

Chapter 6

Distances and Stability

6.1 Distance Metrics on Persistence Diagrams

As we have seen in the previous sections, persistent homology of simplicial filtrations, for example in the form of persistence diagrams or persistence barcodes, can give us a lot of insight into a given point cloud. However, so far we have always been analyzing this information manually. In this section we will show how this can be done on a more mathematical level, by defining some distance metrics that can be used to compare different persistence diagrams, and thus to assess the similarity of point clouds.

6.1.1 Bottleneck Distance

Let \mathcal{F}, \mathcal{G} be two filtrations giving rise to persistence modules $H_p(\mathcal{F}), H_p(\mathcal{G})$, and let $\text{Dgm}_p(\mathcal{F})$ and $\text{Dgm}_p(\mathcal{G})$ be their corresponding persistence diagrams. How can we now compare these two filtrations \mathcal{F} and \mathcal{G} by using only the information stored in these diagrams?

The general idea of the bottleneck distance is to find a matching between the points of the two persistence diagrams, i.e., we consider bijections between the points of $\text{Dgm}_p(\mathcal{F})$ and those of $\text{Dgm}_p(\mathcal{G})$. Since we can only find bijections between sets of the same cardinality, we need the two diagrams to have the same number of points. Recall that the way we defined it, a persistence diagram includes every possible point on the diagonal with infinite multiplicity. This peculiar definition now finally pays off; since both sets of points have the same (infinite) cardinality, and bijections between these sets are thus well-defined.

We do not want to consider any arbitrary bijection, but only the *best possible*. To measure the “quality” or “distance” of such a bijection, we use the L_∞ -norm:

Definition 6.1. *Let $x = (x_1, x_2), y = (y_1, y_2)$ be two points in \mathbb{R}^2 . Then,*

$$\|x - y\|_\infty := \max(|x_1 - y_1|, |x_2 - y_2|),$$

where we say that $\infty - \infty = 0$ for points with coordinates that are ∞ (i.e., points in persistence diagrams that correspond to holes that did not die).

Definition 6.2. Let $\Pi = \{\pi : \text{Dgm}_p(\mathcal{F}) \rightarrow \text{Dgm}_p(\mathcal{G}) \mid \pi \text{ is bijective}\}$ be the set of all bijections between $\text{Dgm}_p(\mathcal{F})$ and $\text{Dgm}_p(\mathcal{G})$. Then, the Bottleneck distance is defined as

$$d_b(\text{Dgm}_p(\mathcal{F}), \text{Dgm}_p(\mathcal{G})) := \inf_{\pi \in \Pi} \sup_{x \in \text{Dgm}_p(\mathcal{F})} \|x - \pi(x)\|_\infty.$$

The Bottleneck distance thus minimizes the maximum L_∞ -norm of any pairing, over all pairings of points.

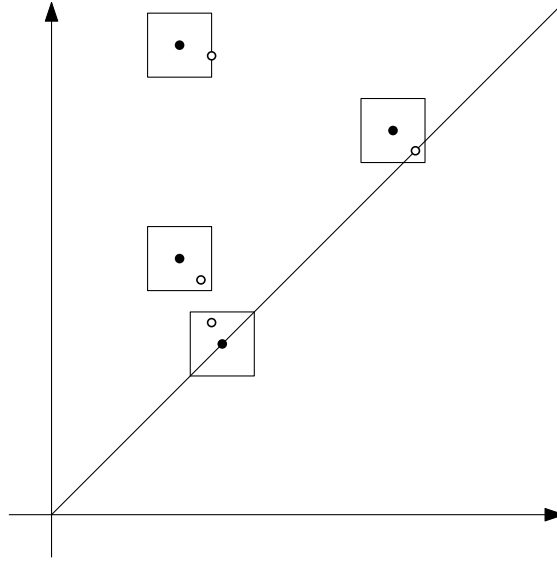


Figure 6.1: An illustration of the idea of bottleneck distance.

Observation 6.3. The Bottleneck distance is a metric on the space of persistence diagrams with finitely many off-diagonal points.

Proof. We check the three properties of metrics:

1. $d_b(X, Y) = 0$ if and only if $X = Y$: This is simple to see, since if $X = Y$, every point can be matched to its copy, and if $X \neq Y$, there exists some point $p \in X \setminus Y \cup Y \setminus X$ which must be matched to some point with positive L_∞ -distance to p .
2. $d_b(X, Y) = d_b(Y, X)$: This is clear by definition.
3. $d_b(X, Y) \leq d_b(X, Z) + d_b(Z, Y)$: Take a bijection π_1 witnessing¹ $d_b(X, Z)$ and a bijection π_2 witnessing $d_b(Z, Y)$, and concatenate the two: $\pi := \pi_2 \circ \pi_1$ is a bijection $X \rightarrow Y$ where for every $x \in X$ we can use the triangle equality of $\|\cdot\|_\infty$ to bound $\|x - \pi(x)\|_\infty \leq \|x - \pi_1(x)\|_\infty + \|\pi_1(x) - \pi_2(\pi_1(x))\|_\infty$. \square

¹Note that since d_b is an infimum and not a minimum, there may not be π_1 and π_2 witnessing d_b . In this case, the same argument can be applied to the converging sequences of bijections witnessing d_b .

Exercise 6.4. Give an algorithm to compute the Bottleneck distance between two persistence diagrams. Your algorithm should be polynomial in n , where n is the total number of off-diagonal points in the two persistence diagrams.

The Bottleneck distance can be used to compare any two filtrations, of possibly wildly different spaces. What if we consider two filtrations of the same simplicial complex? Recall that simplex-wise monotone functions $f, g : K \rightarrow \mathbb{R}$ give rise to simplicial sublevel set filtrations $\mathcal{F}_f, \mathcal{F}_g$. While these two filtrations can be compared using the Bottleneck distance, we can also define a metric that directly compares the two functions f, g :

Definition 6.5 (infinity norm). Let $f, g : X \rightarrow \mathbb{R}$. Then, the infinity norm of $f - g$ is defined as

$$\|f - g\|_\infty := \sup_{x \in X} |f(x) - g(x)|.$$

The following theorem tells us that this infinity norm and the Bottleneck distance are closely related:

Theorem 6.6 (Stability for simplicial filtrations). Let $f, g : K \rightarrow \mathbb{R}$ be simplex-wise monotone functions. Then, $\forall p \geq 0$ we have $d_b(\text{Dgm}_p(\mathcal{F}_f), \text{Dgm}_p(\mathcal{F}_g)) \leq \|f - g\|_\infty$.

Proof. We consider the linear interpolation $f_t := (1 - t)f + tg$ for $t \in [0, 1]$ between f and g . Note that $f_0 = f, f_1 = g$.

We first show that each f_t is a simplex-wise monotone function: Let $\sigma \subseteq \tau$. Since f and g are monotone, we have $f(\sigma) \leq f(\tau)$ and $g(\sigma) \leq g(\tau)$. Thus,

$$f_t(\sigma) = (1 - t)f(\sigma) + tg(\sigma) \leq (1 - t)f(\tau) + tg(\tau) = f_t(\tau).$$

Let $p \geq 0$ be fixed. We now draw the family of persistence diagrams $\text{Dgm}_p(\mathcal{F}_{f_t})$ as a multiset in $\mathbb{R}^2 \times [0, 1]$. Each off-diagonal point of $X_t := \text{Dgm}_p(\mathcal{F}_{f_t})$ is of the form $x(t) = (f_t(\sigma), f_t(\tau), t)$ for σ being the creator and τ being the destructor. Note that the persistence pairings (σ, τ) may only change when the order of simplex insertion changes, which only happens finitely many times when going from $t = 0$ to $t = 1$. Let us call these values $0 = t_0 < t_1 < t_2 < \dots < t_n < t_{n+1} = 1$. For simplicity, we assume that at each of these values t_i exactly two simplices have the same value f_{t_i} .

Within each open interval (t_i, t_{i+1}) the pairings stay constant. Furthermore, every off-diagonal point $x(t)$ is a linear function of t in all three coordinates, meaning that it defines a line segment.

Let us first assume that at some value t_{i+1} , $x(t_{i+1})$ is an off-diagonal point whose creator and destructor are still paired after t_{i+1} . In this case, $x(t)$ continues in the same direction after t_{i+1} .

If on the other hand $x(t_{i+1})$ is an off-diagonal point whose creator and destructor get paired differently, recalling Exercise 4.4, there are exactly two pairs that swap their creators or destructors, and these creators or destructors that are swapped must have the same value in $f_{t_{i+1}}$. In the persistence diagram, this means that two points vertically

or horizontally of each other swap creators/destructors. However, this just means that for both of these line segments going into t_{i+1} , there is a unique continuing line segment.

Note that for $t = 0$ or $t = 1$ we can also have that $x(t)$ lies on the diagonal. This means that its past/future creator and destructor have the same value in f_t .

Every point thus moves along a polygonal path monotone in t . Every such path is called a *vine*, and the multiset of all vines is called a *vineyard*, see Figure 6.2 for an illustration. Based on this vineyard, we now wish to find a good matching giving an upper bound on the Bottleneck distance. We simply take the matching where we match the start point of every vine with its endpoint. To get a bound on the Bottleneck distance, we simply need to get a bound for the distance of each matched pair.

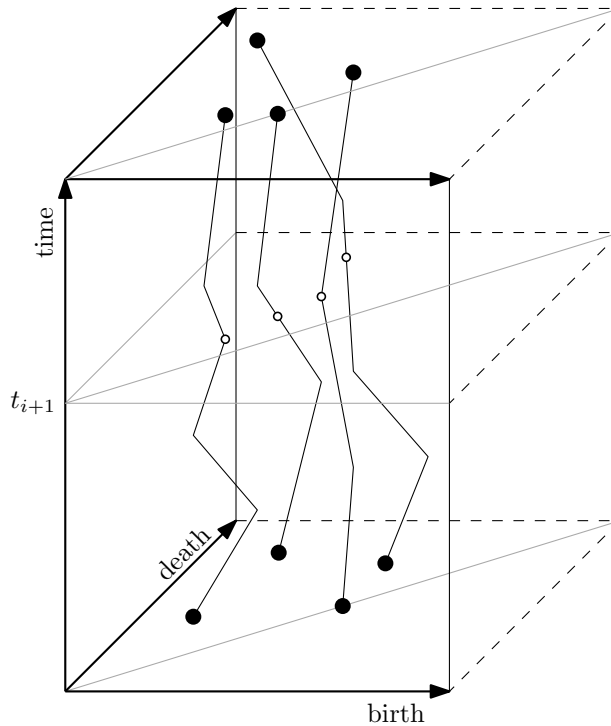


Figure 6.2: The vineyards in the proof of Theorem 6.6.

Between any t_i and t_{i+1} , a point $x(t)$ moves at the rate $\frac{\delta x(t)}{\delta t}$, which we can compute to be

$$\frac{\delta}{\delta t} ((1-t) \cdot (f(\sigma), f(\tau), t)) + t \cdot (g(\sigma), g(\tau), t)) = (g(\sigma) - f(\sigma), g(\tau) - f(\tau), 1).$$

Projecting $x(t_{i+1})$ and $x(t_i)$ to \mathbb{R}^2 we thus get two points y_{i+1}, y_i for which we can see that

$$\|y_{i+1} - y_i\|_\infty = (t_{i+1} - t_i) \cdot \max(|g(\sigma) - f(\sigma)|, |g(\tau) - f(\tau)|) \leq (t_{i+1} - t_i) \cdot \|f - g\|_\infty.$$

Thus, since $\|\cdot\|_\infty$ is a norm and fulfills the triangle inequality, we also have that from $t = 0$ to $t = 1$, the point can move at most $\|f - g\|_\infty$. We thus have the desired bound on the Bottleneck distance. \square

Exercise 6.7. Show that Theorem 6.6 above can be tight for all $p \geq 0$ and all values of $\|f - g\|_\infty$.

Exercise 6.8. Let P and Q be subsets of a metric space. We say that $P \cup Q$ is in δ -separated position if for any $p \in P$ and $q_1, q_2 \in Q$ we have that $|d(p, q_1) - d(p, q_2)| > \delta$. Assume that $P \cup Q$ is in δ -separated position and let Q' be an ε -perturbation of Q , that is, there is a bijection between Q and Q' such that for every original point $q \in Q$ and its image $q' \in Q'$ we have $d(q, q') \leq \varepsilon$. Let \mathcal{D}_p and \mathcal{D}'_p be the persistence diagrams for the p -dimensional persistent homology of the filtration induced by the parameterized witness complexes $\mathbb{W}^r(Q, P)$ and $\mathbb{W}^r(Q', P)$, respectively. Show that if $\varepsilon < \delta/2$ then

$$d_b(\mathcal{D}_p, \mathcal{D}'_p) \leq \varepsilon.$$

Further, show an example where this fails for $\varepsilon > \delta/2$.

We wish to generalize the stability result above from simplicial filtrations to filtrations of general topological spaces. To this end we consider some topological space X and a function $f : X \rightarrow \mathbb{R}$, which induces a sublevel set filtration for every $r \in \mathbb{R}$. We only want to consider *tame* functions: A function f is *tame* if all homology groups of sublevel sets have finite rank, and the homology groups only change at finitely many values, called *critical values*.

Theorem 6.9. Let X be a triangulable topological space, and $f, g : X \rightarrow \mathbb{R}$ be two tame functions. Then $\forall p \geq 0$, we have

$$d_b(\text{Dgm}_p(\mathcal{F}_f), \text{Dgm}_p(\mathcal{F}_g)) \leq \|f - g\|_\infty.$$

We do not prove this theorem at this point, but with additional tools that we will develop in Section 6.2, the proof of this (and of Theorem 6.6) will follow quite easily.

6.1.2 Wasserstein Distance

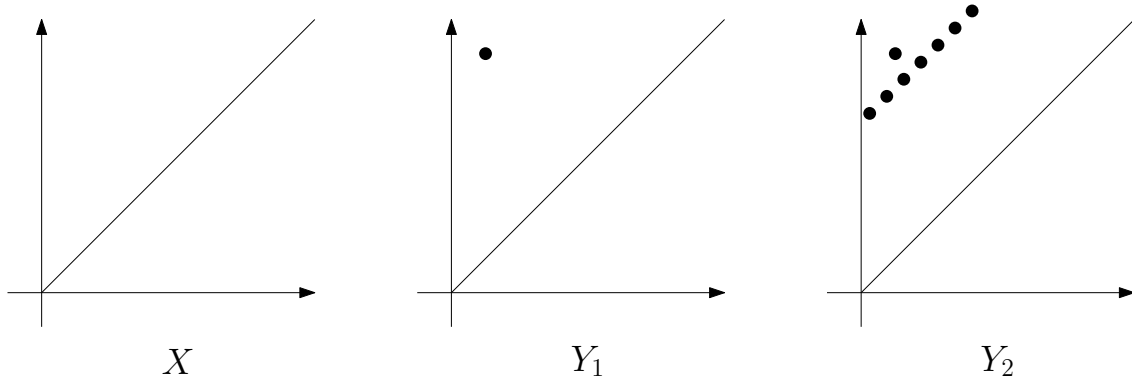
Consider the following three diagrams:

Which of Y_1 and Y_2 is X closer to? Intuitively, one clearly says Y_1 : There are simply fewer features in Y_1 that are not present in X . In terms of Bottleneck distance, there is only one reasonable matching between X and Y_1 , and also only one between X and Y_2 : We simply match each off-diagonal point with its closest point on the diagonal. Since we only look at the longest edge in this matching, the Bottleneck distance is actually the same for both pairs of diagrams, i.e., $d_b(X, Y_1) = d_b(X, Y_2)$.

We can get rid of this counter-intuitive behavior of the Bottleneck distance by using the Wasserstein distance.

Definition 6.10 (Wasserstein distance). For $p \geq 0$, and $q \geq 1$, the q -Wasserstein distance is defined as

$$d_{W,q}(\text{Dgm}_p(\mathcal{F}), \text{Dgm}_p(\mathcal{G})) := \left[\inf_{\pi \in \Pi} \left(\sum_{x \in \text{Dgm}_p(\mathcal{F})} (\|x - \pi(x)\|_\infty)^q \right) \right]^{1/q}$$



Intuitively, we now consider the length of all edges in the matching induced by the bijection, as opposed to just the longest one, but the longer ones get more weight. Note that for $q = \infty$, we retrieve the bottleneck distance, that is, $d_{W,\infty} = d_b$.

We can see that the stability theorem we proved for Bottleneck distance does not hold for Wasserstein distance: consider two simplex-wise monotone functions f and g on a path, as illustrated in Figure 6.3. In both f and g the first vertex on the path is mapped to 1 and the edges along the path are mapped to increasing odd numbers. In f the remaining vertices along the path get mapped to increasing even numbers, and in g to increasing odd numbers. In particular, $\|f - g\|_\infty = 1$. In the filtration defined by f , at every even step we add a vertex, creating a new connected component, which gets connected to the rest of the path at the next step. Thus, each vertex of the path will give an off-diagonal point in the 0-persistence diagram, where all of them except the first one have a lifespan of 1. On the other hand, in the filtration defined by g , we always add the new vertices and their connecting edge in the same step, thus the 0-persistence diagram only has a single off-diagonal point with infinite lifespan. In particular, we have that for arbitrarily long paths we get arbitrarily large Wasserstein distances between the diagrams for all $q < \infty$.

A similar counterexample can also be found for topological spaces. Consider the topological space $[0, 1]$ and the two functions depicted by the curves in Figure 6.4. Here we again have that $\|f - g\|_\infty \leq \epsilon$, but the Wasserstein distance between the two diagrams can be made arbitrarily big.

To avoid these types of counterexamples, we want to consider even nicer functions:

Definition 6.11 (Lipschitz). *Let (X, d) be a metric space. A function $f : X \rightarrow \mathbb{R}$ is Lipschitz if there exists a constant C such that $|f(x) - f(y)| \leq c \cdot d(x, y)$ for all $x, y \in X$.*

For these functions we again get stability theorems, however we will not prove these here.

Theorem 6.12. *Let X be a triangulable, compact metric space. Let $f, g : X \rightarrow \mathbb{R}$ be tame Lipschitz functions. Then there exist constants C and k (that may only depend on*

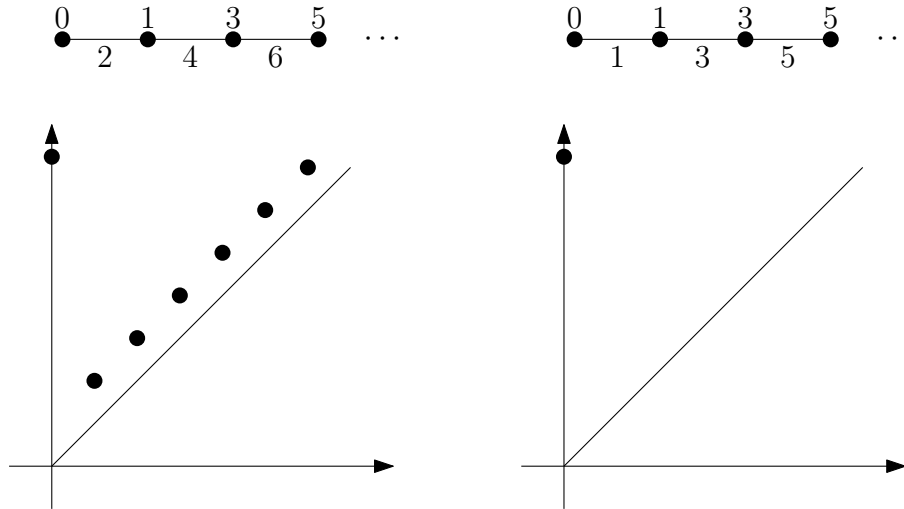


Figure 6.3: Two simplex-wise monotone functions with bounded infinity norm whose persistence diagrams have unbounded Wasserstein distance.

X and on the Lipschitz constants of f, g) such that for every $p \geq 0$ and every $q \geq k$,

$$d_{W,q}(Dgm_p(\mathcal{F}_f), Dgm_p(\mathcal{F}_g)) \leq C \cdot \|f - g\|_\infty^{1-k/q}.$$

Theorem 6.13. Let $f, g : K \rightarrow \mathbb{R}$ be simplex-wise monotone functions. Then for all $p \geq 0$ and all $q \geq 1$,

$$d_{W,q}(Dgm_p(\mathcal{F}_f), Dgm_p(\mathcal{F}_g)) \leq \|f - g\|_q = \left(\sum_{\sigma \in K} |f(\sigma) - g(\sigma)|^q \right)^{1/q}.$$

6.2 Interleaving of Persistence Modules

6.2.1 Interleaving Distance

Until now, we only compared persistence diagrams. We will now introduce the interleaving distance, which instead compares persistence modules directly. Let us begin with a formal definition of persistence modules.

Definition 6.14. A persistence module \mathbb{V} over \mathbb{R} is a collection $\mathbb{V} = \{V_a\}_{a \in \mathbb{R}}$ of vector spaces V_a together with linear maps $v_{a,a'} : V_a \rightarrow V_{a'}$ for $a \leq a'$, such that $v_{a,a} = \text{id}$ and $v_{b,c} \circ v_{a,b} = v_{a,c}$ for all $a \leq b \leq c$.

You already know a few examples of persistence modules, e.g., the persistent homology of sublevel set filtrations or of Čech or Vietoris-Rips complexes (here one simply defines $V_a = 0$ for $a < 0$).

To be able to define distances between persistence modules, we first need to figure out when we want to call two persistence modules “the same”, or more formally speaking, we want to define a notion of isomorphism.

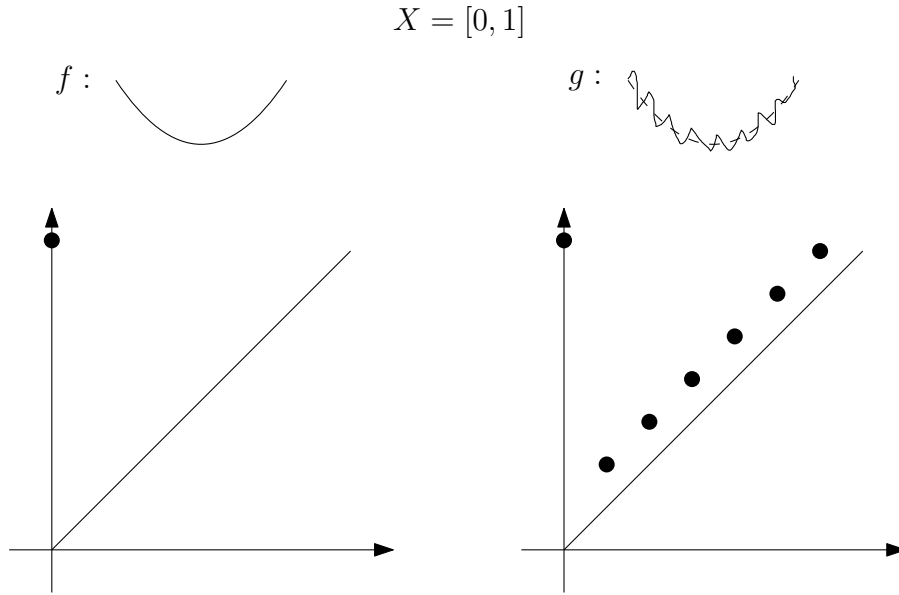


Figure 6.4: Two functions $[0, 1] \rightarrow \mathbb{R}$ with bounded infinity norm whose persistence diagrams have unbounded Wasserstein distance.

Definition 6.15. We say that two persistence modules \mathbb{U} and \mathbb{V} are isomorphic if there are isomorphisms $f_a : \mathbb{U}_a \rightarrow \mathbb{V}_a$ for all $a \in \mathbb{R}$ such that

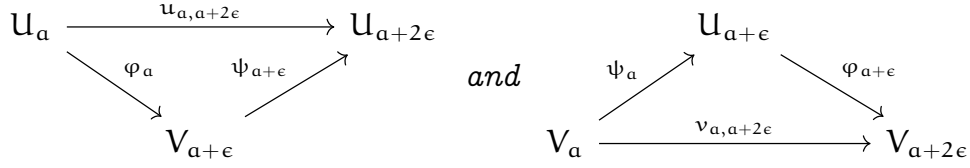
$$\begin{array}{ccc} \mathbb{U}_a & \xrightarrow{u_{a,a'}} & \mathbb{U}_{a'} \\ \uparrow f_a & & \uparrow f_{a'} \\ \mathbb{V}_a & \xrightarrow{v_{a,a'}} & \mathbb{V}_{a'} \end{array}$$

commutes both ways, i.e., $f_{a'} \circ u_{a,a'} = v_{a,a'} \circ f_a$, and $u_{a,a'} \circ f_a^{-1} = f_{a'}^{-1} \circ v_{a,a'}$.

The basic idea of interleaving distance is to measure how close two persistence modules are to being isomorphic. For this, we allow ourselves some slack, in the sense that \mathbb{U}_a does not need to map to \mathbb{V}_a , but it can instead map to $\mathbb{V}_{a+\epsilon}$, as long as all the relevant maps still behave like they would for an isomorphism. We make this formal in the next definition.

Definition 6.16 (ϵ -interleaving persistence modules). Let \mathbb{U} and \mathbb{V} be persistence modules over \mathbb{R} . We say that \mathbb{U} and \mathbb{V} are ϵ -interleaved if there exist two families of linear maps, $\varphi_a : \mathbb{U}_a \rightarrow \mathbb{V}_{a+\epsilon}$ and $\psi_a : \mathbb{V}_a \rightarrow \mathbb{U}_{a+\epsilon}$ such that the following four diagrams are commutative:

$$\begin{array}{ccc} \mathbb{U}_a & \xrightarrow{u_{a,a'}} & \mathbb{U}_{a'} \\ \searrow \varphi_a & & \searrow \varphi_{a'} \\ & \mathbb{V}_{a+\epsilon} & \xrightarrow{v_{a+\epsilon,a'+\epsilon}} & \mathbb{V}_{a'+\epsilon} \end{array} \quad \text{and} \quad \begin{array}{ccc} & \mathbb{U}_{a+\epsilon} & \xrightarrow{u_{a+\epsilon,a'+\epsilon}} & \mathbb{U}_{a'+\epsilon} \\ \nearrow \psi_a & & \nearrow \psi_{a'} & \\ \mathbb{V}_a & \xrightarrow{v_{a,a'}} & \mathbb{V}_{a'} & \end{array}$$

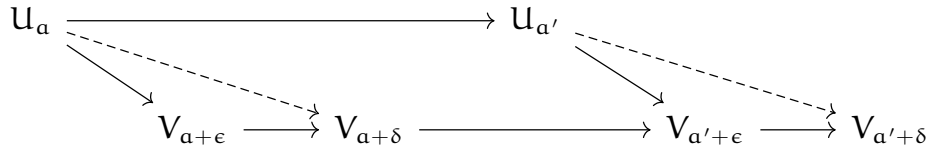


Note that if \mathbb{U} and \mathbb{V} are isomorphic, then they are 0-interleaved: the first type of diagrams (the *square diagrams*) are the commutative diagrams in the definition of isomorphic persistence modules and the the second type of diagrams (the *triangular diagrams*) collapse to two arrows that say that the maps φ_a are isomorphisms with inverses ψ_a .

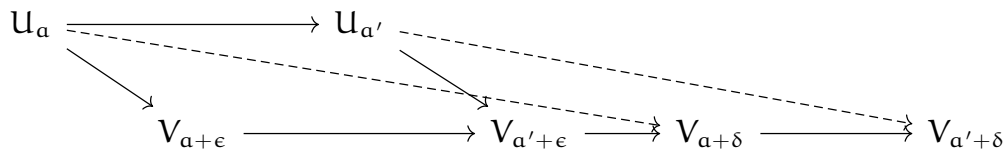
Theorem 6.17. *Assume \mathbb{U} and \mathbb{V} are ϵ -interleaved. Let $\delta > \epsilon$. Then \mathbb{U} and \mathbb{V} are also δ -interleaved.*

Proof. Given $\varphi'_a : \mathbb{U}_a \rightarrow \mathbb{V}_{a+\epsilon}$ we define $\varphi_a : \mathbb{U}_a \rightarrow \mathbb{V}_{a+\delta}$ simply as $\varphi_a := v_{a+\epsilon, a+\delta} \circ \varphi'_a$. Symmetrically, we define $\psi_a := u_{a+\epsilon, a+\delta} \circ \psi'_a$. To check that the correct diagrams commute, we only check the right diagram of every pair of symmetric diagrams shown above. We have to distinguish two cases for the first diagram, $a + \delta < a' + \epsilon$ and $a + \delta > a' + \epsilon$.

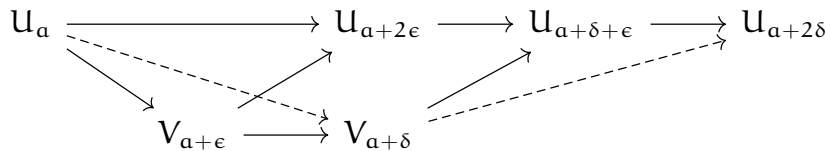
For the first case, we get the following diagram:



For the second case we get the diagram:



And finally, for the triangular diagram we get:



One can now verify that in all of these diagrams the correct paths commute. □

Thus, the following definition makes sense:

Definition 6.18 (Interleaving distance). $d_I(\mathbb{U}, \mathbb{V}) := \inf\{\epsilon \mid \mathbb{U} \text{ and } \mathbb{V} \text{ are } \epsilon\text{-interleaved}\}$.

Exercise 6.19. Show that interleaving distance is a pseudo-metric for persistence modules (up to isomorphism), i.e., prove that (i) the interleaving distance between isomorphic persistence modules is 0, (ii) the interleaving distance is non-negative, and (iii) the interleaving distance fulfills the triangle inequality.

Also show that it is not a metric by showing that there exist non-isomorphic persistence modules with interleaving distance 0.

Exercise 6.20. Let W_1 and W_2 be two arbitrary vector spaces. Let \mathbb{U} be the persistence module such that $U_a = W_1$ for $a \in [w, x)$, and $U_a = 0$, otherwise. For $a, a' \in [w, x)$ we have $u_{a,a'}$ being the identity map. For $a < w$ or $a' \geq x$ (or both), we have $u_{a,a'}$ being the zero map. Similarly, we define the persistence module \mathbb{V} which is W_2 in $a \in [y, z)$ and 0 otherwise.

Show that $d_I(\mathbb{U}, \mathbb{V}) \leq \max(\frac{w-x}{2}, \frac{z-y}{2})$.

The underlying ideas that allowed us to define the interleaving distance of persistence modules can also be applied to filtrations.

Definition 6.21 (Interleaving for Filtrations). Let \mathcal{F}, \mathcal{G} be filtrations over \mathbb{R} . \mathcal{F} and \mathcal{G} are ϵ -interleaved if there exist maps $\varphi_a : F_a \rightarrow G_{a+\epsilon}$ and $\psi_a : G_a \rightarrow F_{a+\epsilon}$ such that the same type of diagrams commute up to homotopy, that is, for example $\varphi_{a'} \circ \iota_{a,a'}^F \simeq \iota_{a+\epsilon, a'+\epsilon}^G \circ \varphi_a$ are homotopic (contiguous).

We again define the interleaving distance (now between filtrations):

$$d_I(\mathcal{F}, \mathcal{G}) = \inf\{\epsilon \mid \mathcal{F} \text{ and } \mathcal{G} \text{ are } \epsilon\text{-interleaved}\}.$$

From induced homology, we immediately get the following observation:

Observation 6.22. For all $p \geq 0$, $d_I(H_p \mathcal{F}, H_p \mathcal{G}) \leq d_I(\mathcal{F}, \mathcal{G})$.

As a first application of interleaving distance, we can quantify how different the Čech and Vietoris-Rips filtrations are. Recall that for a point cloud P and a radius r , we have the relationship between the Čech and Vietoris-Rips complexes as follows: $C^r(P) \subseteq \mathbb{V}R^r(P) \subseteq C^{2r}(P)$. Since this factor 2 is multiplicative, and we need an additive ϵ for interleaving, let us just take the logarithmic scale (base 2) for the radius, i.e., we define $C_{\log}^r = C^{2^r}$ and similarly $\mathbb{V}R_{\log}^r = \mathbb{V}R^{2^r}$. Since $2^{(r+1)} = 2 \cdot 2^r$, we have $C_{\log}^r(P) \subseteq \mathbb{V}R_{\log}^r(P) \subseteq C_{\log}^{r+1}(P)$.

We thus have the following inclusions:

$$\begin{array}{ccccc} C_{\log}^r & \longrightarrow & C_{\log}^{r+1} & \longrightarrow & C_{\log}^{r+2} \\ & \searrow & \nearrow & \searrow & \nearrow \\ \mathbb{V}R_{\log}^r & \longrightarrow & \mathbb{V}R_{\log}^{r+1} & \longrightarrow & \mathbb{V}R_{\log}^{r+2} \end{array}$$

Since these are all inclusions, all relevant diagrams must commute, and thus we get that $d_I(C_{\log}, \mathbb{V}R_{\log}) \leq 1$.

6.2.2 Stability with Respect to Interleaving Distance

The main motivation for interleaving distance is that it can be used to prove stability results, at least under some tameness conditions.

Definition 6.23. *A persistence module \mathbb{V} is q -tame if the linear maps have finite rank.*

Note that in this definition, the q is not a parameter, just a name. All persistence modules that show up in the context of persistent homology on point clouds are q -tame, so this condition is not restrictive.

Theorem 6.24. *If \mathbb{U}, \mathbb{V} are q -tame persistence modules over \mathbb{R} , then*

$$d_b(\text{Dgm}\mathbb{U}, \text{Dgm}\mathbb{V}) = d_I(\mathbb{U}, \mathbb{V}).$$

Thus, for every interleaving one can find between two persistence modules or between filtrations, one immediately gets an upper bound on the Bottleneck distance. This is a very powerful result, and the proof of it is out of scope for these lecture notes. One direction of the proof however follows from a decomposition result of persistence modules that we will discuss in Section 6.3. But first, we will consider some examples of how Theorem 6.24 can be used to prove stability theorems.

Exercise 6.25. *Prove Theorem 6.9.*

6.2.3 Stability for Čech Complexes

So far, we have only seen stability results comparing filtrations induced by different functions on a fixed space. However, in applications in data analysis, we consider complexes on point clouds, and two different point clouds might not have the same size. Thus, the simplicial complexes on which we get filtrations are generally different. Using interleaving distance, we can however still give stability results. In this section, we will do this for Čech complexes.

Consider two point clouds P, Q in the same metric space X . Let us first consider the really simple case, where $P = \{p\}$, and $Q = \{q\}$ with $d(p, q) = d$. Then, $B(p, r) \subseteq B(q, r + d)$. Now, how does this generalize to larger point sets? To get the same kind of behavior, we need that for every point in P , there exists some point in Q with distance at most d . This motivates the following distance measure:

Definition 6.26 (Hausdorff distance). *Let $A, B \subseteq X$ be compact sets. Then the Hausdorff distance between A and B is defined as*

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

Exercise 6.27. *Show that Hausdorff distance is a metric.*

We can now see that if $d_H(P, Q) = d$, then $\bigcup_{p \in P} B(p, r) \subseteq \bigcup_{q \in Q} B(q, r + d)$. From this, we get the following lemma:

Lemma 6.28. *For P, Q with $d_H(P, Q) = d$, the (filtrations given by) the Čech complexes of P and Q are d -interleaved.*

Proof. Consider the following diagram:

$$\begin{array}{ccccc}
 \mathbb{C}^r(P) & \longrightarrow & \mathbb{C}^{r+d}(P) & \longrightarrow & \mathbb{C}^{r+2d}(P) \\
 \downarrow \simeq & & \downarrow \simeq & & \downarrow \simeq \\
 \bigcup_{p \in P} B(p, r) & & \bigcup_{p \in P} B(p, r+d) & & \bigcup_{p \in P} B(p, r+2d) \\
 & \searrow & \nearrow & \searrow & \nearrow \\
 \bigcup_{q \in Q} B(q, r) & & \bigcup_{q \in Q} B(q, r+d) & & \bigcup_{q \in Q} B(q, r+2d) \\
 \downarrow \simeq & & \downarrow \simeq & & \downarrow \simeq \\
 \mathbb{C}^r(Q) & \longrightarrow & \mathbb{C}^{r+d}(Q) & \longrightarrow & \mathbb{C}^{r+2d}(Q)
 \end{array}$$

The relevant sub-diagrams commute up to homotopy, since we only chain together homotopies and inclusion maps. \square

We can conclude the following

Theorem 6.29. $d_b(\text{Dgm}_p(\mathbb{C}(P)), \text{Dgm}_p(\mathbb{C}(Q))) \leq d_H(P, Q)$ for all $p \geq 0$.

Proof. By Theorem 6.24, Observation 6.22, and finally Lemma 6.28, we have

$$d_b(\dots) = d_I(H_p \mathbb{C}(P), H_p \mathbb{C}(Q)) \leq d_I(\mathbb{C}(P), \mathbb{C}(Q)) \leq d_H(P, Q). \quad \square$$

6.3 Interval Decomposition of Persistence Modules

In this section, we again look at persistence modules, this time as algebraic structures. We consider persistence modules over \mathbb{R} of vector spaces over some field \mathbb{F} . We start by looking at some special persistence modules, called *interval modules*.

Definition 6.30. *A interval module $\mathbb{I}[b, d]$ is an persistence module*

$$V_a = \begin{cases} \mathbb{F} & \text{if } a \in [b, d], \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad v_{a, a'} = \begin{cases} \text{id} & b \leq a \leq a' \leq d, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we can define interval modules on open and clopen intervals, denoted by $\mathbb{I}(b, d)$, $\mathbb{I}(b, d]$, and $\mathbb{I}[b, d)$. We write $\mathbb{I}\langle b, d \rangle$ to include all four of these types.

For an interval module we can easily talk about birth and death as we did in persistent homology. If we have a persistent homology module that is (isomorphic to) an interval module, the birth and death correspond to the boundaries b, d of the interval.

Definition 6.31. A persistence module \mathbb{U} is called pointwise finite dimensional (p.f.d.) if for all $a \in \mathbb{R}$, \mathbb{U}_a has finite dimension.

Note that all p.f.d. persistence modules are also q-tame.

Definition 6.32. Given two persistence modules \mathbb{U}, \mathbb{V} , we define their direct sum $\mathbb{U} \oplus \mathbb{V}$ by $(\mathbb{U} \oplus \mathbb{V})_a = \mathbb{U}_a \oplus \mathbb{V}_a$ and $(u \oplus v)_{a,a'} = u_{a,a'} \oplus v_{a,a'}$.

Here, the direct sum of maps just means applying the respective maps component-wise.

Proposition 6.33. If $\mathbb{U}_1, \mathbb{U}_2$ are ϵ -interleaved, and $\mathbb{V}_1, \mathbb{V}_2$ are δ -interleaved, then $\mathbb{U}_1 \oplus \mathbb{V}_1$ and $\mathbb{U}_2 \oplus \mathbb{V}_2$ are $\max\{\epsilon, \delta\}$ -interleaved.

Proof. Without loss of generality, let $\epsilon \geq \delta$, so we need to show that they are ϵ -interleaved. Recall that if two persistence modules are δ -interleaved, they are also ϵ -interleaved. Let φ^u, ψ^u be (series of) functions showing that $\mathbb{U}_1, \mathbb{U}_2$ are ϵ -interleaved. Similarly, let φ^v, ψ^v be (series of) functions showing that $\mathbb{V}_1, \mathbb{V}_2$ are ϵ -interleaved. Then, $\varphi^u \oplus \varphi^v, \psi^u \oplus \psi^v$ show that $\mathbb{U}_1 \oplus \mathbb{V}_1$ and $\mathbb{U}_2 \oplus \mathbb{V}_2$ are ϵ -interleaved. \square

If we now have a direct sum of interval modules, we can still nicely talk about birth and death: we just look at each interval module in isolation. The following theorem shows that surprisingly most persistence modules can be expressed as direct sums of interval modules.

Theorem 6.34 (Structure theorem). Any p.f.d. persistence module decomposes uniquely into interval modules, i.e., we have

$$\mathbb{U} \cong \bigoplus_{i \in I} \mathbb{I}\langle b_i, d_i \rangle.$$

The intervals $\langle b_i, d_i \rangle$ are exactly the barcodes if \mathbb{U} is a persistent homology module.

Note that unless we have some additional tame-ness condition on \mathbb{U} , I is not guaranteed to be finite.

Recall that when we talked about persistent homology, we said that there is some consistent global choice of basis for persistent homology groups. That is a consequence of the structure theorem. The structure theorem also allows us to prove one direction of Theorem 6.24, which we will do in the following.

Proposition 6.35. Consider two interval modules $\mathbb{I}_1 = \mathbb{I}\langle b_1, d_1 \rangle$ and $\mathbb{I}_2 = \mathbb{I}\langle b_2, d_2 \rangle$. Then, $d_I(\mathbb{I}_1, \mathbb{I}_2) = d_b(\text{Dgm}\mathbb{I}_1, \text{Dgm}\mathbb{I}_2)$.

Proof. To prove that $d_I(\mathbb{I}_1, \mathbb{I}_2) \geq d_b(\text{Dgm}\mathbb{I}_1, \text{Dgm}\mathbb{I}_2)$, we show that every upper bound on d_I is also an upper bound on d_b : assume that we have maps φ, ψ showing that the two modules are ϵ -interleaved. Then, consider $\psi_{a+\epsilon} \circ \varphi_a = v_{a, a+2\epsilon}^1$, equality holding because φ, ψ certify ϵ -interleaving. Consider $a \in \langle b_1, d_1 \rangle$.

Case 1: $v_{\alpha, \alpha+2\epsilon}^1 = 0$ for all $\alpha \in \langle b_1, d_1 \rangle$. Then, $d_1 - b_1 < 2\epsilon$, and the (infinity-norm) distance of (b_1, d_1) to the diagonal is less than ϵ .

Case 2: $v_{\alpha, \alpha+2\epsilon}^1 = \text{id}$ for some $\alpha \in \langle b_1, d_1 \rangle$. Then, $d_1 - b_1 \geq 2\epsilon$. Furthermore, we have $\varphi_\alpha(\mathbb{F}) = \mathbb{F}$ for all $\alpha \in \langle b_1, d_1 - 2\epsilon \rangle$. So, for these α , we must also have $\alpha + \epsilon \in \langle b_2, d_2 \rangle$. This tells us that $\langle b_2, d_2 \rangle$ must “cover” a large part of $\langle b_1, d_1 \rangle$, namely we get $b_2 \leq b_1 + \epsilon$, and $d_2 \geq d_1 - \epsilon$. We can now see that $|b_2 - b_1| \leq \epsilon$ and $|d_2 - d_1| \leq \epsilon$: to violate this, $\langle b_2, d_2 \rangle$ would have to be a larger interval than $\langle b_1, d_1 \rangle$ (in particular, it would be longer than 2ϵ), and we could thus exchange their roles and get that $b_1 \leq b_2 + \epsilon$ and $d_1 \geq d_2 - \epsilon$. From this, we get that $d_\infty((b_1, d_1), (b_2, d_2)) \leq \epsilon$, and thus get the bound on d_b .

We now prove the other direction, $d_1(\mathbb{I}_1, \mathbb{I}_2) \leq d_b(\text{Dgm}\mathbb{I}_1, \text{Dgm}\mathbb{I}_2)$. To see this, we show that from every matching whose longest edge is ϵ , we get an ϵ -interleaving.

Case 1: The two off-diagonal points are matched to the diagonal. Then, we get that $d_i - b_i \leq 2\epsilon$ for both of them, and thus for all $\epsilon' > \epsilon$, \mathbb{I}_1 and \mathbb{I}_2 are ϵ' -interleaved with $\varphi, \psi = 0$. Thus, $d_1 \leq \epsilon$.

Case 2: The points are matched with each other. Then, $|b_2 - b_1| \leq \epsilon$ and $|d_2 - d_1| \leq \epsilon$. Taking $\varphi, \psi = \text{id}$ we can see that \mathbb{I}_1 and \mathbb{I}_2 are ϵ -interleaved. Thus, $d_1 \leq \epsilon$. \square

Corollary 6.36. *Let \mathbb{U}, \mathbb{V} be p.f.d. persistence modules. Then, $d_1(\mathbb{U}, \mathbb{V}) \leq d_b(\text{Dgm}\mathbb{U}, \text{Dgm}\mathbb{V})$.*

Proof. We apply the structure theorem to write $\mathbb{U} = \bigoplus_{i \in I} \mathbb{I}\langle b_i, d_i \rangle \oplus \bigoplus_{j \in J} 0$ and $\mathbb{V} = \bigoplus_{j \in J} \mathbb{I}\langle b_j, d_j \rangle \oplus \bigoplus_{i \in I} 0$. From the Bottleneck matching we get a matching between parts making up \mathbb{U} and \mathbb{V} . Since the Bottleneck distance is the maximum length of any edge, we have $d_b(\text{Dgm}\mathbb{U}, \text{Dgm}\mathbb{V}) \geq d_b(\text{Dgm}\mathbb{I}_1, \text{Dgm}\mathbb{I}_2) = d_1(\mathbb{I}_1, \mathbb{I}_2)$ for every two interval modules that were matched together, where we used Proposition 6.35. Finally, we use Proposition 6.33 to get the desired statement. \square

Questions

20. *How can we measure distances between persistence diagrams?* Discuss Bottleneck and Wasserstein distance.
21. *How stable are filtrations derived from simplex-wise monotone functions with respect to Bottleneck distance?* State, illustrate and prove the stability theorem (Theorem 6.6).
22. *How can we measure distances between persistence diagrams?* Define interleaving distance and discuss its relation to Bottleneck distance.
23. *How stable are Čech complexes to perturbations of the underlying point set?* Define Hausdorff distance, state and prove the stability theorem for Čech complexes (Theorem 6.29).