Topological Data Analysis

2022–2023

Lecture 10

Proof of the Stability Theorem

12 December 2022

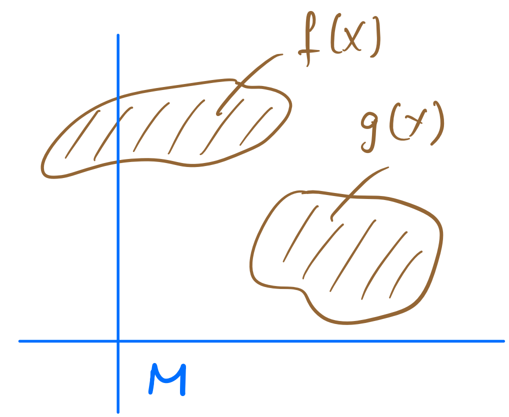
Proof of the Kalton-Ostrovskii Theorem

Let X and Y be compact metric spaces with respective distances d^X and d^Y. For simplicity we will assume that X and Y are <u>finite</u>.

<u>Step 1</u>: Proving that $d_{GH}(X,Y)$ is attained with a metric on the disjoint union $X \amalg Y$ extending d^X and d^Y.

Pick any $\varepsilon > 0$ and choose isometric embeddings $f: X \hookrightarrow M$ and $g: Y \hookrightarrow M$ for some metric space M with distance d^M such that

$$d_H^M(f(X), g(Y)) \leq d_{GH}(X,Y) + \varepsilon.$$



Then d^M restricts to a metric on $f(X) \cup g(Y)$.

- If $f(X) \cap g(Y) = \emptyset$, then d^M yields a metric $d^{\amalg}$ on $X \amalg Y$ pulling back along f and g.

  This metric satisfies

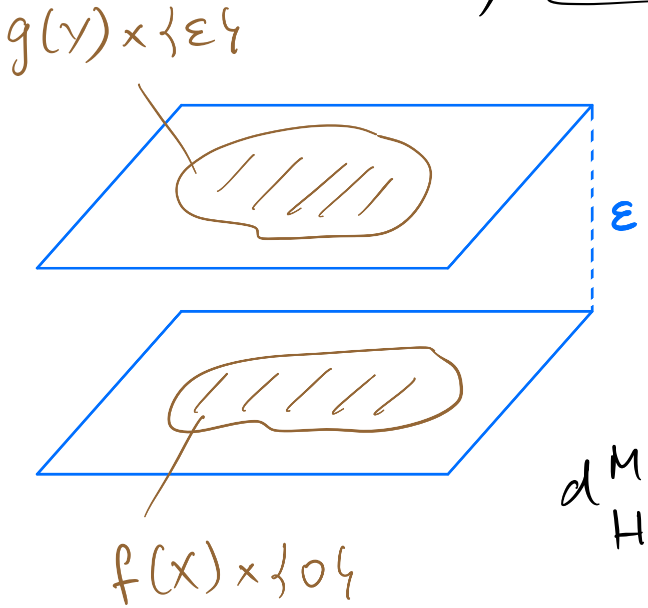
$$d_H^{\amalg}(X,Y) = d_H^M(f(X), g(Y)) \leq d_{GH}(X,Y) + \varepsilon.$$

• If $f(X) \cap g(Y) \neq \emptyset$, consider $M \times \mathbb{R}$ with the distance

$$d^{M\times\mathbb{R}}((p,s),(q,t)) = d^M(p,q) + |s-t|,$$

and embed $X \hookrightarrow M \times \mathbb{R}$ as $x \mapsto (f(x), 0)$

$Y \hookrightarrow M \times \mathbb{R}$ as $y \mapsto (g(y), \varepsilon)$.



Then we have, for all $x \in X$ and $y \in Y$,

$$d^{M\times\mathbb{R}}((f(x),0),(g(y),\varepsilon)) = d^M(f(x), g(y)) + \varepsilon.$$

Therefore

$$d_H^{M\times\mathbb{R}}(f(X)\times\{0\}, g(Y)\times\{\varepsilon\}) = d_H^M(f(X), g(Y)) + \varepsilon$$

$$\leq d_{GH}(X,Y) + 2\varepsilon.$$

Hence $d^{M\times\mathbb{R}}$ induces a metric $d^{\amalg}$ on $X \amalg Y$ such that

$$d_H^{\amalg}(X,Y) = d_H^{M\times\mathbb{R}}(f(X)\times\{0\}, g(Y)\times\{\varepsilon\}) \leq d_{GH}(X,Y) + 2\varepsilon.$$

Since ε is arbitrary, we conclude that the infimum of $d_H(X,Y)$ among metrics on $X \amalg Y$ extending d^X and d^Y is $\leq d_{GH}(X,Y)$.

Conversely, the infimum of $d_H(X,Y)$ among metrics on $X \amalg Y$ extending d^X and d^Y is $\geq d_{GH}(X,Y)$, since the latter is the infimum of $d_H^M(X,Y)$ among all isometric embeddings of X and Y into a common metric space.

In conclusion,

$$d_{GH}(X,Y) = \inf\{ d_H^{\amalg}(X,Y) \text{ for metrics on } X \amalg Y \text{ extending } d^X \text{ and } d^Y \}.$$

Next we show that this infimum is <u>attained</u> by a <u>unique</u> metric.

Let $\mathcal{F} = \{ d : X \times Y \longrightarrow \mathbb{R} \mid d \text{ yields a distance } d^{\amalg} \text{ on } X \amalg Y \text{ extending } d^X \text{ and } d^Y \}$.

Then $\mathcal{F}$ is a <u>convex</u> subset of the Euclidean space $\mathbb{R}^{X \times Y}$, since

$$\left.\begin{array}{l} d_1 \in \mathcal{F} \\ d_2 \in \mathcal{F} \end{array}\right\} \Rightarrow \quad t d_1 + (1-t) d_2 \in \mathcal{F} \quad \text{for all } 0 \leq t \leq 1.$$

Define $f : \mathcal{F} \longrightarrow \mathbb{R}$ by $f(d) = d_H^{\amalg}(X,Y)$.

If X and Y are isometric, pick an isometry $\varphi: Y \to X$ and consider $d(x,y) = d^X(x, \varphi(y))$. Then $d \in \mathcal{F}$ and $d_H^{\amalg}(X,Y) = 0$.

Otherwise, $d_{GH}(X,Y) \neq 0$ and hence $f(d) \neq 0$ for all $d \in \mathcal{F}$.

Since $\mathcal{F}$ is closed in $\mathbb{R}^{X \times Y}$, there exists $d_0 \in \mathcal{F}$ attaining a minimum value of $f(d)$ for $d \in \mathcal{F}$. Moreover, d_0 is unique since $\mathcal{F}$ is convex.

This completes step 1.

Step 2: Considering distortion.

Let $d_{GH}(X,Y) = r$. Pick a metric $d^{\amalg}$ on $X \amalg Y$ extending d^X and d^Y such that $d_H^{\amalg}(X,Y) = r$. Then the following is a correspondence from X to Y:

$$C_0 = \{(x,y) \mid d^{\amalg}(x,y) \leq r\}.$$

Indeed, for every $x \in X$ there is some $y \in Y$ with $d^{\amalg}(x,y) \leq r$, since otherwise $d^{\amalg}(x,Y) > r$ since Y is compact, and this contradicts the assumption that $d_H^{\amalg}(X,Y) = r$. The same is true for $y \in Y$.

If (x,y) and (x',y') are in C_0, we have

$$|d^X(x,x') - d^Y(y,y')| \leq d^{\amalg}(x,y) + d^{\amalg}(x',y') \leq 2r.$$

<u>Proof</u>: Suppose that $d^X(x,x') \geq d^Y(y,y')$ without loss of generality. Then

$$d^X(x,x') - d^Y(y,y') \leq d^{\amalg}(x,y) + \cancel{d^{\amalg}(y,y')} + d^{\amalg}(y',x') - \cancel{d^{\amalg}(y,y')}.$$

Therefore $\mathrm{dis}(C_0) \leq 2r$, and this implies that

$$\inf\{\mathrm{dis}(C)\} \leq \mathrm{dis}(C_0) \leq 2r = 2\, d_{GH}(X,Y).$$

Next we aim to obtain the converse inequality.

Let C_1 be a correspondence from X to Y attaining the minimum distortion. Existence follows from the assumption that X and Y are finite. Define a metric $d^{\amalg}$ on $X \amalg Y$ as follows:

$$d^{\amalg}(x,y) = \inf\{d^X(x,x') + d^Y(y,y')\} + \frac{1}{2}\mathrm{dis}(C_1)$$

for $x \in X$ and $y \in Y$, and extending d^X and d^Y, where the infimum is taken over all $(x',y') \in C_1$. We postpone checking the triangle inequality.

For every $x \in X$ there is some $y \in Y$ such that $(x,y) \in C_1$.

Consequently, $$d^{\amalg}(x,y) = \tfrac{1}{2}\operatorname{dis}(C_1).$$

Hence $d^{\amalg}(x,Y) = \frac{1}{2}\operatorname{dis}(C_1)$. By symmetry, $d^{\amalg}_H(X,Y) = \frac{1}{2}\operatorname{dis}(C_1)$.

Therefore $$d_{GH}(X,Y) \le d^{\amalg}_H(X,Y) = \tfrac{1}{2}\operatorname{dis}(C_1).$$

Since C_1 attained the minimum distortion, the proof is complete.

Now we check the triangle inequality for $d^{\amalg}$.

If (x',y') and (x'',y'') are in C_1, then

$$\operatorname{dis}(C_1) \ge |d^X(x',x'') - d^Y(y',y'')| \ge d^X(x',x'') - d^Y(y',y'')$$

$$\Rightarrow \quad d^Y(y',y'') \ge d^X(x',x'') - \operatorname{dis}(C_1). \qquad (*)$$

Hence $d^Y(y,y') + d^Y(y,y'') \ge d^Y(y',y'') \ge d^X(x',x'') - \operatorname{dis}(C_1)$ for all y.

For $x_1, x_2 \in X$ and $y \in Y$, if $(x',y') \in C_1$ and $(x'',y'') \in C_1$ attain the respective infima for (x_1,y) and (x_2,y), then

$$
\begin{cases}
d^{\amalg}(x_1, y) = d^X(x_1, x') + d^Y(y, y') + \frac{1}{2}\operatorname{dis}(C_1) \\
d^{\amalg}(x_2, y) = d^X(x_2, x'') + d^Y(y, y'') + \frac{1}{2}\operatorname{dis}(C_1).
\end{cases}
$$

Consequently,

$$
\begin{aligned}
d^{\amalg}(x_1, y) + d^{\amalg}(y, x_2) &= \\
&= d^X(x_1, x') + d^X(x_2, x'') + d^Y(y, y') + d^Y(y, y'') + \operatorname{dis}(C_1) \overset{\text{by } (*)}{\geq} \\
&\geq d^X(x_1, x') + d^X(x', x'') + d^X(x'', x_2) \geq d^X(x_1, x_2).
\end{aligned}
$$

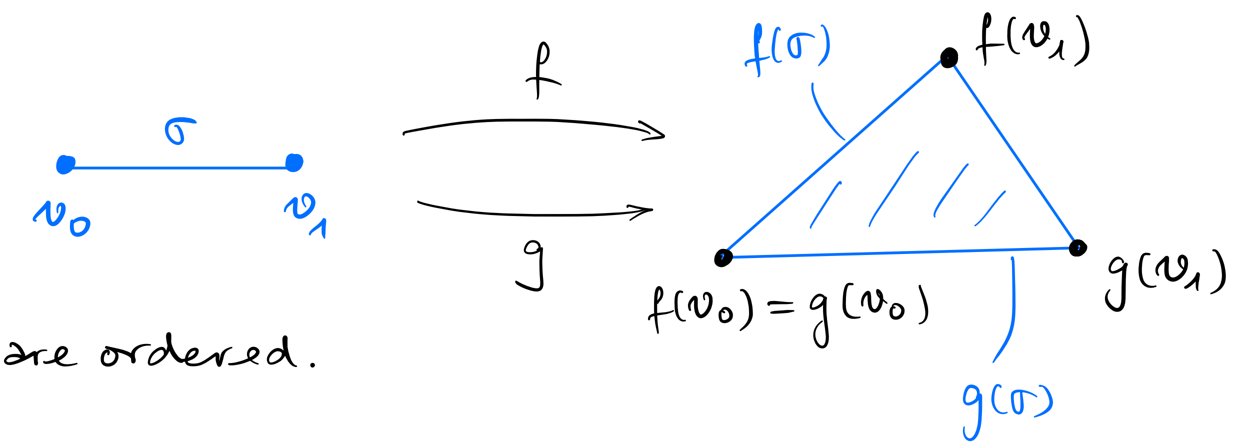
The rest follows by symmetry. ✓

Proof of the Stability Theorem

Let $f, g : K \to L$ be simplicial maps between abstract simplicial complexes. We say that f and g are <u>contiguous</u> if, for each face $\{v_{i_0}, \dots, v_{i_n}\}$ of K, the points (not necessarily distinct!)

$$f(v_{i_0}), \dots, f(v_{i_n}), g(v_{i_0}), \dots, g(v_{i_n})$$

form a face of L.



Now suppose K and L are ordered.

Lemma: If f and g are contiguous maps, then the homomorphisms $f_*, g_* : H_n(K) \to H_n(L)$ are equal for all n.

The proof of this fact is postponed.

Let X and Y be point clouds in $\mathbb{R}^N$ for some N.

Let $R_t(X)$ and $R_t(Y)$ be their Vietoris-Rips complexes, and let

$$V_t(X) = H_*(R_t(X)), \quad V_t(Y) = H_*(R_t(Y))$$

for $t \in \mathbb{R}$, with coefficients in a field $\mathbb{F}$.

Our goal is to prove the inequality

$$\boxed{d_{int}(V(X), V(Y)) \leq 2\, d_{GH}(X, Y).}$$

For this, we need to prove that $V(X)$ and $V(Y)$ are δ-interleaved if $\delta = 2\, d_{GH}(X, Y)$.

Since X and Y are finite, there is a correspondence $C \subseteq X \times Y$ such that $\mathrm{dis}(C) = \delta = 2\, d_{GH}(X, Y)$.

A function $f: X \to Y$ is called <u>subordinate</u> to C if

$$\{(x, f(x)) \mid x \in X\} \subseteq C.$$

A subordinate f <u>exists</u>, since for every x there is some y with $(x, y) \in C$.

If $f : X \to Y$ is subordinate to C, then for each face $\sigma = \{v_{i_0}, \dots, v_{i_n}\}$ of $R_t(X)$, since $d(v_{i_k}, v_{i_l}) \leq t$ and $(v_{i_k}, f(v_{i_k})) \in C$ for all k,

$$\mathrm{dis}(C) \geq |d(v_{i_k}, v_{i_l}) - d(f(v_{i_k}), f(v_{i_l}))| \quad \text{for all } k, l.$$

Hence $$d(f(v_{i_k}), f(v_{i_l})) \leq d(v_{i_k}, v_{i_l}) + \mathrm{dis}(C) \leq t + \delta$$

for all k, l. This implies that $f(\sigma)$ is a <u>face</u> of $R_{t+\delta}(Y)$.

Hence f induces a simplicial map

$$f_t : R_t(X) \longrightarrow R_{t+\delta}(Y) \quad \text{for each } t.$$

Let $F_t = (f_t)_*$ be the induced homomorphism in homology:

$$F_t : V_t(X) \longrightarrow V_{t+\delta}(Y) = V(Y)[\delta]_t .$$

Similarly, we can choose a function $g : Y \to X$ subordinate to C^{-1} and obtain

$$G_t : V_t(Y) \longrightarrow V_{t+\delta}(X) = V(X)[\delta]_t .$$

The collections $\{F_t\}$ and $\{G_t\}$ define morphisms of persistence modules

$$V(X) \xrightarrow{F} V(Y)[\delta], \qquad V(Y) \xrightarrow{G} V(X)[\delta].$$

The theorem will be proved if we show that $V(X)$ and $V(Y)$ are δ-interleaved by means of F and G.

The shift morphism $\sigma_{2\delta} : V(X) \to V(X)[2\delta]$ is induced by the inclusions $R_t(X) \hookrightarrow R_{t+2\delta}(X)$ for all t.

Proving that $G[\delta] \circ F = \sigma_{2\delta}$ amounts to proving that $(g \circ f)_t$ is <u>contiguous</u> to the inclusion $R_t(X) \hookrightarrow R_{t+2\delta}(X)$ for every t.

For this, let $\{v_{i_0}, \dots, v_{i_n}\}$ be any face of $R_t(X)$. Then, for all k, ℓ,

$$\underbrace{(v_{i_\ell}, f(v_{i_\ell})) \in C}_{\text{since } f \text{ is subordinate to } C} \quad \text{and} \quad \underbrace{(g(f(v_{i_k})), f(v_{i_k})) \in C}_{\text{since } g \text{ is subordinate to } C^{-1}}$$

Hence,

$$d(g(f(v_{i_k})), v_{i_\ell}) \underbrace{\leq}_{\text{since } \operatorname{dis}(C) = \delta} d(f(v_{i_k}), f(v_{i_\ell})) + \delta.$$

Similarly,

$$\left.\begin{array}{l}(v_{i_k}, f(v_{i_k})) \in C \\ (v_{i_l}, f(v_{i_l})) \in C\end{array}\right\} \Rightarrow d(f(v_{i_k}), f(v_{i_l})) + \delta \leq$$

$$\leq d(v_{i_k}, v_{i_l}) + 2\delta \leq t + 2\delta.$$

$$\left.\begin{array}{l}(g(f(v_{i_k})), f(v_{i_k})) \in C \\ (g(f(v_{i_l})), f(v_{i_l})) \in C\end{array}\right\} \Rightarrow d(g(f(v_{i_k})), g(f(v_{i_l}))) \leq$$

$$\leq d(f(v_{i_k}), f(v_{i_l})) + \delta \leq$$

$$\leq d(v_{i_k}, v_{i_l}) + 2\delta \leq t + 2\delta.$$

This proves that the points

$$g(f(v_{i_0})), \dots, g(f(v_{i_n})), v_{i_0}, \dots, v_{i_n}$$

form a face of $R_{t+2\delta}(X)$, as needed.

The argument with $f \circ g$ is analogous. ✓

Proof of the Lemma

Let $f, g : K \to L$ be simplicial maps between ordered complexes.

A homotopy from f to g is a collection of homomorphisms

$$h_n : C_n(K) \longrightarrow C_{n+1}(L)$$

such that $\boxed{\partial^L_{n+1} \circ h_n + h_{n-1} \circ \partial^K_n = g_n - f_n}$ for all n,

where $f_n, g_n : C_n(K) \to C_n(L)$ are induced by f and g.

If there is a homotopy from f to g, then $f_* = g_*$ in homology, since, if z is an n-cycle of K, then

$$g_*([z]) = [g_n(z)] = [f_n(z) + \partial^L_{n+1} h_n(z) + \cancel{h_{n-1}(\partial^K_n z)}] =$$

$$= [f_n(z)] = f_*([z]).$$

Now assume that f and g are contiguous. We need to prove that f and g are homotopic. For this, construct $h_0 : C_0(K) \to C_1(L)$ such that $\partial^L_1 \circ h_0 = g_0 - f_0$ as follows.

By assumption, for each vertex $v \in K$, $\{f(v), g(v)\}$ is a face σ of L. Define $h_0(v) = \sigma$ with the usual sign conventions.

$h_0(v) = 0$ if $f(v) = g(v)$; $h_0(v) = \sigma$ if $f(v) < g(v)$; $h_0(v) = -\sigma$ otherwise.

Then $\partial_1^L h_0(v) = g_0(v) - f_0(v)$ for all v, as needed.

Next we define $h_1 : C_1(K) \longrightarrow C_2(L)$ as follows.

Let $\{v_0, v_1\}$ be a 1-face of K with $v_0 < v_1$. Since f and g are contiguous, $\{f(v_0), f(v_1), g(v_0), g(v_1)\}$ is a face of L.

Let $\sigma_0 = \{f(v_0), g(v_0), g(v_1)\}$

$\sigma_1 = \{f(v_0), f(v_1), g(v_1)\}$ and $h_1(\{v_0, v_1\}) = \sigma_0 - \sigma_1$.

Then

$$\partial_2^L h_1(\{v_0, v_1\}) = \{g(v_0), g(v_1)\} - \{f(v_0), g(v_1)\} + \{f(v_0), g(v_0)\} - \{f(v_1), g(v_1)\} + \{f(v_0), g(v_1)\} - \{f(v_0), f(v_1)\}$$

$$= g_1(\{v_0, v_1\}) - f_1(\{v_0, v_1\}) + h_0(v_0) - h_0(v_1) =$$

$$= g_1(\{v_0, v_1\}) - f_1(\{v_0, v_1\}) - h_0 \partial_1^K(\{v_0, v_1\}),$$ as needed.

The general case is given by

$$h_n(\{v_0,\dots,v_n\}) = \sum_{i=0}^{n} (-1)^i \{f(v_0),\dots,f(v_i),g(v_i),\dots,g(v_n)\},$$

which is the prism operator formula from algebraic topology. ✓

References

- N. J. Kalton, M. I. Ostrovskii, Distances between Banach spaces, Forum Math. 11 (1999), 17–48.
- F. Chazal, V. de Silva, S. Oudot, Persistence stability for geometric complexes, Geom. Dedicata 173 (2014), 193–214.
- L. Polterovich, D. Rosen, K. Samvelyan, J. Zhang, Topological Persistence in Geometry and Analysis, University Lecture Series 74, Amer. Math. Soc., 2020.
- D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, Disc. Comput. Geom. 37 (2007), 103–120.