Wasserstein stability for the p-Dowker nerve

May 4, 2023

1 Setting

We fix a real number $p \geq 1$ and endow the interval $[0, \infty]$ with the binary operation $\otimes \colon [0, \infty]^2 \to [0, \infty]$ defined by $t \otimes s := (t^p + s^p)^{\frac{1}{p}}$.

Definition 1.1. Given a Dowker dissimilarity $\Lambda \colon L \times W \to [0, \infty]$, define the *$p$-Dowker nerve* $D\Lambda$ of Λ by letting

$$D\Lambda_t := \left\{ \sigma \subseteq L \mid \exists w \in W \text{ such that } \bigotimes_{l \in \sigma} \Lambda(l, w) \leq t \right\} \text{ for } t \in [0, \infty).$$

Let $W = \mathbb{R}^d$ be equipped with the Euclidean distance $d(x, y) = \|x - y\|_2$, and let L and L' be finite subsets of W with $|L| = |L'|$[1]. Define the Dowker dissimilarity $\Gamma \colon W \times W \to [0, \infty]$ by letting $\Gamma(w, w') = d(w, w')$, and let $\Lambda = \Gamma \mid_{L \times W}$ and $\Lambda' = \Gamma \mid_{L' \times W}$. Setting $p = \infty$ we have that $D\Lambda$ and $D\Lambda'$ are the ambient Čech filtrations of the point clouds L and L', respectively. In this case, we have classical stability, i.e., the bottleneck distance between the persistence diagrams are upper-bounded by the Hausdorff distance between the two point clouds. In this note, we apply the Cellular Wasserstein Stability Theorem of [3] to establish an analogous stability result for the case $1 \leq p < \infty$.

2 Definitions

Let D and D' be two persistence diagrams. A *matching* M between D and D' is a set of pairs (x, y) with $x \in X \cup \Delta$, $y \in Y \cup \Delta$ and every $x \in X$ and $y \in Y$ is used exactly once. Here Δ denotes the diagonal.

[1]It would be nice if we could drop this assumption.

Definition 2.1 (p-Wasserstein distance)**.** Given $p \geq 1$, the *p-Wasserstein distance* between D and D' is defined as

$$\mathbf{W}_p(D, D') := \inf_M \left(\sum_{(x,y)\in M} \|x - y\|_p^p \right)^{\frac{1}{p}}$$

where M ranges over all matchings between the two diagrams. If $D = (D_k)_k$ and $D' = (D'_k)_k$, which is the case when we consider different homological dimensions, we define the *total p-Wasserstein distance* by

$$\mathbf{W}_p(D, D') := \left(\sum_k W_p(D_k, D'_k)^p \right)^{\frac{1}{p}}.$$

Definition 2.2 (L^p norm)**.** Let K be a finite CW-complex. The *L^p norm* of a function $f\colon K \to \mathbb{R}$ is given by $\|f\|_p = (\sum_{\sigma\in K} |f(\sigma)|^p)^{\frac{1}{p}}$. Given two monotone functions $f, g\colon K \to \mathbb{R}$, the *$L^p$ distance* between f and g is the distance induced by the L^p norm. That is,

$$\|f - g\|_p = \left(\sum_{\sigma\in K} |f(\sigma) - g(\sigma)|^p \right)^{\frac{1}{p}}.$$

Let L and L' be finite subsets of $\mathbb{R}^d$. A subset $C \subseteq L \times L'$ is called a *correspondence* between L and L' if for each $l \in L$ there exists an $l' \in L$ with $(l, l') \in C$, and for each $l' \in L$ there exists an $l \in L$ such that $(l, l') \in C$. Let $\mathbf{C}(L, L')$ denote the set of all correspondences between L and L'. The following definition of the p-Hausdorff distance between point sets appears in [2].

Definition 2.3 (p-Hausdorff distance)**.** For $p \geq 1$, we define the *p-Hausdorff* distance between L and L' by

$$\mathbf{H}_p(L, L') := \inf_{C\in\mathbf{C}(L,L')} \left(\sum_{(l,l')\in C} \|l - l'\|_2^p \right)^{\frac{1}{p}}$$

Note that if $|L| = |L'|$ and $C \in \mathbf{C}(L, L')$ is the correspondence minimizing the expression $\left(\sum_{(l,l')\in C} \|l - l'\|_2^p\right)^{\frac{1}{p}}$, then C is necessarily a bijection.

3 Stability

Theorem 3.1 (Cellular Wasserstein Stability Theorem [3])**.** Let K be a finite CW complex and let $f, g\colon K \to \mathbb{R}$ be monotone functions. Then

$$\mathbf{W}_p(\mathrm{Dgm}(f), \mathrm{Dgm}(g)) \leq \|f - g\|_p.$$

For a fixed homological dimension k, we have

$$\mathbf{W}_p(\mathrm{Dgm}_k(f), \mathrm{Dgm}_k(g))^p \leq \sum_{k \leq \dim(\sigma) \leq k+1} |f(\sigma) - g(\sigma)|^p.$$

For the rest of this section, let L and L' be finite subsets of $W = \mathbb{R}^d$ with $n = |L| = |L'|$. Define the Dowker dissimilarity $\Gamma\colon W \times W \to [0, \infty]$ by letting $\Gamma(w, w') = d(w, w')$, and let $\Lambda = \Gamma\mid_{L \times W}$ and $\Lambda' = \Gamma\mid_{L' \times W}$. Let $\mathrm{Dgm}(\Lambda)$ denote the persistence diagram corresponding to the sublevel set homology of the filtration function $D\Lambda \to \mathbb{R}$ defined by

$$\sigma \mapsto \inf\{t \in [0, \infty] \mid \sigma \in D\Lambda_t\} = \inf_{w \in W} \bigotimes_{l \in \sigma} \Lambda(l, w) = \inf_{w \in W} \left(\sum_{l \in \sigma} \|l - w\|_2^p \right)^{\frac{1}{p}}. \quad (1)$$

Corollary 3.1.1 (Stability of the p-Dowker nerve)**.** Let Λ and Λ' be defined as above. Then

$$\mathbf{W}_p(\mathrm{Dgm}(\Lambda), \mathrm{Dgm}(\Lambda')) \leq C \cdot \mathbf{H}_p(L, L').$$

where $C = (2^n - 1)^{\frac{1}{p}}$.

Proof. Let C be a minimizing correspondence in the p-Hausdorff distance $\mathbf{H}_p(L, L')$. Since $|L| = |L'|$ this defines a bijection $\varphi\colon L \to L'$. Let $f\colon D\Lambda \to \mathbb{R}$ be the filtration function defined in Equation (1). Similarly, let $f'\colon D\Lambda' \to \mathbb{R}$ be the filtration function for $D\Lambda'$ and define $g := f' \circ \varphi\colon D\Lambda \to \mathbb{R}$. For $\sigma \in D\Lambda$, we have

$$
\begin{aligned}
|f(\sigma)-g(\sigma)| &= \left|\inf_{w\in W}\bigotimes_{l\in\sigma}\Lambda(l,w)-\inf_{w\in W}\bigotimes_{l\in\sigma}\Lambda(\varphi(l),w)\right| \\
&\overset{(1)}{\leq} \sup_{w\in W}\left|\bigotimes_{l\in\sigma}\Lambda(l,w)-\bigotimes_{l\in\sigma}\Lambda(\varphi(l),w)\right| \\
&= \sup_{w\in W}\left|\left(\sum_{l\in\sigma}\|l-w\|_2^p\right)^{\frac{1}{p}}-\left(\sum_{l\in\sigma}\|\varphi(l)-w\|_2^p\right)^{\frac{1}{p}}\right| \\
&\overset{(2)}{\leq} \sup_{w\in W}\left|\left(\sum_{l\in\sigma}|\|l-w\|_2-\|\varphi(l)-w\|_2|^p\right)^{\frac{1}{p}}\right| \\
&\overset{(3)}{\leq} \sup_{w\in W}\left(\sum_{l\in\sigma}\|l-\varphi(l)\|_2^p\right)^{\frac{1}{p}} = \left(\sum_{l\in\sigma}\|l-\varphi(l)\|_2^p\right)^{\frac{1}{p}} \leq \mathbf{H}_p(L,L')
\end{aligned}
$$

In (1) we used the inequality $|\inf f-\inf g|\leq \sup|f-g|$. In (2) and (3) we used the reverse triangle inequality $|\|x\|-\|y\||\leq\|x-y\|$ for the L^p norm and the L^2 norm, respectively. Applying Theorem 3.1 we then get

$$
\begin{aligned}
\mathbf{W}_p(\mathrm{Dgm}(\Lambda),\mathrm{Dgm}(\Lambda')) \leq \|f-g\|_p &= \left(\sum_{\sigma\in D\Lambda}|f(\sigma)-g(\sigma)|^p\right)^{\frac{1}{p}} \\
&\leq \left(\sum_{\sigma\in D\Lambda}\mathbf{H}_p(L,L')^p\right)^{\frac{1}{p}} \\
&= \left(\sum_{k=1}^{n}\binom{n}{k}\right)^{\frac{1}{p}}\mathbf{H}_p(L,L') \\
&= (2^n-1)^{\frac{1}{p}}\mathbf{H}_p(L,L')
\end{aligned}
$$

Similarly, if we fix the homological dimension $k\geq 0$, we get that

$$
\mathbf{W}_p(\mathrm{Dgm}_k(\Lambda),\mathrm{Dgm}_k(\Lambda')) \leq \left(\binom{n}{k}+\binom{n}{k+1}\right)^{\frac{1}{p}}\mathbf{H}_p(L,L') = \binom{n+1}{k+1}^{\frac{1}{p}}\mathbf{H}_p(L,L').
$$

□

4 Comments

1. We use $q = p$ in the Wasserstein distance. Can we generalize for other choices of q? (See definition in [3] for example.)
2. In the definition of the p-Hausdorff distance and the dissimilarities we use $q = 2$. Other choices of q are probably interleaved with ours.
3. What about the case when $|L| \neq |L'|$?
4. Any natural alternatives for the p-Hausdorff distance as it is defined in this note?
5. Could be interesting to look at the part about stability in [1][p. 14-16].

References

[1] Morten Brun and Nello Blaser. "Sparse Dowker nerves". In: *Journal of Applied and Computational Topology* 3.1-2 (2019), pp. 1–28. DOI: 10.1007/s41468-019-00028-9. URL: https://doi.org/10.1007%2Fs41468-019-00028-9.

[2] Yueqi Cao and Anthea Monod. *Approximating Persistent Homology for Large Datasets*. 2022. arXiv: 2204.09155 [stat.ML].

[3] Primoz Skraba and Katharine Turner. *Wasserstein Stability for Persistence Diagrams*. 2022. arXiv: 2006.16824 [math.AT].