Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute of Theoretical Computer Science
Patrick Schnider

Introduction to Topological Data Analysis **Scribe Notes 16** **FS23**

Scribe notes by Simon Weber. Please contact me for corrections.
Lecture date: April 27, 2023
Last update: Thursday 27th April, 2023, 14:04

We wish to slightly generalize the stability result from the last lecture to general topological spaces.

Consider some topological space X and a function $f : X \to \mathbb{R}$, which induces a sublevel set filtration for every $r \in \mathbb{R}$. We only want to consider *tame* functions: A function f is *tame* if all homology groups of sublevel sets have finite rank, and the homology groups only change at finitely many values, called *critical values*.

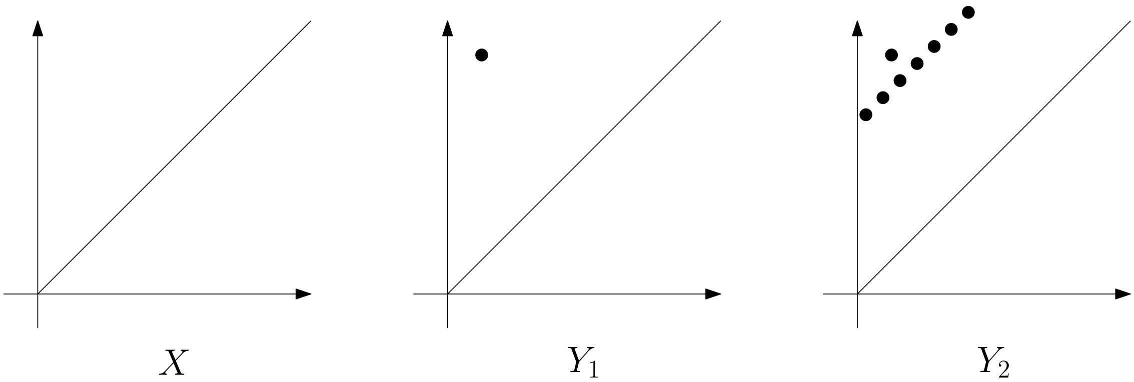
Theorem 1. *Let X be a triangulable topological space, and $f, g : X \to \mathbb{R}$ be two tame functions, then $\forall p \geq 0$, we have*

$$d_b(\mathrm{Dgm}_p(\mathcal{F}_f), \mathrm{Dgm}_p(\mathcal{F}_g)) \leq \|f - g\|_\infty.$$

To prove this theorem, we need some more tools that we will develop in the next few lectures.

Wasserstein Distance

Consider the following three diagrams:



Which of Y_1 and Y_2 is X closer to? Intuitively, one clearly says Y_1: There are simply fewer features in Y_1 that are not present in X. In terms of Bottleneck distance, there is only one reasonable matching between X and Y_1, and also only one between X and Y_2: We simply match each off-diagonal point with its closest point on the diagonal. Since we only look at the longest edge in this matching, the Bottleneck distance $d_b(X, Y_1) = d_b(X, Y_2)$.

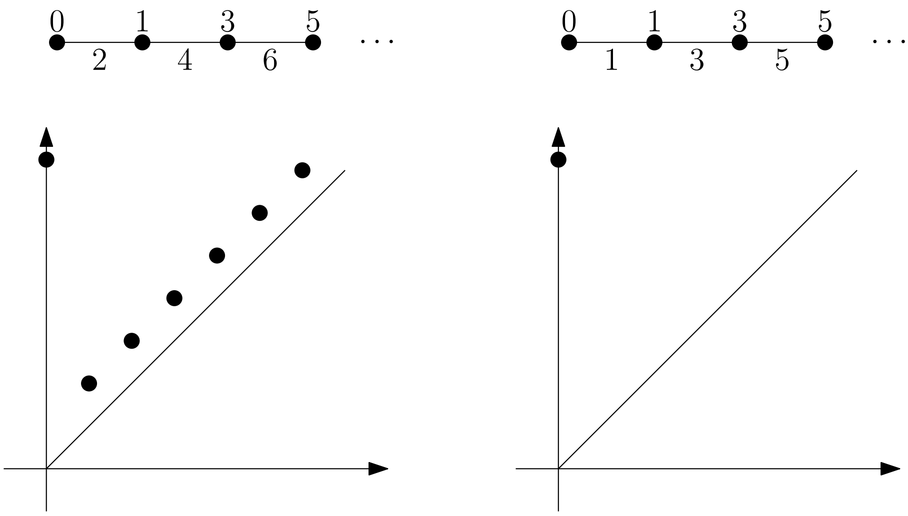
We hope to get rid of this counterintuitive behavior of the Bottleneck distance by using the Wasserstein distance.

Definition 2 (Wasserstein distance). *For* $p \geq 0$*, and* $q \geq 1$*, the* q*-Wasserstein distance is defined as*

$$d_{W,q}(\mathrm{Dgm}_p(\mathcal{F}), \mathrm{Dgm}_p(\mathcal{G})) := \left[\inf_{\pi \in \Pi} \left(\sum_{x \in \mathrm{Dgm}_p(\mathcal{F})} (\|x - \pi(x)\|_\infty)^q \right) \right]^{1/q}$$

Note that for $q = \infty$, $d_{W,q} = d_b$.

We can see that the stability theorem we proved for Bottleneck distance does not hold for Wasserstein distance:



The infinity norm between the two simplex-wise monotone functions is 1, but the Wasserstein distance is unbounded for all $q < \infty$.

A similar counterexample can also be found for topological spaces, see Figure 1.

Note again that $\|f - g\|_\infty \leq \epsilon$, but the Wasserstein distance between the two diagrams can be made arbitrarily big.

To avoid this, we only want to consider even nicer functions:

Definition 3 (Lipschitz). *Let* (X, d) *be a metric space. A function* $f : X \to \mathbb{R}$ *is* Lipschitz *if there exists a constant* C *such that* $|f(x) - f(y)| \leq c \cdot d(x, y)$ *for all* $x, y \in X$.

$X = [0, 1]$

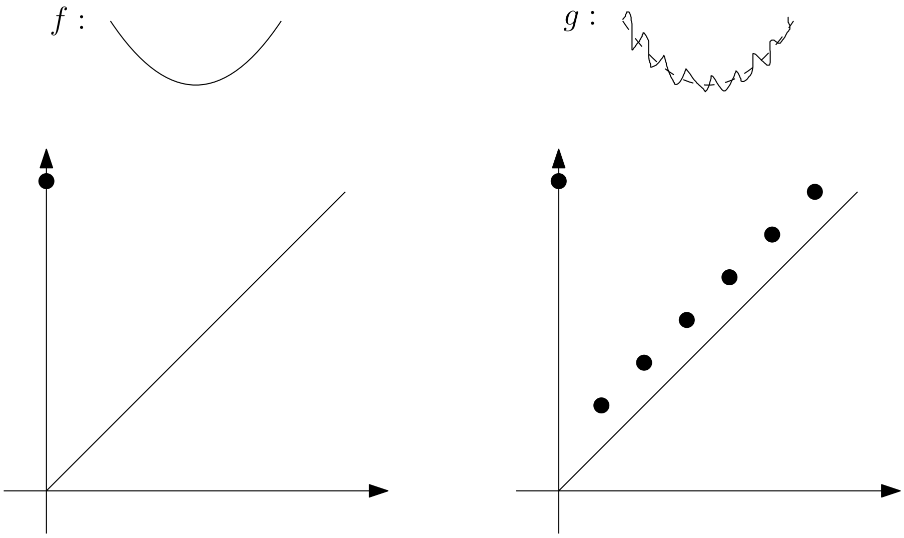


Figure 1:

Theorem 4. *Let* X *be a triangulable, compact metric space. Let* $f, g : X \to \mathbb{R}$ *be Lipschitz functions. Then there exist constants* C *and* k *(that may only depend on* X *and on the Lipschitz constants of* f, g*) such that for every* $p \geq 0$ *and every* $q \geq k$*,*

$$d_{W,q}(\mathrm{Dgm}_p(\mathcal{F}_f), \mathrm{Dgm}_p(\mathcal{F}_g)) \leq C \cdot ||f - g||_{\infty}^{1-k/q}.$$

Theorem 5. *Let* $f, g : K \to \mathbb{R}$ *be simplex-wise monotone functions. Then for all* $p \geq 0$ *and all* $q \geq 1$*,*

$$d_{W,q}(\mathrm{Dgm}_p(\mathcal{F}_f), \mathrm{Dgm}_p(\mathcal{F}_g)) \leq ||f - g||_q = \left(\sum_{\sigma \in K} |f(\sigma) - g(\sigma)|^q \right)^{1/q}.$$