Topological Data Analysis

17 December 2019

12 Further about stability

12.1 Stability for functions

Let M be any topological space, and let $f\colon M \to \mathbb{R}$ and $g\colon M \to \mathbb{R}$ be two functions. For each $t \in \mathbb{R}$, consider the sublevel sets

$$L_t(f) = \{x \in M \mid f(x) \le t\}, \qquad L_t(g) = \{x \in M \mid g(x) \le t\}.$$

Define

$$V_t(f) = H_*(L_t(f); \mathbb{R}), \qquad V_t(g) = H_*(L_t(g); \mathbb{R}),$$

where singular homology is meant. Let $\pi_{s,t}\colon V_s(f) \to V_t(f)$ and $\pi'_{s,t}\colon V_s(g) \to V_t(g)$ be induced by the inclusions $L_s(f) \subseteq L_t(f)$ and $L_s(g) \subseteq L_t(g)$ if $s \le t$.

In what follows, we suppose that $V(f)$ and $V(g)$ are persistence modules. This happens in the following cases, among other situations:

- If M is a closed interval $[a,b] \subset \mathbb{R}$, and f and g are differentiable functions with finitely many critical points.
- If M is a closed smooth manifold, and f and g are smooth functions with finitely many critical points.
- If $M = \mathbb{R}^N$, and f and g are defined as

  $$f(x) = d(x, X), \qquad g(x) = d(x, Y),$$

  where X and Y are point clouds in $\mathbb{R}^N$.

Our aim is to prove the following inequality in all such cases:

$$\boxed{d_{\mathrm{int}}(V(f), V(g)) \le \|f - g\|_\infty}$$

where $\|f - g\|_\infty = \sup\{|f(x) - g(x)| : x \in M\}$.

For this, we pick $\delta = \|f - g\|_\infty$ and prove that $V(f)$ and $V(g)$ are δ-interleaved. Note that $V(f)[\delta] = V(f - \delta)$ and $V(g)[\delta] = V(g - \delta)$. By our choice of δ, we have $|f(x) - g(x)| \le \delta$ for all $x \in M$. This implies that

$$g(x) - \delta \le f(x) \le g(x) + \delta \quad \text{and} \quad f(x) - \delta \le g(x) \le f(x) + \delta$$

for all $x \in M$. Therefore we also have

$$f(x) - 2\delta \le g(x) - \delta \le f(x) \quad \text{and} \quad g(x) - 2\delta \le f(x) - \delta \le g(x)$$

for all $x \in M$. Now the inclusions $L_t(f) \subseteq L_{t+\delta}(g)$ and $L_t(g) \subseteq L_{t+\delta}(f)$ for all t yield morphisms of persistence modules

$$F\colon V(f) \longrightarrow V(g)[\delta] \quad \text{and} \quad G\colon V(g) \longrightarrow V(f)[\delta],$$

and $F[\delta] \circ G$ is equal to the morphism induced by the inclusion $L_t(f) \subseteq L_{t+2\delta}(f)$, which is precisely the shift morphism $\sigma_{2\delta}$ for $V(f)$. The argument for $G[\delta] \circ F$ is the same, by symmetry. Hence we conclude that $V(f)$ and $V(g)$ are δ-interleaved by means of F and G, as needed.

12.2 Stability for Čech complexes

We apply the previous result to the case when f and g are the distance functions on $\mathbb{R}^N$ to point clouds X and Y. In this section we consider the *Čech persistence modules* $\tilde{V}_t(X) = H_*(C_t(X); \mathbb{R})$ and $\tilde{V}_t(Y) = H_*(C_t(Y); \mathbb{R})$, where $C_t(X)$ and $C_t(Y)$ are the Čech complexes of X and Y (assumed empty if $t \leq 0$).

Note that, if $X = \{x_i\}_{i \in I}$, then

$$L_t(f) = \{x \in \mathbb{R}^n \mid d(x, X) \leq t\} = \bigcup_{i \in I} \bar{B}_t(x_i).$$

Hence $L_t(f)$ is homotopy equivalent to the geometric realization of the Čech complex $C_{2t}(X)$. Consequently,

$$V_t(f) = H_*(L_t(f); \mathbb{R}) = H_*(C_{2t}(X); \mathbb{R}) = \tilde{V}_{2t}(X),$$

or, equivalently, $\tilde{V}_t(X) = V_{t/2}(f) = V_t(2f)$. Similarly, $\tilde{V}_t(Y) = V_t(2g)$.

On the other hand,

$$\|f - g\|_\infty = \sup\{|d(x, X) - d(x, Y)| : x \in \mathbb{R}^N\} = d_H(X, Y),$$

where d_H denotes the Hausdorff distance. Thus we obtain that

$$d_{\mathrm{int}}(\tilde{V}(X), \tilde{V}(Y)) = d_{\mathrm{int}}(V(2f), V(2g)) \leq \|2f - 2g\|_\infty = 2\, d_H(X, Y),$$

as with the Vietoris–Rips persistence modules. However, there does not seem to be a similar inequality relating the Gromov–Hausdorff distance between X and Y (which is intrinsic) with the Čech persistence modules $\tilde{V}(X)$ and $\tilde{V}(Y)$. The difficulty with Čech complexes is that they are not solely determined by the table of distances between the points in the given point cloud, but they depend on the topology of the ambient space $\mathbb{R}^N$. For this reason, the interleaving distance between the persistence modules $\tilde{V}(X)$ and $\tilde{V}(Y)$ also depends on the ambient space.

Reference: Stability was first proved in the article [D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, *Disc. Comput. Geom.* 37 (2007), 103–120].