Stability of persistence diagrams

Essay to a Journal Club contribution held by Daniel Spitz, the 30th of October, 2019

1 Introduction

Given a point cloud X, it is a natural question to ask for the stability of the persistence diagram of the filtration of alpha complexes of X against perturbations of the latter. This is precisely what persistence theorems accomplish, making persistent homology a useful notion in the analysis of potentially noisy data.

In this Journal Club contribution we describe two of the first such stability results, for two different types of metrics on persistence diagrams: the Bottleneck and the Wasserstein distance. Actually, both do not describe persistence diagrams of inter alia alpha complexes of point clouds, but, instead, those of the filtration of sublevel sets of functions with preimage a triangulable topological space.

Throughout this section, a persistence diagram can be regarded as a multiset of points in the plane $\bar{\mathbb{R}}^2$, all laying above the diagonal. We may add arbitrarily many points on the diagonal, which have zero persistence, making them not essential to the diagram but simplifying definitions. We work with homology groups having $\mathbb{Z}_2$-coefficients.

2 Bottleneck distance stability

Let X, Y be two persistence diagrams and $\eta : X \to Y$ a bijection between them, possibly adding points to the diagonal to be able to define a bijection. Measuring the distance between two points $x = (x_1, x_2)$ and $y = (y_1, y_2)$ as $||x - y||_\infty = \max\{|x_1 - y_1, x_2 - y_2|\}$ and taking the infimum over all bijections, we define the *bottleneck distance* between X and Y as

$$W_\infty(X, Y) = \inf_{\eta: X \to Y} \sup_{x \in X} ||x - \eta(x)||_\infty. \tag{1}$$

Indeed, with $W_\infty(X, Y) = 0$ if and only if $X = Y$, $W_\infty(X, Y) = W_\infty(Y, X)$ and $W(X, Z) \leq W_\infty(X, Y) + W_\infty(Y, Z)$ we find that W_∞ is a metric on the space of persistence diagrams.

We note that homology groups can be defined not only for simplicial complexes, but also for any topological space, for example via singular homology. For details on this we refer to Refs. [1, 2].

Definition 1. Let T be a topological space and f a real function on T. A *homological critical value* of f is a real number a for which there exists an integer k such that for all sufficiently small $\epsilon > 0$ the map $H_k(f^{-1}(-\infty, a - \epsilon]) \to H_k(f^{-1}(-\infty, a + \epsilon])$ induced by inclusion is not an isomorphism.

Definition 2. A function $f : T \to \mathbb{R}$ is *tame* if it has a finite number of homological critical values and the homology groups $H_k(f^{-1}(-\infty, a])$ are finite-dimensional for all $k \in \mathbb{Z}$ and $a \in \mathbb{R}$.

Let $f : T \to \mathbb{R}$ be such a tame function and set $X_a := f^{-1}((-\infty, a])$. Noting that $X_a \subseteq X_b$ whenever $a \leq b$, we obtain the filtration of sublevel sets. To this end, for any $a \leq b$ there exists a map $\iota_\ell^{a,b} : H_\ell(X_a) \to H_\ell(X_b)$ induced by the inclusion. We say that a class $\alpha \in H_\ell(X_a)$ is born at X_a if $\alpha \notin \mathrm{im}(\iota_\ell^{a-\delta,a})$ for any $\delta > 0$, setting $b(\alpha) = a$. A class α born at X_a dies entering X_b if $\iota_\ell^{a,b-\delta}(\alpha) \notin \mathrm{im}(\iota_\ell^{a-\delta,b-\delta})$ for all $\delta > 0$ but $\iota_\ell^{a,b}(\alpha) \in \mathrm{im}(\iota_\ell^{a-\delta,b})$, setting $d(\alpha) = b$. We define its persistence as $\mathrm{pers}(\alpha) = d(\alpha) - b(\alpha)$. By $\mathrm{Dgm}_\ell(f)$ we denote the corresponding persistence diagram, consisting of all points $(b(\alpha), d(\alpha))$ for ℓ-dimensional persistent homology classes α.

We recall that a topological space is triangulable if there is a (finite) simplicial complex with homeomorphic underlying space.

Following Ref. [3], the Bottleneck stability theorem finally reads as follows.

Theorem 1 (Bottleneck stability theorem)**.** *Let T be a triangulable space with continuous tame functions $f, g : T \to \mathbb{R}$. Then the persistence diagrams satisfy for all $\ell \in \mathbb{N}$*

$$W_\infty(\mathrm{Dgm}_\ell(f), \mathrm{Dgm}_\ell(g)) \leq ||f - g||_\infty = \sup_x |f(x) - g(x)|. \tag{2}$$

To this extend, under mild assumptions on the function, the persistence diagram is stable. Small changes in the function imply only small changes in the diagram. The proof of the theorem proceeds via diagram chasing and an intermediate upper bound on the Hausdorff distance between the persistence diagrams.

3 Wasserstein distance stability

The Wasserstein stability theorem we deduce in somewhat more detail than the Bottleneck stability theorem. Derivations proceed along the lines of Ref. [4]. We begin by stating preliminary technicalities.

Let X be a triangulable, compact n-dimensional metric space, $d : X \times X \to \mathbb{R}$ its metric. As stated previously, a triangulation of X is a finite simplicial complex K with homeomorphism $\vartheta : |K| \to X$. We define the diameter of a simplex $\sigma \in K$ as $\mathrm{diam}(\sigma) := \max_{x,y\in\sigma} d(\vartheta(x), \vartheta(y))$ and the mesh of a triangulation K as $\mathrm{mesh}(K) := \max_{\sigma\in K} \mathrm{diam}(\sigma)$. For all $0 \leq \ell \leq n$ we denote the ℓ-skeleton of K by $K^{(\ell)}$. We are interested in the smallest triangulation with mesh at most r,

$$N(r) := \min_{\mathrm{mesh}(K)\leq r} \mathrm{card}(K), \qquad N_\ell(r) := \min_{\mathrm{mesh}(K)\leq r} (\mathrm{card}(K^{(\ell)}) - \mathrm{card}(K^{\ell-1})). \tag{3}$$

As an example consider X a compact Riemannian manifold. Then, for sufficiently small r there exist $c, C > 0$, such that $c/r^n \leq N(r) \leq C/r^n$.

Additionally, for any subset $z \subseteq X$, we define

$$z^r := \{x \in X \,|\, \exists\, y \in z : d(x, y) \leq r\}. \tag{4}$$

A series of lemmas brings us to the main results.

Lemma 1 (Snapping Lemma)**.** Let K be a triangulation of a compact metric space X with $\mathrm{mesh}(K) = r$. Then for each cycle z of dimension ℓ in X there is a cycle $\bar{z}$ in the ℓ-skeleton of K that is homologous to z inside z^r.

A crucial ingredient of Wasserstein stability is the involved functions being Lipschitz. A function $f : X \to \mathbb{R}$ is Lipschitz on X, if there exists a positive constant c, such that $|f(x) - f(y)| \leq c\, d(x, y)$ for any $x, y \in X$. The infimum of such c is called Lipschitz constant and denoted by $\mathrm{Lip}(f)$.

A useful lemma follows, including persistent homology notions.

Lemma 2 (Persistent Cycle Lemma)**.** Let X be a triangulable, compact metric space, $f : X \to \mathbb{R}$ a tame Lipschitz function. Then the number of points in the persistence diagrams of f whose persistence exceeds ϵ is at most $N(\epsilon/\mathrm{Lip}(f))$.

We define the *degree-k total persistence*,

$$\mathrm{Pers}_k(f,t) := \sum_{\mathrm{pers}(x)>t} \mathrm{pers}(x)^k, \qquad \mathrm{Pers}_k(f) := \mathrm{Pers}_k(f,0). \tag{5}$$

Lemma 3 (Moment Lemma)**.** Let X be a triangulable, compact metric space, $f : X \to \mathbb{R}$ a tame Lipschitz function. Then,

$$\mathrm{Pers}_k(f,t) \leq t^k N\left(\frac{t}{\mathrm{Lip}(f)}\right) + k \int_{\epsilon=t}^{\mathrm{Amp}(f)} N\left(\frac{\epsilon}{\mathrm{Lip}(f)}\right) \epsilon^{k-1}\, d\epsilon, \tag{6}$$

setting $\mathrm{Amp}(f) = \max_{x\in X} f(x) - \min_{y \in X} f(y)$.

The first term on the right-hand side of Eq. (6) we denote by A, the second by B.

We define the notion of polynomial growth and bounded total persistence in what follows. Assume that the size of the smallest triangulation *grows polynomially* with one over the mesh, that is, there exist $C_0, M > 0$, such that $N(r) \leq C_0/r^M$ for all $r > 0$. Let $\delta > 0$ and $k = M + \delta$. We then find upper bounds for A and B:

$$A \leq C_0 \mathrm{Lip}(f)^M \mathrm{Amp}(f)^\delta, \qquad B \leq C_0 \mathrm{Lip}(f)^M \mathrm{Amp}(f)^\delta \frac{M+\delta}{\delta}, \tag{7}$$

which motivates the introduction of the following concept.

Definition 3. A metric space X *implies bounded degree-k total persistence*, if there exists a constant $C_X > 0$ depending only on X, such that $\mathrm{Pers}_k(f) \leq C_X$ for every tame function $f : X \to \mathbb{R}$ with $\mathrm{Lip}(f) \leq 1$.

As an example consider $X = S^n$. One finds a $C_0 > 0$, such that $N(r) \leq C_0/r^n$. Thus, a $C > 0$ exists with $\mathrm{Pers}_k(f) \leq C$ for some C and every $k = n + \delta$, $\delta > 0$.

Let $f, g : X \to \mathbb{R}$ be two tame functions with persistence diagrams $\mathrm{Dgm}_\ell(f)$ and $\mathrm{Dgm}_\ell(g)$, respectively, $\ell \in \mathbb{N}$. The *degree-p Wasserstein distance* between the persistence diagrams of f and g is defined as

$$W_p(f,g) = \left[\sum_\ell \inf_{\gamma_\ell} \sum_x ||x - \gamma_\ell(x)||_\infty^p\right]^{1/p}, \tag{8}$$

where the first sum runs of all dimensions ℓ, the infimum is taken over all bijections $\gamma_\ell : \mathrm{Dgm}_\ell(f) \to \mathrm{Dgm}_\ell(g)$, adding zero-persistence points to render this well-defined, and the second sum is over all $x \in \mathrm{Dgm}_\ell(f)$.

Theorem 2 (Wasserstein Stability Theorem)**.** *Let X be a triangulable, compact metric space that implies bounded degree-k total persistence for $k \geq 1$, and let $f, g : X \to \mathbb{R}$ be two tame Lipschitz functions. Then,*

$$W_p(f,g) \leq C_k^{1/p} \cdot ||f-g||_\infty^{1-k/p} \tag{9}$$

for all $p \geq k$ and $C_k = C_X \max\{\mathrm{Lip}(f)^k, \mathrm{Lip}(g)^k\}$.

Another result follows quickly.

Theorem 3 (Total Persistence Stability Theorem)**.** *Let X be a triangulable, compact metric space that implices bounded degree-k total persistence for $k \geq 0$, and let $f, g : X \to \mathbb{R}$ be two tame Lipschitz functions. Then,*

$$|\mathrm{Pers}_p(f) - \mathrm{Pers}_p(g)| \leq 4p\, w^{p-1-k}\, C_k \cdot ||f-g||_\infty, \tag{10}$$

for every real $p \geq k+1$, $C_k := C_X \max\{\mathrm{Lip}(f)^k, \mathrm{Lip}(g)^k\}$ and w is bounded from above by $\max\{\mathrm{Amp}(f), \mathrm{Amp}(g)\}$.

As the notation suggests, the Bottleneck distance, W_∞, arises as the limit of the Wasserstein distance, W_p, for $p \to \infty$ [5]. To this end, under the assumptions of the Wasserstein stability theorem the Bottleneck stability theorem follows.

References

[1] J. Munkres, *Elements of algebraic topology.* Advanced book program, Addison-Wesley, 1984.

[2] A. Hatcher, *Algebraic topology.* Cambridge: Cambridge Univ. Press, 2000.

[3] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, "Stability of persistence diagrams," *Discrete & Computational Geometry*, vol. 37, no. 1, pp. 103–120, 2007.

[4] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko, "Lipschitz functions have l p-stable persistence," *Foundations of computational mathematics*, vol. 10, no. 2, pp. 127–139, 2010.

[5] H. Edelsbrunner, "Alpha shapes—a survey," *Tessellations in the Sciences*, 01 2010.