Topological Data Analysis

2022–2023

Lecture 8

Bottleneck Distance

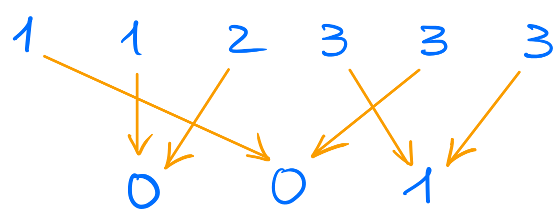
28 November 2022

Suppose given persistence modules V and V' of finite type with respective persistence diagrams D and D'. Suppose, in addition, that

$$\dim V_\infty = \dim V'_\infty .$$

Recall that persistence diagrams are <u>multisets</u>, which means that their points may have a multiplicity. If we choose an <u>order</u> within each multiple point, then we can treat multisets as ordinary sets.

<u>Example</u>: A function $f : \{1,1,2,3,3,3\} \longrightarrow \{0,0,1\}$ is given by

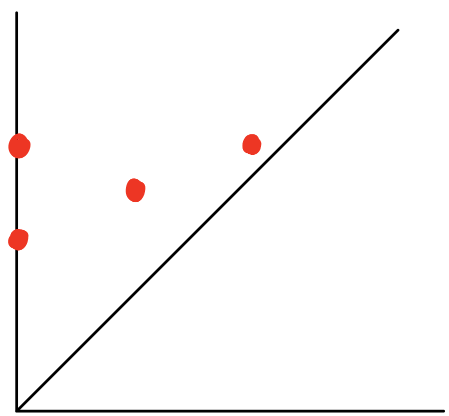


We view it as $\{0_1, 0_2, 1\}$

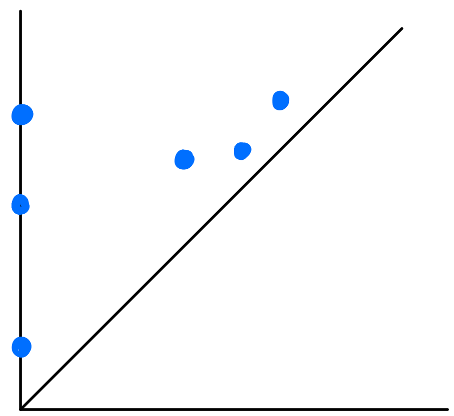
<u>Convention</u>: If Δ denotes the diagonal $b=d$, then points in Δ are members of every persistence diagram with countably infinite multiplicity.

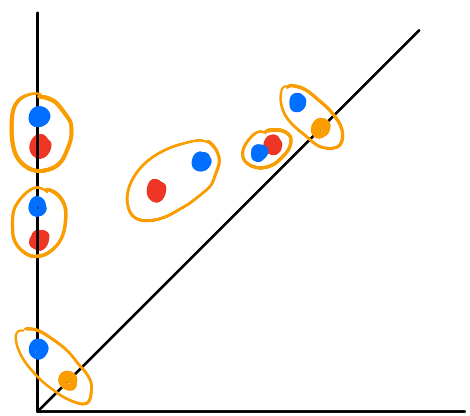
A <u>matching</u> between D and D' is a bijective function $\varphi: D \to D'$ such that, for every $(x,x) \in \Delta$, either $\varphi(x,x) = (x,x)$ or $\varphi(x,x) = (b,d)$ with $b \neq d$.

<u>Example</u>:



$D \xrightarrow{\varphi} D'$

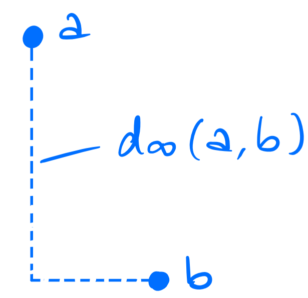




A matching between D and D'

For each matching $\varphi : D \to D'$, define

$$\|\varphi\| = \max\{ d_\infty((x,y), \varphi(x,y)) \mid (x,y) \in D \}$$



where d_∞ is the l_∞-distance on $\mathbb{R}^2$, namely

$$d_\infty((x,y),(x',y')) = \max\{ |x-x'|, |y-y'| \}.$$

The <u>bottleneck distance</u> between two persistence diagrams is defined as

$$W_\infty(D, D') = \min\{ \|\varphi\| \mid \varphi : D \to D' \text{ matching} \}.$$

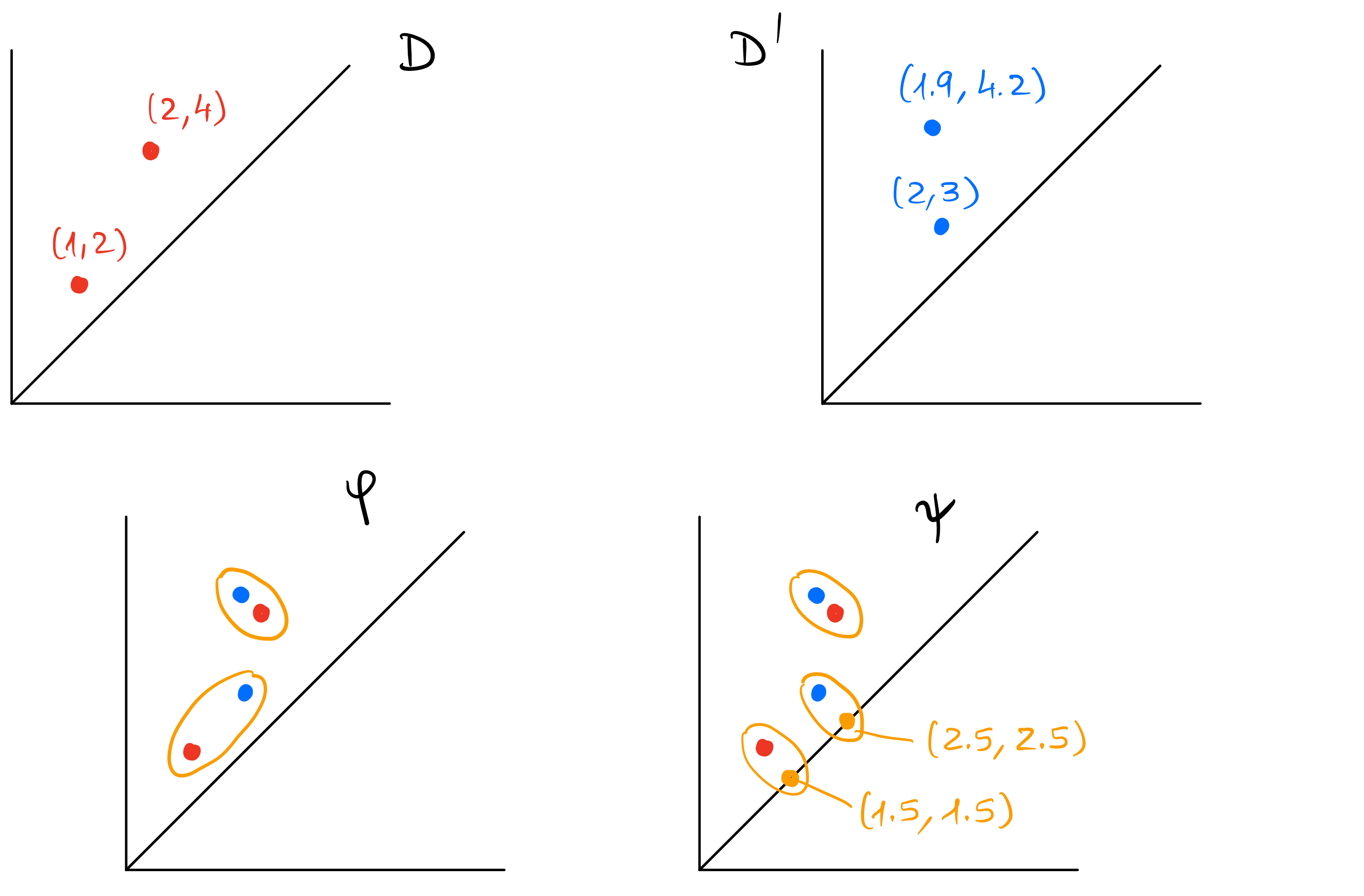
Hence $W_\infty(D,D')$ is the smallest $\varepsilon \geq 0$ for which there exists a matching $\varphi : D \to D'$ for which $d_\infty((x,y), \varphi(x,y)) \leq \varepsilon$ for all $(x,y) \in D$.

More generally, the <u>Wasserstein distances</u> are defined for $p, q \geq 1$ as

$$W_p[q](D, D') = \min_{\varphi : D \to D'} \left(\sum_{(x,y)\in D} d_q((x,y), \varphi(x,y))^p \right)^{1/p}$$

where $d_q((x,y),(x',y')) = (|x-x'|^q + |y-y'|^q)^{1/q}.$

Example:



$\|\varphi\| = \max\{1.0, 0.2\} = 1$ $\qquad$ $\|\psi\| = \max\{0.5, 0.5, 0.2\} = 0.5$

$$W_\infty(D, D') = 0.5$$

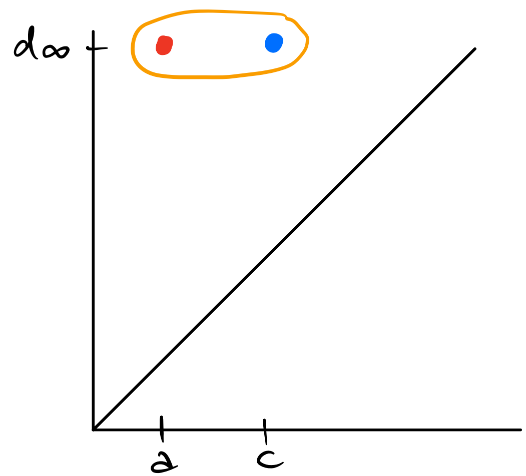
Isometry Theorem:

If V and V' are persistence modules of finite type with $\dim V_\infty = \dim V'_\infty$ and $D(V)$, $D(V')$ denote their respective persistence diagrams, then

$$\boxed{W_\infty(D(V), D(V')) = d_{int}(V, V').}$$

Proof:

① $V = \mathbb{F}[a, \infty)$
$V' = \mathbb{F}[c, \infty)$

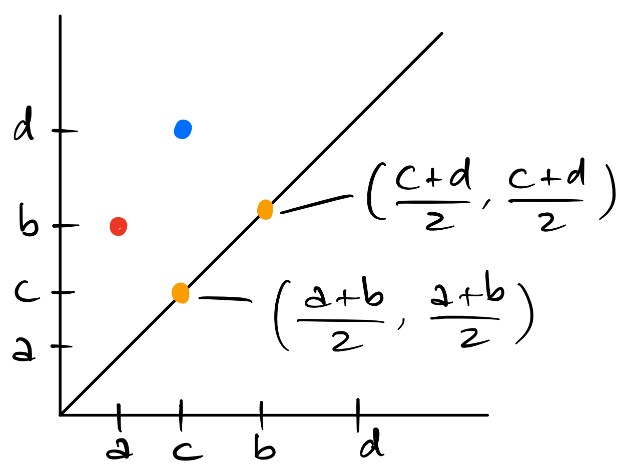


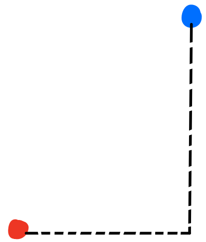
$$W_\infty(D(V), D(V')) = |a - c| =$$
$$= d_{int}(V, V'). \quad \checkmark$$

② $V = \mathbb{F}[a, b)$
$V' = \mathbb{F}[c, d)$

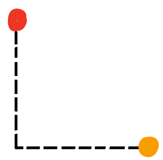
First case:

$a \leq c < b \leq d$





$$d_\infty((a,b),(c,d)) = \max\{c-a,\ d-b\}$$



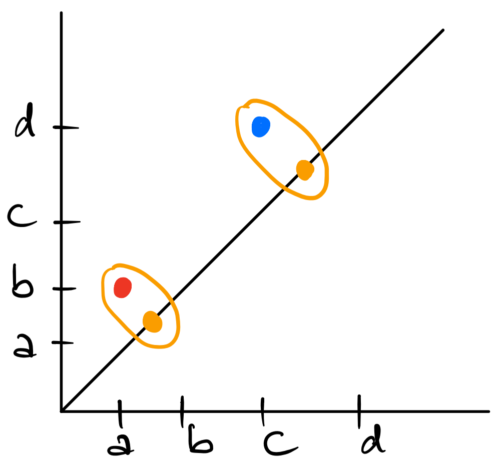
$$d_\infty\left((a,b),\left(\frac{a+b}{2},\frac{a+b}{2}\right)\right) = \frac{a+b}{2} - a = \frac{b-a}{2}$$

$$W_\infty(D(V),D(V')) = \min\left\{\max\{c-a,\ d-b\},\ \max\left\{\frac{b-a}{2},\ \frac{d-c}{2}\right\}\right\} =$$

$$= d_{int}(V,V')$$

Second case:

$a < b \leq c < d$



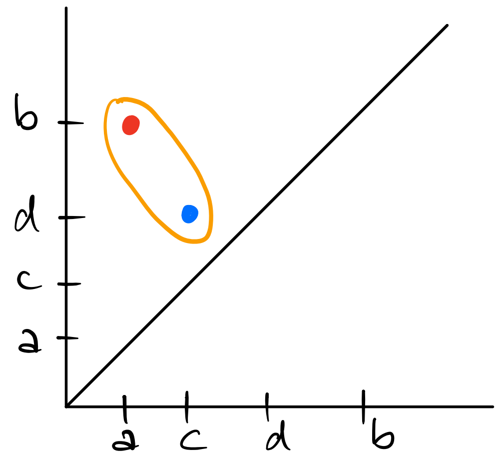
Note that $\begin{cases} d-b \geq d-c > \frac{1}{2}(d-c) \\ c-a \geq b-a > \frac{1}{2}(b-a) \end{cases}$

Hence $d_\infty((a,b),(c,d))$ is larger than $\|\varphi\|$ if φ is the matching with diagonal points.

$$W_\infty(D(V),D(V')) = \max\left\{\frac{b-a}{2},\ \frac{d-c}{2}\right\} = d_{int}(V,V').$$

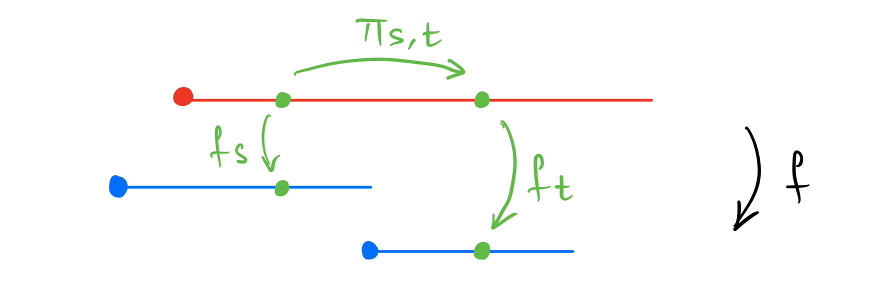
Third case:

$a \leq c < d \leq b$



$W_\infty(D(V), D(V')) = \max\{c-a,\ b-d\} = d_{int}(V, V')$. ✓

For the general case, note that V and V' are δ-interleaved if and only if their intervals $\mathbb{F}[a,b)$ that are not δ-short can be pairwise matched in such a way that each pair of matched intervals are δ-interleaved.

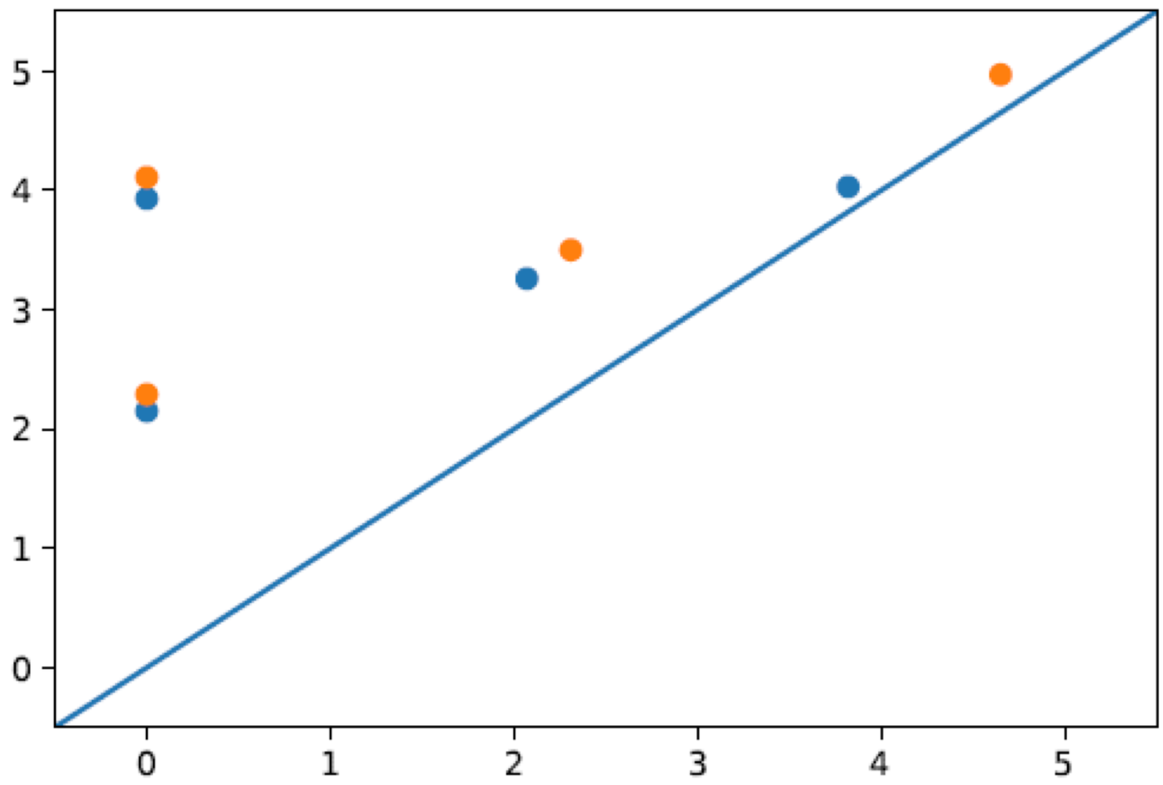


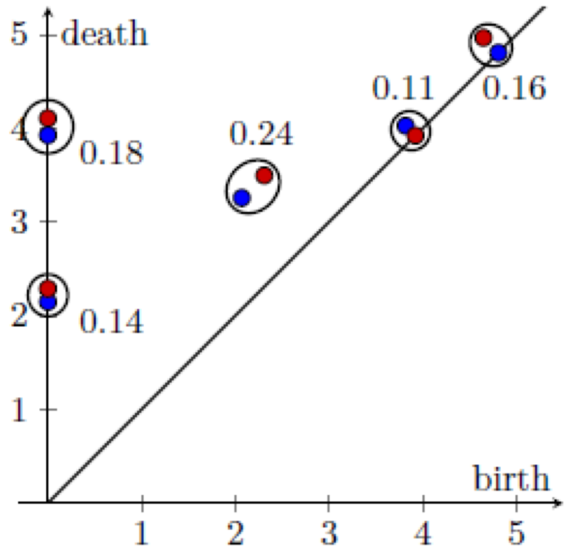
This cannot occur

For this, we use the fact that every morphism between persistence modules must send intervals to intervals.

$$V = \mathbb{F}[0, 2.15) \oplus \mathbb{F}[0, 3.93) \oplus \mathbb{F}[2.07, 3.26) \oplus \mathbb{F}[3.82, 4.03)$$

$$V' = \mathbb{F}[0, 2.29) \oplus \mathbb{F}[0, 4.11) \oplus \mathbb{F}[2.31, 3.50) \oplus \mathbb{F}[4.65, 4.97)$$





$$d_{\text{int}}(V, V') = W_{\infty}(D(V), D(V')) = 0.24$$

R Console

```
> library("TDA")
> Diag1 <- matrix(c(0, 0, 2.15, 0, 0, 3.93, 0, 2.07, 3.26, 0, 3.82, 4.03), ncol = 3, byrow = TRUE)
> Diag2 <- matrix(c(0, 0, 2.29, 0, 0, 4.11, 0, 2.31, 3.50, 0, 4.65, 4.97), ncol = 3, byrow = TRUE)
> par(mfrow = c(1,2))
> plot.diagram(Diag1, diagLim=c(0,6))
> plot.diagram(Diag2, diagLim=c(0,6))
> bottleneckDist <- bottleneck(Diag1, Diag2, dimension = c(0, 1))
> print(bottleneckDist)
[1] 0.24
>
```