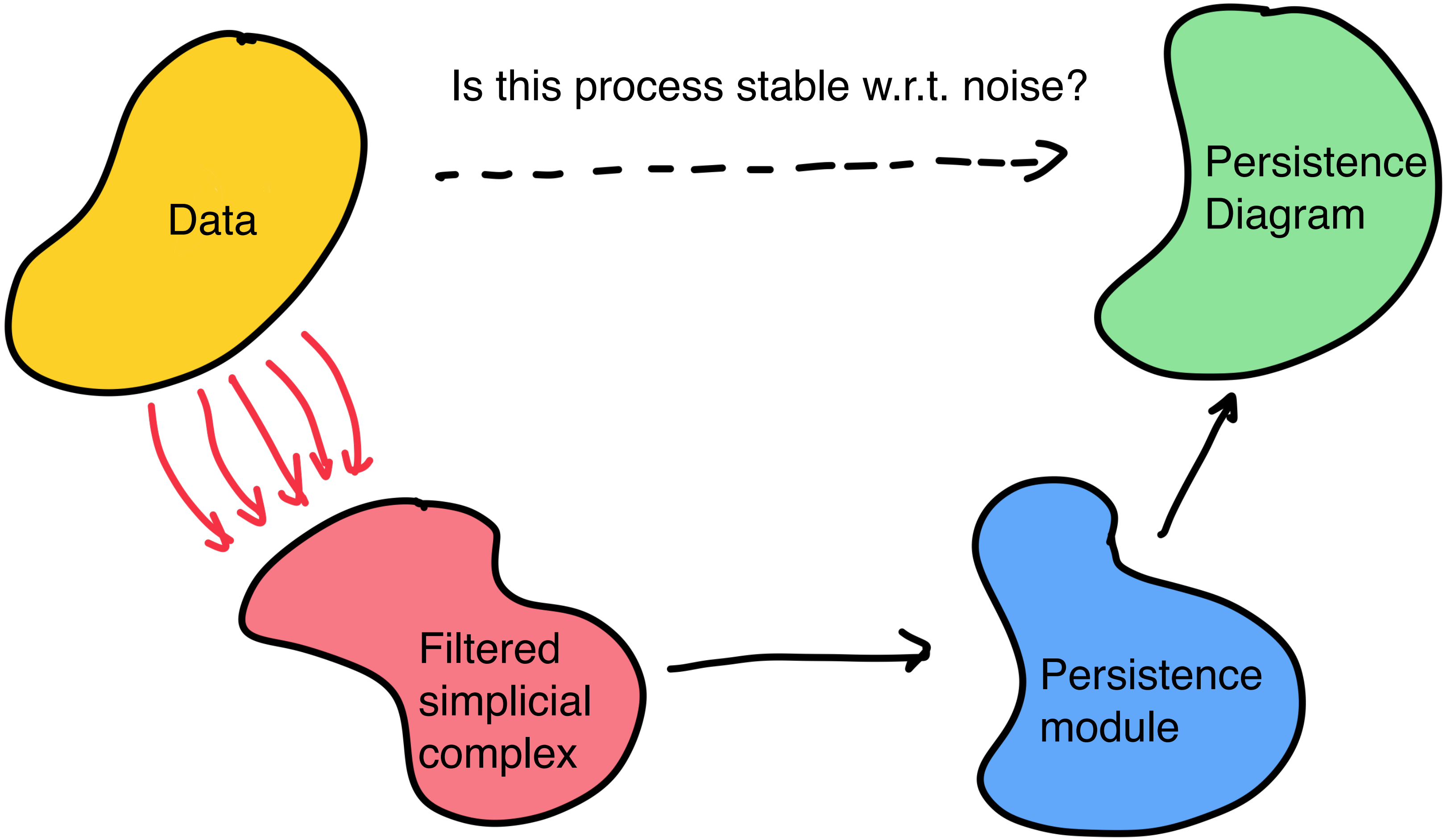
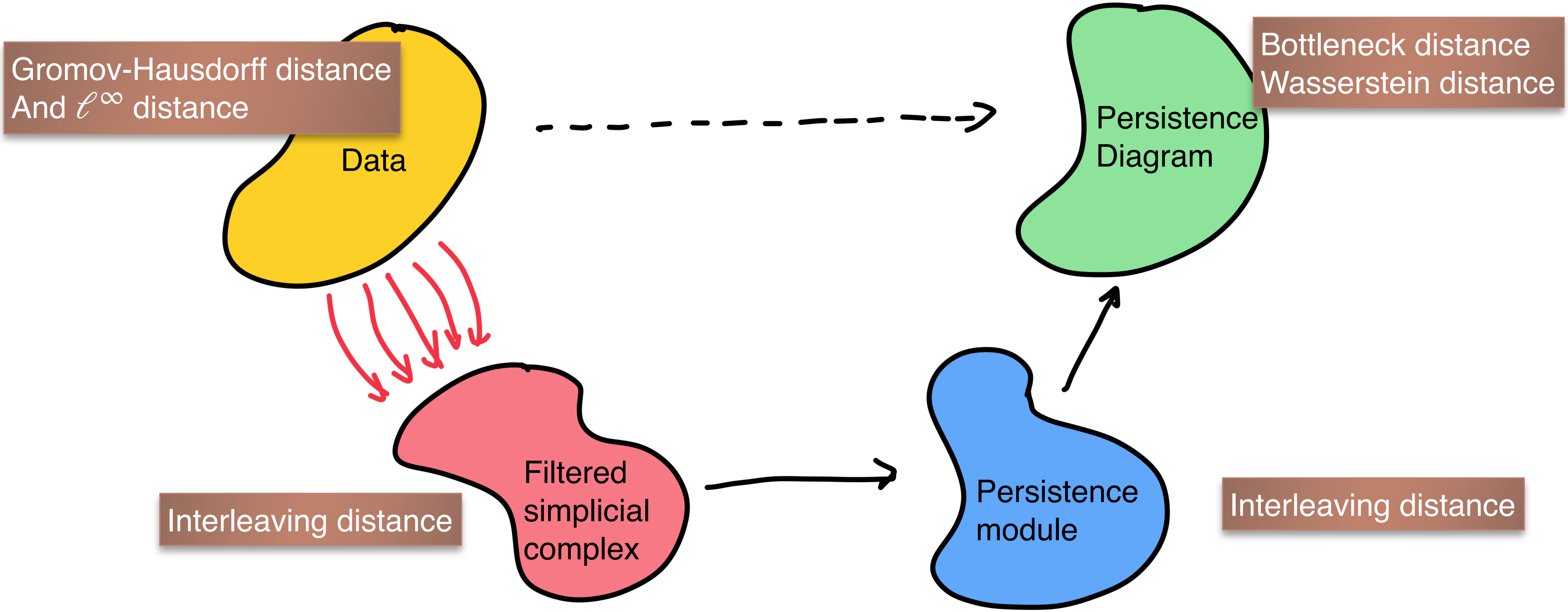
DSC 214
Topological Data Analysis

Topic 5: Stability of PD

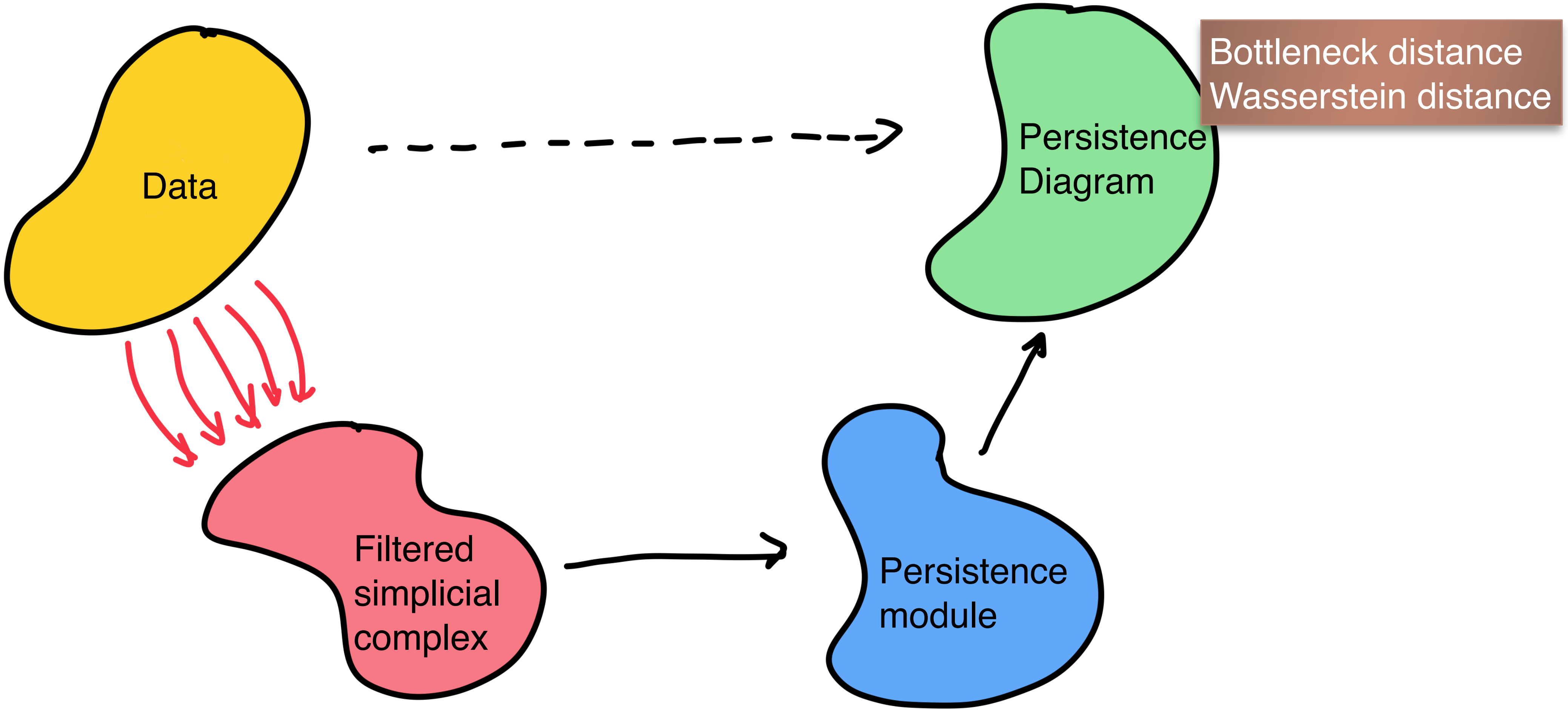
Instructor: Zhengchao Wan

Persistence-based Framework




Gromov-Hausdorff distance
And ℓ^∞ distance
Data
Persistence Diagram
Bottleneck distance
Wasserstein distance
Filtered simplicial complex
Interleaving distance
Persistence module
Interleaving distance

Section 1:
Distances between persistence Diagrams


Data
Filtered simplicial complex
Persistence module
Persistence Diagram
Bottleneck distance
Wasserstein distance

Recall: Persistence Diagram

- $V \cong I[b_1, d_1) \oplus I[b_2, d_2) \oplus \cdots \oplus I[b_M, d_M)$
- Each (b_j, d_j) is called a **persistence pairing**
- The multiset $D = \{(b_j, d_j)\}_{j=1,\ldots,M} \subseteq (\mathbb{R} \cup \infty)^2$ is called the **persistence diagram** of V

Persistence Diagram

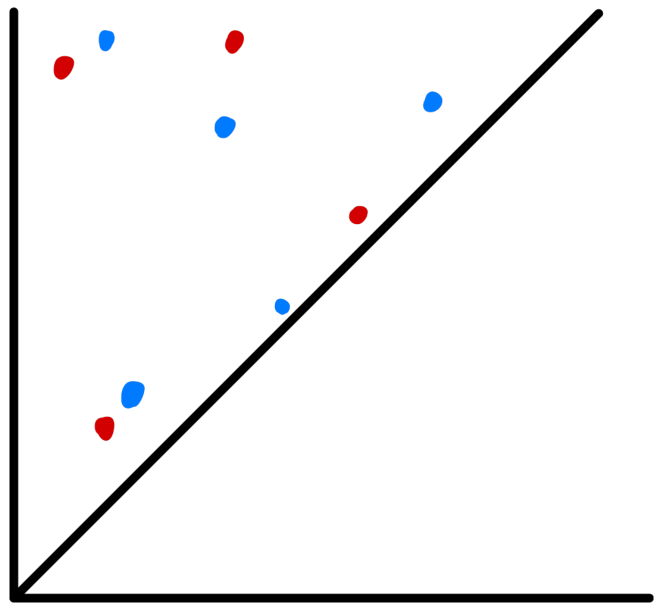
- Any finite multiset $D = \{(b_j, d_j)\}_{j=1,\ldots,M} \subseteq (\mathbb{R} \cup \infty)^2$ is called the **persistence diagram,** where $0 \leq b_i < d_i \leq \infty$ for each $i = 1,\ldots,M$

- How do you compare two different persistence diagrams?

Persistence Diagram

- Any finite multiset $D = \{(b_j, d_j)\}_{j=1,\ldots,M} \subseteq (\mathbb{R} \cup \infty)^2$ is called the **persistence diagram**, where $0 \leq b_i < d_i \leq \infty$ for each $i = 1,\ldots, M$

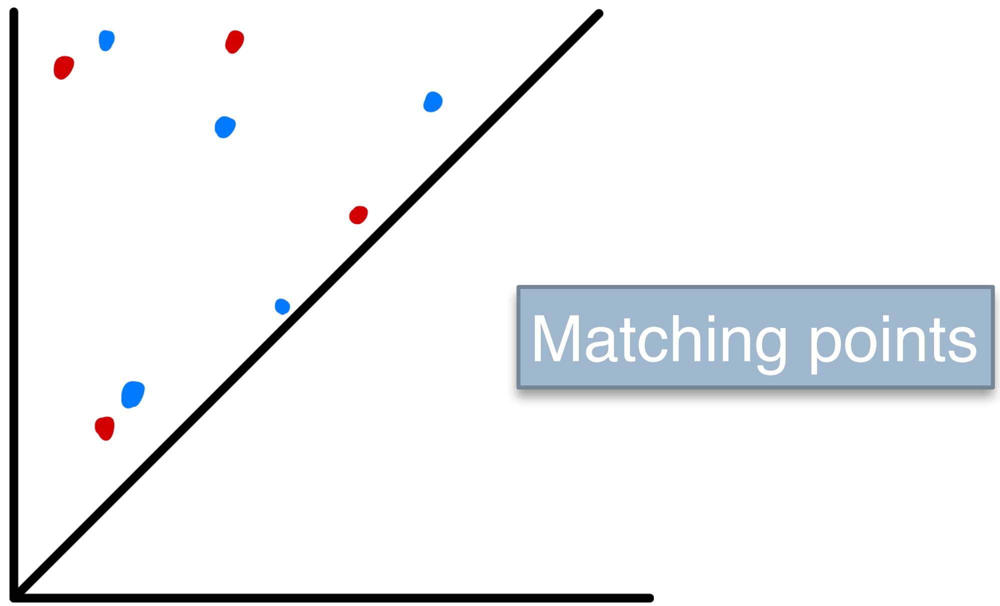
- How do you compare two different persistence diagrams?



Persistence Diagram

- Any finite multiset $D = \{(b_j, d_j)\}_{j=1,\ldots,M} \subseteq (\mathbb{R} \cup \infty)^2$ is called the **persistence diagram**, where $0 \leq b_i < d_i \leq \infty$ for each $i = 1,\ldots,M$

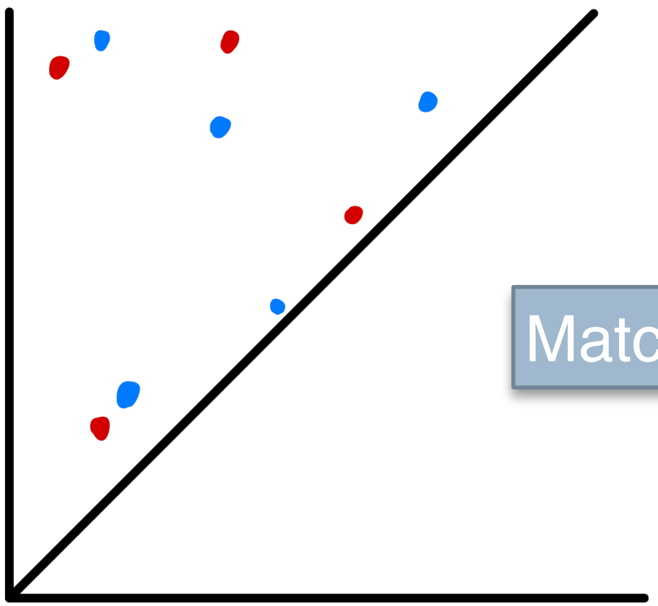
- How do you compare two different persistence diagrams?



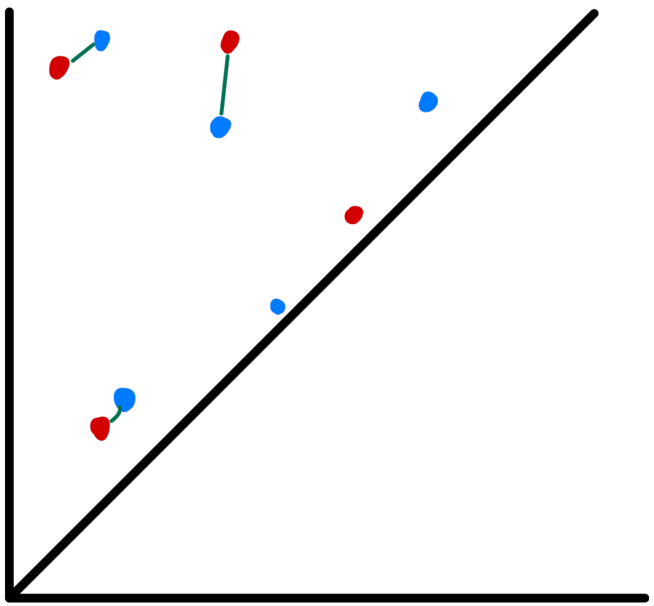
Persistence Diagram

- Any finite multiset $D = \{(b_j, d_j)\}_{j=1,\ldots,M} \subseteq (\mathbb{R} \cup \infty)^2$ is called the **persistence diagram**, where $0 \leq b_i < d_i \leq \infty$ for each $i = 1,\ldots,M$

- How do you compare two different persistence diagrams?

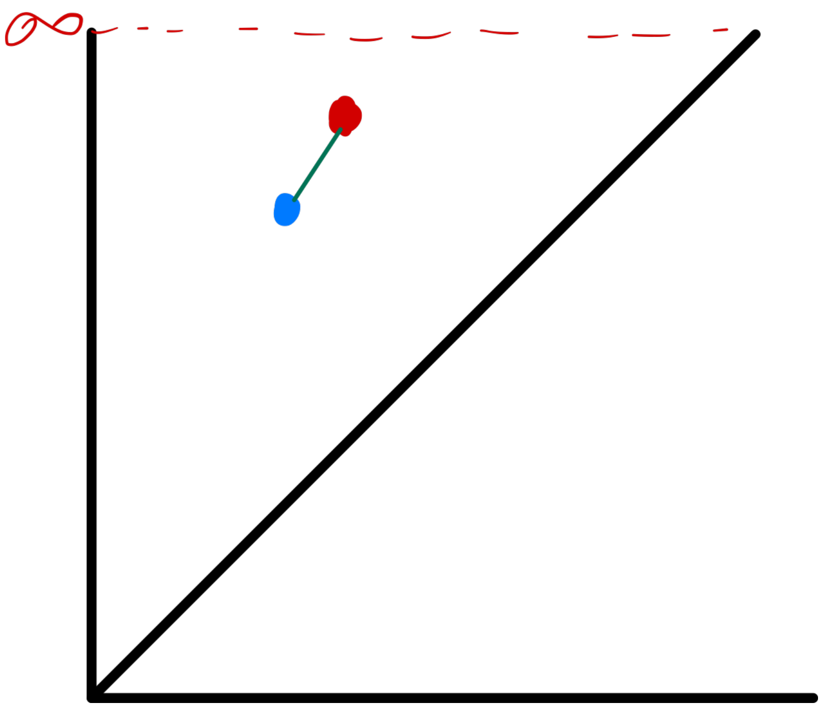


Matching points



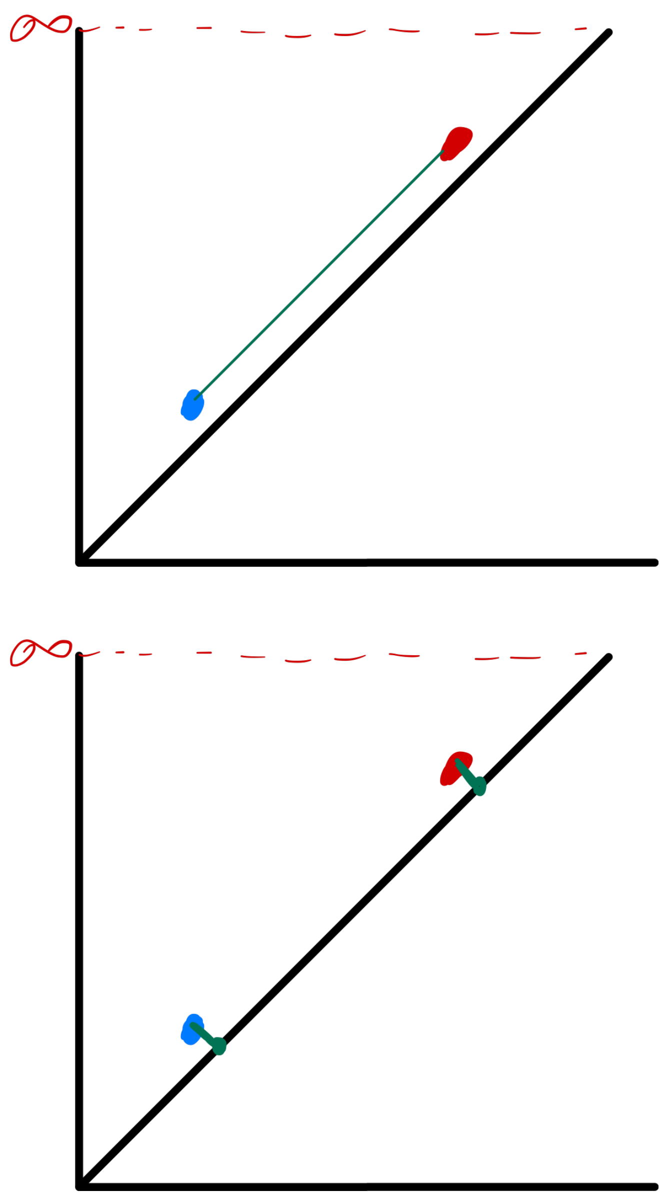
Motivating examples

- Given two points $p = (b, d)$ and $q = (b', d') \in (\mathbb{R} \cup \infty)^2$
- $||p - q||_\infty = \max(|b - b'|, |d - d'|)$
- $\infty - \infty = 0$



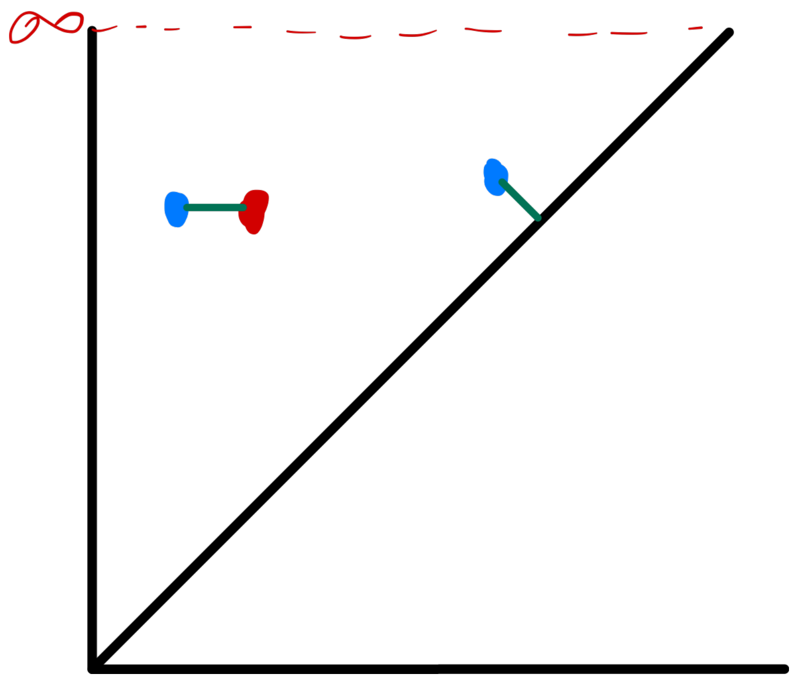
Motivating examples

- Points close to the diagonal $\Delta = \{(x, y) \mid x = y\}$ are not important
- We don't want to match points too far away from each other especially when they are not important
- Note that $\|p - \Delta\|_\infty = \dfrac{|b - d|}{2}$
- We are matching points to the closest points on the diagonal!



Motivating examples

- Two persistence diagrams D and D' may have different number of points
- There is no matching (or bijection) between D and D'
- Match part of D and part of D'
- Compute ℓ^∞ between matched pairs
- Record "importance" of unmatched points; i.e., distances to Δ



Bottleneck distance

- Given two persistence-diagrams (multiset of points in $(\mathbb{R} \cup \{\infty\})^2$)
  - $D_1 = \{p_1, p_2, \ldots, p_s\}$ and $D_2 = \{q_1, q_2, \ldots, q_t\}$
- A partial-matching (partial bijection) between D_1 and D_2 is
  - $M \subseteq D_1 \times D_2 \quad s.t.$
    - $\forall p \in D_1,\ \exists\ at\ most\ one\ (p,\ x) \in M$
    - $\forall q \in D_2,\ \exists\ at\ most\ one\ (x,\ q) \in M$
- The cost of a partial matching $M \subseteq D_1 \times D_2$, denoted by $cost(M)$ is the smallest δ such that
    - $\|p - q\|_\infty \leq \delta$ for $\forall (\mathrm{p}, \mathrm{q}) \in M$ (we assume that $\infty - \infty = 0$)
    - If $p \in D_1 \cup D_2$ is unmatched, then $\|p - \Delta\|_\infty \leq \delta$
      - where Δ is the diagonal

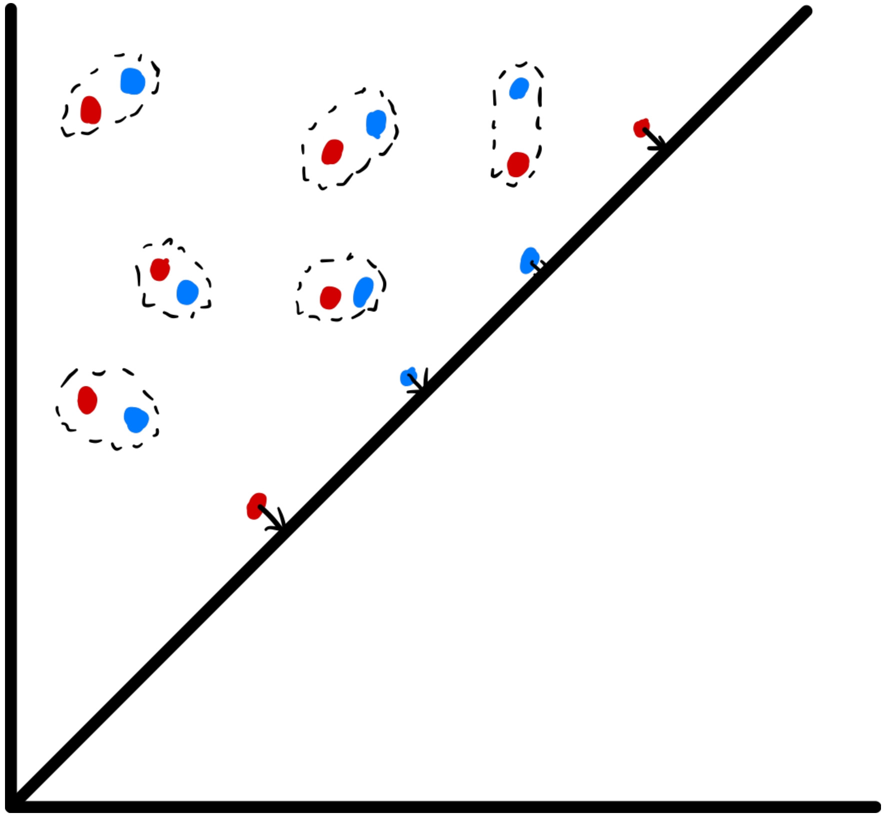
Bottleneck distance

- Given two persistence-diagrams (multiset of points in $(\mathbb{R} \cup \{\infty\})^2$)
  - $D_1 = \{p_1, p_2, \ldots, p_s\}$ and $D_2 = \{q_1, q_2, \ldots, q_t\}$
- A partial-matching (partial bijection) between D_1 and D_2 is
  - $M \subseteq D_1 \times D_2 \quad s.t.$
    - $\forall p \in D_1, \ \exists \ at\ most\ one\ (p, x) \in M$
    - $\forall q \in D_2, \ \exists \ at\ most\ one\ (x, q) \in M$
- The cost of a partial matching $M \subseteq D_1 \times D_2$ can be computed as follows
- $cost(M) = \max\left(\max_{(p,q)\in M} \|p - q\|_\infty, \max_{p \text{ unmatched}} \|p - \Delta\|_\infty \right)$

Bottleneck distance

- *[Cohen-Steiner, Edelsbrunner, Harer, DCG 2007]*

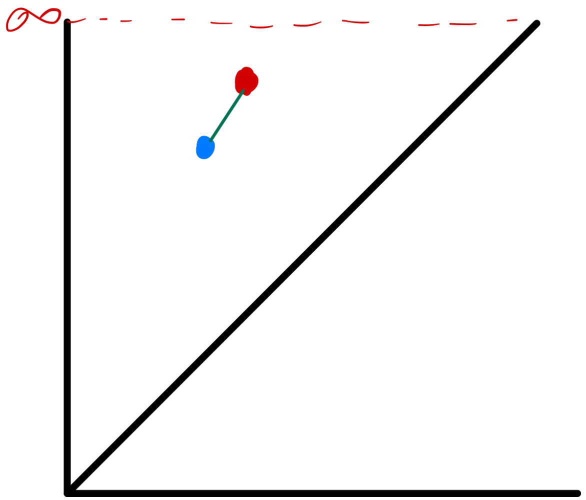
- The bottleneck distance between D_1 and D_2 is
  - $d_B(D_1, D_2) = \min_{M} cost(M)$

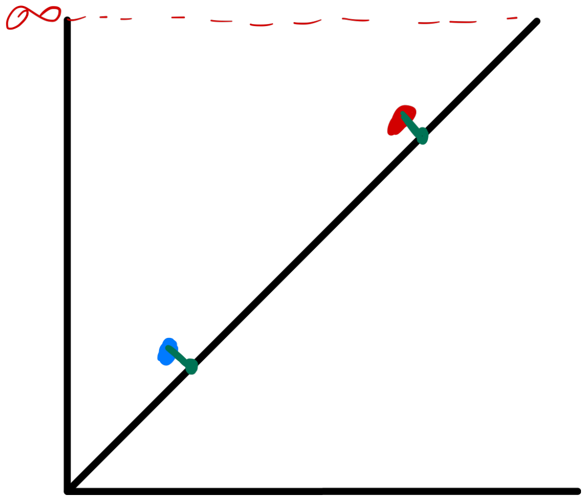


Bottleneck distance between 1-point PDs

- Assume that $D = \{p\}$ and $D' = \{q\}$
  - There are only two possible partial matchings:
    - $M_1 = \{(p,q)\}$ with $cost(M_1) = \|p - q\|_\infty$
    - $M_2 = \emptyset$ with $cost(M) = \max(\|p - \Delta\|_\infty, \|q - \Delta\|_\infty)$
- In conclusion,

$$d_B(D, D') = \min\left(\max(|b - b'|, |d - d'|), \max(\frac{|b-d|}{2}, \frac{|b'-d'|}{2})\right)$$





Alternative formulation

- Given two persistence-diagrams (multiset of points in $(\mathbb{R} \cup \{\infty\})^2$)
  - $D_1 = \{p_1, p_2, \ldots, p_s\}$ and $D_2 = \{q_1, q_2, \ldots, q_t\}$
- Augment $\bar{D}_1 = D_1 \cup \Delta$ and $\bar{D}_2 = D_2 \cup \Delta$
  - where $\Delta = \{(x, x) \in R^2\}$ is diagonal and each point in Δ is added with infinite multiplicity
- A partial-matching between D_1 and D_2 is
  - a bijection $\bar{M} \subseteq \bar{D}_1 \times \bar{D}_2$.
- The bottleneck distance between D_1 and D_2
  - $d_B\left(D_1,\ D_2\right) := \inf_{\bar{M}} \max_{(x,y)\in\bar{M}} ||x - y||_\infty$

p-th Wasserstein distance

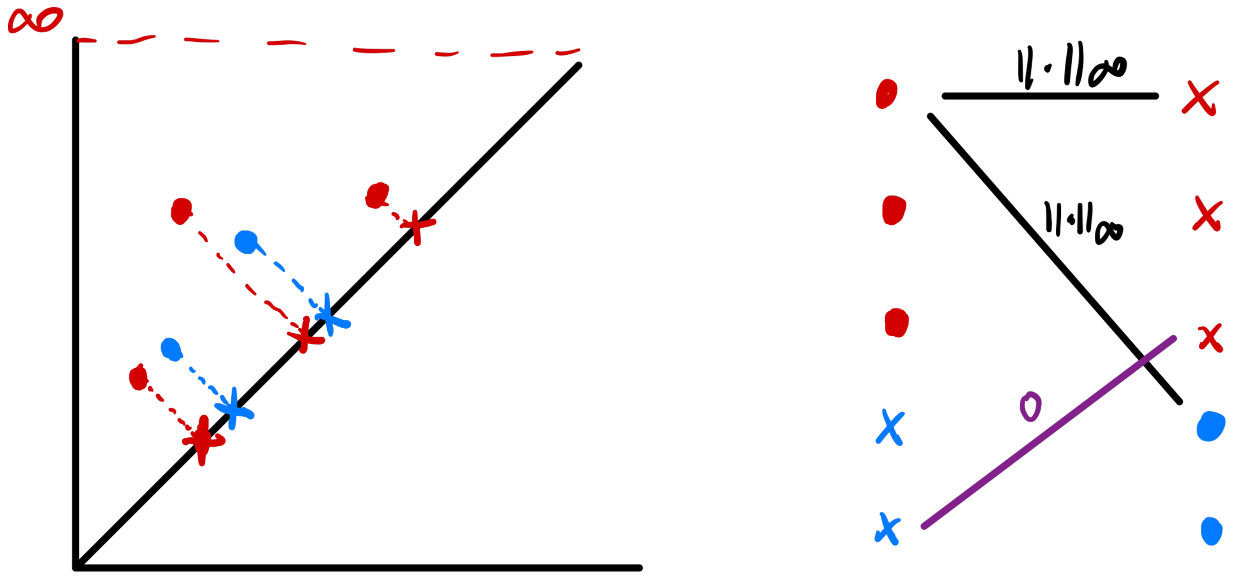
- Given two persistence-diagrams (multiset of points in $(\mathbb{R} \cup \{\infty\})^2$)
  - $D_1 = \{p_1, p_2, \ldots, p_s\}$ and $D_2 = \{q_1, q_2, \ldots, q_t\}$
- Augment $\bar{D}_1 = D_1 \cup \Delta$ and $\bar{D}_2 = D_2 \cup \Delta$
  - where $\Delta = \{(x, x) \in R^2\}$ is diagonal and each point in Δ is added with infinite multiplicity
- A partial-matching between D_1 and D_2 is
  - a bijection $\bar{M} \subseteq \bar{D}_1 \times \bar{D}_2$.
- The p-th Wasserstein distance distance between D_1 and D_2
  - $d_{W,p}\left(D_1,\ D_2\right) := \inf_{\bar{M}} \left[\sum_{(x,y)\in\bar{M}} ||x - y||_\infty^p \right]^{\frac{1}{p}}$
  - $d_{W,\infty}(D_1, D_2) = d_B(D_1, D_2)$

Bottleneck (Wasserstein) distance vs Matching Problem

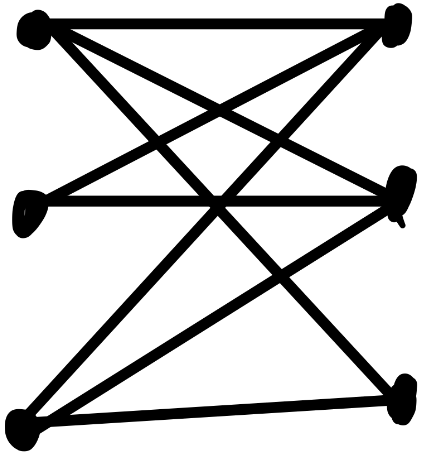
- Let $D_1 = \{x_1, \ldots, x_n\}$ and $D_2 = \{y_1, \ldots, y_m\}$ be two persistence diagrams
- $D_1' = \{x_1', \ldots, x_n'\}$: projections of x_i on to $x_i' \in \Delta$
- Same for D_2'
- $U = D_1 \cup D_2'$ and $V = D_1' \cup D_2$
- Construct a fully connected bipartite graph $G = (U \cup V, E, w)$
- $w(u, v) = \begin{cases} \|u - v\|_\infty, & u \in D_1 \text{ or } v \in D_2 \\ 0, & \text{otherwise} \end{cases}$

Bottleneck (Wasserstein) distance vs Matching Problem

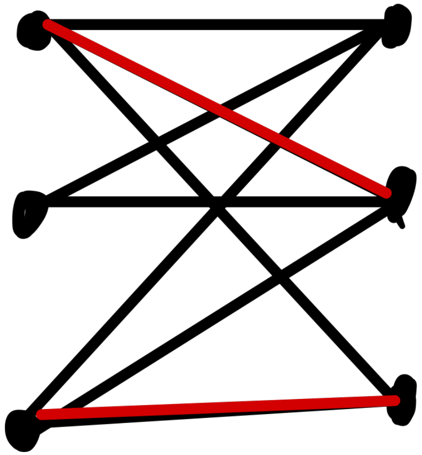
- $U = D_1 \cup D_2'$ and $V = D_1' \cup D_2$
- Construct a fully connected bipartite graph $G = (U \cup V, E, w)$
- $w(u, v) = \begin{cases} \|u - v\|_\infty, & u \in D_1 \text{ or } v \in D_2 \\ 0, & \text{otherwise} \end{cases}$



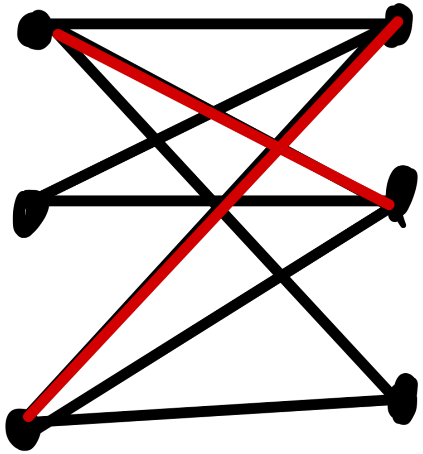
Matching



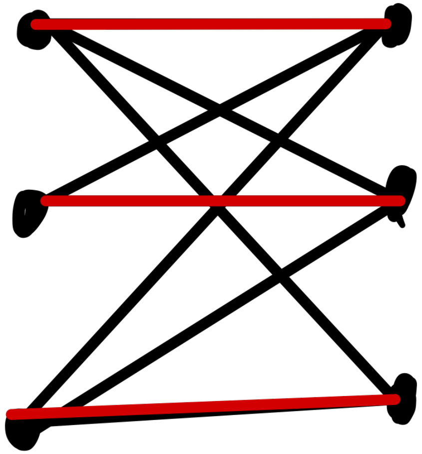
A bipartite graph



Matching



Maximal matching

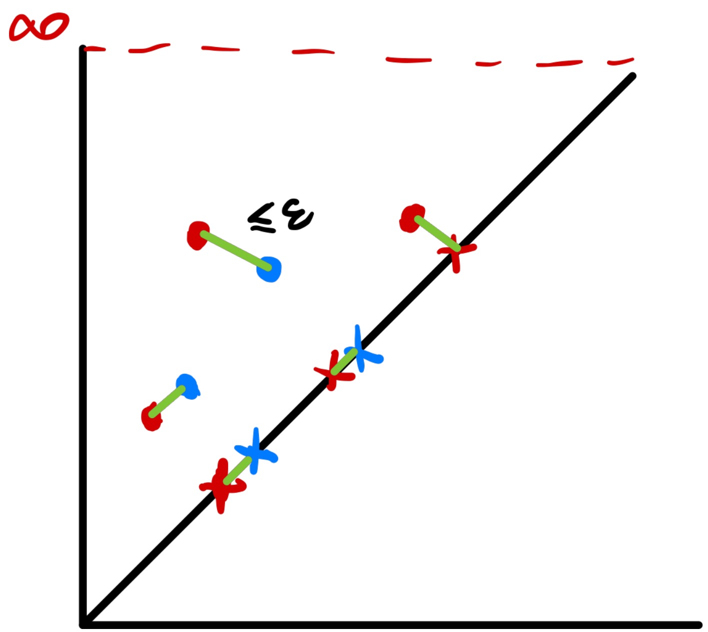


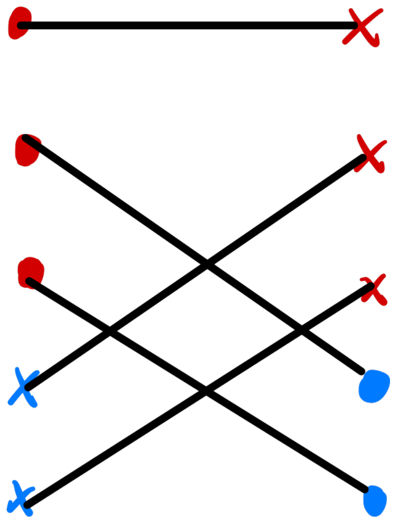
Perfect matching

Bottleneck (Wasserstein) distance vs Matching Problem

- Let $G_\epsilon = (U \cup V, E_\epsilon, w)$ where E_ϵ contains edges with cost $\leq \epsilon$

> - (Reduction Lemma)
>
> $d_B(D_1, D_2) = \inf\{\epsilon : G_\epsilon \text{ has a perfect matching}\}$





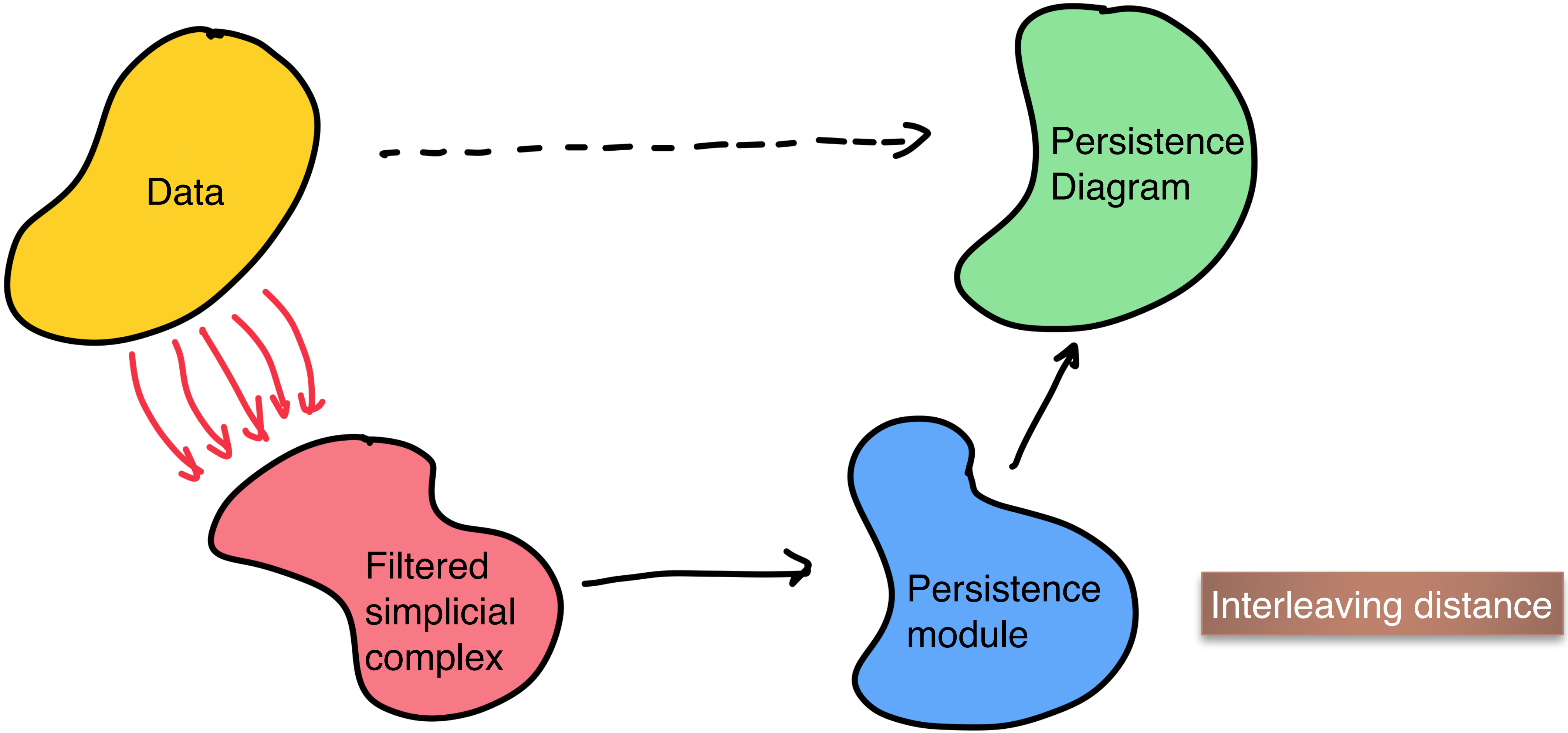
Bottleneck (Wasserstein) distance vs Matching Problem

- Let $G_\epsilon = (U \cup V, E_\epsilon, w)$ where E_ϵ contains edges with cost $> \epsilon$

> - (Reduction Lemma)
>
> $d_B(D_1, D_2) = \inf\{\epsilon : G_\epsilon \text{ has a perfect matching}\}$

- The computation of the bottleneck distance reduces to matching problems for bipartite graphs
  - Ford Fulkerson Algorithm
  - Hungarian Algorithm
  - Hopcroft-Karp Algorithm

Section 2:
Interleaving distance between Persistence Modules


Data
Persistence Diagram
Filtered simplicial complex
Persistence module
Interleaving distance

Interleaving Distance

- A general way to measure distance between two arbitrary persistence modules
  - Interleaving distance
  - First introduced in [Chazal, Cohen-Steiner, Gliss, Guibas, Oudot, 2009]
  - [Lesnick PhD Thesis]
  - [Chazal, de Silva, Gliss and Oudot, 2016] (available on arXiv)

- Two persistence modules (**indexed by** $[0,\infty)$)
  - $U = \{u_{r,s} : U_r \to U_s\}_{r \le s}$
  - $V = \{v_{r,s} : V_r \to V_s\}_{r \le s}$
- Goal: define a distance between them depending on how they interconnect (interleaving) to each other

Intuition

U:	$\cdots \to U_a \to \cdots \to U_b \to \cdots \to \cdots$
V:	$\cdots \to V_a \to \cdots \to V_b \to \cdots \to \cdots$

Intuition

- Isomorphic persistence modules

$$
\begin{tikzcd}
U: & \cdots \arrow[r] & U_a \arrow[r] \arrow[d, leftrightarrow, "\cong"] & \cdots \arrow[r] & U_b \arrow[r] \arrow[d, leftrightarrow, "\cong"] & \cdots \arrow[r] & \cdots \\
V: & \cdots \arrow[r] & V_a \arrow[r] & \cdots \arrow[r] & V_b \arrow[r] & \cdots \arrow[r] & \cdots
\end{tikzcd}
$$

- Vertical maps also have to commute with horizontal maps (in all possible combinations)

Intuition

- Isomorphic persistence modules

$$
\begin{tikzcd}
U: & \cdots \arrow[r] & U_a \arrow[r] \arrow[d, shift right, "\varphi_a"'] & \cdots \arrow[r] & U_b \arrow[r] \arrow[d, shift right, "\varphi_b"'] & \cdots \arrow[r] & \cdots \\
 & \cdots & V_a \arrow[u, shift right, "\phi_a"'] & \cdots & V_b \arrow[u, shift right, "\phi_b"'] & \cdots & \\
V: & \cdots \arrow[r] & V_a \arrow[r] & \cdots \arrow[r] & V_b \arrow[r] & \cdots \arrow[r] & \cdots
\end{tikzcd}
$$

- Vertical maps also have to commute with horizontal maps (in all possible combinations)

Intuition

- Isomorphic persistence modules

$$
\begin{tikzcd}
U: & \cdots \arrow[r] & U_a \arrow[r] \arrow[rr, bend left, "u_{a,b}"] \arrow[d, shift right, "\varphi_a"'] & \cdots \arrow[r] & U_b \arrow[r] \arrow[d, shift right, "\varphi_b"'] & \cdots \arrow[r] & \cdots \\
V: & \cdots \arrow[r] & V_a \arrow[r] \arrow[rr, bend right, "v_{a,b}"'] \arrow[u, shift right, "\phi_a"'] & \cdots \arrow[r] & V_b \arrow[r] \arrow[u, shift right, "\phi_b"'] & \cdots \arrow[r] & \cdots
\end{tikzcd}
$$

- Vertical maps also have to commute with horizontal maps (in all possible combinations)

ϵ-Interleaving

- U and V are ϵ-interleaving if there exists maps
  - $\varphi_a : U_a \to V_{a+\epsilon}$ and $\phi_a : V_a \to U_{a+\epsilon}$ for any $a \in \mathbb{R}$
  - s.t. these maps commute with horizontal maps u's and v's

$$
\begin{tikzcd}
U: & \cdots \arrow[r] & U_a \arrow[r] \arrow[ddrr, "\varphi_a"] & \cdots \arrow[r] & U_{a+\epsilon} \arrow[r] \arrow[ddrr, "\varphi_{a+\epsilon}"] & \cdots \arrow[r] & U_{a+2\epsilon} \arrow[r] & \cdots \\
 & \cdots & & & \cdots & & \cdots & \\
V: & \cdots \arrow[r] & V_a \arrow[r] \arrow[uurr, "\phi_a"] & \cdots \arrow[r] & V_{a+\epsilon} \arrow[r] \arrow[uurr, "\phi_{a+\epsilon}"] & \cdots \arrow[r] & V_{a+2\epsilon} \arrow[r] & \cdots
\end{tikzcd}
$$

ϵ-Interleaving

- U and V are ϵ-interleaving if there exists maps
  - $\varphi_a : U_a \to V_{a+\epsilon}$ and $\phi_a : V_a \to U_{a+\epsilon}$ for any $a \in \mathbb{R}$
  - s.t. these maps commute with horizontal maps u's and v's
- To verify commutativity of maps, only need to check four configurations)

$$
\begin{tikzcd}
U_a \arrow[r, "u_{a,a'}"] \arrow[dr, "\varphi_a"'] & U_{a'} \arrow[dr, "\varphi_{a'}"] & \\
& V_{a+\varepsilon} \arrow[r, "v_{a+\varepsilon,a'+\varepsilon}"'] & V_{a'+\varepsilon}
\end{tikzcd}
\qquad
\begin{tikzcd}
& U_{a+\varepsilon} \arrow[r, "u_{a+\varepsilon,a'+\varepsilon}"] & V_{a'+\varepsilon} \\
V_a \arrow[ur, "\psi_a"] \arrow[r, "v_{a,a'}"'] & V_{a'} \arrow[ur, "\psi_{a'}"] &
\end{tikzcd}
$$

$$
\begin{tikzcd}
U_a \arrow[rr, "u_{a,a+2\varepsilon}"] \arrow[dr, "\varphi_a"'] & & U_{a+2\varepsilon} \\
& V_{a+\varepsilon} \arrow[ur, "\psi_{a+\varepsilon}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& U_{a+\varepsilon} \arrow[dr, "\varphi_{a+\varepsilon}"] & \\
V_a \arrow[ur, "\psi_a"] \arrow[rr, "v_{a,a+2\varepsilon}"'] & & V_{a+2\varepsilon}
\end{tikzcd}
$$

ϵ-Interleaving

- U and V are ϵ-interleaving if there exists maps
  - $\varphi_a : U_a \to V_{a+\epsilon}$ and $\phi_a : V_a \to U_{a+\epsilon}$ for any $a \in [0,\infty)$
  - s.t. these maps commute with horizontal maps u's and v's

- If U and V are 0-interleaving, then they are isomorphic

Interleaving Distance

- $d_I(V, U) = \inf_{\epsilon \geq 0} \{V \text{ and } U \text{ are } \epsilon\text{-interleaved}\}$

  - It is an extended pseudo-metric
    - Satisfying triangle inequality
      - $d_I(U, W) \leq d_I(U, V) + d_I(V, W)$
    - Can take value ∞
    - Non isomorphic persistance modules can have 0 distance

Examples

- A closed interval module $I[1,2]$
- A half-closed interval module $I[1,2)$

Examples

- An infinitely long interval module $I[1,\infty)$
- A finite interval module $I[1,2)$

Examples

- $I[1,2)$ vs $I[1.1,2.1)$

Examples

- $I[0.1, 0.2)$ vs $I[10.1, 10.2)$

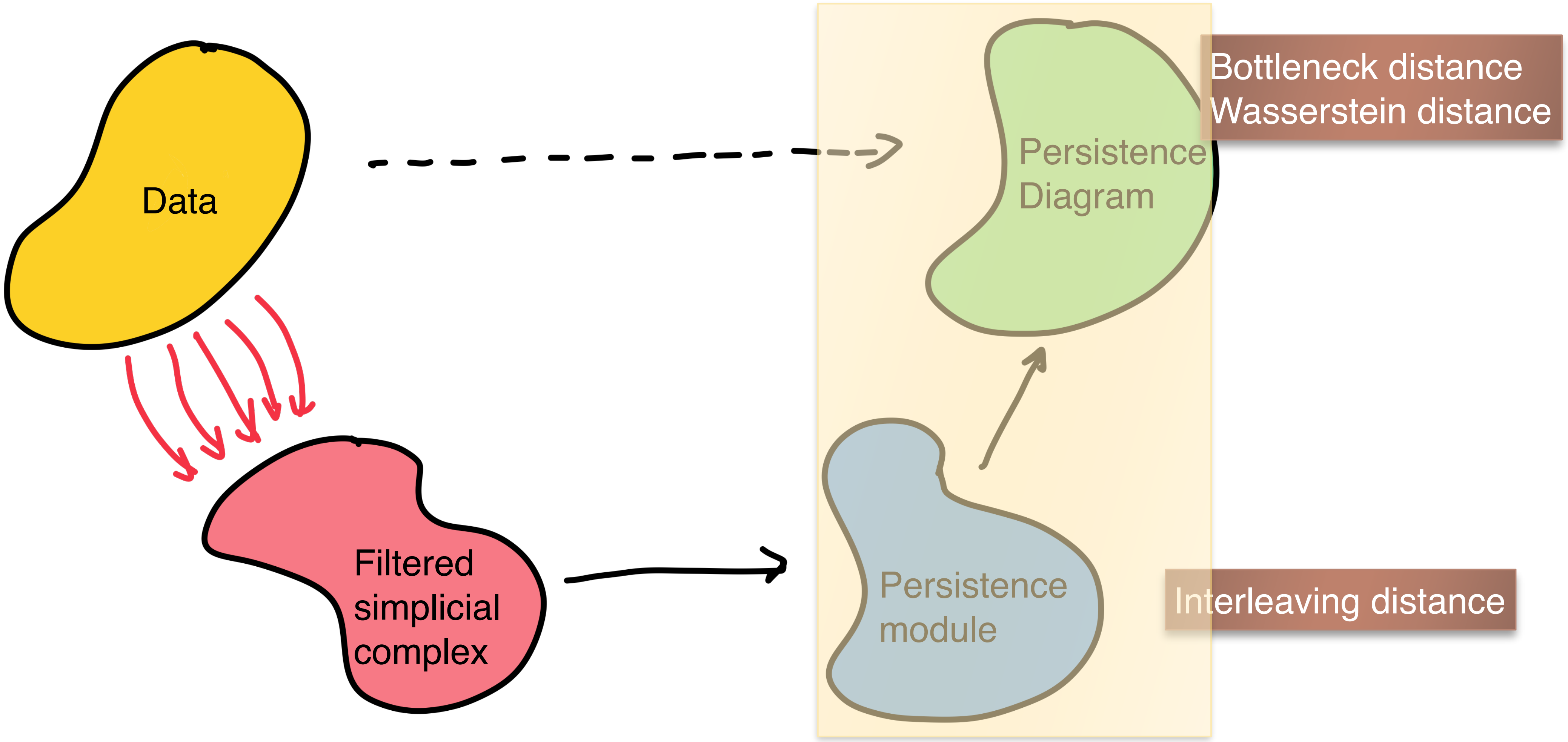
Interleaving distance between interval modules

- For two interval modules $I = I[b, d)$ and $I' = I[b', d')$

- $$d_I(I, I') = \min\left(\max(|b - b'|, |d - d'|), \max(\frac{|b - d|}{2}, \frac{|b' - d'|}{2})\right)$$

- So $d_I(I, I') = d_B(Dgm(I), Dgm(I'))$!

Bottleneck distance vs interleaving distance



Recall: Finitely presented filtration

Recall: Finitely presented filtration

- A filtration $(K_t)_{t \in [0,\infty)}$ is called **finitely represented** if

Recall: Finitely presented filtration

- A filtration $(K_t)_{t \in [0,\infty)}$ is called **finitely represented** if
  - There exist $0 = t_0 < t_1 < \cdots < t_n$ such that

Recall: Finitely presented filtration

- A filtration $(K_t)_{t \in [0,\infty)}$ is called **finitely represented** if
  - There exist $0 = t_0 < t_1 < \cdots < t_n$ such that
  - $K_t = K_{t'}, \quad \forall t_i \leq t < t' < t_{i+1}$ and $i = 0,\ldots,n$ $(t_{n+1} := \infty)$

Recall: Finitely presented filtration

- A filtration $(K_t)_{t \in [0,\infty)}$ is called **finitely represented** if
  - There exist $0 = t_0 < t_1 < \cdots < t_n$ such that
  - $K_t = K_{t'}, \quad \forall t_i \leq t < t' < t_{i+1}$ and $i = 0,\ldots,n$ $(t_{n+1} := \infty)$

Recall: Finitely presented filtration

- A filtration $(K_t)_{t \in [0,\infty)}$ is called **finitely represented** if
  - There exist $0 = t_0 < t_1 < \cdots < t_n$ such that
  - $K_t = K_{t'}, \quad \forall t_i \leq t < t' < t_{i+1}$ and $i = 0,\ldots,n$ $(t_{n+1} := \infty)$
- Both Čech and Rips filtrations are finitely represented

Interleaving Distance

- $d_I(V, U) = \inf_{\epsilon \geq 0} \{V$ and U are ϵ-interleaved$\}$

Interleaving Distance

- $d_I(V, U) = \inf_{\epsilon \geq 0} \{V \text{ and } U \text{ are } \epsilon\text{-interleaved}\}$

General Stability Theorem [Chazal, Cohen-Steiner, Gliss, Guibas, Oudot 2009]

Given two finitely represented persistence modules U and V, let D_U and D_V be their corresponding persistence diagrams. We then have:

$$d_B(D_U, D_V) \leq d_I(U, V)$$

A More General Result

- $d_I(V, U) = \inf_{\epsilon \geq 0} \{V \text{ and } U \text{ are } \epsilon\text{-interleaved}\}$

A More General Result

- $d_I(V, U) = \inf_{\epsilon \geq 0} \{V$ and U are ϵ-interleaved$\}$

Isometry Theorem [Lesnick 2015], [Chazal, de Silva, Gliss and Oudot, 2016]

Given two finitely represented persistence modules U and V, let D_U and D_V be their corresponding persistence diagrams. We then have:

$$d_B(D_U, D_V) = d_I(U, V)$$

A More General Result

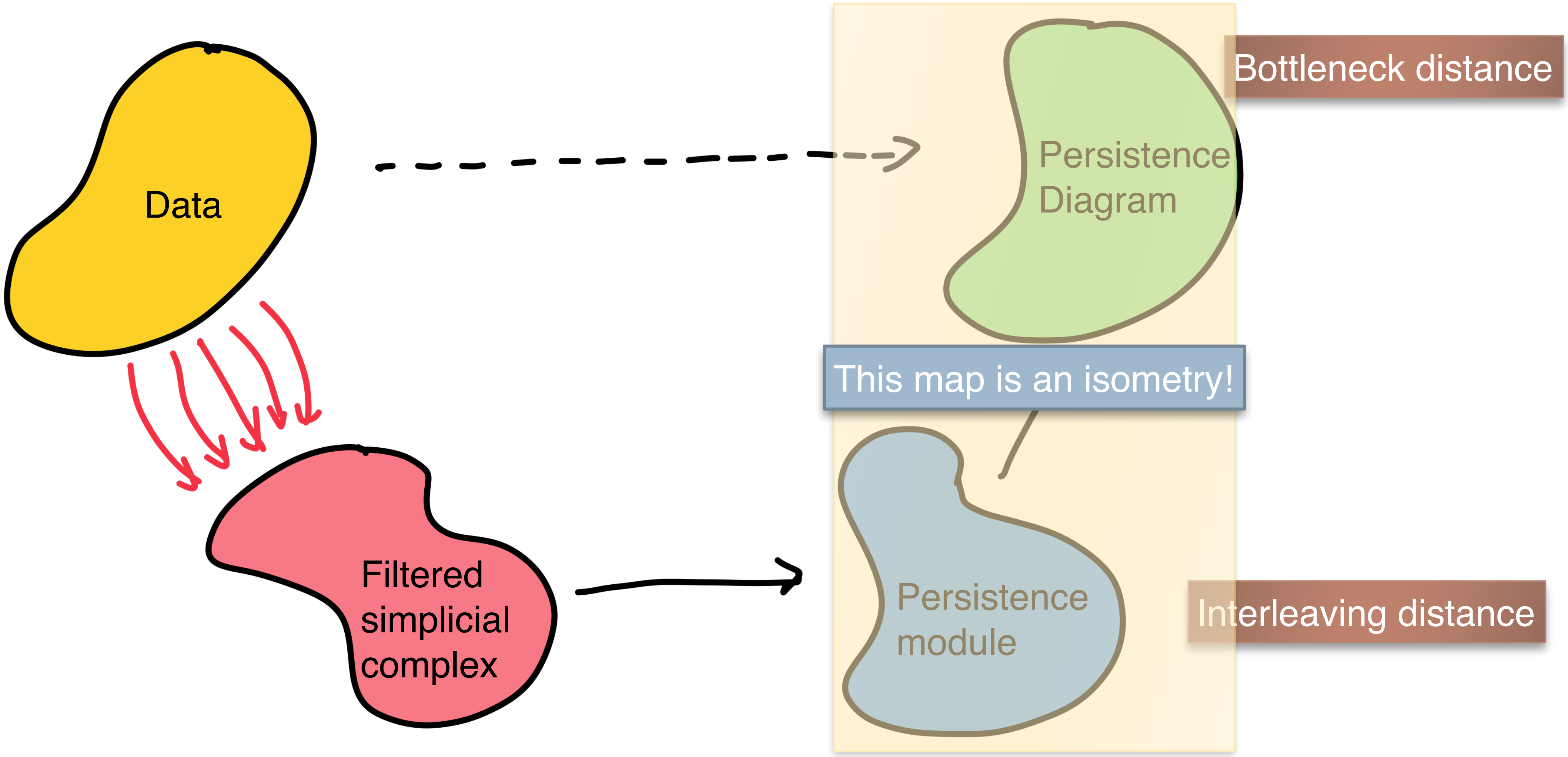
- $d_I(V, U) = \inf\limits_{\epsilon \geq 0} \{V \text{ and } U \text{ are } \epsilon\text{-interleaved}\}$

Isometry Theorem [Lesnick 2015], [Chazal, de Silva, Gliss and Oudot, 2016]

Giv Holds for more general persistence modules modules U and V, let D_U and D_V be their corresponding persistence diagrams. We then have:

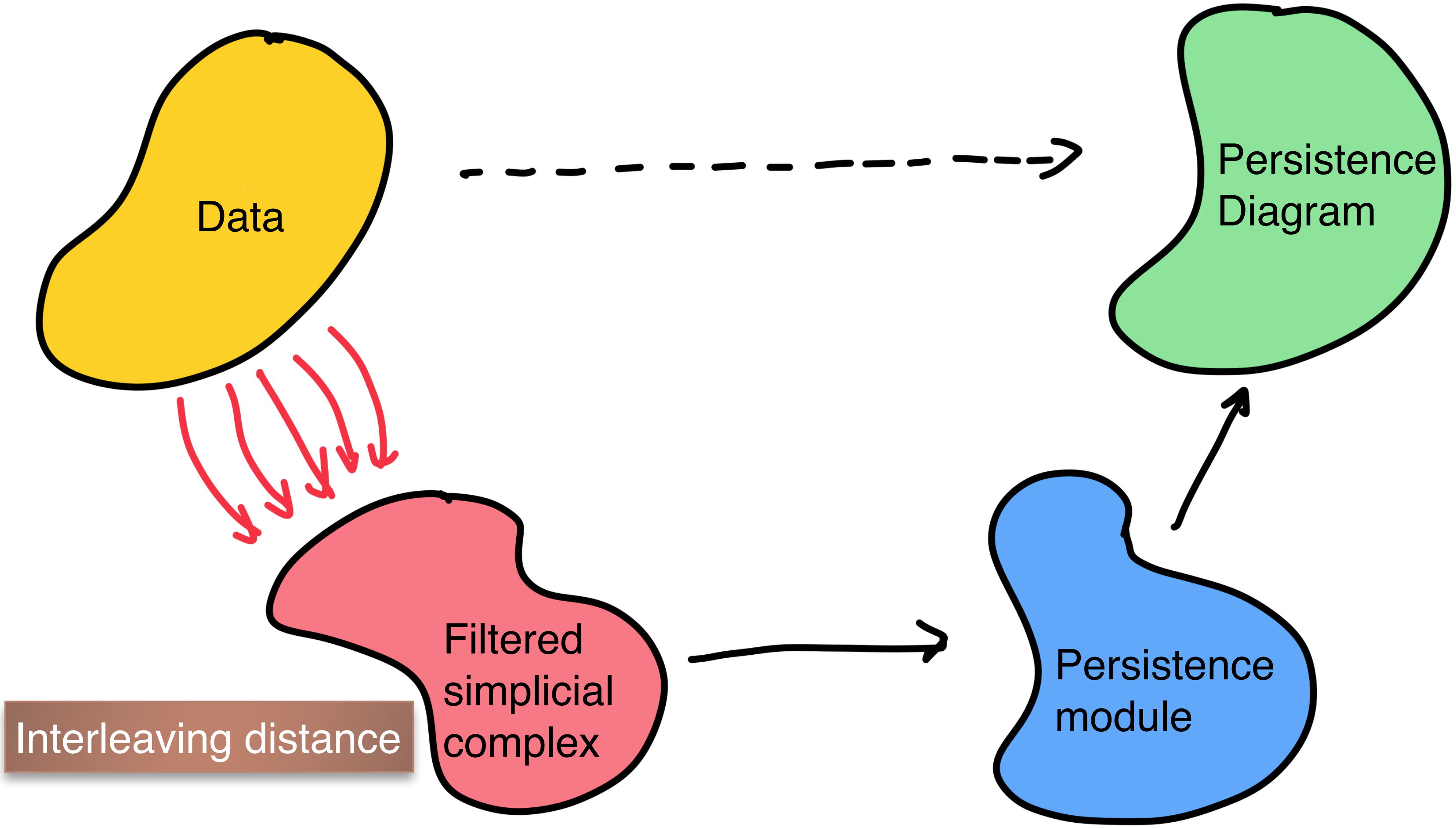
$$d_B(D_U, D_V) = d_I(U, V)$$

Bottleneck distance vs interleaving distance



Section 3:
Interleaving distance between filtrations

Bottleneck distance vs interleaving distance



Filtered simplicial complexes over the same vertex set

- Given two simplicial filtrations $\mathscr{X}$ and $\mathscr{Y}$ over the "same" vertex set V
- We say they are ϵ-interleaved if there exist **inclusion** maps $\varphi_a : X_a \hookrightarrow Y_{a+\epsilon}$ and $\phi_{a'} : Y_{a'} \hookrightarrow X_{a'+\epsilon}$ such that the following diagram commutes

$$
\begin{tikzcd}
\mathscr{X}: & \cdots \arrow[r, phantom, "\subseteq"] & X_a \arrow[r, phantom, "\subseteq"] \arrow[drr, "\varphi_a"] & \cdots \arrow[r, phantom, "\subseteq"] & X_{a+\epsilon} \arrow[r, phantom, "\subseteq"] \arrow[drr, "\varphi_{a+\epsilon}"] & \cdots \arrow[r, phantom, "\subseteq"] & X_{a+2\epsilon} \arrow[r, phantom, "\subseteq"] & \cdots \\
\mathscr{Y}: & \cdots \arrow[r, phantom, "\subseteq"] & Y_a \arrow[r, phantom, "\subseteq"] \arrow[urr, "\phi_a"'] & \cdots \arrow[r, phantom, "\subseteq"] & Y_{a+\epsilon} \arrow[r, phantom, "\subseteq"] \arrow[urr, "\phi_{a+\epsilon}"'] & \cdots \arrow[r, phantom, "\subseteq"] & Y_{a+2\epsilon} \arrow[r, phantom, "\subseteq"] & \cdots
\end{tikzcd}
$$

Filtered topological spaces over the same ambient space

- Given two topological filtrations $\mathcal{X}$ and $\mathcal{Y}$ of subspaces in a common ambient space Z
- We say they are ϵ-interleaved if there exist **inclusion** maps $\varphi_a : X_a \hookrightarrow Y_{a+\epsilon}$ and $\phi_{a'} : Y_{a'} \hookrightarrow X_{a'+\epsilon}$ such that the following diagram commutes

$$
\begin{tikzcd}
\mathcal{X}: & \cdots \subseteq & X_a \arrow[ddrr, "\varphi_a"] & \subseteq \cdots \subseteq & X_{a+\epsilon} \arrow[ddrr, "\varphi_{a+\epsilon}"] & \subseteq \cdots \subseteq & X_{a+2\epsilon} & \subseteq \cdots \\
 & \cdots & & & \cdots & & & \cdots \\
\mathcal{Y}: & \cdots \subseteq & Y_a \arrow[uurr, "\phi_a"'] & \subseteq \cdots \subseteq & Y_{a+\epsilon} \arrow[uurr, "\phi_{a+\epsilon}"'] & \subseteq \cdots \subseteq & Y_{a+2\epsilon} & \subseteq \cdots
\end{tikzcd}
$$

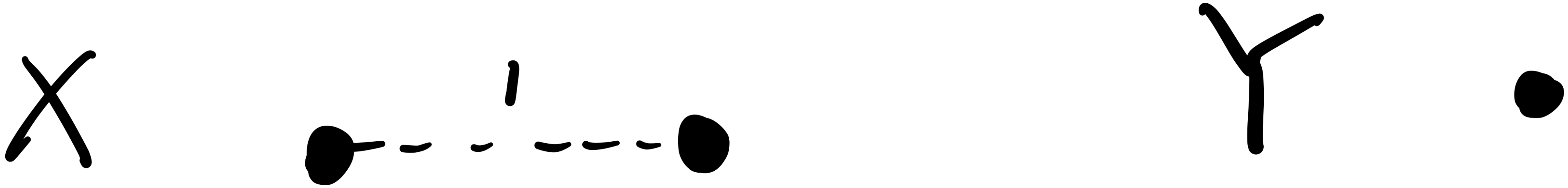
A first Interleaving distance

- Let $\mathcal{X}$ and $\mathcal{Y}$
  - Be two simplicial filtrations over the "same" vertex set V or
  - two topological filtrations of subspaces in a common ambient space Z
- $d_I(\mathcal{X}, \mathcal{Y}) = \inf_{\epsilon \geq 0} \{\mathcal{X} \text{ and } \mathcal{Y} \text{ are } \epsilon\text{-interleaved}\}$

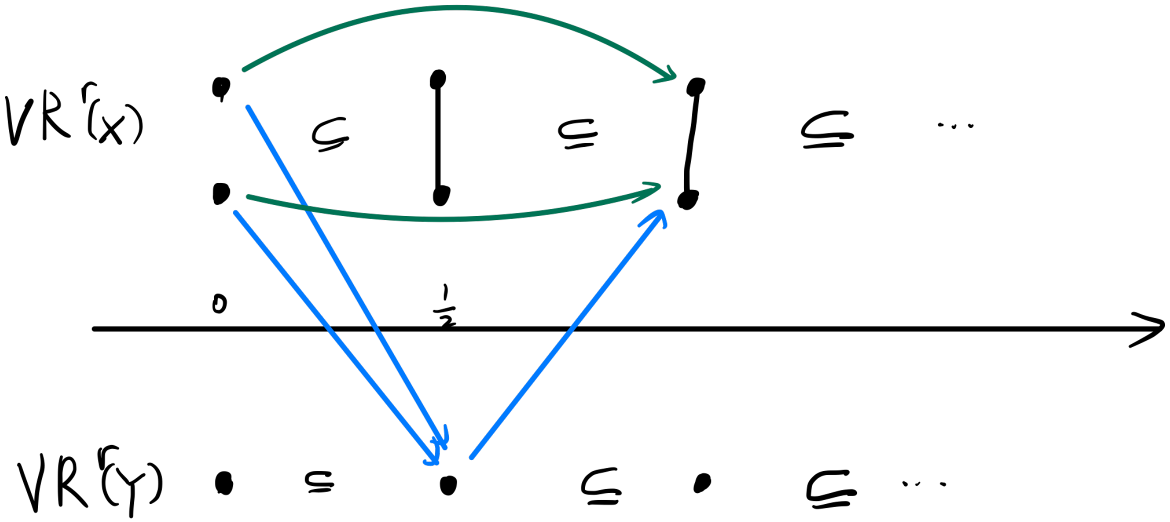
General filtered simplicial complexes - an educated guess

- Given two simplicial filtrations $\mathscr{X}$ and $\mathscr{Y}$
- We say they are ϵ-interleaved if there exist **simplicial** maps $\varphi_a : X_a \to Y_{a+\epsilon}$ and $\phi_{a'} : Y_{a'} \to X_{a'+\epsilon}$ such that the following diagram commutes

$$
\begin{tikzcd}
\mathscr{X}: & \cdots \arrow[r, phantom, "\subseteq"] & X_a \arrow[r, phantom, "\subseteq"] \arrow[ddrr, "\varphi_a"] & \cdots \arrow[r, phantom, "\subseteq"] & X_{a+\epsilon} \arrow[r, phantom, "\subseteq"] \arrow[ddrr, "\varphi_{a+\epsilon}"] & \cdots \arrow[r, phantom, "\subseteq"] & X_{a+2\epsilon} \arrow[r, phantom, "\subseteq"] & \cdots \\
 & \cdots & & & \cdots & & & \cdots \\
\mathscr{Y}: & \cdots \arrow[r, phantom, "\subseteq"] & Y_a \arrow[r, phantom, "\subseteq"] \arrow[uurr, "\phi_a"'] & \cdots \arrow[r, phantom, "\subseteq"] & Y_{a+\epsilon} \arrow[r, phantom, "\subseteq"] \arrow[uurr, "\phi_{a+\epsilon}"'] & \cdots \arrow[r, phantom, "\subseteq"] & Y_{a+2\epsilon} \arrow[r, phantom, "\subseteq"] & \cdots
\end{tikzcd}
$$



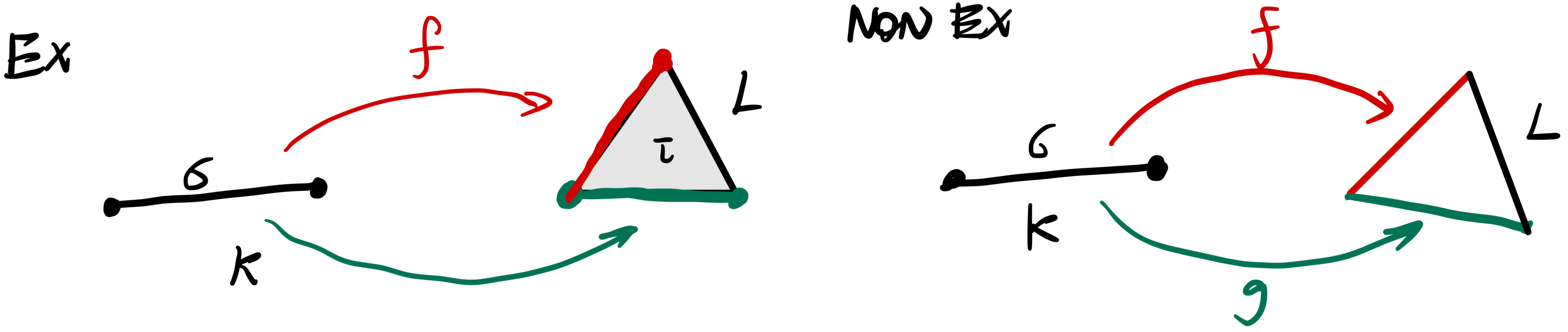
- Rips filtration



- $d_I(VR(X), VR(Y)) = \infty$! Definitely larger than any reasonable distance between the data sets X and Y. This makes Data $\to$ filtration unstable!

Contiguity

- Two simplicial maps $f, g : K \to L$ are contiguous if for any $\sigma \in \Sigma_K$ there exists a simplex $\tau \in \Sigma_L$ such that $f(\sigma) \cup g(\sigma) \subseteq \tau$



- $f, g : |K| \to |L|$ are homotopic
- $f_* : H_*(K) \to H_*(L)$ is the same map as $g_* : H_*(K) \to H_*(L)$

General filtered simplicial complexes

- Given two simplicial filtrations $\mathcal{X}$ and $\mathcal{Y}$
- We say they are ϵ-interleaved if there exist **simplicial** maps $\varphi_a : X_a \to Y_{a+\epsilon}$ and $\phi_{a'} : Y_{a'} \to X_{a'+\epsilon}$ such that the following diagram commutes up to **contiguity**

$$
\begin{tikzcd}
\mathcal{X}: & \cdots \subseteq & X_a \arrow[ddrr, "\varphi_a"] & \subseteq \cdots \subseteq & X_{a+\epsilon} \arrow[ddrr, "\varphi_{a+\epsilon}"] & \subseteq \cdots \subseteq & X_{a+2\epsilon} & \subseteq \cdots \\
 & \cdots & & & \cdots & & & \cdots \\
\mathcal{Y}: & \cdots \subseteq & Y_a \arrow[uurr, "\phi_a"'] & \subseteq \cdots \subseteq & Y_{a+\epsilon} \arrow[uurr, "\phi_{a+\epsilon}"'] & \subseteq \cdots \subseteq & Y_{a+2\epsilon} & \subseteq \cdots
\end{tikzcd}
$$

- $d_I(\mathcal{X}, \mathcal{Y}) = \inf_{\epsilon \geq 0} \{\mathcal{X}$ and $\mathcal{Y}$ are ϵ-interleaved$\}$

A generalization to simplicial towers

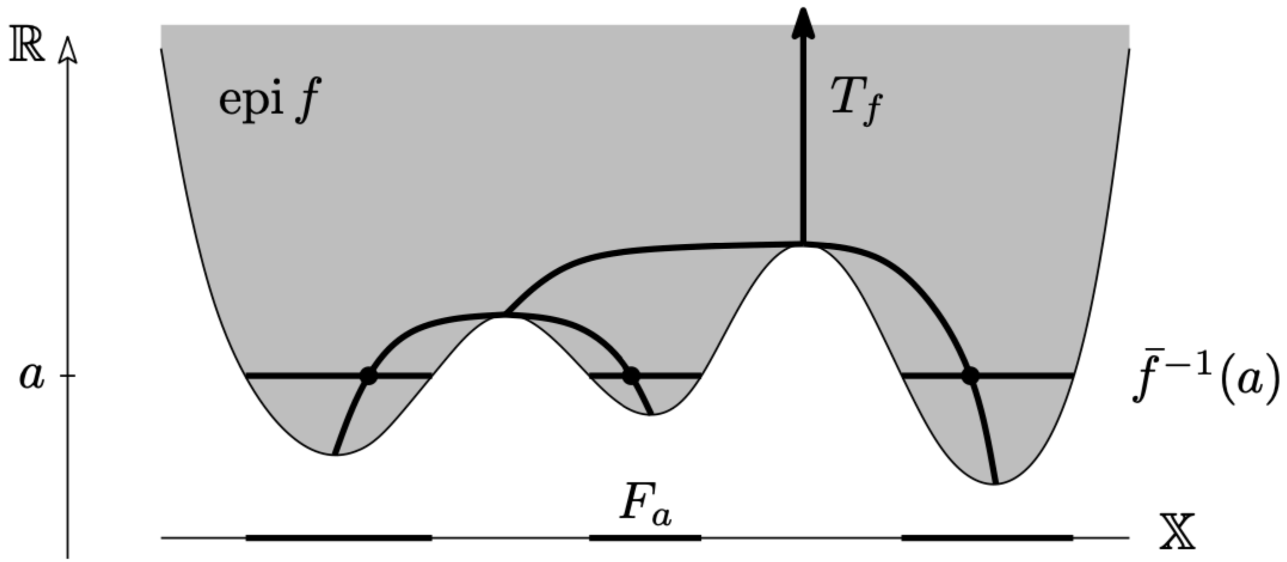
Simplicial maps

- A simplicial tower $\mathscr{X} : \cdots \to X_a \to \cdots \to X_{a+\epsilon} \to \cdots \to X_{a+2\epsilon} \to \cdots$
- We say two simplicial towers $\mathscr{X}$ and $\mathscr{Y}$ are ϵ-interleaved if there exist **simplicial** maps $\varphi_a : X_a \to Y_{a+\epsilon}$ and $\phi_{a'} : Y_{a'} \to X_{a'+\epsilon}$ such that the following diagram commutes up to **contiguity**

$$
\begin{tikzcd}
\mathscr{X}: & \cdots \arrow[r] & X_a \arrow[r] \arrow[ddrr, "\varphi_a"] & \cdots \arrow[r] & X_{a+\epsilon} \arrow[r] \arrow[ddrr, "\varphi_{a+\epsilon}"] & \cdots \arrow[r] & X_{a+2\epsilon} \arrow[r] & \cdots \\
& \cdots & & & \cdots & & \cdots & \\
\mathscr{Y}: & \cdots \arrow[r] & Y_a \arrow[r] \arrow[uurr, "\phi_a"] & \cdots \arrow[r] & Y_{a+\epsilon} \arrow[r] \arrow[uurr, "\phi_{a+\epsilon}"] & \cdots \arrow[r] & Y_{a+2\epsilon} \arrow[r] & \cdots
\end{tikzcd}
$$

A special example - Merge tree

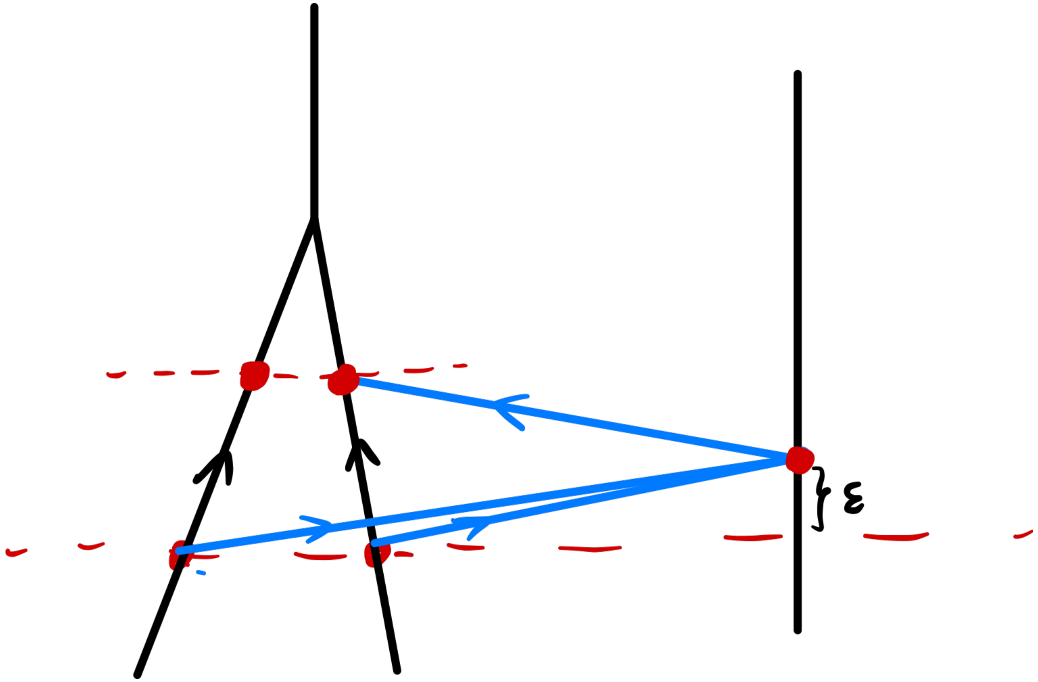
- What if each X_a in $\mathscr{X} = \{X_a\}_a$ is just a finite set (or $\dim(X_a) = 0$)?
- Merge tree: a simplicial tower generated by level sets



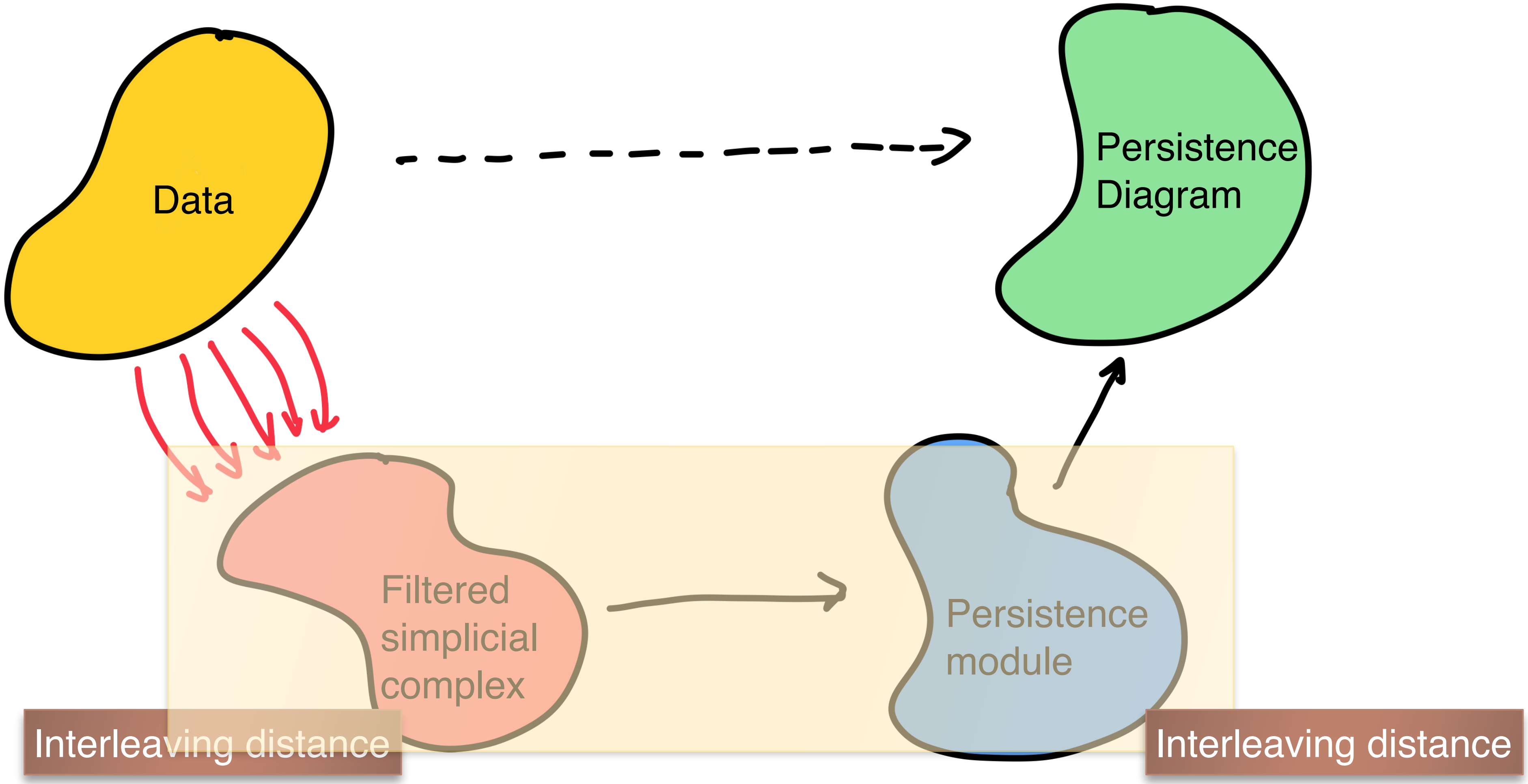
Courtesy of Morozov et al.

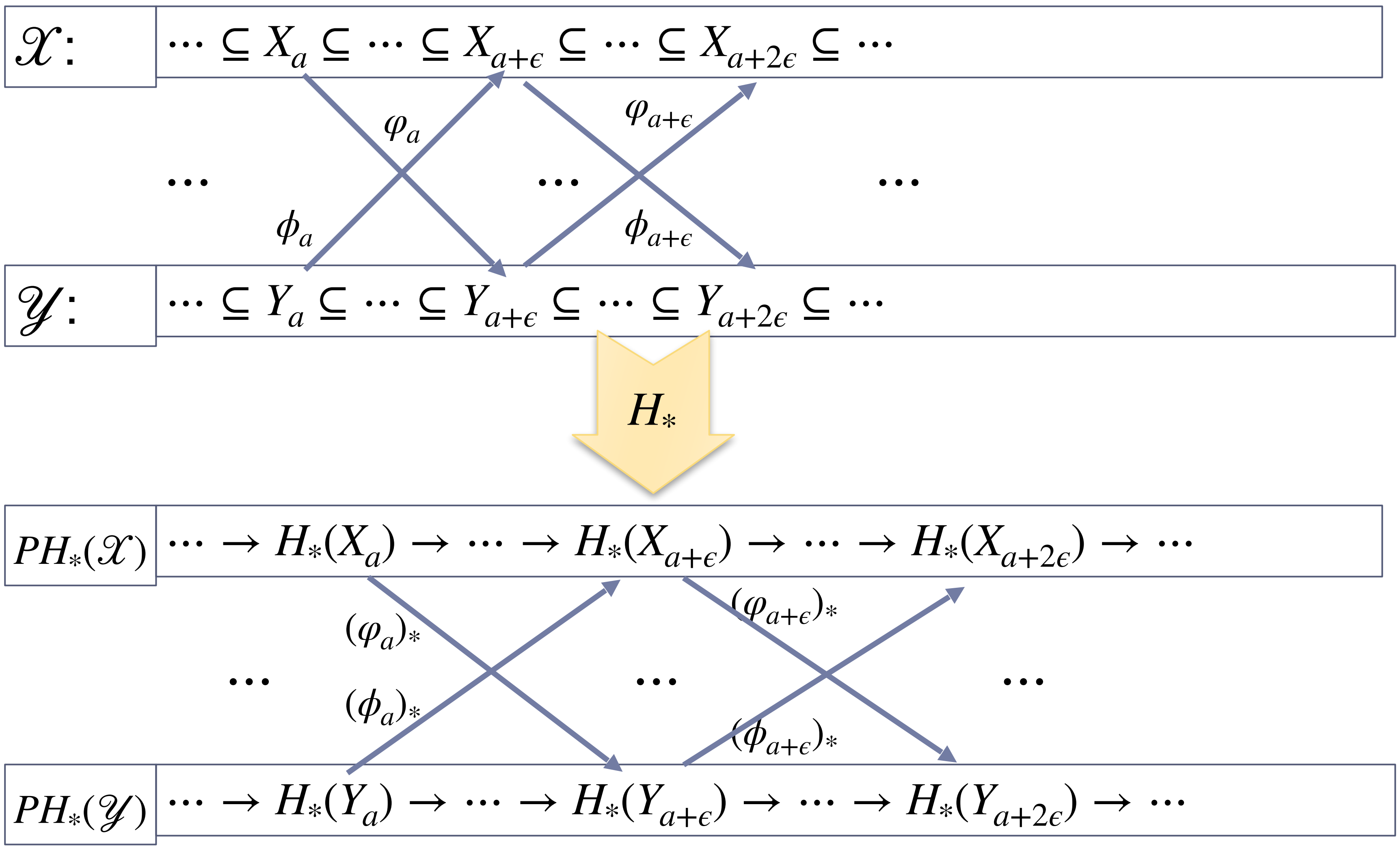
A special example - Merge tree

- What if each X_a in $\mathscr{X} = \{X_a\}_a$ is just a finite set (or $\dim(X_a) = 0$)?
- Merge tree: a simplicial tower generated by level sets
- The contiguity requirement can be replaced by the equality requirement



Interleaving distance vs interleaving distance




$\mathscr{X}$:
$\cdots \subseteq X_a \subseteq \cdots \subseteq X_{a+\epsilon} \subseteq \cdots \subseteq X_{a+2\epsilon} \subseteq \cdots$
φ_a
$\varphi_{a+\epsilon}$
ϕ_a
$\phi_{a+\epsilon}$
$\mathscr{Y}$:
$\cdots \subseteq Y_a \subseteq \cdots \subseteq Y_{a+\epsilon} \subseteq \cdots \subseteq Y_{a+2\epsilon} \subseteq \cdots$
H_*
$PH_*(\mathscr{X})$
$\cdots \to H_*(X_a) \to \cdots \to H_*(X_{a+\epsilon}) \to \cdots \to H_*(X_{a+2\epsilon}) \to \cdots$
$(\varphi_a)_*$
$(\varphi_{a+\epsilon})_*$
$(\phi_a)_*$
$(\phi_{a+\epsilon})_*$
$PH_*(\mathscr{Y})$
$\cdots \to H_*(Y_a) \to \cdots \to H_*(Y_{a+\epsilon}) \to \cdots \to H_*(Y_{a+2\epsilon}) \to \cdots$

- An ϵ-interleaving between simplicial filtrations induces an ϵ-interleaving between persistence modules!

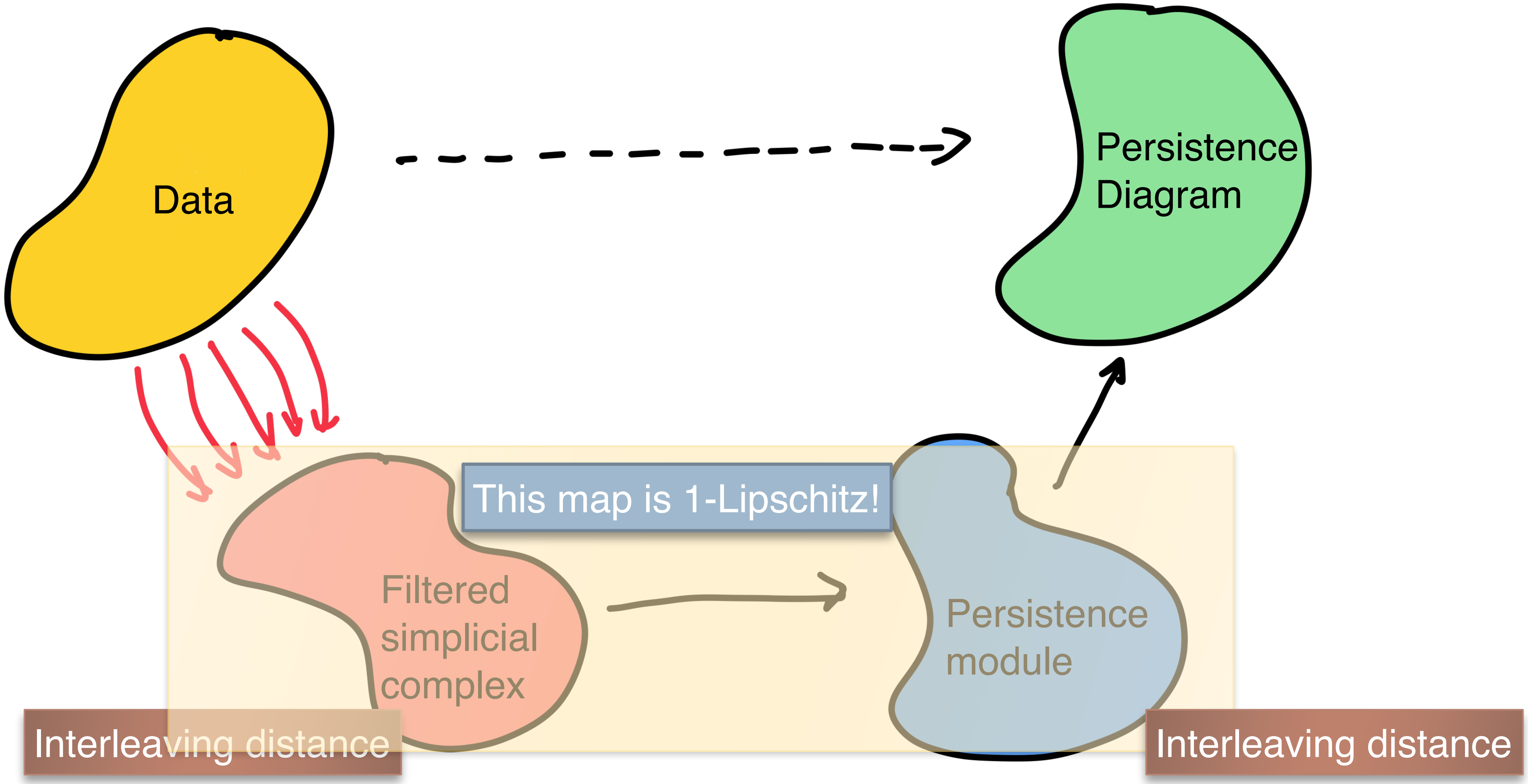
- An ϵ-interleaving between simplicial filtrations induces an ϵ-interleaving between persistence modules!

Theorem

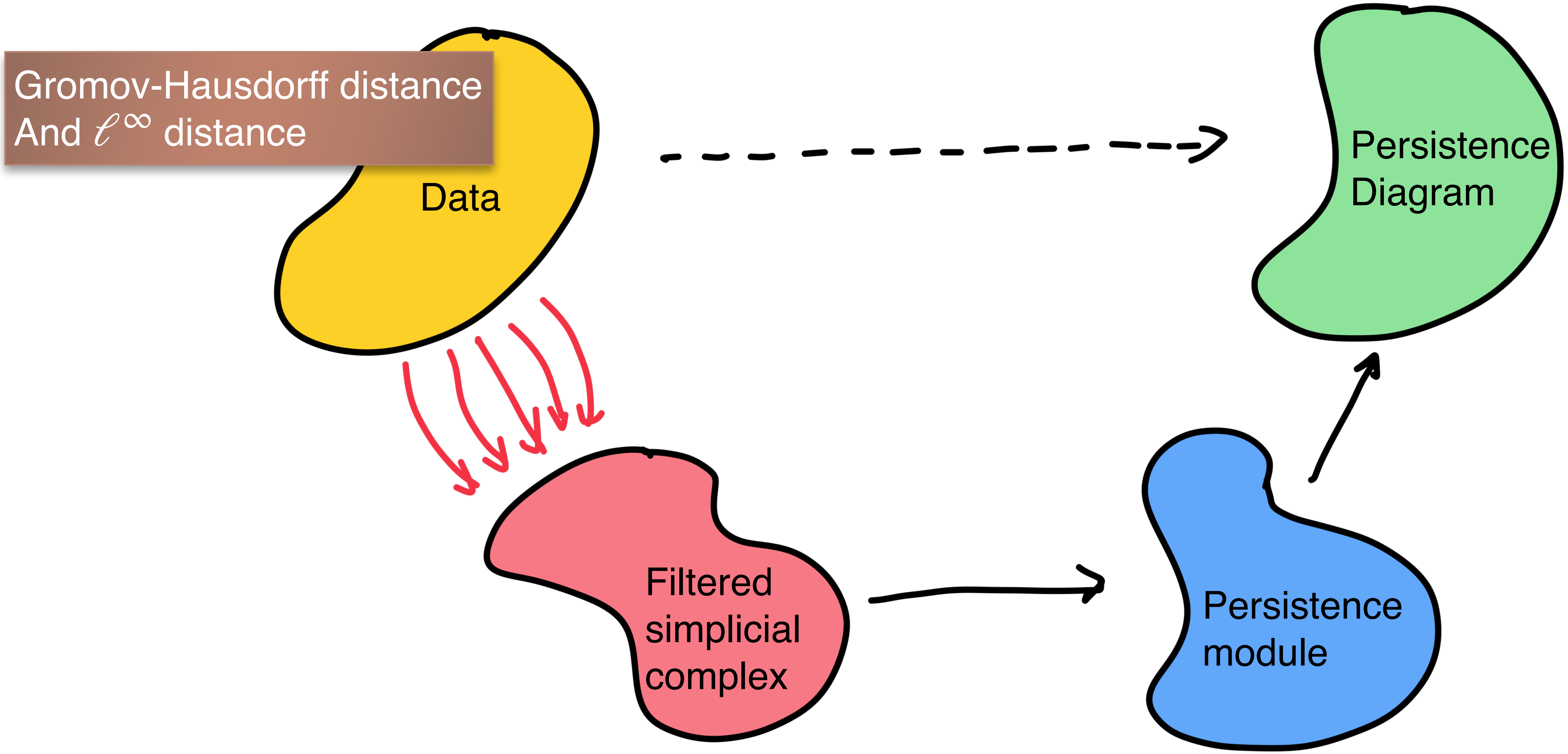
Given two simplicial filtrations $\mathcal{X}$ and $\mathcal{Y}$, let $PH_p(\mathcal{X})$ and $PH_p(\mathcal{Y})$ be the corresponding p-dim persistence modules induced by them. We then have:

$$d_I(PH_p(\mathcal{X}), PH_p(\mathcal{X})) \leq d_I(\mathcal{X}, \mathcal{Y})$$

Interleaving distance vs interleaving distance



Section 4:
Distances for data and stability


Gromov-Hausdorff distance
And ℓ^∞ distance
Data
Persistence
Diagram
Filtered
simplicial
complex
Persistence
module

Functions on a given space

- Let X be a set (e.g., X is a manifold or a subset in $\mathbb{R}^d$)

- Consider the collection of **bounded** functions $f : X \to \mathbb{R}$, i.e.,

  $\sup_{x \in X} |f(x)| < \infty$

- A natural distance between $f, g : X \to \mathbb{R}$ is the ℓ^∞ distance
  - $\|f - g\|_\infty := \sup_{x \in X} |f(x) - g(x)|$

- Given a triangulable space X and two "nice" functions $f, g : X \to \mathbb{R}$
- Let $\epsilon = \|f - g\|_\infty$ and let $X_f^t := f^{-1}(-\infty, t]$
- $f^{-1}(-\infty, t] \subseteq g^{-1}(-\infty, t + \epsilon] \subseteq f^{-1}(-\infty, t + 2\epsilon]$
  - $x \in f^{-1}(-\infty, t]$ means $f(x) \leq t$
  - Since $|f(x) - g(x)| \leq \epsilon$, we have that $g(x) \leq t + \epsilon$

- Given a triangulable space X and two "nice" functions $f, g : X \to \mathbb{R}$
- Let $\epsilon = \|f - g\|_\infty$ and let $X_f^t := f^{-1}(-\infty, t]$
- $f^{-1}(-\infty, t] \subseteq g^{-1}(-\infty, t + \epsilon] \subseteq f^{-1}(-\infty, t + 2\epsilon]$
- So the two sub level set filtrations $X_f = \{X_f^t\}_t$ and $X_g = \{X_g^t\}_t$ are ϵ interleaved

$$
\begin{tikzcd}
X_f: & \cdots \arrow[r, phantom, "\subseteq"] & X_f^a \arrow[r, phantom, "\subseteq"] \arrow[ddrr, "\varphi_a"] & \cdots \arrow[r, phantom, "\subseteq"] & X_f^{a+\epsilon} \arrow[r, phantom, "\subseteq"] \arrow[ddrr, "\varphi_{a+\epsilon}"] & \cdots \arrow[r, phantom, "\subseteq"] & X_f^{a+2\epsilon} \arrow[r, phantom, "\subseteq"] & \cdots \\
 & & & \cdots & & \cdots & & \cdots \\
X_g: & \cdots \arrow[r, phantom, "\subseteq"] & X_g^a \arrow[r, phantom, "\subseteq"] \arrow[uurr, "\phi_a"'] & \cdots \arrow[r, phantom, "\subseteq"] & X_g^{a+\epsilon} \arrow[r, phantom, "\subseteq"] \arrow[uurr, "\phi_{a+\epsilon}"'] & \cdots \arrow[r, phantom, "\subseteq"] & X_g^{a+2\epsilon} \arrow[r, phantom, "\subseteq"] & \cdots
\end{tikzcd}
$$

- Given a triangulable space X and two "nice" functions $f, g : X \to \mathbb{R}$
- Let $\epsilon = \|f - g\|_\infty$ then
- $f^{-1}(-\infty, t] \subseteq g^{-1}(-\infty, t + \epsilon] \subseteq f^{-1}(-\infty, t + 2\epsilon]$
- So the two sub level set filtrations $X_f = \{f^{-1}(-\infty, t]\}_t$ and $X_g = \{g^{-1}(-\infty, t]\}_t$ are ϵ interleaved

- $d_I(PH_*(X_f), PH_*(X_g)) \leq d_I(X_f, X_g) \leq \|f - g\|_\infty$

Stability of persistence diagrams - Function induced persistence

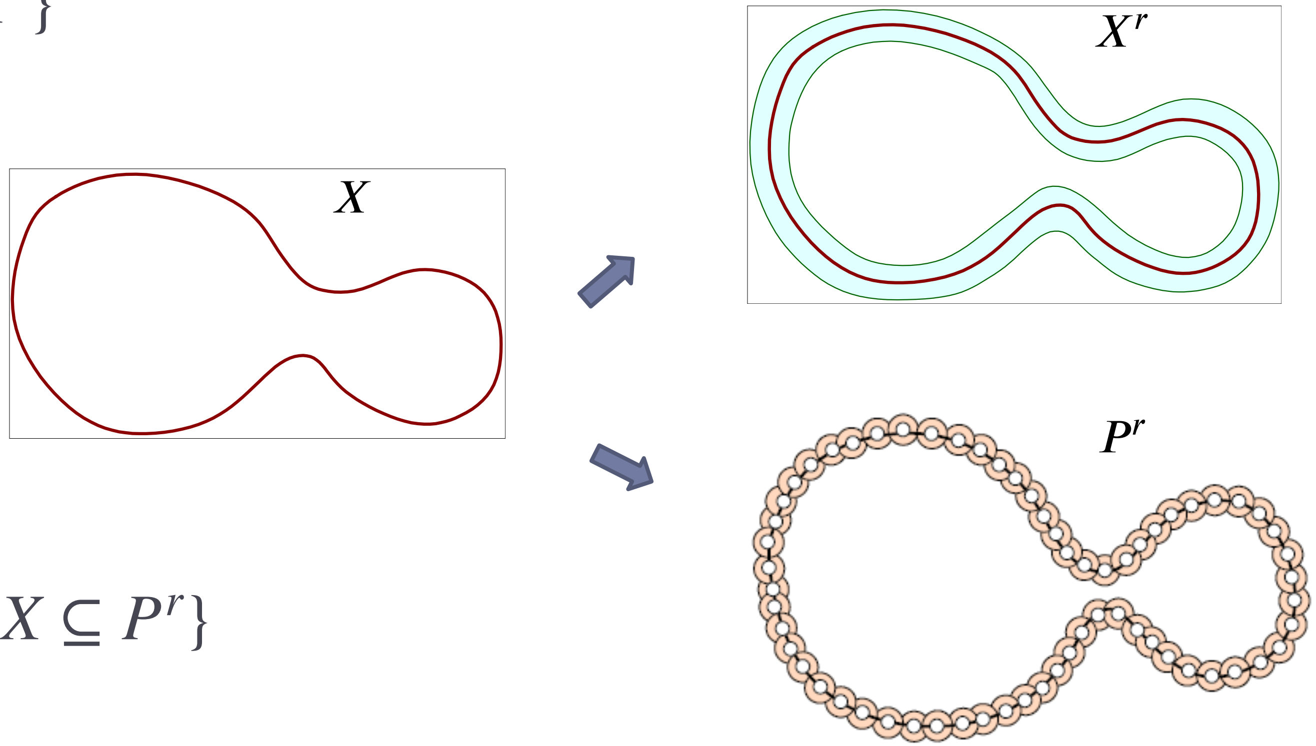
Stability Theorem [Cohen-Steiner et al 2007]

Given two functions $f, g: X \to R$, let D_f and D_g be the persistence diagrams for the persistence modules induced by the sub-level set (resp. super-level set) filtrations w.r.t f and g, respectively. We then have:

$$d_B\left(D_f,\ D_g\right) \leq ||\ f - g\ ||_\infty$$

Hausdorff distance between subsets

- Hausdorff distance between two sets $A,\ B \subset \left(Z,\ d_Z\right)$
  - $d_H(A,\ B) = \max\{\max_{a\in A}\min_{b\in B} d_Z(a,b), \max_{b\in B}\min_{a\in A} d_Z(a,b)\}$
  - $d_H(A,B) = \inf\{r : A \subseteq B^r, B \subseteq A^r\}$



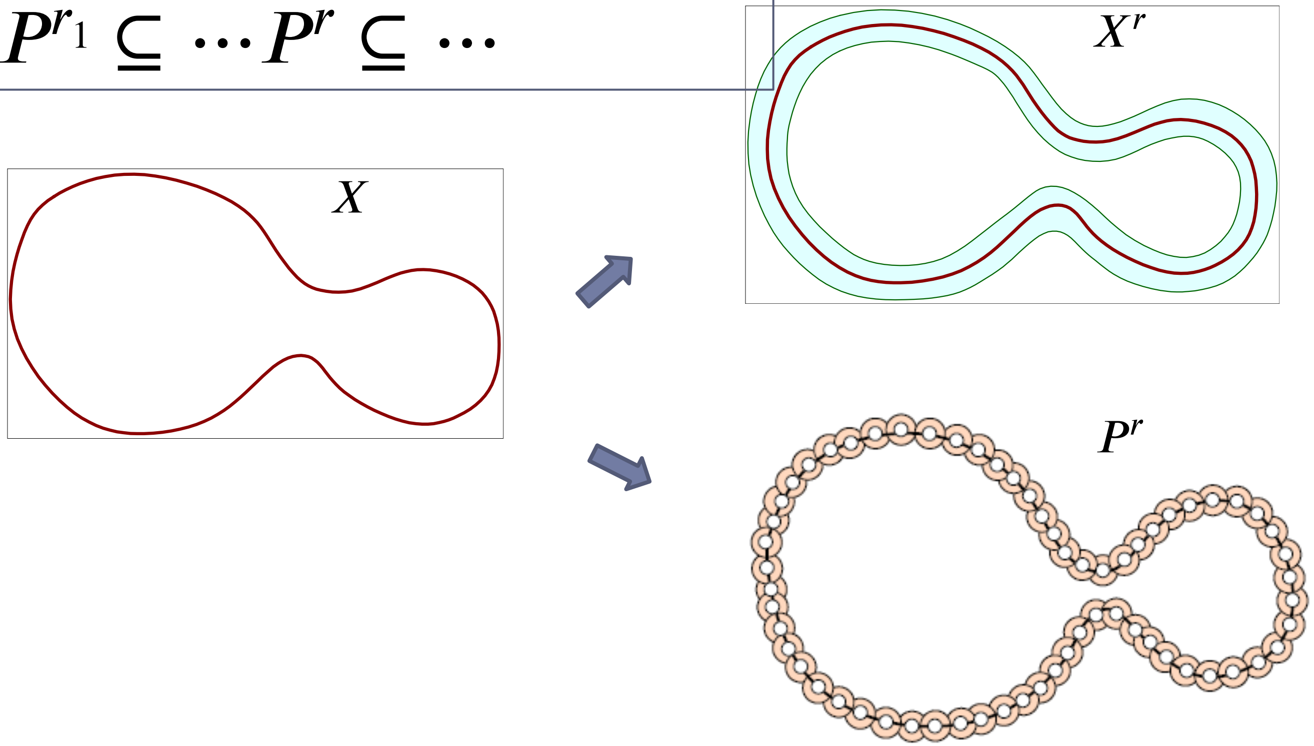
  - If $P \subseteq X$ then $d_H(P,H) = \inf\{r : X \subseteq P^r\}$

Hausdorff distance between subsets

- If $P \subseteq X$ satisfies that $d_H(P, X) = \inf\{r : X \subseteq P^r\} < \epsilon$

Target filtration (F_X):	$X^{r_0} \subseteq X^{r_1} \subseteq \cdots X^r \subseteq \cdots$

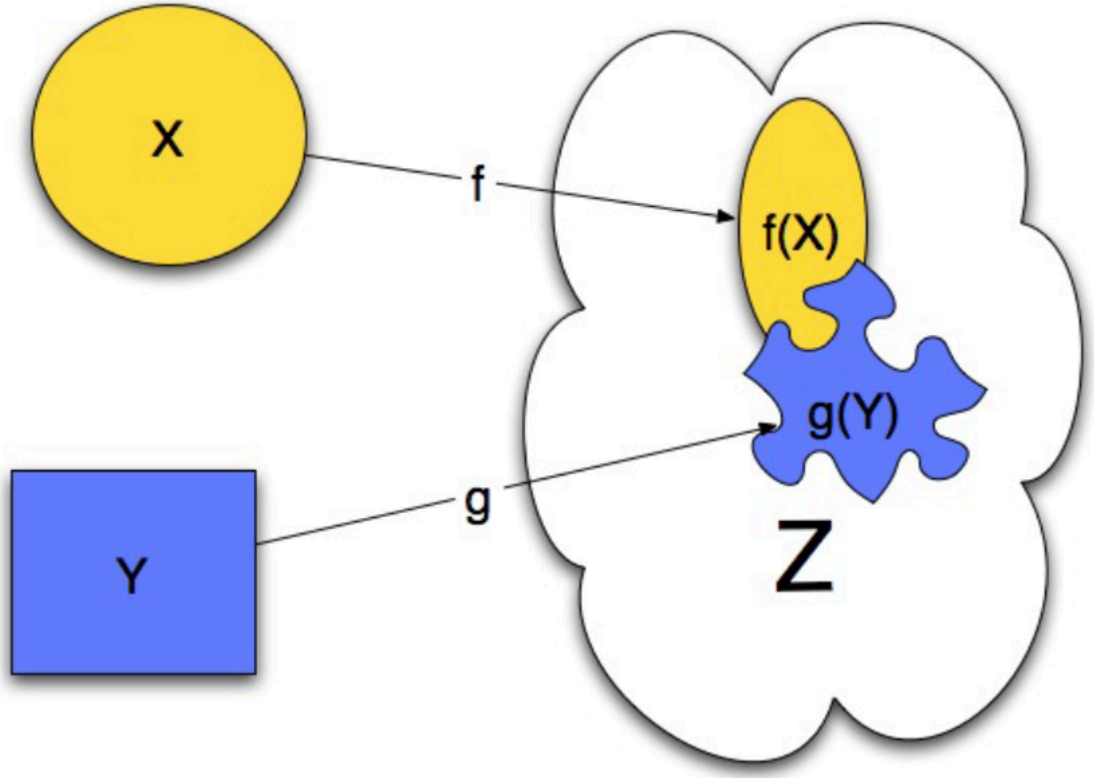
Intermediate filtration:	$P^{r_0} \subseteq P^{r_1} \subseteq \cdots P^r \subseteq \cdots$



- Note that
  - $P^r \subset X^{r+\epsilon}$
  - $X^r \subset P^{r+\epsilon}$
- So $d_I(P, F_X) \leq \epsilon$

Gromov-Hausdorff distance between metric spaces

- Given two metric spaces X and Y, the **Gromov-Hausdorff distance** between them is defined as $d_{GH}(X, Y) := \inf_{X \hookrightarrow Z, Y \hookrightarrow Z} d_H^Z(X, Y)$



Alternative formulation

Alternative formulation

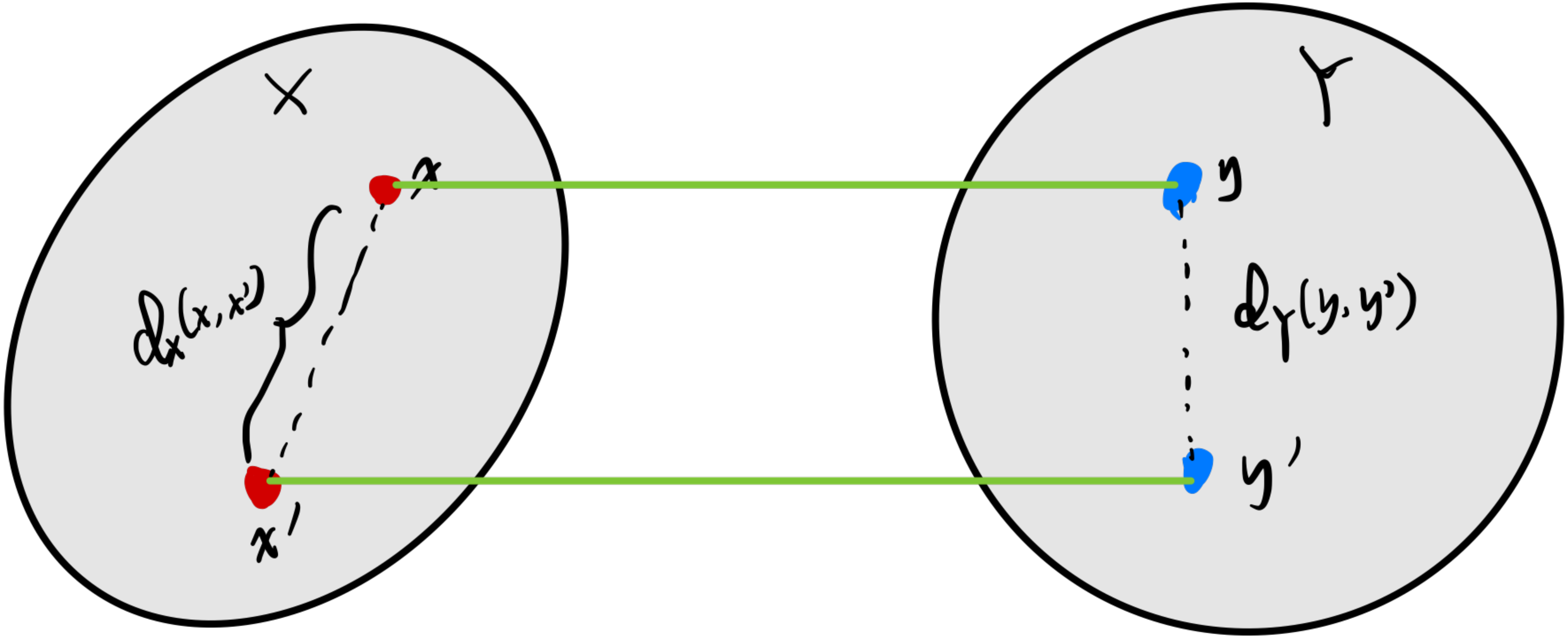
Definition 6.3 (Gromov-Hausdorff distance)**.** Given two metric spaces (X, d_X) and (Y, d_Y), a *correspondence* C is a subset $C \subseteq X \times Y$ so that (i) for every $x \in X$, there exists some $(x, y) \in C$; and (ii) for every $y' \in Y$, there exists some $(x', y') \in C$. The *distortion induced by* C is

$$\mathrm{distort}_C(X, Y) := \frac{1}{2} \sup_{(x,y),(x',y') \in C} |\mathsf{d}_X(x, x') - \mathsf{d}_Y(y, y')|.$$

The *Gromov-Hausdorff distance between* (X, d_X) *and* (Y, d_Y) is the smallest distortion possible by any correspondence; that is,

$$\mathsf{d}_{GH}(X, Y) := \inf_{C \subseteq X \times Y} \mathrm{distort}_C(X, Y).$$

Alternative formulation



Stability of persistence diagrams - metric spaces

- Given two metric spaces X and Y, one has that
- $d_I(VR(X), VR(Y)) \leq d_{GH}(X, Y)$

- Therefore
- $d_B(Dgm_*(VR(X)), Dgm_*(VR(Y))) \leq d_I(VR(X), VR(Y)) \leq d_{GH}(X, Y)$

FIN