ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute of Theoretical Computer Science
Patrick Schnider

Introduction to Topological Data Analysis **Scribe Notes 15** **FS23**

Scribe notes by Simon Weber. Please contact me for corrections.
Lecture date: April 21, 2023
Last update: Monday 24th April, 2023, 14:58

Distance Metrics on Persistence Diagrams

In this section we will define some distance metrics that can be used to compare different persistence diagrams.

Bottleneck Distance

Let $\mathcal{F}, \mathcal{G}$ be two filtrations giving rise to persistence modules $H_p\mathcal{F}, H_p\mathcal{G}$. Let $\mathrm{Dgm}_p(\mathcal{F})$ and $\mathrm{Dgm}_p(\mathcal{G})$ be their corresponding persistence diagrams. These diagrams are the information we want to use to compare $\mathcal{F}$ and $\mathcal{G}$.

The general idea of the bottleneck distance is to pair up points of the two persistence diagrams, i.e., consider bijections between points of $\mathrm{Dgm}_p(\mathcal{F})$ and $\mathrm{Dgm}_p(\mathcal{G})$. Since we can only find bijections between sets of the same cardinality, we need the two diagrams to have the same number of points. This is where the definition of the persistence diagram comes in: recall that a persistence diagram includes every possible point on the diagonal with infinite multiplicity. Thus, both sets of points have the same (infinite) cardinality, and bijections between these sets are thus well-defined.

To measure the "quality" or "distance" of such a bijection, we use the L_∞-norm:

Definition 1. *Let $x = (x_1, x_2), y = (y_1, y_2)$ be two points in $\mathbb{R}^2$. Then,*

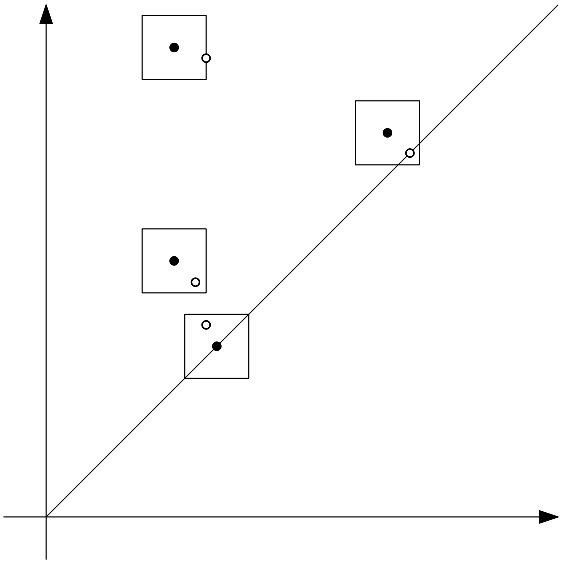
$$||x - y||_\infty := \max(|x_1 - y_1|, |x_2 - y_2|),$$

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

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

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



Observation 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. $d_b(X, Y) = 0$ iff $X = Y$ 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. $d_b(X, Y) = d_b(Y, X)$ is clear by definition.

3. $d_b(X, Y) \leq d_b(X, Z) + d_b(Z, Y)$. Take a bijection π_1 witnessing $d_b(X, Z)$ and a bijection π_2 witnessing $d_b(Z, Y)$, and concatenate the two: $\pi := \pi_2 \circ \pi_1$ is a bijection $X \to Y$ where for every $x \in X$ we have $||x - \pi(x)||_\infty \leq ||x - \pi_1(x)||_\infty + ||\pi_1(x) - \pi_2(x)||_\infty$. Note 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. □

Recall that simplex-wise monotone functions $f, g : K \to \mathbb{R}$ give rise to simplicial sublevel set filtrations $\mathcal{F}_f, \mathcal{F}_g$. We could now compare the persistence diagrams of these two filtrations using the Bottleneck distance, but we wish to define a metric directly between the two functions f, g:

Definition 4 (infinity norm). *Let* $f, g : X \to \mathbb{R}$. *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 5 (Stability for simplicial filtrations). *Let* $f, g : K \to \mathbb{R}$ *be simplex-wise monotone functions. Then,* $\forall p \geq 0$ *we have* $d_b(\mathrm{Dgm}_p(\mathcal{F}_f), \mathrm{Dgm}_p(\mathcal{F}_g)) \leq \|f - g\|_\infty$.

Proof. Let $f_t := (1-t)f + tg$ for $t \in [0, 1]$ be the linear interpolation 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. It is clearly simplex-wise, and we prove that it is also monotone: 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 $\mathrm{Dgm}_p(\mathcal{F}_{f_t})$ as a multiset in $\mathbb{R}^2 \times [0, 1]$. Each off-diagonal point of $X_t := \mathrm{Dgm}_p(\mathcal{F}_{f_t})$ 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 $0 = t_0 < t_1 < t_2 < \ldots < t_n < t_{n+1} = 1$. Without loss of generality, 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.

At t_{i+1}, if $x(t_{i+1})$ is an off-diagonal point whose creator and destructor are still paired after t_{i+1}, $x(t)$ 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, recall by Exercise Sheet 5, Question 3, 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, and there is a unique continuing line segment for both of them.

Lastly, if $x(t_{i+1})$ is on the diagonal, this means that its previous constructor and destructor now have the same value in $f_{t_{i+1}}$. There is no continuation for this point.

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*. 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.

Between t_i and t_{i+1} we get for $\frac{\delta x(t)}{\delta t}$:

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

Projecting $x(t_{i+1})$ and $x(t_i)$ to $\mathbb{R}^2$ we get two points y_{i+1}, y_i such 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. $\square$