Topological Data Analysis

2022–2023

Lecture 9

Stability Theorem

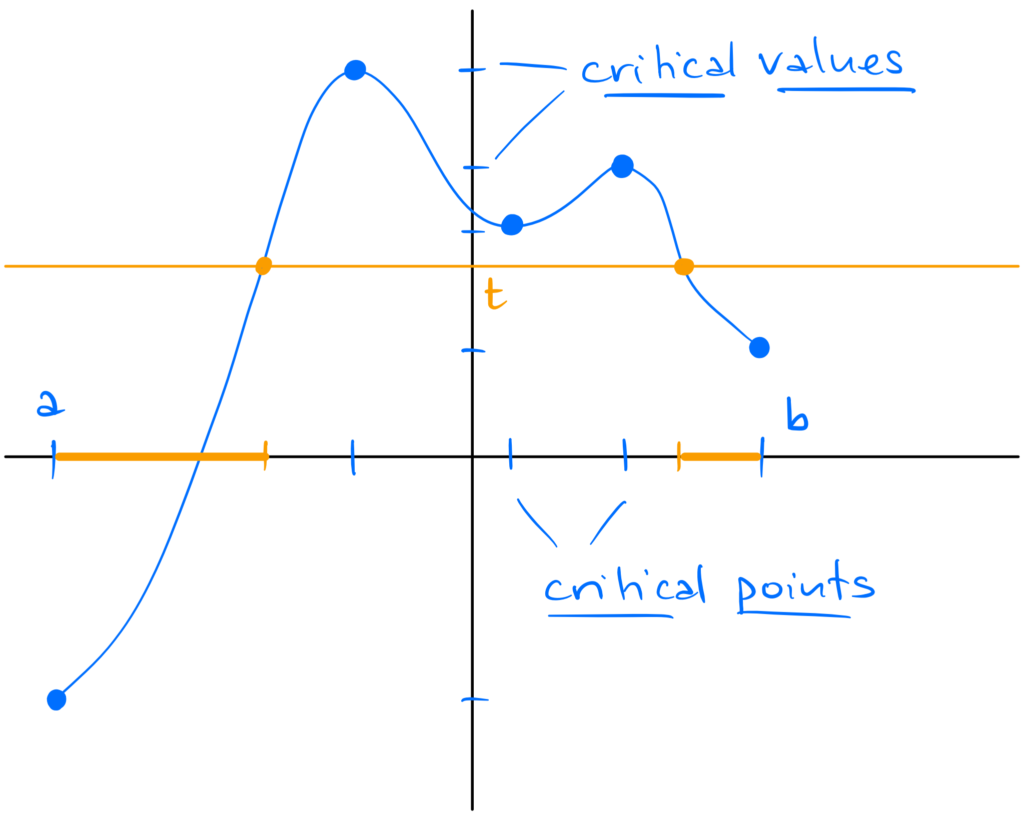
1 December 2022

Sublevel sets

For a continuous function $f : [a,b] \to \mathbb{R}$ denote, for each $t \in \mathbb{R}$,

$$L_t(f) = \{ x \in [a,b] \mid f(x) \le t \},$$ called a <u>sublevel set</u>.

Note that if $s \le t$ then $L_s(f) \subseteq L_t(f)$.



$L_t(f) = \emptyset$ if $t < \inf(f)$

$L_t(f) = [a,b]$ if $t \ge \sup(f)$

We call $x_0 \in [a,b]$ a <u>critical point</u> if it is a local maximum or a local minimum, including $x_0 = a$ and $x_0 = b$.

If x_0 is a critical point, then $f(x_0)$ is a <u>critical value</u>.

If f is differentiable and $x_0 \in (a,b)$ is a critical point, then $f'(x_0) = 0$.

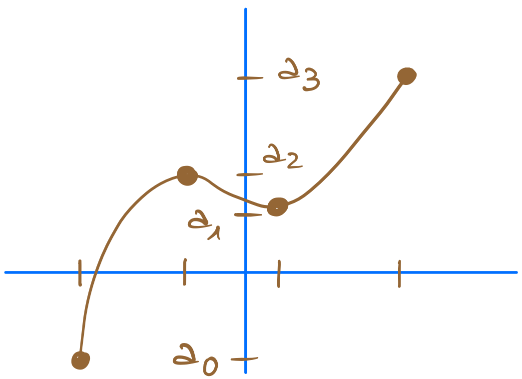
From now on we assume that f has finitely many critical points (hence each critical point is isolated).

Under this assumption, we associate to f a persistence module:

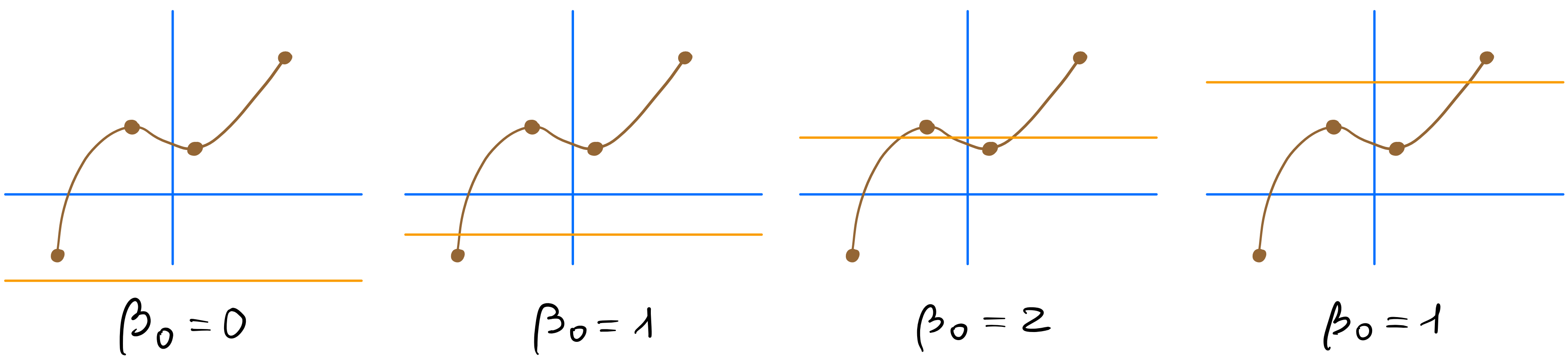
$$V_t(f) = H_0(L_t(f)),$$ where H_0 denotes zero-homology,

and we let $\pi_{s,t} : V_s(f) \to V_t(f)$ be induced by the inclusion $V_s(f) \hookrightarrow V_t(f)$ if $s \leq t$.

The spectrum of (V, π) is contained in the set of critical values of f.



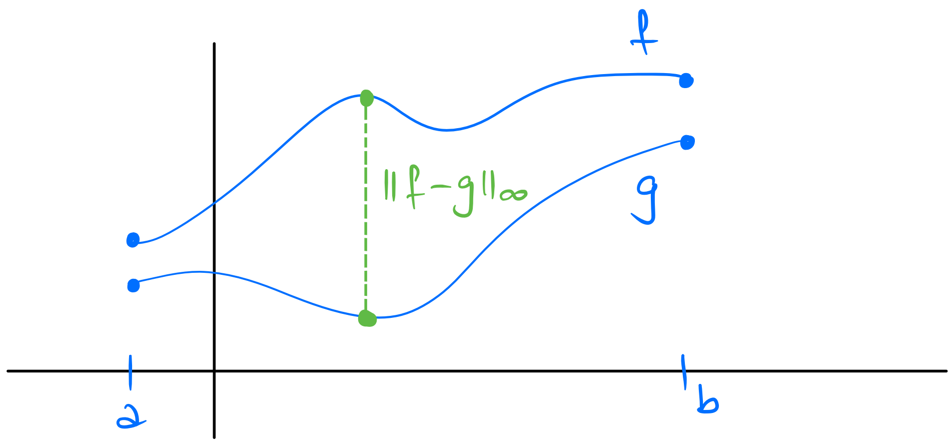
Here a_3 is a critical value of f but the spectrum of $V(f)$ is $\{a_0, a_1, a_2\}$



Stability Theorem:

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

Here $\|f-g\|_\infty = \sup\{ |f(x)-g(x)| : a \leq x \leq b \}$.



Hausdorff distance

Let M be a metric space. The <u>diameter</u> of a subset $X \subseteq M$ is defined as

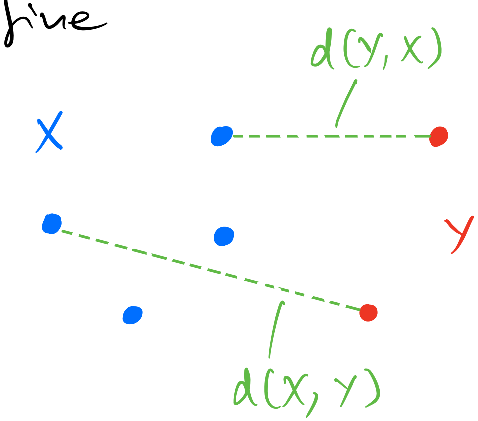
$$\operatorname{diam}(X) = \sup\{d(p,q) \mid p,q \in X\}.$$

We say that X is <u>bounded</u> if $\operatorname{diam}(X)$ is finite.

Suppose given two subsets X, Y of M. Define

$$d(p,Y) = \inf\{d(p,y) \mid y \in Y\},$$

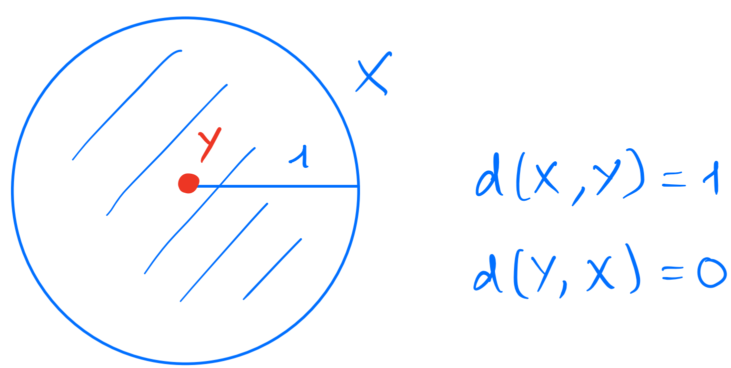
$$d(X,Y) = \sup\{d(p,Y) \mid p \in X\}.$$



Note that $d(X,Y) \neq d(Y,X)$ in general.

Note also that

$$Y \subseteq X \Rightarrow d(Y,X) = 0.$$



If X is compact, then $d(Y,X)=0 \iff Y \subseteq X$.

Proof: Since every metric space is Hausdorff, X is a closed subset of M. If $d(Y,X)=0$ then $d(p,X)=0$ for all $p \in Y$. Since X is closed, this implies that $p \in X$. Hence $Y \subseteq X$. ✓

From now on we assume that X and Y are compact.

The Hausdorff distance between X and Y is defined as

$$d_H(X,Y) = \max\{d(X,Y), d(Y,X)\}.$$

It follows that $d_H(X,Y)=0$ if and only if $X=Y$.

This is false if X, Y are not compact: if $M=\mathbb{R}$, $X=\mathbb{Q}$, $Y=\mathbb{R}\setminus\mathbb{Q}$ then $d_H(X,Y)=0$.

Moreover, d_H satisfies the triangle inequality

$$d_H(X,Z) \leq d_H(X,Y) + d_H(Y,Z).$$

Hence it is indeed a distance on the set of compact subsets of M.

Gromov-Hausdorff distance

Let X and Y be compact metric spaces. The <u>Gromov-Hausdorff distance</u> between X and Y is defined as

$$d_{GH}(X,Y) = \inf\{ d_H^M(f(X), g(Y)) \mid f: X \hookrightarrow M,\ g: Y \hookrightarrow M \text{ isometrically} \}$$

where the infimum is taken over all isometric embeddings of X and Y into metric spaces, and d_H^M denotes Hausdorff distance in M.

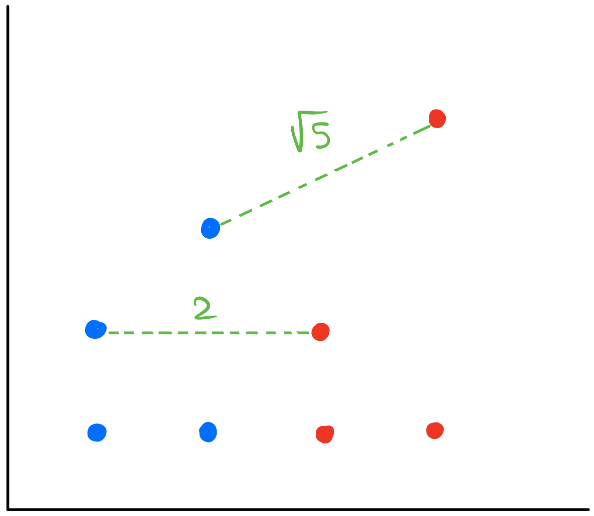
Hence $d_{GH}(X,Y)=0$ if and only if X and Y are isometric.

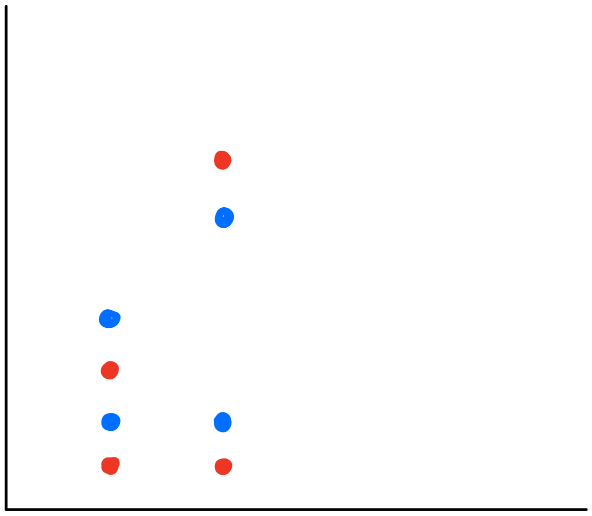
Proof: $d_{GH}(X,Y)=0$ if and only if there are isometric embeddings $f: X \hookrightarrow M$, $g: Y \hookrightarrow M$ such that $f(X)=g(Y)$, since X and Y are compact and hence $d_H^M(f(X), g(Y))=0$ implies that $f(X)=g(Y)$.

Then

$$X \underbrace{\cong}_{\text{isometry}} f(X) = g(Y) \underbrace{\cong}_{\text{isometry}} Y. \quad \checkmark$$

Example:





$$X = \{(1,1),(1,2),(2,1),(2,3)\}$$

$$Y = \{(3,1),(3,2),(4,1),(4,4)\}$$

$$\begin{cases} d(X,Y) = 2 \\ d(Y,X) = \sqrt{5} \end{cases}$$

$$d_H(X,Y) = \boxed{\sqrt{5}}$$

$$f : X \longrightarrow \mathbb{R}^2, \quad f(x,y) = (x,y)$$

$$g : Y \longrightarrow \mathbb{R}^2, \quad g(x,y) = \left(x-2,\ y-\tfrac{1}{2}\right)$$

$$d_{GH}(X,Y) = d_H(f(X), g(Y)) = \boxed{\tfrac{1}{2}}$$

A <u>correspondence</u> between X and Y is a surjective multivalued function from X to Y. That is, a subset $C \subseteq X \times Y$ such that for all $x_0 \in X$ there is some $(x_0, y) \in C$ and for all $y_0 \in Y$ there is some $(x, y_0) \in C$.

If C is a correspondence, then

$$C^{-1} = \{(y, x) \in Y \times X \mid (x, y) \in C\}$$

is also a correspondence.

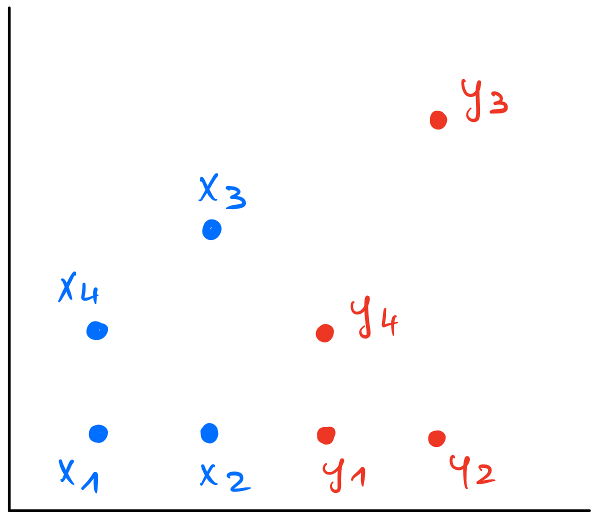
The <u>distortion</u> of a correspondence $C \subseteq X \times Y$ is defined as

$$\operatorname{dis}(C) = \max\{ |d^X(x, x') - d^Y(y, y')| : (x, y), (x', y') \in C\}.$$

<u>Example</u>: Suppose that $C = \{(x, f(x)) \mid x \in X\}$ for some surjective function $f: X \to Y$. Then $\operatorname{dis}(C) = 0$ if and only if f preserves distance.

$$\operatorname{dis}(C) = 0 \iff d^X(x, x') = d^Y(f(x), f(x')) \quad \forall x \, \forall x'$$

Example:



The smallest distortion is achieved with

$$C = \{(x_1,y_1),(x_2,y_2),(x_3,y_3),(x_4,y_4)\}$$

$$|d^X(x_2,x_3) - d^Y(y_2,y_3)| = |2-3| = 1$$

$$|d^X(x_4,x_3) - d^Y(y_4,y_3)| = |\sqrt{2}-\sqrt{5}| = 0.82$$

$$|d^X(x_1,x_3) - d^Y(y_1,y_3)| = |\sqrt{5}-\sqrt{10}| = 0.93$$

and the rest are zero.

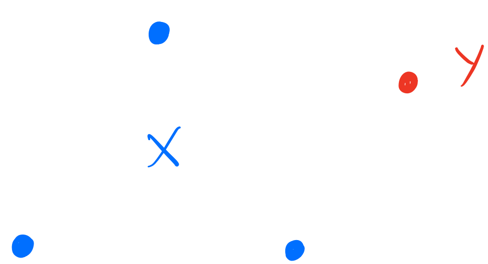
Hence $\mathrm{dis}(C) = 1$.

Theorem (Kalton–Ostrovskii, 1999)

$$\boxed{d_{GH}(X,Y) = \frac{1}{2}\inf\{\mathrm{dis}(C) \mid C \subseteq X \times Y \text{ correspondence}\}.}$$

Example:

X equilateral of side 1

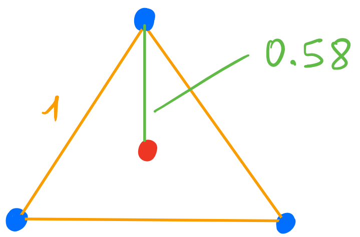


The only correspondence is

$$C = \{(x_1,y),(x_2,y),(x_3,y)\},$$

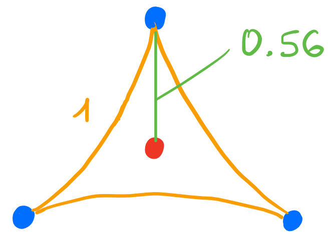
and $\mathrm{dis}(C) = 1$. Hence

$$d_{GH}(X,Y) = \frac{1}{2}.$$



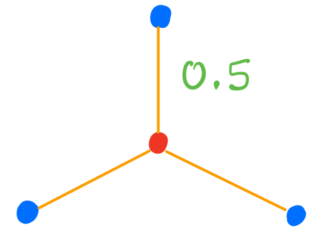
$d_H(X,Y) = \frac{1}{3}\sqrt{3} = 0.58$

in the Euclidean plane



$d_H(X,Y) = 0.56$

in the hyperbolic disk of curvature -1



$d_H(X,Y) = 0.5$

graph distance

The proof of the Kalton–Ostrovskii Theorem is based on the fact that the infimum of the Hausdorff distances $d_H^M(X,Y)$ for all isometric embeddings $X \hookrightarrow M$ and $Y \hookrightarrow M$ is attained with a metric on the disjoint union $X \amalg Y$ extending d^X and d^Y.

<u>Stability Theorem</u>:

Vietoris–Rips complex

Let X and Y be point clouds. If $V_t(X) = H_*(R_t(X))$ then

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

equal to $W_\infty(D(X), D(Y))$ by the Isometry Theorem