Persistent homology: Stability Theorem

Žiga Virk

December 2, 2021

In the previous chapter we introduced persistent homology and its basic method of computation in the discrete setting. However, it turns out that the concept of persistent homology can be treated on a much deeper theoretical level, through which many of its advantages become apparent.

In this chapter we will delve further into the theoretical machinery of persistent homology. We will introduce continuous filtrations and the underlying algebraic structure of persistence modules. These structures will be crucial in the formulation of the Stability theorem, which states that, unlike homology, persistent homology behaves continuously with respect to the underlying filtration. We conclude by mentioning a series of interpretations and examples of our expanded scope of persistence.

1 Continuous filtrations

Recall that a discrete filtration of a simplicial complex K is a sequence of subcomplexes

$$K_1 \leq K_2 \leq \ldots \leq K_m = K.$$

An example of a filtration if given in Figure 1.
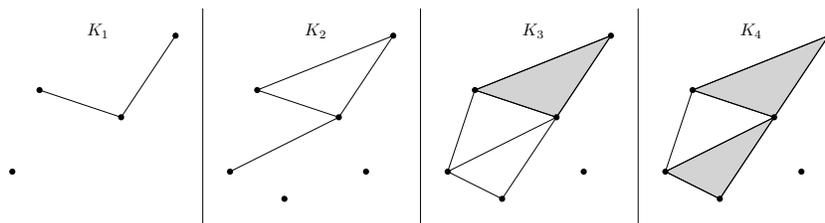
Figure 1: A discrete filtration.

Discrete filtrations[1] formalize finite nested sequences of complexes. While this approach is geometrically intuitive, there is an alternative shorter description of a filtration. Rather than storing a sequence of separate subcomplexes, we **annotate** each simplex $\sigma \in K$ by the index $f(\sigma)$ at which σ first appears. Given such an annotation it is easy to reconstruct $K_i = \{\sigma \in K \mid f(\sigma) \leq i\}$. See Figure 2 for an example. This observation motivates us to expand the scope of filtrations in two ways: by considering continuous filtrations with infinitely many subcomplexes; and by defining filtrations from an appropriate annotation function.
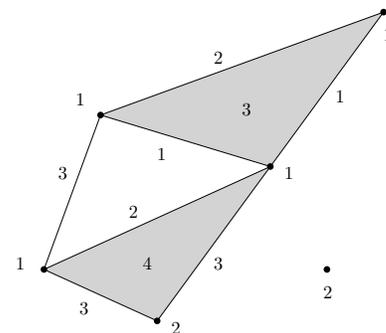
Figure 2: The annotation of simplices encoding the filtration of Figures 1.

[1] I.e., filtrations given by finitely many nested simplicial complexes

1. A **continuous filtration** of a finite simplicial complex K is a collection of subcomplexes[2] $\{K_r\}_{r \geq 0}$ of K such that

$$\forall r < q : K_r \leq K_q \leq K.$$

2. Given a simplicial complex K, let f be a **filtration function**, i.e., an annotation of each of the simplices of K by a non-negative number such that $\sigma \leq \tau \implies f(\sigma) \leq f(\tau)$. The **sublevel filtration** associated to f is a continuous filtrations consisting of sublevel complexes[3] $K_r = \{\sigma \in K \mid f(\sigma) \leq r\} \leq K$ for $r \geq 0$.

There are two motivating reasons for the introduction of continuous filtrations:

- Most of the standard constructions of filtrations actually yield continuous filtrations: Cech filtration[4], Rips filtration[5], filtration by alpha complexes, sublevel filtration, etc.

- The interleaving structure and the resulting stability theorem depend[6] on the continuous choice of parameter.

The definition of persistent homology groups for continuous filtrations is the same, with the only difference being the continuous range of the indices $0 \leq s \leq t$, which results in nominally infinitely many persistent homology groups. The matrix-reduction based computation of persistent homology is unhindered by the expansion to continuous filtrations as the computations depend only on filtration function defined on (finitely many) simplices, and do not require all (infinitely many) complexes of filtration separately. In a nutshell, we can compute the same barcodes using the same procedure for continuous or discrete filtrations.

[2] Throughout the book we will **additionally require** that for each simplex $\sigma \in K$ the minimum $\mathrm{argmin}_r\{\sigma \in K_r\}$ exists, i.e., there exists the smallest scale r at which σ appears. An equivalent condition is the following: for each r there exists $r' > r$ such that $K_r = K_{r'}$, i.e., if a simplex is absent at scale r, it is also absent at slightly larger scales.

Under this condition each continuous filtration is the sublevel filtration of its associated annotation function. In particular, each sublevel filtration is a continuous filtration and vice versa. The Rips and Cech filtrations as defined in Chapter 5 are continuous filtrations of this sort.

[3] In this setting parameter r is often referred to as the scale, a notion arising from Rips and Cech filtrations, or the level, a notion arising from the filtration function.

[4] The corresponding filtration function on simplices being the radius of the smallest enclosing ball of the vertices.

[5] The corresponding filtration function on simplices being the diameter of the set of vertices.

[6] To be discussed in detail later. In a nutshell, the continuous choice of the parameter eventually results in the continuity of persistent homology.
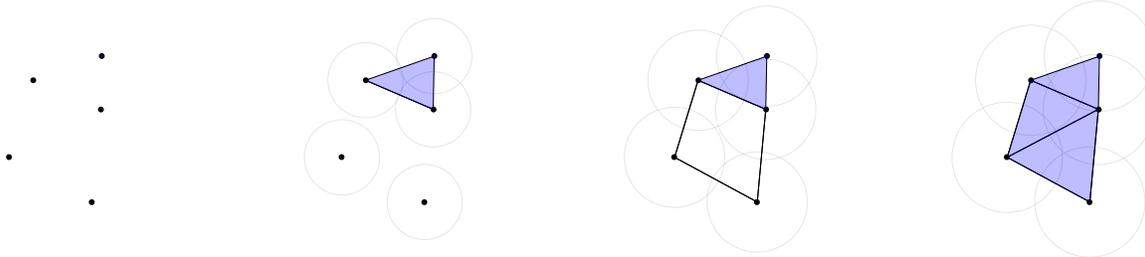
Figure 3: An excerpt from the Rips filtration on the five points on the left.

We next discuss the interpaly between discrete and continuous filtrations:

1. Given a discrete filtration, there is an obvious extension of it as the sublevel filtration of the annotation function.

2. Given a continuous sublevel filtration $\{K_r\}_{r \geq 0}$ associated to a filtration function f there are two ways of generating a discrete filtration:

   (a) By restriction to $K_1 \leq K_2 \leq \ldots \leq K_{\lceil \max f \rceil}$. While mathematically convenient, this approach has many drawbacks[7] and is mostly avoided.

   (b) A more beneficial way of thinking about the index i of a discrete filtration is not as the scale parameter[8] but rather as the index of the **critical scale**[9] of the continuous filtration. Formally speaking we define critical scales $r_1 < r_2 < \ldots < r_k$ as the enumeration[10] of the image of f and define

   $$K_i = \{\sigma \in K \mid f(\sigma) \leq r_i\}.$$

   The corresponding finite filtration contains the information about all changes in the original continuous filtration.

Continuous filtrations conveniently model the geometric setup of the standard filtrations. On the other hand, discrete filtrations are a convenient finite description on which we may develop algorithmic approaches.

$r = 0$ $r = 1/2$ $r = 1/\sqrt{3}$
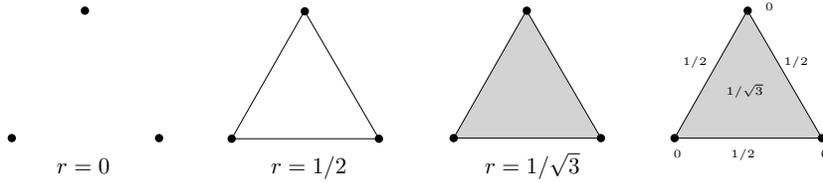
Example 1.1. *[Topology of offsets] Given a finite collection of points $S \subset \mathbb{R}^n$ we have already mentioned that the Nerve Theorem implies that for each $r > 0$ the Cech complex $\mathrm{Cech}(X, r)$ is homotopy equivalent to the r-neighborhood[11] of X:*

$$\mathrm{Cech}(S, r) = \mathcal{N}(\{B(s, r)\}_{s \in S}) \simeq \bigcup_{s \in S} B(s, r) = N(S, r).$$

It turns out that the Nerve Theorem behaves consistently[12] with the maps, which results in the following fact: if for $r_1 < r_2$ the inclusion $N(S, r_1) \hookrightarrow N(S, r_2)$ is a homotopy equivalence, then[13] so is the inclusion $\mathrm{Cech}(S, r_1) \hookrightarrow \mathrm{Cech}(S, r_2)$.

The Cech filtration thus models the homotopy type of growing offsets: if on some interval the growth of r results in homotopy equivalent growth of offsets, then it also results in a homotopy equivalent growth of the Cech complexes.

[7] The corresponding continuous filtration as defined by 1. may be significantly different from $\{K_r\}_{r \geq 0}$. The information about the sequence of changes between each pair of integer scales is lost.

[8] An interpretation prevalent in the context of continuous filtrations.

[9] A scale r of a continuous filtration is **critical**, if at least one simplex appears at r.

[10] $\{r_1, r_2, \ldots, r_k\} = \mathrm{Im} f$.

☞ From this point on, whenever we mention an unspecified filtration, or consider a transition from a finite to continuous filtration or vice versa, the underlying interplays we have in mind are 1. and 2. (b). For an example see Figure 4.

Figure 4: The Cech filtration on three vertices forming an equilateral triangle of side length 1 nominally consists of infinitely many simplices. However, only at scales $0, 1/2$ and $1/\sqrt{3}$ do the changes occur and hence the corresponding discrete filtration (according to 2 (b) above) consists of simplices at those scales, depicted by the first three complexes in the figure. The annotation function is provided on the right, its image consists of the mentioned scales.

[11] Also called the r-offset of S.

[12] The formal term corresponding to this consistency is "functoriality" and the relevant extension of the Nerve Theorem is referred to as "Persistent Nerve Theorem" or "Functorial Nerve Theorem".

[13] At this point we crucially use the fact the Euclidean balls always form a good cover as required by the Nerve Theorem.

Interleaving distance for filtrations

We conclude this section by recalling the interleaving distance between filtrations. The concepts has already been defined in Chapter 5 for Rips and Cech filtrations. With the established general notions we can use the same definition for filtrations in general.

Definition 1.2. *Choose $\varepsilon > 0$. Continuous filtrations $\{K_r\}_{r\geq 0}$ and $\{L_r\}_{r\geq 0}$ are ε-**interleaved** if there exist simplicial maps $\varphi_r\colon K_r \to L_{r+\varepsilon}$ and $\psi_r\colon L_r \to K_{r+\varepsilon}$ such that $\varphi_{r+\varepsilon}\circ\psi_r\colon L_r \to L_{r+2\varepsilon}$ and $\psi_{r+\varepsilon}\circ \varphi_r\colon K_r \to K_{r+2\varepsilon}$ are equal to the corresponding inclusions.*

☞ Suppose filtrations $\{K_r\}_{r\geq 0}$ and $\{L_r\}_{r\geq 0}$ consist of subcomplexes of a simplicial complex K and let $\varepsilon \geq 0$. It is easy to see that if for each r we have $K_r \leq L_{r+\varepsilon}$ and $L_r \leq K_{r+\varepsilon}$, then the filtrations are ε-interleaved. The argument of this sort is used in the proof of Proposition 1.3.

$$\cdots \longrightarrow K_r \longrightarrow K_{r+\varepsilon} \longrightarrow K_{r+2\varepsilon} \longrightarrow \cdots$$
$$\varphi_r \qquad \psi_r$$
$$\cdots \longrightarrow L_r \longrightarrow L_{r+\varepsilon} \longrightarrow L_{r+2\varepsilon} \longrightarrow \cdots$$

Given two filtrations their **interleaving distance** is defined as the minimum[14] of all values $\varepsilon > 0$, for which the filtrations are ε-interleaved. It turns out that the interleaving distance is a metric[15].

In Chapter 5 we proved that Rips and Cech filtrations equipped with the interleaving distance are continuous (stable) with respect to perturbations of the underlying points. Generalizing this result we now prove the sublevel filtrations are continuous with respect to perturbations of the filtration function in the max metric[16].

[14] It is not hard to prove that the minimum exists due to the additional requirement imposed on our filtrations.

[15] In order to maintain this view we declare two filtrations to be **isomorphic** if they are 0-interleaved. The interleaving distance is a metric on the isomorphy classes of filtrations.

[16] Given two functions $f,g\colon K \to \mathbb{R}$ defined on all simplices of a finite simplicial complex K, the **max distance** between them is

$$||f - g||_\infty = \max_{\sigma\in K} |f(\sigma) - g(\sigma)|.$$

Proposition 1.3. *Let K be a simplicial complex. Assume $f,g\colon K \to [0,\infty)$ are filtration functions. Then the sublevel filtrations of K corresponding to K and L are $||f - g||_\infty$ interleaved.*

Proof. In order to align our notation with the diagram above for $\varepsilon = ||f - g||_\infty$ define $K_r = \{\sigma \in K \mid f(\sigma) \leq r\} \leq K$ and $L_r = \{\sigma \in K \mid g(\sigma) \leq r\} \leq K$. The interleaving maps φ, ψ are defined to be identities on vertices. The maps are well defined by the following argument:

- For each vertex $v \in K$: if $v \in K_r$ then $v \in L_{r+||f-g||_\infty}$ by the definition of the max distance. In a similar fashion, if $v \in L_r$ then $v \in K_{r+||f-g||_\infty}$. Hence maps φ, ψ are well defined on vertices.

- The same argument for simplices[17] implies maps φ, ψ are simplicial. □

[17] For example, if a simplex $\sigma \in K$ is contained in K_r, it is also contained in $L_{r+||f-g||_\infty}$.

2 Persistence modules

Persistent homology is obtained by applying homology to a filtration. In this section we present the properties of the resulting algebraic

objects (persistence modules) which model persistent homology. Just as filtrations model the growth of simplicial complexes, persistence modules model the evolution[18] of vector spaces[19].

Throughout the rest of this chapter we fix a field $\mathbb{F}$, which will provide coefficients to all mentioned vector spaces, including homology groups.

Persistence modules

Definition 2.1. *A **persistence module** is a collection of (finite dimensional) vector spaces $\{V_r\}_{r \geq 0}$ along with linear maps*

$$h_{r,q} \colon V_r \to V_q, \quad \forall r < q$$

satisfying $h_{r,q} = h_{r,s} \circ h_{s,q}$ for all $r < q < s$.

*Scale $r \geq 0$ is said to be **regular** if there exists $\varepsilon > 0$ such that maps $h_{p,q}$ are isomorphisms for all $p, q \in (r - \varepsilon, r + \varepsilon)$ or (in the case $r = 0$) or all $p, q \in [0, \varepsilon)$, i.e., the maps h are isomorphisms close to r. Scale r is **critical** if it is not regular.*

☞ In general literature persistence modules may consist of infinite dimensional vector spaces.

☞ Critical scales of a continuous filtration are a supset of critical scales of its persistent homology as any change in homology requires a change of the underlying complex, but not vice versa.

Our interest in persistence modules stems from the fact that they are the underlying algebraic structure of persistent homology of continuous filtrations. In order to simplify our treatment we thus restrict to persistence modules that appear as persistent homology of continuous filtrations as defined above. In particular, each persistence module treated here will be **assumed to have the following properties**:

1. There exists $R > 0$ such that for each $R \leq r < q$ maps $h_{r,q}$ are isomorphisms[20], i.e., eventually all maps h are isomorphisms.

2. For each $r > 0$ there exists $r' > r$ such that for all $q \in [r, r')$ the maps $h_{r,q}$ are isomorphisms[21].

3. There exist finitely[22] many critical scales.

Definition 2.2. *Persistence modules $\{V_r\}_{r \geq 0}$ and $\{W_r\}_{r \geq 0}$ are **isomorphic** if for each $r \geq 0$ there exist isomorphisms $V_r \to W_r$ such that for each $0 \leq r_1 < r_2 < \dots$ the following diagram commutes*

$$\cdots \longrightarrow V_{r_j} \longrightarrow V_{r_{j+1}} \longrightarrow V_{r_{j+2}} \longrightarrow \cdots$$
$$\Big\downarrow \cong \qquad \Big\downarrow \cong \qquad \Big\downarrow \cong$$
$$\cdots \longrightarrow W_{r_j} \longrightarrow W_{r_{j+1}} \longrightarrow W_{r_{j+2}} \longrightarrow \cdots$$

☞ Properties 2. and 3. imply that the interval $[0, \infty)$ can be decomposed into finitely many intervals of the form $[*_1, *_2)$ on which all maps h are isomorphisms.

Decomposition

It is often advantageous to decompose[23] mathematical objects into simple pieces and thus obtain a canonical form. In the previous chapter we decomposed persistent homology into pieces represented by bars. In this subsection we will formalize such a decomposition for persistence modules.

We first explain what we mean by a "decomposition".

Definition 2.3. *The **direct sum** of persistence module $\{V_r\}_{r\geq 0}$ and $\{V_r'\}_{r\geq 0}$ along with respective linear maps $h_{r,q}$ and $h_{r,q}'$, is a persistence module consisting of:*

- *spaces $W_r = V_r \oplus V_r$ and*

- *maps $\tilde{h}_{r,q} = (h_{r,q}, h_{r,q}')$*

for all $0 \leq r < q$.

We next present elementary intervals, which are the pieces represented by bars.

Definition 2.4. *Let $0 \leq p < q$. An **elementary interval** $\mathbb{F}_{p,q}$ corresponding to the pair (p, q) is a persistence module $\{V_r\}_{r\geq 0}$ defined as follows:*

- *$V_r = \mathbb{F}$ for $r \in [p, q)$ and $V_r = 0$ else.*

- *Maps $h_{s,s'}$ are isomorphisms whenever possible.*

☞ Rewriting condition on maps in Definition 2.4:

- for $p \leq s < s' < q$ map $h_{s,s'}$ is the identity on $\mathbb{F}$.

- else $h_{s,s'}$ is the zero map.

☞ It is easy to verify that elementary intervals $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ are isomorphic iff $p = p'$ and $q = q'$.

Theorem 2.5. *[Structure Theorem for persistent homology] Each persistence module is isomorphic to a direct sum of elementary intervals. The decomposition is unique up to the permutation of the intervals.*

Barcodes and bars introduced in the previous chapter correspond to this decomposition and elementary intervals. Theorem 2.5 is an algebraic expression of the existence of barcodes. It states that the persistence module can be decomposed into the intervals and is completely determined[24] by the elementary intervals of its decomposition.

Interleaving distance for persistence modules

The interleaving distance has already been defined for filtrations. Conceptually the same definition applies to persistence modules.

---

[23] Functions are decomposed into monomials (Taylor series) or trigonometric functions (Fourier series). Closed connected surfaces other than the sphere can be decomposed as a direct sum of tori or projective planes. Every n-dimensional vector space is of a form $\mathbb{F}^n$ and in one of the previous appendices we mentioned how finitely generated Abelian groups can be decomposed into smallest indecomposable groups: groups of the form $\mathbb{Z}_p$ and $\mathbb{Z}$.

[24] And as a result, barcodes and persistence diagrams are complete descriptions of persistence modules.

Definition 2.6. *Choose $\varepsilon > 0$. Persistence modules $\{V_r\}_{r\geq 0}$ and $\{W_r\}_{r\geq 0}$ along with respective linear maps $h_{r,q}$ and $h'_{r,q}$ are ε-**interleaved** if there exist linear maps $\varphi_r\colon V_r \to W_{r+\varepsilon}$ and $\psi_r\colon W_r \to V_{r+\varepsilon}$ such that $\varphi_{r+\varepsilon} \circ \psi_r\colon W_r \to W_{r+2\varepsilon}$ and $\psi_{r+\varepsilon} \circ \varphi_r\colon V_r \to V_{r+2\varepsilon}$ are equal $h'_{r,r+\varepsilon}$ and $h_{r,r+\varepsilon}$ correspondingly.*

*Given two persistence modules their **interleaving distance** d_I is defined as the minimum of all values $\varepsilon > 0$, for which the filtrations are ε-interleaved.*

$$\cdots \longrightarrow V_r \longrightarrow V_{r+\varepsilon} \longrightarrow V_{r+2\varepsilon} \longrightarrow \cdots$$
$$\varphi_r \qquad \psi_r$$
$$\cdots \longrightarrow W_r \longrightarrow W_{r+\varepsilon} \longrightarrow W_{r+2\varepsilon} \longrightarrow \cdots$$

It is not hard to prove that the minimum in the definition of the interleaving distance exists due to the additional requirement imposed on persistence modules. It is easy to verify that the interleaving distance is a metric on the isometry classes of persistence modules. As such the interleaving distance is the metric[25] of choice on persistent homologies.

The functoriality of homology implies that ε-interleaved filtrations[26] induce ε-interleaved persistence modules. Another setting in which ε-interleaved persistence modules (but not necessarily ε-interleaved filtrations) are obtained is that of spaces, which are "close" to each other. Let us first define closeness.

Definition 2.7. *Let (X, d) be a metric space and assume $A, B \subset X$ are finite subsets. The **Hausdorff distance** $d_H(A, B)$ is defined as*

$$d_H(A, B) = \max \left\{ \max_{b\in B} \min_{a\in A} d(a,b),\ \max_{a\in A} \min_{b\in B} d(a,b) \right\}.$$

The Hausdorff distance is a metric on all finite subspaces of a metric space X. It has a natural geometric meaning. Given the setting of Definition 2.7 find:

- The minimal r_A such that $N(A, r_A) \supset B$, i.e., the r_A-neighborhood of A contains B.

- The minimal r_B such that $N(B, r_B) \supset A$.

We conclude that $d_H(A, B) = \min\{r_A, r_B\}$. Note that for each $a \in A$ there exists $b \in B$ such that $d(a, b) \leq d_H(a, b)$, and vice versa. An example is given in Figure 5, where a black set A and a red set B are displayed on top, while their respective neighborhoods are displayed in the middle and on the bottom.

[25] At this point it should be clear that continuity and small perturbations of persistent homology depend on the ability to perform continuous and small steps in the index set. An interleaving distance defined on persistent homology of discrete filtrations or a single complex would have been, in the best of cases, restricted to the integer values, that do not accommodate the idea of continuity.

[26] We have already discussed how these appear by perturbing points when using Rips or Cech complexes, and by perturbing the filtration function when using the sublevel filtration.
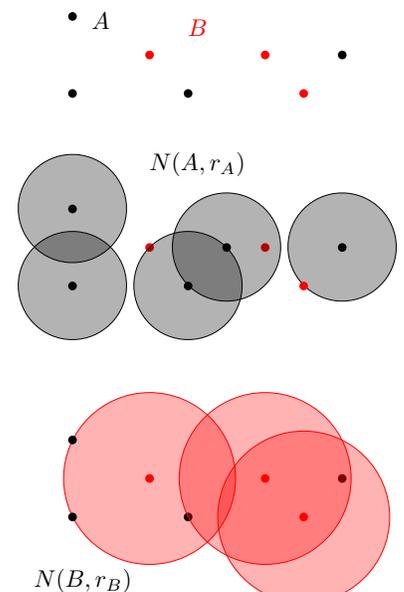
Figure 5: $d_H(A, B) = r_B > r_A$.

Hausdorff distance measures the distances between finite subspaces of a metric space and heavily depends on a way in which these subspaces are embedded. For example, different isometric subspaces will be at a positive Hausdorff distance.

Similar to Hausdorff distance is the Gromov-Hausdorff distance.

Definition 2.8. *Suppose A, B are finite metric spaces. The* **Gromov-Hausdorff distance** $d_{GH}(A, B)$ *is defined as*

$$d_{GH}(A, B) = \inf_{\mu, \nu} \{d_H(\mu(A), \nu(B))\},$$

where the infimum is over all isometric embeddings $\mu \colon A \to X$ and $\nu \colon B \to X$ into a metric space X.

It turns out that the infimum in Definition 2.8 is always attained and that d_{GH} is a metric on the isometry classes[27] of finite metric spaces.

Proposition 2.9. *Let A, B be finite metric spaces with $\varepsilon = d_{GH}(A, B)$. Then for each $q \in \{0, 1, \dots\}$:*

1. *$\{H_q(\mathrm{Rips}(A, r))\}_{r \geq 0}$ and $\{H_q(\mathrm{Rips}(B, r))\}_{r \geq 0}$ are 2ε-interleaved.*

2. *$\{H_q(\mathrm{Cech}(A, r))\}_{r \geq 0}$ and $\{H_q(\mathrm{Cech}(B, r))\}_{r \geq 0}$ are ε-interleaved.*

Proof. We will only sketch the proof for $q = 1$ and Rips filtrations. The proof of other cases follows the same idea but requires some technical diligence. Without loss of generality we may assume A and B are subspaces of a metric space X and $\varepsilon = d_H(A, B)$.

We aim to define maps φ and ψ that constitute a commutative diagram:

$$\cdots \longrightarrow H_1(\mathrm{Rips}(A, r)) \longrightarrow H_1(\mathrm{Rips}(A, r + 2\varepsilon)) \longrightarrow H_1(\mathrm{Rips}(A, r + 4\varepsilon)) \longrightarrow \cdots$$

φ_r

ψ_r

$$\cdots \longrightarrow H_1(\mathrm{Rips}(B, r)) \longrightarrow H_1(\mathrm{Rips}(B, r + 2\varepsilon)) \longrightarrow H_1(\mathrm{Rips}(B, r + 4\varepsilon)) \longrightarrow \cdots$$

We first define maps on the vertices of the Rips complexes:

- For each $a \in A$ choose $b_a \in B$ such that $d(a, b_a) \leq \varepsilon$ and define $\varphi_r(a) = b_a, \forall r$.

- For each $b \in B$ choose $a_b \in A$ such that $d(b, a_b) \leq \varepsilon$ and define[28] $\psi_r(b) = a_b, \forall r$.

Given a 1-cycle $\alpha = \sum_i \langle a_i, a_{i+1} \rangle$ in $\mathrm{Rips}(A, r)$ define $\varphi_r([\alpha]) = [\sum_i \langle b_{a_i}, b_{a_{i+1}} \rangle]$. This gives well defined maps φ_r (and also ψ_r) by the following arguments:

☞ Observe that $d_{GH}(A, B) \leq d_H(A, B)$ for finite subspaces of a metric space X. As a result Proposition 2.9 also holds for d_H. However, Gromov-Hausdorff distance is typically harder to compute and thus it is occasionally more convenient to use d_H as the easily computable parameter of interleaving.

[27] In particular, $d_{GH}(A, B) = 0$ iff the spaces are isometric.

[28] As defined, maps φ_r and ψ_r do not define an interleaving between the Rips filtrations as in general $a_{b_a} \neq a$.

- $\sum_i \langle b_{a_i}, b_{a_{i+1}} \rangle$ is a cycle in $\mathrm{Rips}(B, r+2\varepsilon)$ as $d(a_i, a_{i+1}) \leq r$ implies $d(b_{a_i}, b_{a_{i+1}}) \leq r + 2\varepsilon$.

- If $[\sum_i \langle a_i, a_{i+1} \rangle] = [\sum_i \langle a'_i, a'_{i+1} \rangle]$ holds[29], then $\varphi([\sum_i \langle a_i, a_{i+1} \rangle]) = \varphi([\sum_i \langle a'_i, a'_{i+1} \rangle])$ as well[30].

  At last we need to show that

  $$\left[\sum_i \langle a_i, a_{i+1} \rangle \right] = \left[\sum_i \langle a''_i, a''_{i+1} \rangle \right]$$

  in $H_1(\mathrm{Rips}(A, r+4\varepsilon))$ where $a''_i = a_{b_{a_i}}$. First note that $d(a_i, a''_i) \leq 2\varepsilon, \forall i$. The difference $\sum_i \langle a_i, a_{i+1} \rangle - \sum_i \langle a''_i, a''_{i+1} \rangle$ is a boundary as demonstrated by the blue 2-chain in Figure 6.

  $\square$

3 Bottleneck distance and Stability theorem

The many versions of the Stability Theorem for persistent homology state that persistent homology is continuous with respect to continuous change of the input parameters[31]. We have already seen examples of this sort: through Propositions 2.9 and 1.3 we can conclude that persistent homology behaves "continuously" in the interleaving distance. One of the main advantages of persistent homology though is its visualization and so the final step towards a geometrically convenient form of the Stability Theorem is to interpret[32] the interleaving distance in geometric terms as a distance on persistence diagrams[33]. The resulting distance on persistence diagrams is called the bottleneck distance.

Bottleneck distance

We start by explaining notions and setting needed to define the bottleneck distance. Suppose $\mathcal{A} = (a_1, a_2, \ldots, a_m)$ and $\mathcal{B} = (b_1, b_2, \ldots, b_n)$ are persistence diagrams, i.e.:

- each a_i and b_i is a point above the diagonal in the first quadrant in the plane, and

- each point may appear multiple times in any of the diagrams.

For a point $v = (x, y) \in \mathbb{R}^2$ let[34] $\bar{v} = ((x+y)/2, (x+y)/2) \in \mathbb{R}^2$. A **partial matching** between $\mathcal{A}$ and $\mathcal{B}$ is a bijective map $\varphi \colon \mathcal{A}' \to \mathcal{B}'$ where[35] $\mathcal{A}' \subseteq \mathcal{A}$ and $\mathcal{B}' \subseteq \mathcal{B}$. The **matching distance** of such a φ is defined as

$$d_M(\varphi) = \max \left\{ \max_{v \in \mathcal{A}'} \{ d_\infty(v, \varphi(v)) \}, \max_{v \in \mathcal{A} \setminus \mathcal{A}'} \{ d_\infty(v, \bar{v}) \}, \max_{v \in \mathcal{B} \setminus \mathcal{B}'} \{ d_\infty(v, \bar{v}) \} \right\}.$$

Let $\mu(\mathcal{A}, \mathcal{B})$ denote the collection of all partial matchings between $\mathcal{A}$ and $\mathcal{B}$.

[29] This means $\sum_i \langle a_i, a_{i+1} \rangle - \sum_i \langle a'_i, a'_{i+1} \rangle = \partial \sum_j \langle x_j, y_j, z_j \rangle$.

[30] This holds as $\sum_i \langle b_{a_i}, b_{a_{i+1}} \rangle - \sum_i \langle b_{a'_i}, b_{a'_{i+1}} \rangle = \partial \sum_j \langle b_{x_j}, b_{y_j}, b_{z_j} \rangle$ and $\langle b_{x_j}, b_{y_j}, b_{z_j} \rangle$ are triangles in $\mathrm{Rips}(B, r+2\varepsilon)$.
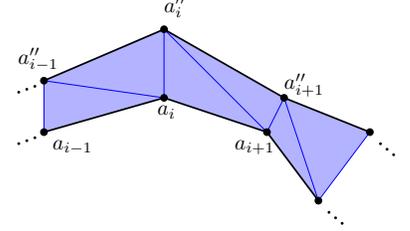
Figure 6: An excerpt from the proof of Proposition 2.9. Each edge connects points at distance at most $r + 2\varepsilon$.

[31] With various versions discussing various forms of input.

[32] A brief idea about a transition from the interleaving distance to the bottleneck distance is provided in appendix.

[33] For this setting, the visualization with persistence diagrams is much preferred to the visualization with the barcodes.

[34] $\bar{v}$ represents the point on the diagonal $\Delta = \{ (z, z) \mid z \in \mathbb{R} \}$ which is the closest to v in d_∞ (and also in d_2) metric.

[35] Again, a point can appear in $\mathcal{A}'$ or $\mathcal{B}'$ multiple times but not more times than in $\mathcal{A}$ or $\mathcal{B}$ respectively.

☞ Recall that the max distance $d_\infty((x_1, y_1), (x_2, y_2))$ between points in the plane is defined as

$$\max\{ |x_1 - x_2|, |y_1 - y_2| \}.$$

Definition 3.1. *The **bottleneck distance** between persistence diagrams $\mathcal{A}$ and $\mathcal{B}$ is the minimal matching distance between them, i.e.,*

$$d_B(\mathcal{A}, \mathcal{B}) = \min_{\varphi \in \mu(\mathcal{A}, \mathcal{B})} d_M(\varphi).$$

Figure 7: Examples of partial matchings between the red and the blue persistence diagrams with with points unmatched by φ being matched to the closest diagonal point.

Examples of partial matchings are given in Figure 7. In order to demonstrate the additional pairs used in the definition of the bottleneck distance, the unmatched points are connected to the closest point on the diagonal. The matching with the smallest matching distance is the second from the left, a fact that can be verified in Figure 8, which illustrates the matching distances for matchings of three diagrams of Figure 7. The $d_\infty(a, b)$ distance between points a and b can be thought to represent one half of the side-length of the square centered at a which has b on its boundary. The maximal length of such sides is the smallest in the second case and the resulting quantity is the bottleneck distance d_B.

☞ At this point it should become apparent why it is geometrically convenient to consider points on the diagonal represent the trivial persistence module. A side effect of this approach is that any two points on the diagonal represent the same trivial persistence module. In a way, the entire diagonal should thus be treated as a single point.

Figure 8: The distances between the matched pairs are demonstrated by the squares arising as the balls of the d_∞ metric. The diagram with the smallest maximal square amongst all matchings (even the ones not displayed here) is the middle one. Hence the resulting bottleneck distance d_B arises from the middle diagram.

Theorem 3.2 (Isometry Theorem)**.** *The interleaving distance between persistence modules equals the bottleneck distance between the corresponding persistence diagrams.*
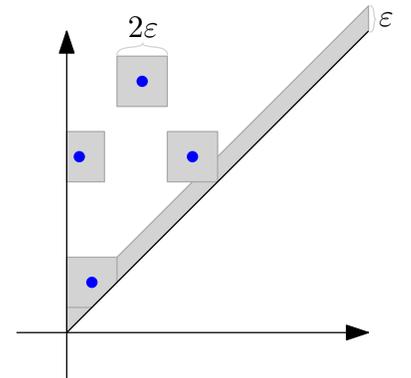
Figure 9: A schematic representation of the ε-neighborhood of the diagram consisting of blue points as discusses in Example 3.3.

Example 3.3. *Let $\mathcal{A}$ be the persistence diagram presented by the four blue points in Figure 9. If persistence diagram $\mathcal{B}$ satisfies $d_B(\mathcal{A}, \mathcal{B}) \leq \varepsilon$, then $\mathcal{B}$ consists of the following:*

- *For each blue point there exists one designated[36] red point within the grey square (i.e., the ε-ball in d_∞) around it.*

- *Arbitrarily many points within the grey ε-band[37] along the diagonal.*

Stability Theorem

Theorem 3.4. *[Stability Theorem] Assume persistence diagrams $\mathcal{A}$ and $\mathcal{B}$ represent persistent homologies of filtrations $\mathcal{V}$ and $\mathcal{W}$ obtained by one of the following procedures:*

1. *As the sublevel filtrations of filtration functions f and g satisfying condition $||f - g||_\infty \leq \varepsilon$, see Proposition 1.3.*

2. *As the Rips filtrations of metric spaces X and Y satisfying condition $d_{GH}(X, Y) \leq \varepsilon/2$, see 1. of Proposition 2.9.*

3. *As the Cech filtrations of metric spaces X and Y satisfying condition $d_{GH}(X, Y) \leq \varepsilon$, see 2. of Proposition 2.9.*

Then $d_B(\mathcal{A}, \mathcal{B}) \leq \varepsilon$.

Figure 10 is a schematic representation of the discussion leading to the Stability Theorem as presented here.
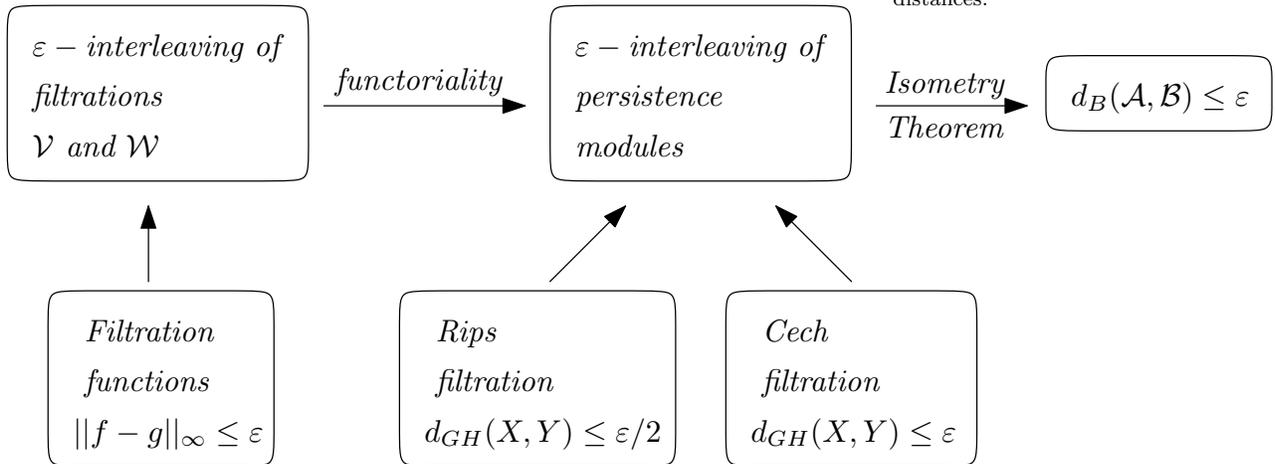
The moral of the theorem is that small perturbations of the input lead to small changes in persistence diagrams. On the other hand, critical simplices and homology representatives may be unstable.

[36] If some of the squares had non-empty intersection, then within that intersection there might be more points of $\mathcal{B}$, so a single square might contain more red points. However, one of them, potentially a diagonal point, has to be the designated one, i.e., the point to which the blue point in question is matched in an optimal matching.

[37] This band is actually the ε-neighborhood of the trivial (empty) persistence diagram. Again, this implies that the squares intersecting the band may contain more points.

☞ The stated version combines several separate version of the Stability Theorem found throughout the literature by stating several different initial assumptions.

☞ While the presented results explain stability in terms of the bottleneck distance, there is another family of distances on persistence diagrams called the **Wasserstein distances**. For example, the 1-Wasserstein distance is obtained by defining the matching distance as the sum (rather than max) of individual terms. Under appropriate assumptions the persistence diagrams are also stable when using Wasserstein distances.

Figure 10: The diagram summarizing the Stability Theorem and strategy of its proof that have been discussed.

4 Interpretations and examples

With the Stability Theorem, persistent homology may be thought of as a stable description of a geometric shape. The stability itself justifies the following observations:

- Given a geometric shape, ever better approximating point-clouds induce persistent diagrams ever closer (converging) the the persistence diagram of the shape.

- The points of higher persistence[38] represent more stable[39] features and are thus typically deemed to be of higher importance, leading to simplification[40] schemes on data.

In this section we will present several examples[41] of persistent homology arising via Rips complexes and comment on their structure.

1-dimensional persistence of geodesic spaces

Let X be a closed[42] geodesic manifold or, more generally, the body of a finite simplicial complex equipped with a geodesic metric. Assume S_n is a sequence[43] of finite metric spaces converging towards X in the Gromov-Hausdorff metric. Let $\mathcal{A}_n$ denote the 1-dimensional persistence diagram obtained from S_n via Rips filtration and coefficients in $\mathbb{F}$. It turns out that the limiting diagram[44] $\mathcal{A} = \lim_{n \to \infty} \mathcal{A}_n$ encodes a shortest base of $H_1(X; \mathbb{F})$: for each member[45] α of a shortest homology base of X we obtain a bar $[0, |\alpha|/3)$, where $|\alpha|$ is the length of α, see Figure 11. Without going through all the details let us demonstrate the situation through examples.

Figures 12, 13, and 14 represent three surfaces in $\mathbb{R}^3$ approximated by a finite collection of points. An approximation of a geodesic metric is induced on the points and used to compute 1-dimensional persistent homology. The right part od the figures represents the longest one or two bars obtained from each of the computations. Starting with a discrete set of points a multitude of short bars is also generated but are the artefact of a finite approximation rather than topologically significant features. By the Stability Theorem their lifespans decrease (although their numbers increase) as we improve the approximation density.

Let us interpret the results:

1. Our chosen samples are dense enough for the longest bars to detect the shortest 1-dimensional homology bases, which in this case coincide for all coefficients.

2. The bars would ideally be born at 0 and run until one-third of the lengths of the corresponding homology generators. With increased density the resulting barcode would approach this scenario.

[38] I.e., the longer bars in the barcode.

[39] I.e., they remain non-trivial under larger perturbations.

[40] I.e., denoising.

[41] Generated by Ripserer.jl, with coefficients in $\mathbb{Z}_2$.

[42] This implies it admits a finite triangulation.

[43] Such a sequence may be, roughly speaking, obtained by constructing ever finer finite approximations of X and inducing an approximation of a geodesic metric on them.

[44] $\mathcal{A}$ can be obtained as persistence diagram of the Rips filtration of X, a construction which involves infinite simplicial complexes and is formally beyond the scope of this book.

[45] Members are formally cycles whose length in this case is the length of the corresponding loop in X. One can choose a triangulation for which these simplicial loops are shortest possible.
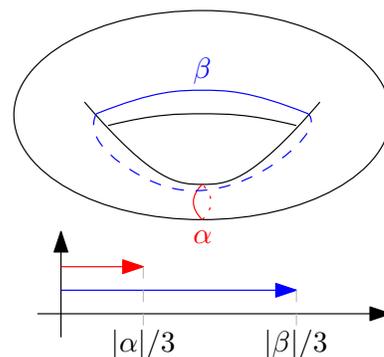
Figure 11: The 1-dimensional persistent homology of a torus detects its shortest homology basis.

3. The visualizations of approximating points also contain a loop or two: these are obtained by connecting the vertices of the critical triangles by the shortest paths through our points. Note that for bars corresponding to the basis, this gives an approximation of the shortest homology bases. Again, in the spirit of Stability Theorem, the finer the approximation by points, the closer approximation of the loops we obtain.

4. Going beyond the basis, we see that the next bar in Figure 14 detected a hole in our approximating points. The lifespan of this bar would decrease towards zero with ever better approximations.
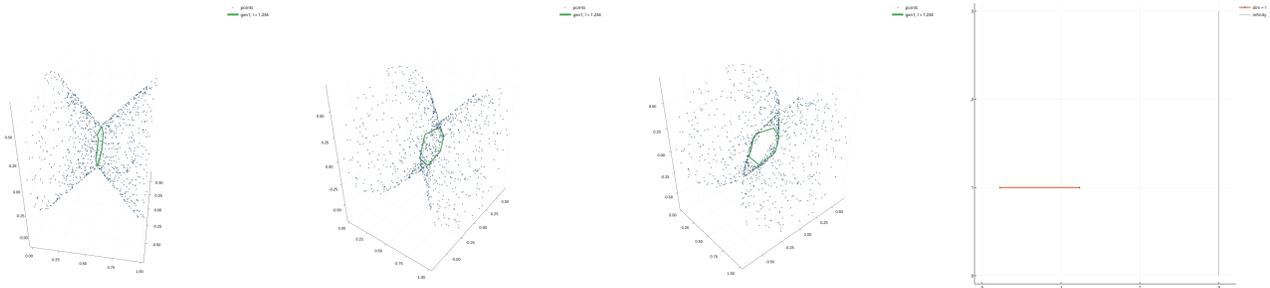
Figure 12: The longest bar of 1-dimensional persistent homology.
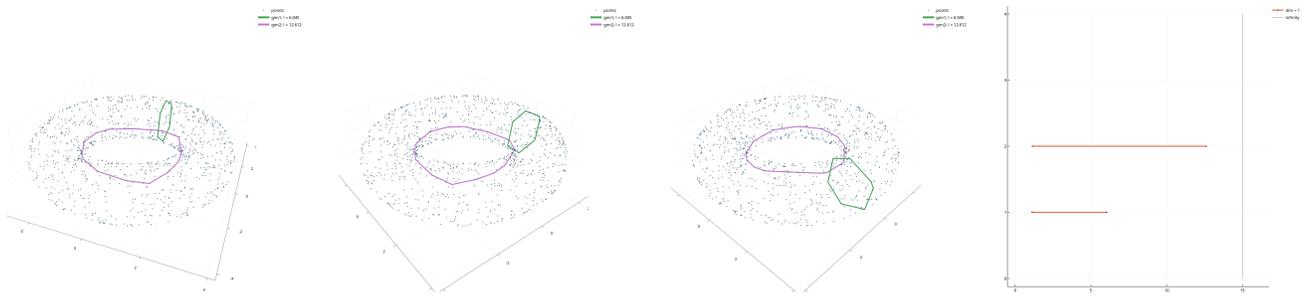
Figure 13: The longest two bars of 1-dimensional persistent homology.
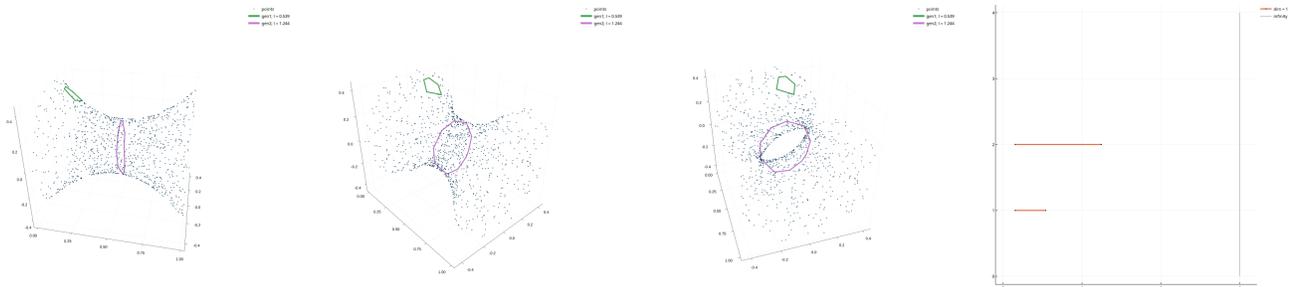
Figure 14: The longest two bars of 1-dimensional persistent homology.

Stability demonstrated
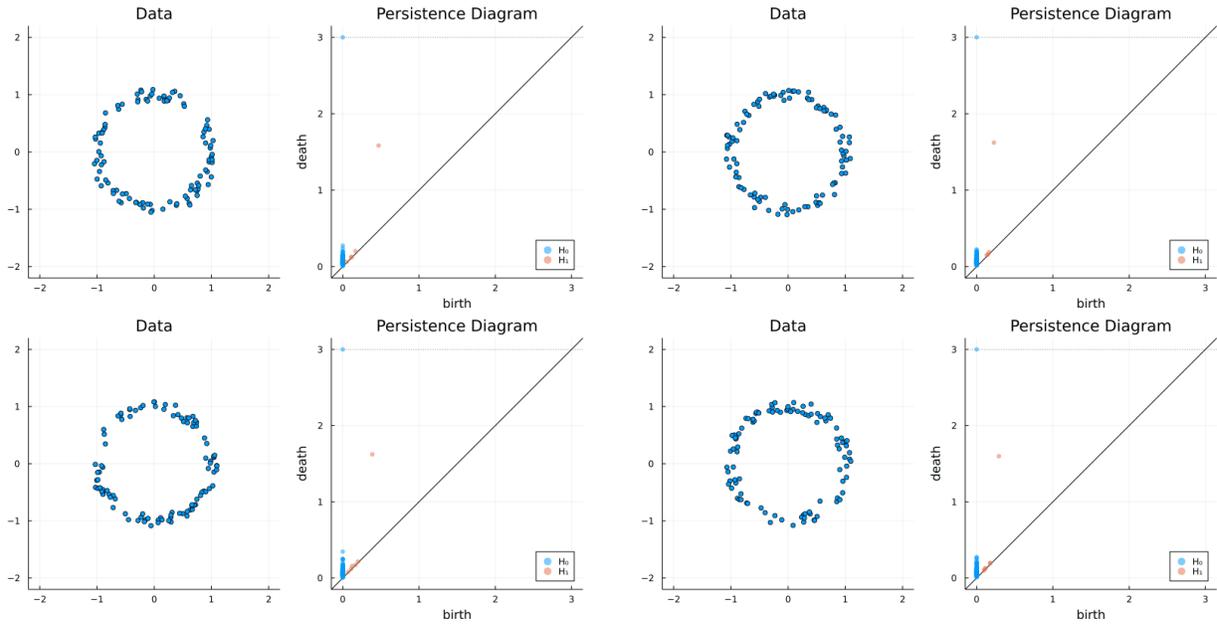
Figure 15: Four approximations of a circle and the corresponding persistence diagrams.

Figure 15 contains four approximations of a circle by discrete sets and the corresponding persistence diagrams arising from the Rips filtration. The Stability Theorem states that as the induced persistence diagrams should be close to each other, and the figure demonstrated this is indeed the case. A few comments on the diagrams:

- There is a point $(0,3)$ representing $(0,\infty)$ indicating the one persistent component.

- The main dominating feature is the very persistent point terminating at around 1.5. It represents the 1-dimensional hole, i.e., the homology of S^1. Its precise coordinates tell us more about the geometry of the sample.

  - The birth is between 0 and $.5$. The precise birth depends on the edges of the Rips complex going "around the circle"[46]. The maximal gap needed for such circumcision is the birth time. We can see that such a gap is smallest in the upper-right case resulting in early birth. On the other hand, the gap is largest in the upper-left case[47] and results in a later birth.

  - The terminal scale of this feature is the minimal diameter of an "almost equilateral triangle" reaching "around the circle".

- The other points on persistence diagrams are of low persistence and appear[48] as an artefact of discretization.

[46] ...and thus generating the 1-cycle.

[47] The gap of this sample appears in the upper-right part

[48] For example, as the Rips complex on n points is a discrete collection of n points at small scales, each such diagram will have n many points indicating persistent 0-dimensional homology.

Spheres

We next present examples approximating spheres. On Figure 16 we present persistence diagrams via Rips complexes of a sample of 100 points on unit spheres: S^2 (on left) and S^3 (on right); using Euclidean (top) or geodesic distance (bottom).
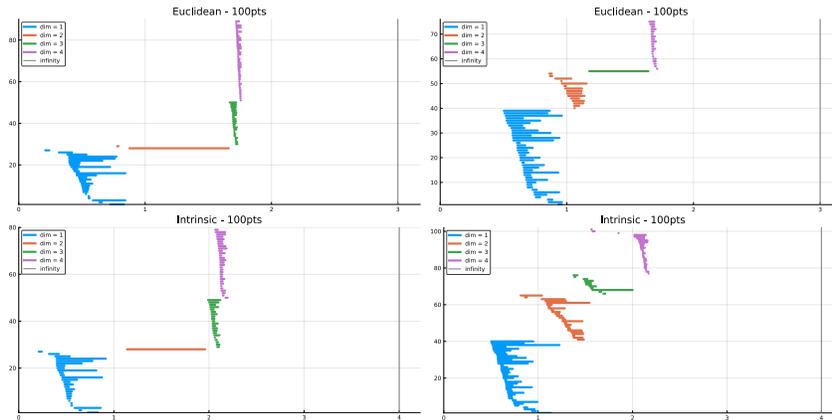
Figure 16: Persistence diagrams via Rips complexes of samples of one hundred points sampled from unit S^2 (on left) and S^3 (on right) using Euclidean or intrinsic (geodesic) distance.

As the only non-trivial homology (except for dimension 0) of S^2 is $H_2(S^2; \mathbb{F}) \cong \mathbb{F}$, we expect a long persistent line in that dimension, which is indeed the case. In fact, the long 2-dimensional bar clearly indicates that in both cases the most prominent homology is of rank 1 in dimension 2. The same holds for S^3 although a denser sample would have made the same observation easier in the geodesic case.

To demonstrate the improvement induced by larger density we present in Figure 18 a sequence of diagrams with increasing density. The underlying space is a cut-off sphere, i.e., a 2-dimensional sphere with a cap above the parallel of circumference approximately 1.5 removed, see Figure 17. We take a sample of $100, 200,$ and 400 points, generate a geodesic distance, and generate persistence diagrams via Rips filtrations.

Here is what we would expect from resulting persistence diagrams:

1. The cut-off sphere is contractible for small scales and thus initial clutter of 1-dimensional bars should be decreasing in size as we increase density.

2. As certain scale an offset of the cut-off sphere will fill in the top and create a void and hence[49] a 2-dimensional homology in the Cech complex. As Rips complexes are interleaved with Cech complexes[50], we might hope that the same 2-dimensional bar might appear in our case. That is indeed the case and the mentioned long bar grows with increasing density of the sample.

☞ In persistence diagrams in Figure 16 there are a lot of short bars. Some of these are artefacts of discretization, other indicate a more complex structure of persistent homology reaching beyond the interpretation of the size of homology representatives of the underlying space. Interpreting such bars is a very active research topic.
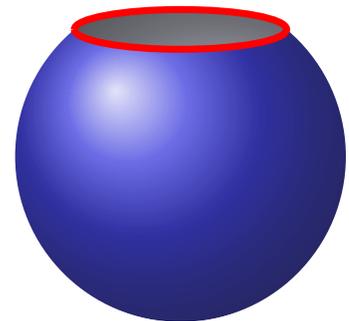
Figure 17: A cut-off sphere as an underlying space leading to diagrams in Figure 18.

[49] See Example 1.1.

[50] ...and hence the persistence diagrams are not too different.

3. As is frequently the case, there appear multitudes of short bars we choose to ignore at this point.
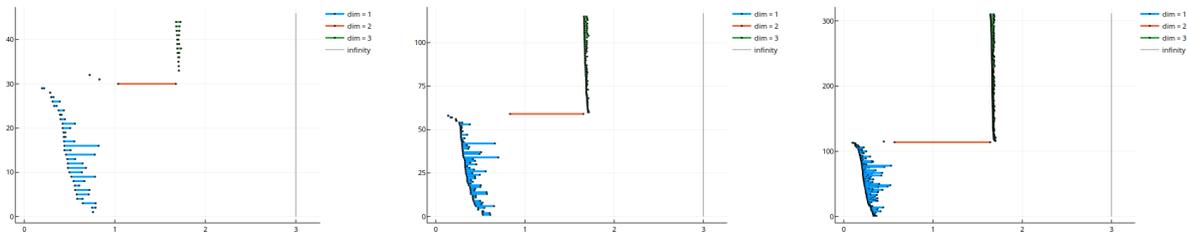
Figure 18: Four persistence diagrams via Rips complexes of a cut-off sphere, based on samples of 100, 200, and 400 points.

De-noising a function

Suppose we are given an approximation of a function f in the form of a discrete set of equally spaced measurements. We can connect the resulting points by edges and obtain a graph G representing our measurements, see the left side of Figure 19. Suppose we want to extract the global behaviour of f as in the center of Figure 19 by removing the local oscillations we consider to be noise. A way to achieve it would be to construct the sublevel[51] filtration of the simplicial complex G and choose the threshold ε for the noise level. We would then draw the corresponding 0-dimensional persistence diagram and ignore the points in the shaded ε-neighborhood[52] of the diagonal, see the right side of Figure 19. As a result we obtain two prominent points $\bullet, \square$ in the persistence diagram. The de-noised function presented in the center of Figure 19 can now be obtained by connecting the critical simplices corresponding to these points:

1. Blue birth simplices get connected to the higher endpoint of the red terminal simplex.

2. The only exception is $\square$, which is not a terminal simplex, but gets added as the highest point in the graph in order to finalize[53] our approximation by connecting it to $\square$.

[51] A vertex of G appears at the function value it represents. An edge of G appears as soon as both of its vertices appear.

[52] Each local minimum in our approximation except for $\bullet$ and $\square$ generates a point in this neighborhood.

☞ By adjusting the threshold ε we can adjust the level of details we want to preserve.

[53] The component represented by $\square$ can not get terminated as a homology element.
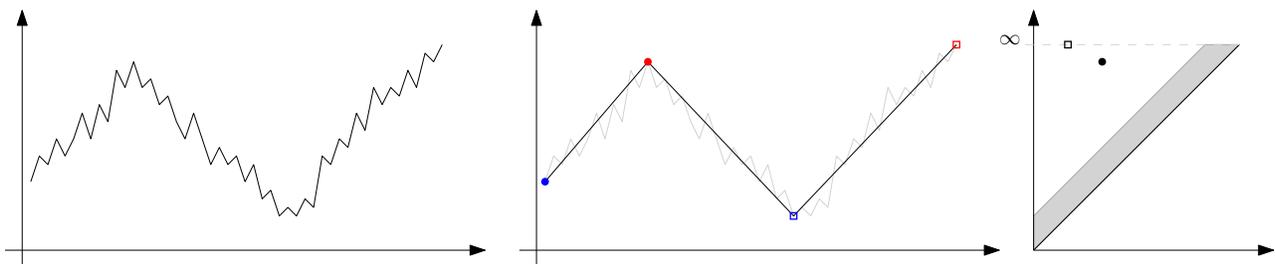
Figure 19: A noisy function, its reconstruction and the corresponding persistence diagram. The shaded region contains a multitude of points we choose to ignore in our reconstruction.

5 Concluding remarks

Recap (highlights) of this chapter

- continuous filtrations;

- persistence modules;

- interleaving;

- Stability Theorem;

Background and applications

There are three different proofs of the Stability Theorem in the literature. The initial proof was combinatorial[54] and did not include the Isometry theorem or interleavings but rather just the continuity of persistence diagrams. It motivated a more algebraically oriented proof using interleavings[55]. The third and most direct proof uses explicit matching[56]. There is also a recent treatment simplifying some of these ideas[57].

The existence of a decomposition of a persistence module into indecomposable parts is a particular case of a standard approach referred to as the Krull Remak Schmidt Principle. The fact that the indecomposable are precisely the elementary intervals is a special case of the Gabriel's Theorem. The fact that the indecomposable parts of multiparameter persistent homology are not as simple as the elementary intervals is the major obstacle to exhaustive applications of multiparameter persistence.

The material presented up to this point represents the core ideas and properties of persistent homology. From this point on the topics diverge, with some of the major motivations being:

- Theoretical treatment: persistent homology represents a parameterized version of homology and as such there are many ways to explore the structure further, either by generalizing the framework[58], determining what geometrical properties it encodes[59] and expanding the ideas into other theoretical contexts.

- Practical treatment, mostly associated with data analysis: in this context persistent homology is often viewed as a stable shape descriptor. As a result considerable effort is being invested to incorporate persistent homology into the flow of data analysis, either by adjusting it to specific data types[60], establishing meaningful probabilistic[61] and statistical[62] analysis, and to extract relevant features.

- Computational treatment: the aim of this context is to optimize the computational resources required to obtain persistence (or at least

[54] By Cohen-Steiner, Edelsbrunner, and Harer.

[55] By Chazal, Cohen-Steiner, Glisse, Guibas, and Oudot.

[56] By Bauer and Lesnick.

[57] By Skraba and Turner

[58] For example, with Zig-Zag persistence, multi-parameter persistence, introduction of new constructions of complexes (Witness, selective Rips complex), etc.

[59] For example it encodes, at least to some degree, shortest homology bases, intrinsic volumes, geometric shapes, curvatures, dimension, etc.

[60] Besides point-clouds, these include time series, high-dimensional data, dynamical systems, sensor networks, etc.

[61] It turns out there are significant phase transitions in persistent homology of random processes, etc.

[62] This is typically done by mapping persistence diagrams into Hilbert space via any of the multitude maps available, for example persistence landscapes, persistence silhouette, persistence images, etc.

a part of it) by developing faster algorithms often incorporating various shortcuts[63] or additional structure[64]. Currently available software for computing persistent homology includes (but is not restricted to) Ripser and related Ripserer.jl, Ripser.py, and Cubical Ripser, Dionysus, PHAT, GUDHI, javaPlex, Perseus, Eirene, etc.

[63] For example, the twist.

[64] For example, when computing one-dimensional persistence of geodesic spaces.

This list of topics and software is by no means exhaustive.

Appendix: From the interleaving distance to the bottleneck distance

In this appendix we will provide an explanation that leads to the bottleneck distance by determining the interleaving distance between pairs of elementary intervals.

Case 1: distance between an elementary interval and the zero persistence module. The situation is presented in Figures 20 and 21. For $0 \leq p < q$ let us discuss the interleaving of the elementary interval $\mathbb{F}_{p,q}$ (the bold portion of the figures) and the trivial persistence module (the grey portion below).

Figure 20: Elementary interval $\mathbb{F}_{p,q}$ is not ε-interleaved with the trivial interval if $\varepsilon < (q-p)/2$.

- If the interleaving parameter was $\varepsilon < (q-p)/2$ as in Figure 20, the composition of the red diagonal maps:

  - is the trivial map as it factors through the trivial vector space below;

  - should have been identity on $\mathbb{F}$ by the interleaving condition, as its domain and target are in $[p,q)$.
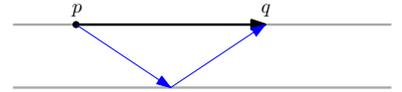
Figure 21: Elementary interval $\mathbb{F}_{p,q}$ is ε-interleaved with the trivial interval if $\varepsilon \geq (q-p)/2$.

  These two observations contradict each other hence the interleaving parameter is at least $(q-p)/2$.

- For $\varepsilon = (p-q)/2$ though, the composition of the diagonal maps increases the scale parameter by $p-q$, and any such structure map of $\mathbb{F}_{p,q}$ is trivial, hence the ε-interleaving consisting of trivial maps exists, see Figure 21.

We conclude that the interleaving distance is $\varepsilon = (p-q)/2$.

Case 2: general case. From Case 1 we can conclude that for $0 \leq p' < q'$ the following holds: If $\mathbb{F}_{p,q}$ is ε-interleaved with $\mathbb{F}_{p',q'}$ for $\varepsilon < (p-q)/2$, then $[p',q') \supset [p+\varepsilon, q-\varepsilon)$ see Figure 22. By symmetry the opposite also holds: $[p,q) \supset [p'+\varepsilon, q'-\varepsilon)$. It is easy to see these conditions are also sufficient. For minimal ε for which these two conditions hold we obtain the ε-interleaving by mapping the designated generator of $\mathbb{F}_{p,q}$ to the designated generator of $\mathbb{F}_{p',q'}$ whenever possible, with other maps being trivial, see Figure 24. It should be apparent from Figure 23 that the ε in question is $\max\{|p-q|, |p'-q'|\}$.

Figure 22: ε-interleaving implies the orange part is non-trivial as the interleaving maps in the blue region have to be non-trivial.

We conclude that $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ are $\max\{|p-q|, |p'-q'|\}$ interleaved. However, since $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ are also $\max\{(p-q)/2, (p'-$
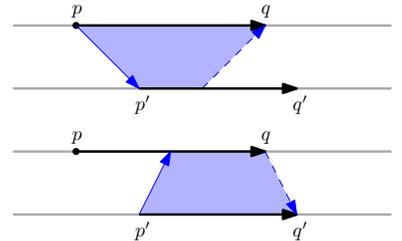
Figure 23: Condition of Figure 22 induces two shapes. The interleaving distance is the larger ε parameter these shapes induce.

$q')/2$}-interleaved by the trivial maps[65], the interleaving distance between $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ is

$$\min\{\max\{|p-q|,|p'-q'|\},\max\{(p-q)/2,(p'-q')/2\}\}.$$

We now interpret the obtained distanced in the context of persistence diagrams representing elementary intervals. First note that

$$(p-q)/2 = d_\infty\left((p,q),\left(\frac{p+q}{2},\frac{p+q}{2}\right)\right)$$

is the d_∞ distance between (p_q) and the diagonal Δ.

Case 1. The interleaving distance between $\mathbb{F}_{p,q}$ and the zero persistence module is realized by matching point (p,q) to the closest point on the diagonal and computing the resulting d_∞ distance, see Figure 25.

Case 2. The distance between $\mathbb{F}_{p,q}$ and $\mathbb{F}_{p',q'}$ is the smaller of the following two:

1. Either $\max\{|p-q|,|p'-q'|\} = d_\infty((p,q),(p',q'))$, which is the d_∞ distance between the points, see Figure 26.

2. Or $\max\{(p-q)/2,(p'-q')/2\}$ which can be interpreted as follows: match each of the points to the closest point on the diagonal Δ and take the maximal d_∞ distance, see Figure 27.

We have thus interpreted the interleaving distance between elementary intervals in the context of persistence modules and obtained the bottleneck distance for diagrams containing at most one point. The crucial ingredients of the interpretation are the matching and d_∞. Theorem 3.2 essentially states that an optimal interleaving between any pair of persistence modules essentially consists of such matchings: match some pairs of elementary intervals from both persistence modules, and then match the remaining elementary intervals to Δ. The interleaving distance (and thus the bottleneck distance) corresponds to the matching whose d_∞-distance of its maximal matching is minimal.
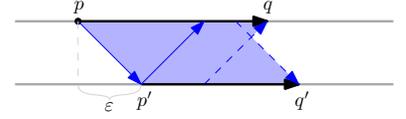
Figure 24: The interleaving for $\varepsilon = \max\{|p-q|,|p'-q'|\}$. The non-trivial maps are in the shaded region.
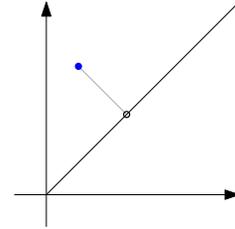
Figure 25: Matching a point to Δ.
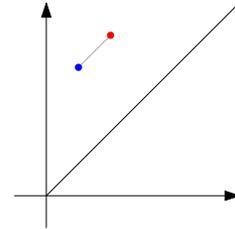
Figure 26: Matching two points.
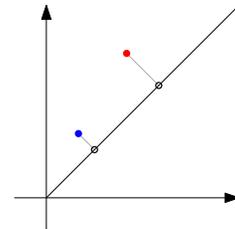
Figure 27: Matching each of the two points to Δ.