Vietoris-Rips filtration for example is constructed. This filtration systematically explores the data at different spatial scales. It involves forming simplicial complexes by connecting points that are within a certain distance threshold. As the threshold increases, the complexes grow, revealing topological structures such as clusters, loops, and voids.

5.3 Homology and Persistence Modules:

The homology of these simplicial complexes is then computed. Homology is a mathematical tool used to quantify topological features. It identifies connected components, loops, and higher-dimensional holes in the data. By tracking how these features persist across the filtration scales, persistence modules are created. These modules encode the lifespan of topological features as they emerge and vanish during the filtration.

5.4 Barcode Extraction:

To represent the persistence modules visually, a barcode diagram is typically generated. Each bar in the diagram corresponds to a topological feature, and its length indicates how long that feature persists in the data. An algorithm processes the persistence modules to extract this barcode, providing a concise summary of the data's topological structure.

5.5 Feature Map Neural Network:

Machine learning techniques are then applied to make sense of these topological features. A feature map neural network is employed, which takes the barcode information as input. The network learns to extract relevant features from the barcodes and encode them into a format suitable for machine learning.

5.6 Model Training and Predictions:

The trained neural network model boasts versatility in handling tasks ranging from classification and regression to image analysis, natural language processing, and recommendation systems. Given barcode-derived features as input, it leverages its learned patterns to provide predictions or classifications across diverse applications. Its adaptability extends to complex domains like anomaly detection, generative tasks, and reinforcement learning, making it a versatile and valuable tool in the field of machine learning and artificial intelligence.