

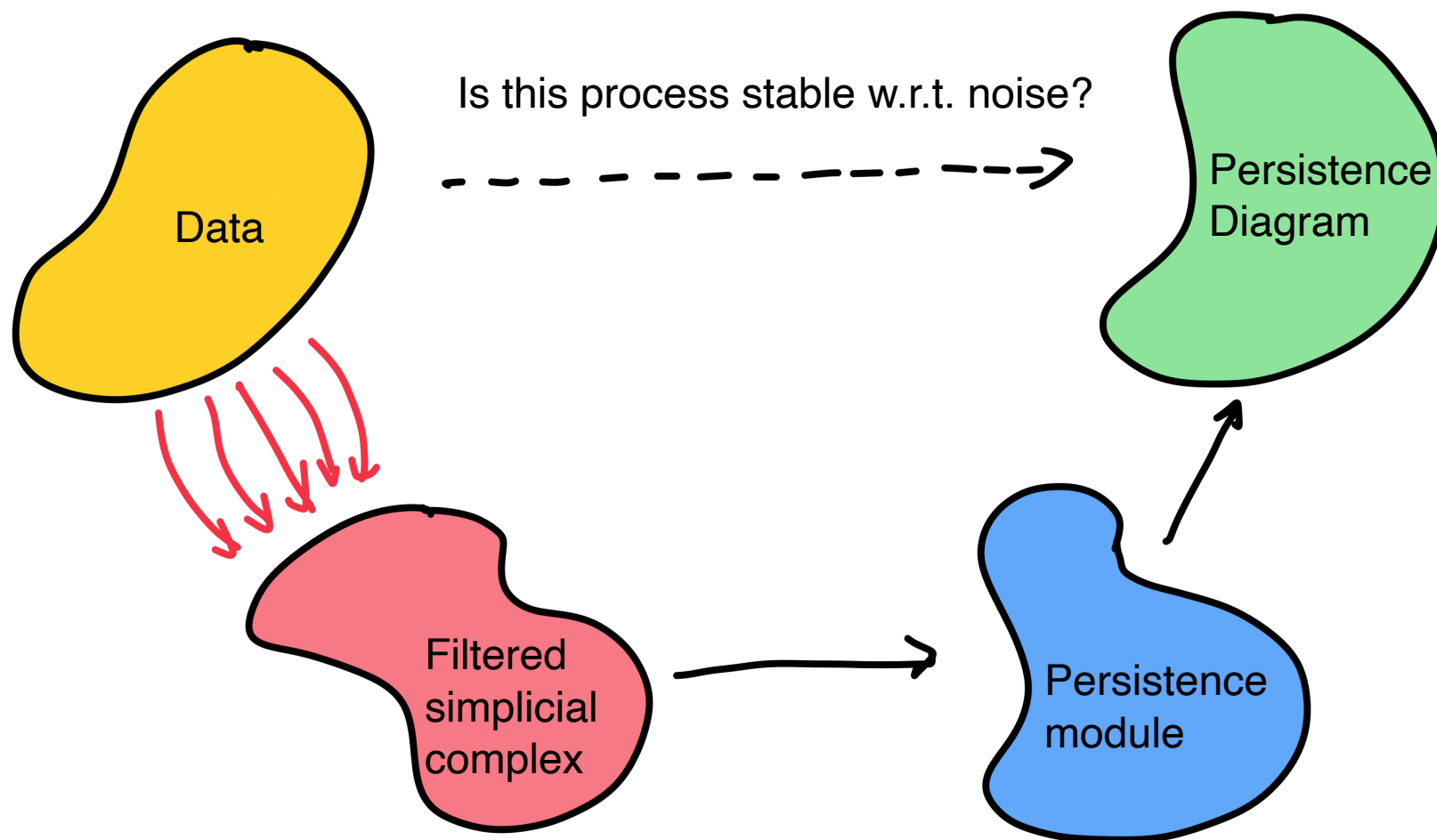
DSC 214

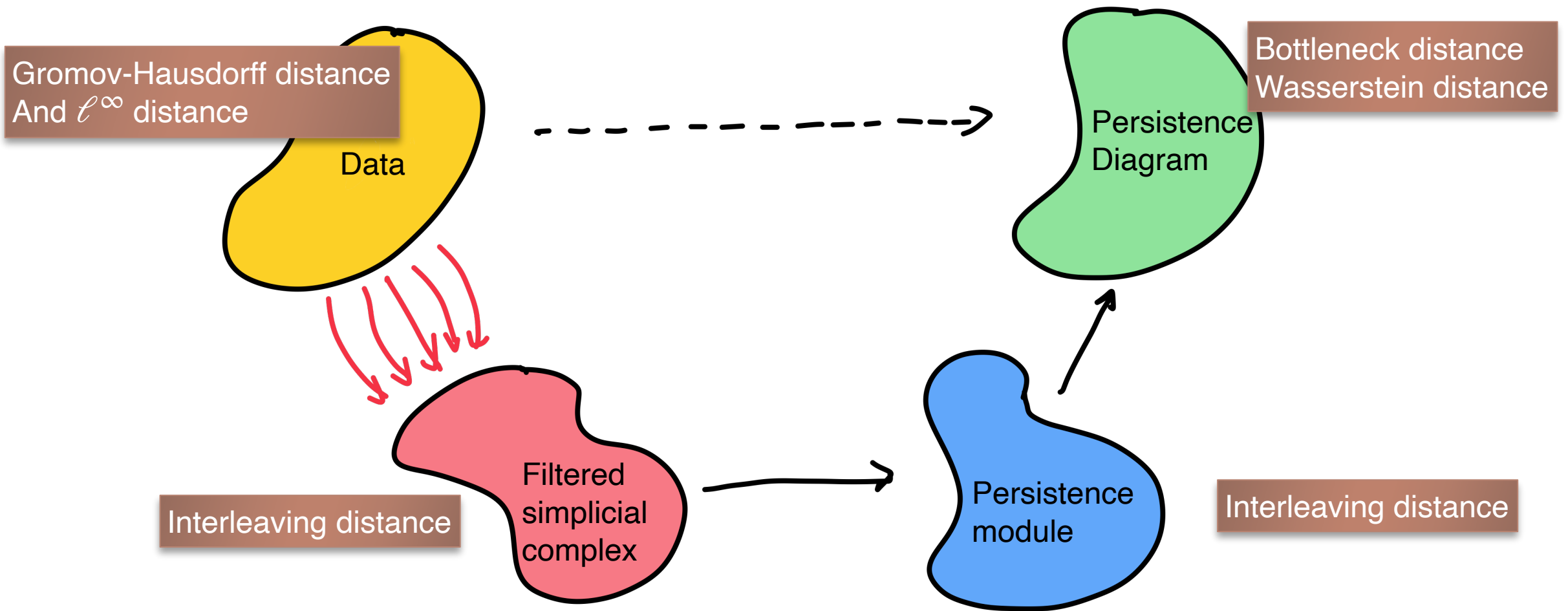
Topological Data Analysis

Topic 5: Stability of PD

Instructor: Zhengchao Wan

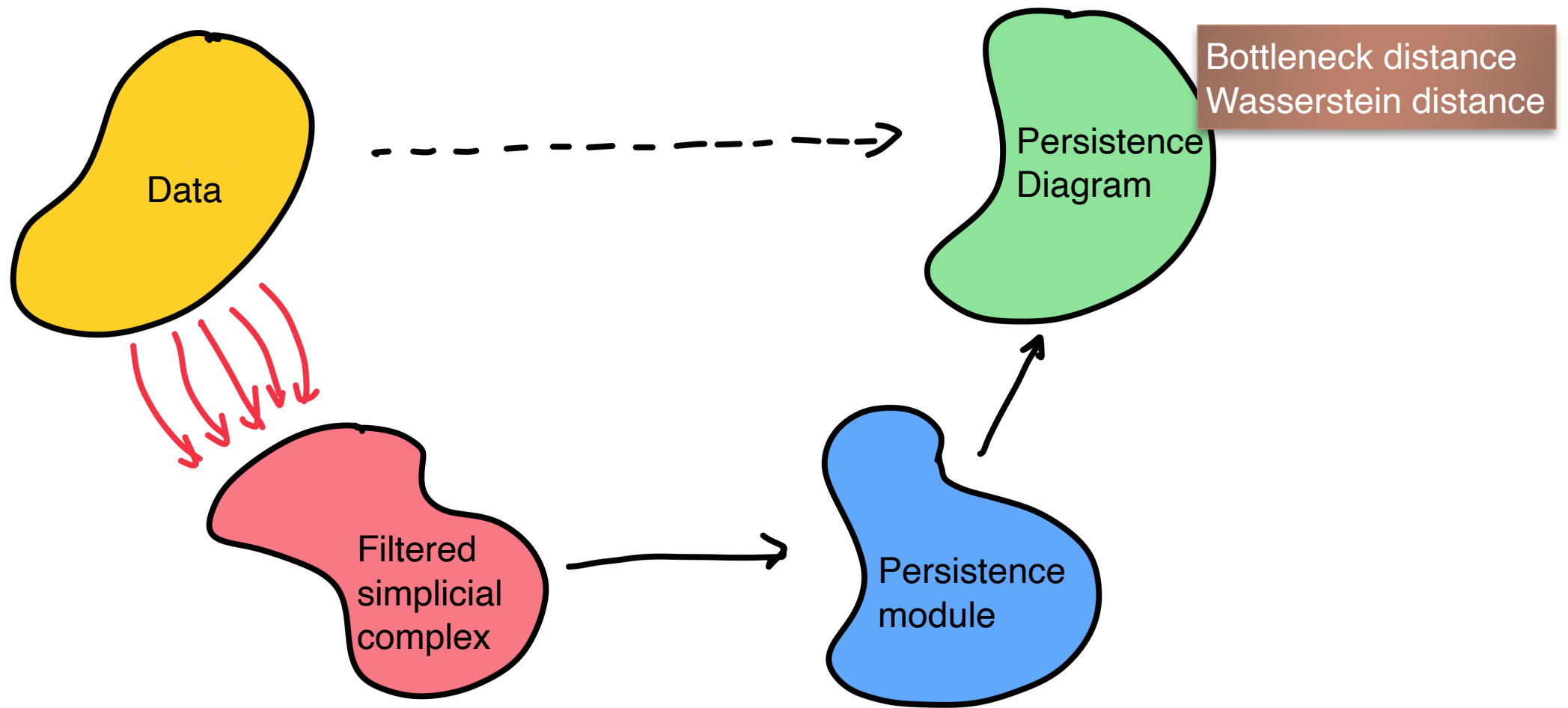
Persistence-based Framework





Section 1:

Distances between persistence Diagrams

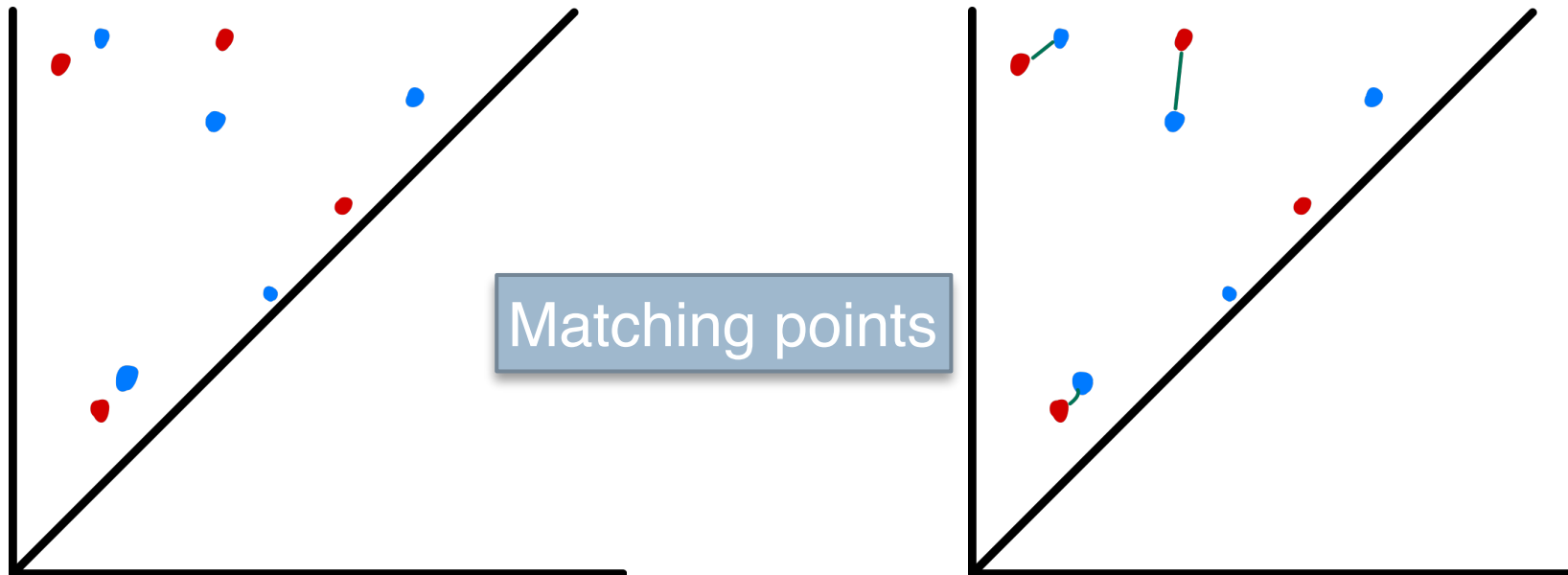


Recall: Persistence Diagram

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

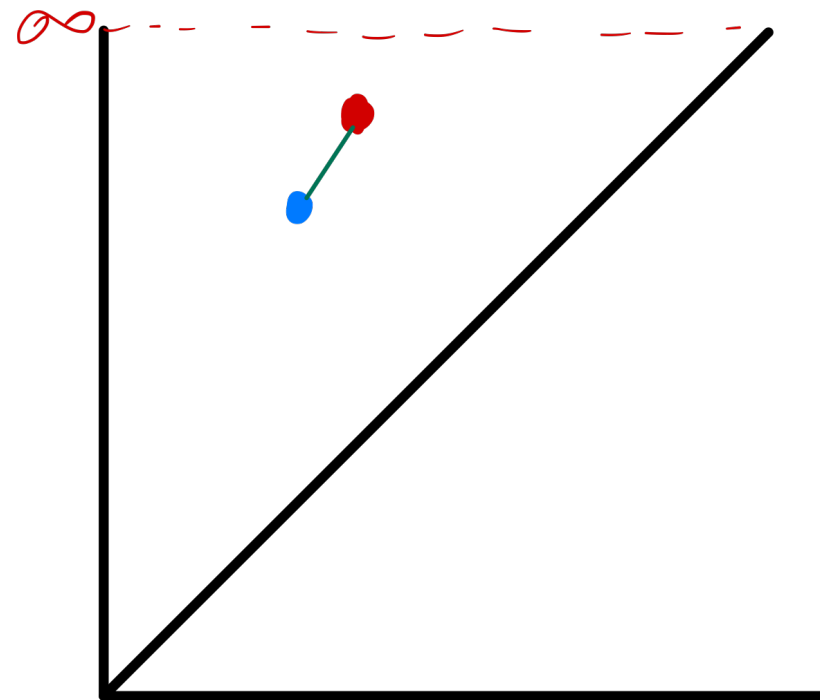
Persistence Diagram

- ▶ Any finite multiset $D = \{(b_j, d_j)\}_{j=1,\dots,M} \subseteq (\mathbb{R} \cup \infty)^2$ is called the **persistence diagram**, where $0 \leq b_i < d_i \leq \infty$ for each $i = 1, \dots, M$
- ▶ How do you compare two different persistence diagrams?



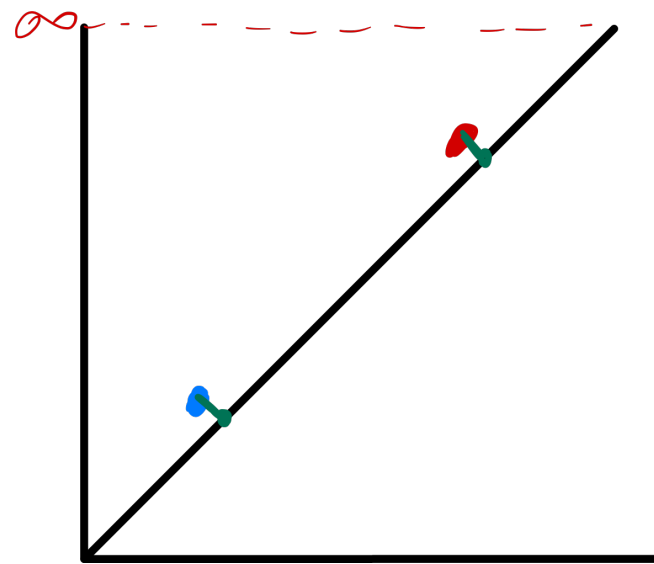
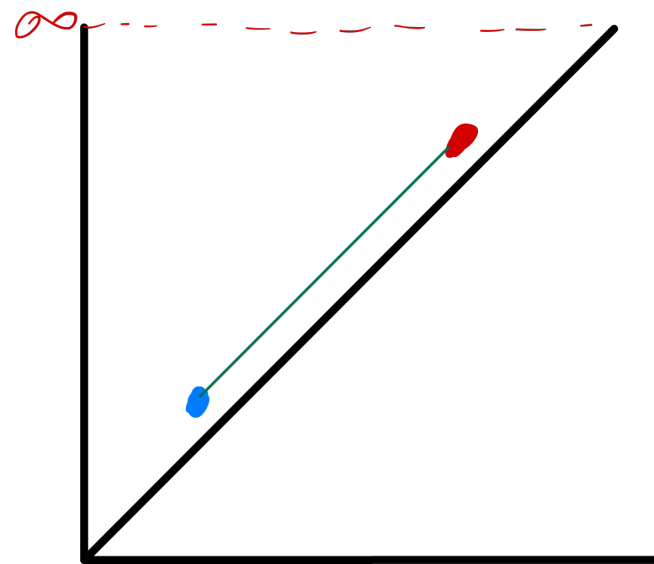
Motivating examples

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



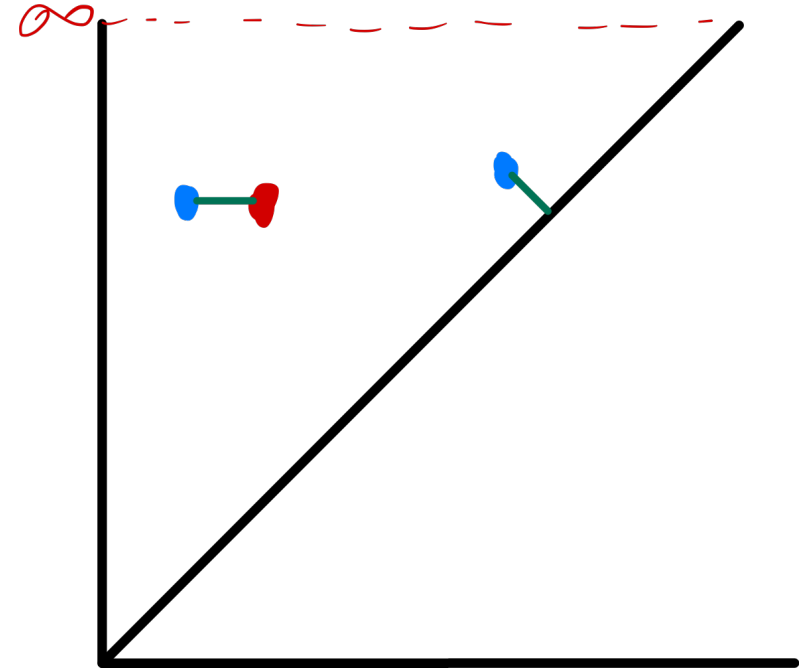
Motivating examples

- ▶ 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 $\|p - \Delta\|_\infty = \frac{|b - d|}{2}$
- ▶ 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

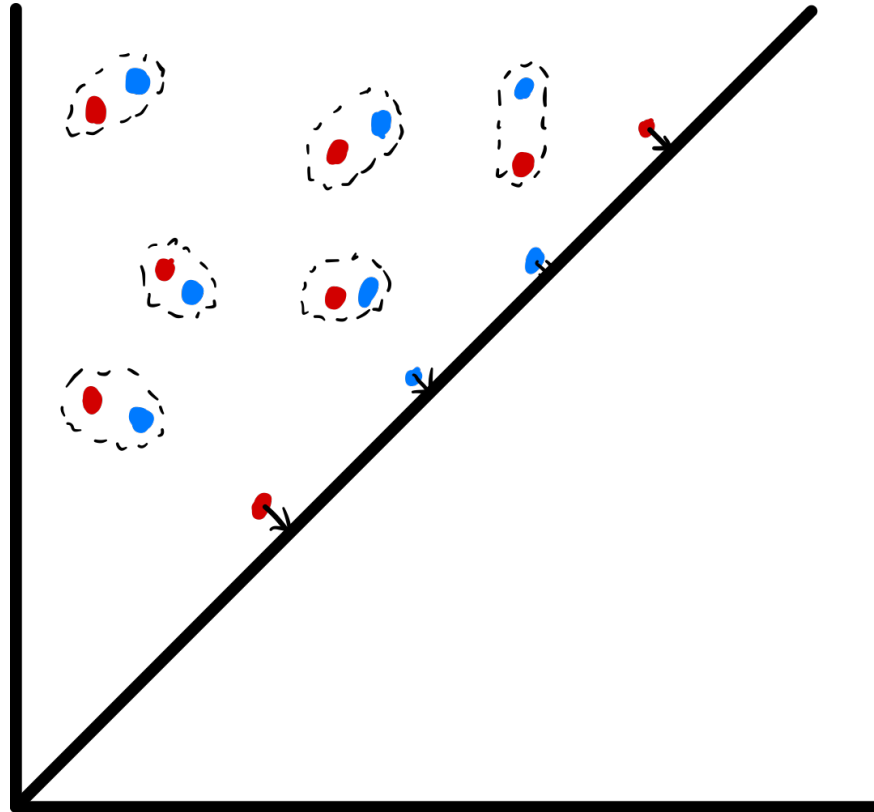
- ▶ Given two persistence-diagrams (multiset of points in $(\mathbb{R} \cup \{\infty\})^2$)
 - ▶ $D_1 = \{p_1, p_2, \dots, p_s\}$ and $D_2 = \{q_1, q_2, \dots, q_t\}$
- ▶ A **partial-matching (partial bijection)** between D_1 and D_2 is
 - ▶ $M \subseteq D_1 \times D_2$ s.t.
 - ▶ $\forall p \in D_1, \exists$ 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 $\forall (p, q) \in M$ (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 $(\mathbb{R} \cup \{\infty\})^2$)
 - ▶ $D_1 = \{p_1, p_2, \dots, p_s\}$ and $D_2 = \{q_1, q_2, \dots, q_t\}$
- ▶ A **partial-matching (partial bijection)** between D_1 and D_2 is
 - ▶ $M \subseteq D_1 \times D_2$ 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
 - ▶ $d_B(D_1, D_2) = \min_M \text{cost}(M)$



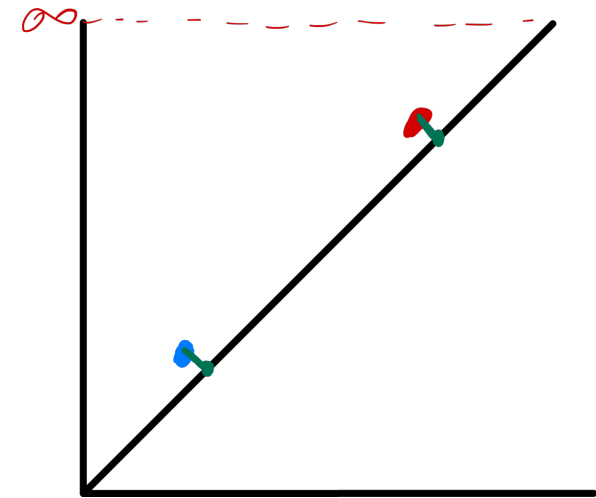
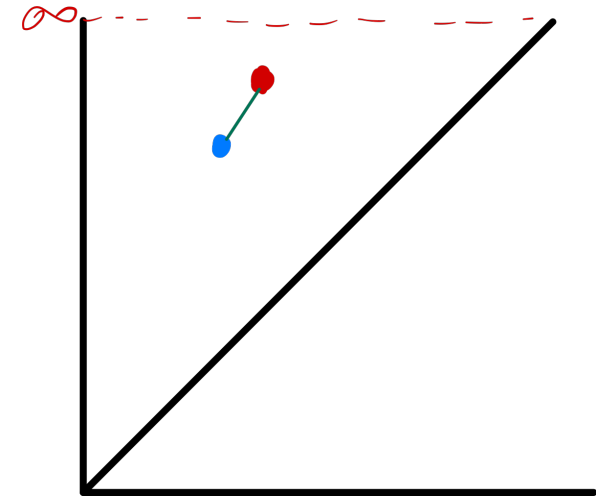
Bottleneck distance between 1-point PDs

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

$$\text{cost}(M) = \max(\|p - \Delta\|_\infty, \|q - \Delta\|_\infty)$$

- ▶ In conclusion,

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



Alternative formulation

- ▶ Given two persistence-diagrams (multiset of points in $(\mathbb{R} \cup \{\infty\})^2$)
 - ▶ $D_1 = \{p_1, p_2, \dots, p_s\}$ and $D_2 = \{q_1, q_2, \dots, q_t\}$
- ▶ Augment $\bar{D}_1 = D_1 \cup \Delta$ and $\bar{D}_2 = D_2 \cup \Delta$
 - ▶ where $\Delta = \{(x, x) \in \mathbb{R}^2\}$ 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
 - ▶ $d_B(D_1, D_2) := \inf_{\bar{M}} \max_{(x,y) \in \bar{M}} ||x - y||_\infty$

p -th Wasserstein distance

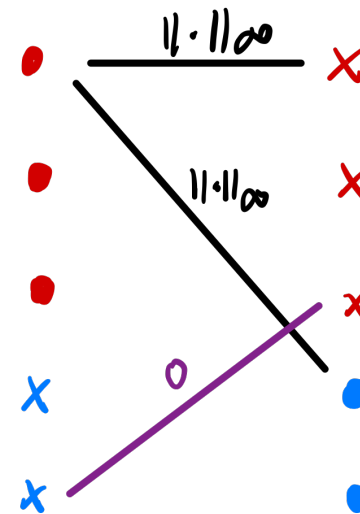
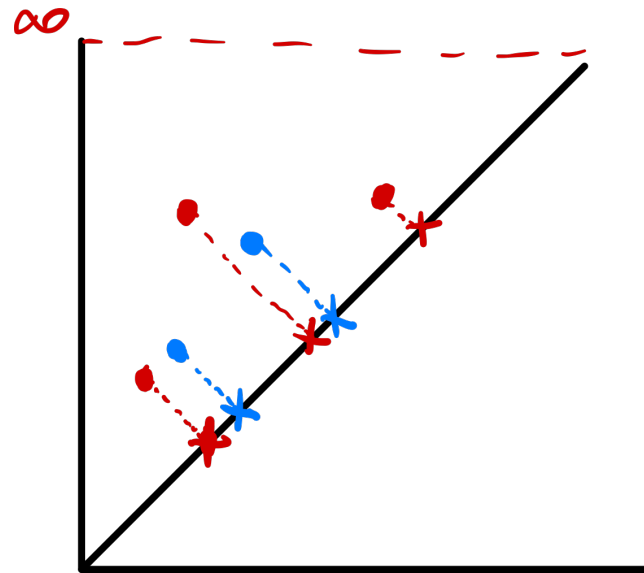
- ▶ Given two persistence-diagrams (multiset of points in $(\mathbb{R} \cup \{\infty\})^2$)
 - ▶ $D_1 = \{p_1, p_2, \dots, p_s\}$ and $D_2 = \{q_1, q_2, \dots, q_t\}$
- ▶ Augment $\bar{D}_1 = D_1 \cup \Delta$ and $\bar{D}_2 = D_2 \cup \Delta$
 - ▶ where $\Delta = \{(x, x) \in \mathbb{R}^2\}$ is diagonal 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 between D_1 and D_2
 - ▶
$$d_{W,p}(D_1, D_2) := \inf_{\bar{M}} \left[\sum_{(x,y) \in \bar{M}} ||x - y||_{\infty}^p \right]^{\frac{1}{p}}$$
 - ▶ $d_{W,\infty}(D_1, D_2) = d_B(D_1, D_2)$

Bottleneck (Wasserstein) distance vs Matching Problem

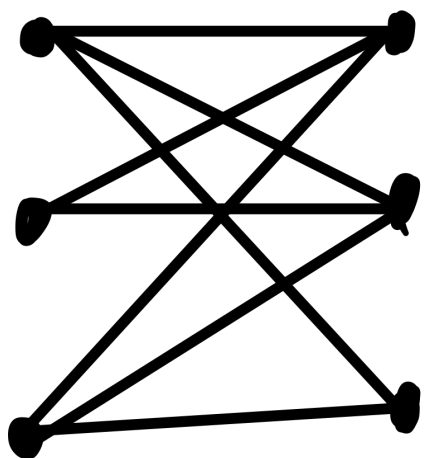
- ▶ Let $D_1 = \{x_1, \dots, x_n\}$ and $D_2 = \{y_1, \dots, y_m\}$ be two persistence diagrams
- ▶ $D'_1 = \{x'_1, \dots, x'_n\}$: projections of x_i on to $x'_i \in \Delta$
- ▶ Same for D'_2
- ▶ $U = D_1 \cup D'_2$ and $V = D'_1 \cup D_2$
- ▶ Construct a fully connected bipartite graph $G = (U \cup V, E, w)$
- ▶ $w(u, v) = \begin{cases} \|u - v\|_\infty, & u \in D_1 \text{ or } v \in D_2 \\ 0, & \text{otherwise} \end{cases}$

Bottleneck (Wasserstein) distance vs Matching Problem

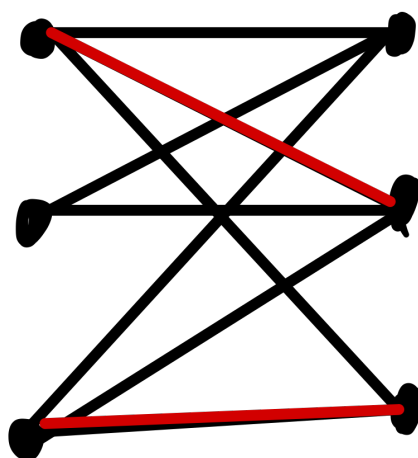
- ▶ $U = D_1 \cup D_2'$ and $V = D_1' \cup D_2$
- ▶ Construct a fully connected bipartite graph $G = (U \cup V, E, w)$
- ▶ $w(u, v) = \begin{cases} \|u - v\|_\infty, & u \in D_1 \text{ or } v \in D_2 \\ 0, & \text{otherwise} \end{cases}$



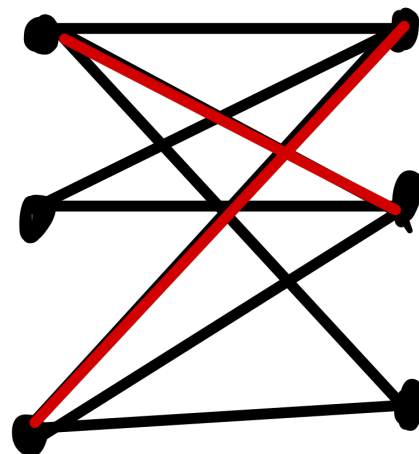
Matching



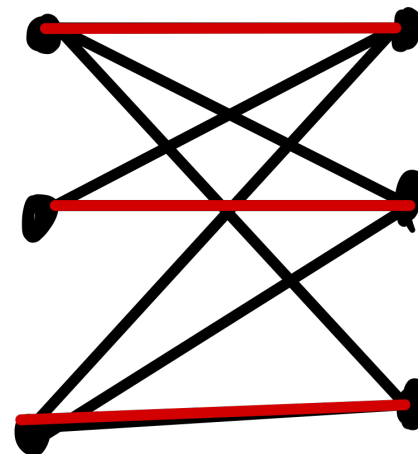
A bipartite graph



Matching



Maximal matching



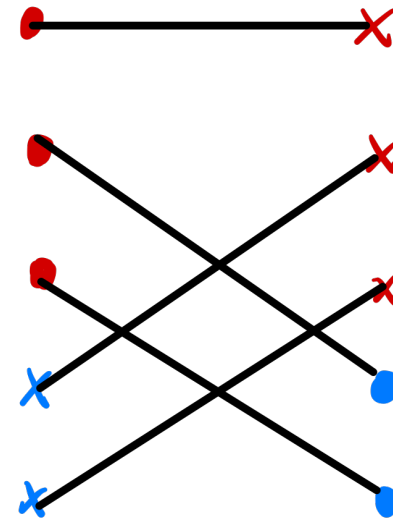
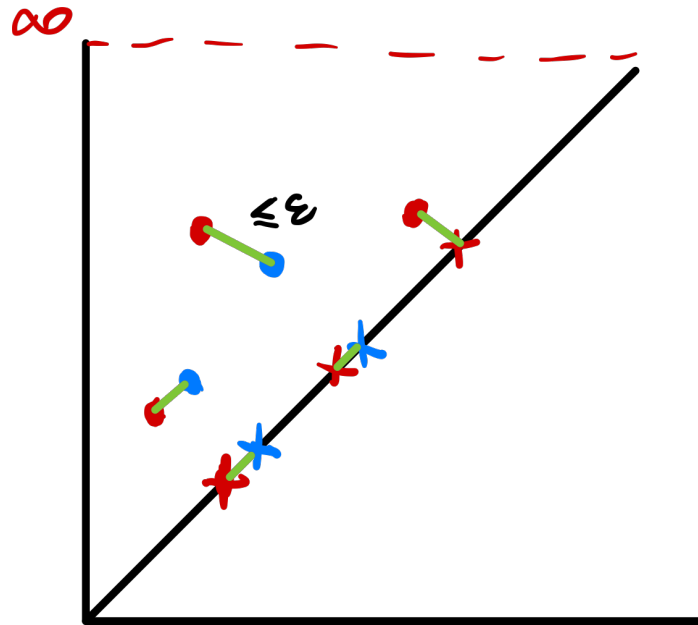
Perfect matching

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 $> \epsilon$

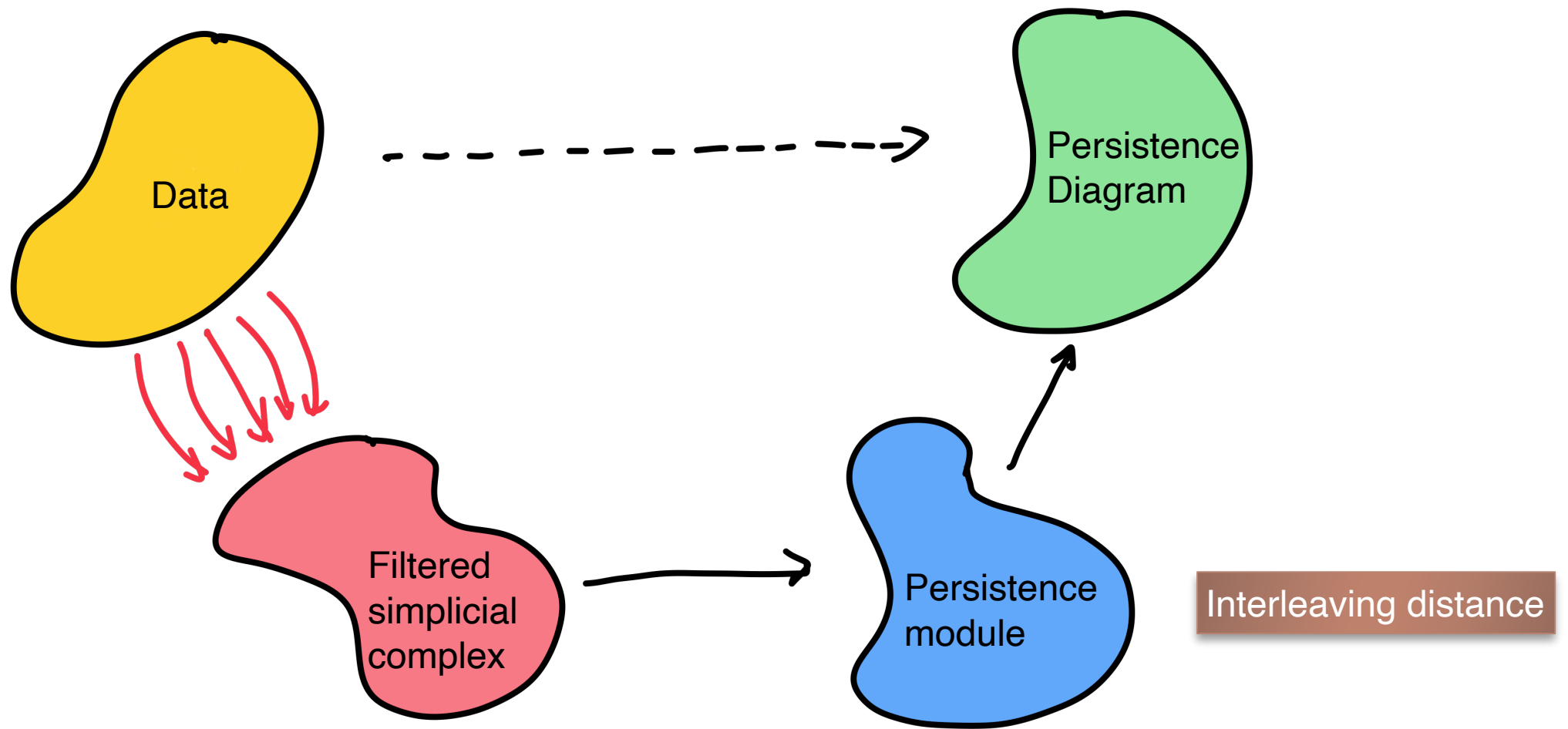
▶ (Reduction Lemma)

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

- ▶ 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



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

Intuition

U :

$$\dots \rightarrow U_a \rightarrow \dots \rightarrow U_b \rightarrow \dots \rightarrow \dots$$

V :

$$\dots \rightarrow V_a \rightarrow \dots \rightarrow V_b \rightarrow \dots \rightarrow \dots$$

Intuition

- Isomorphic persistence modules

$$\begin{array}{c} \boxed{U:} \quad \boxed{\cdots \rightarrow U_a \rightarrow \cdots \rightarrow U_b \rightarrow \cdots \rightarrow \cdots} \\ \quad \quad \quad \begin{array}{ccc} \cdots & \updownarrow \cong & \cdots & \updownarrow \cong & \cdots \end{array} \\ \boxed{V:} \quad \boxed{\cdots \rightarrow V_a \rightarrow \cdots \rightarrow V_b \rightarrow \cdots \rightarrow \cdots} \end{array}$$

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

Intuition

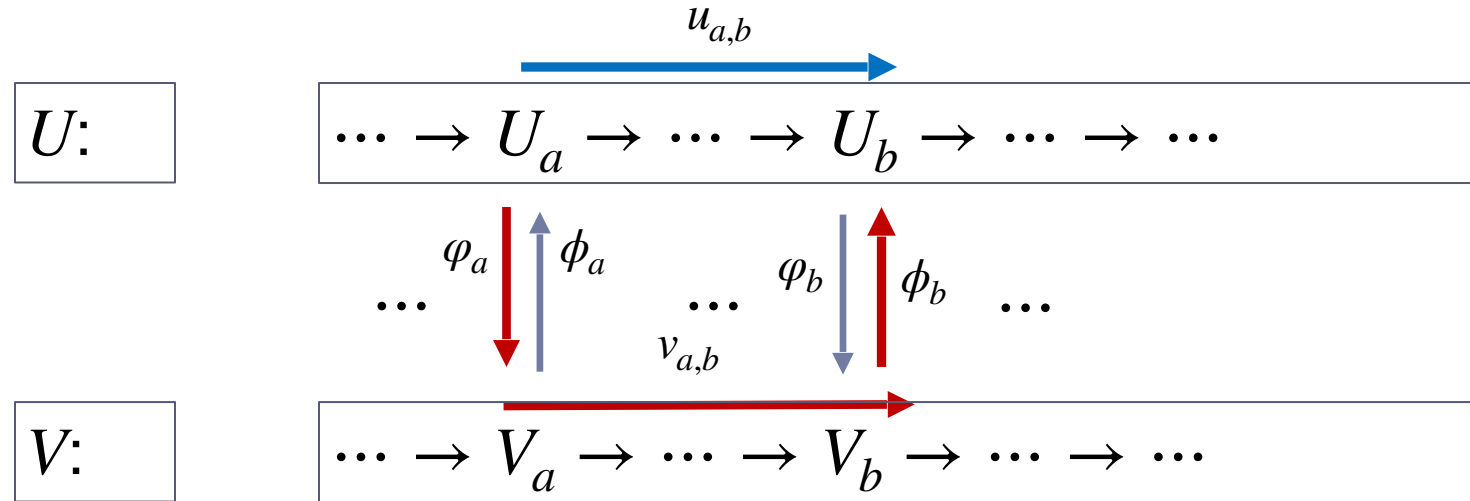
- ▶ Isomorphic persistence modules

$U:$	$\cdots \rightarrow U_a \rightarrow \cdots \rightarrow U_b \rightarrow \cdots \rightarrow \cdots$
	$\begin{array}{ccccc} \cdots & \varphi_a \downarrow \uparrow \phi_a & \cdots & \varphi_b \downarrow \uparrow \phi_b & \cdots \end{array}$
$V:$	$\cdots \rightarrow V_a \rightarrow \cdots \rightarrow V_b \rightarrow \cdots \rightarrow \cdots$

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

Intuition

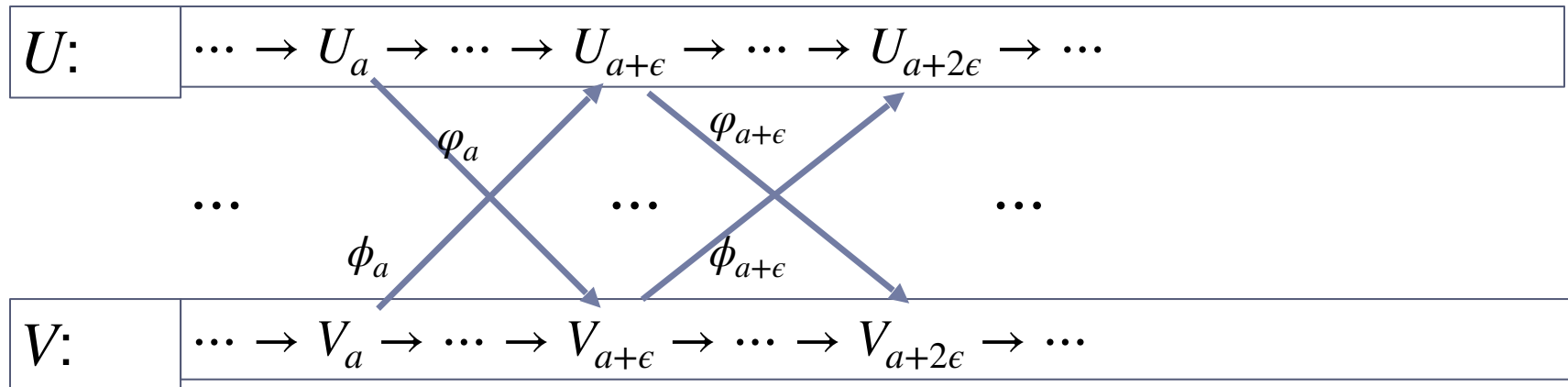
- Isomorphic persistence modules



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

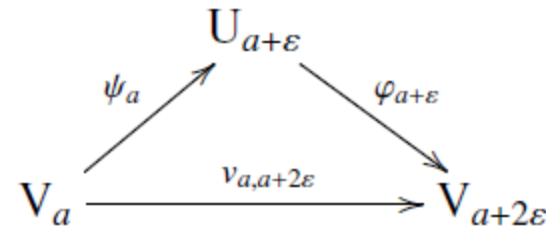
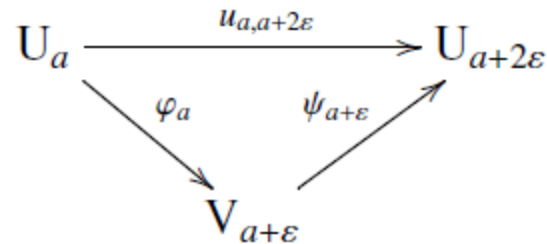
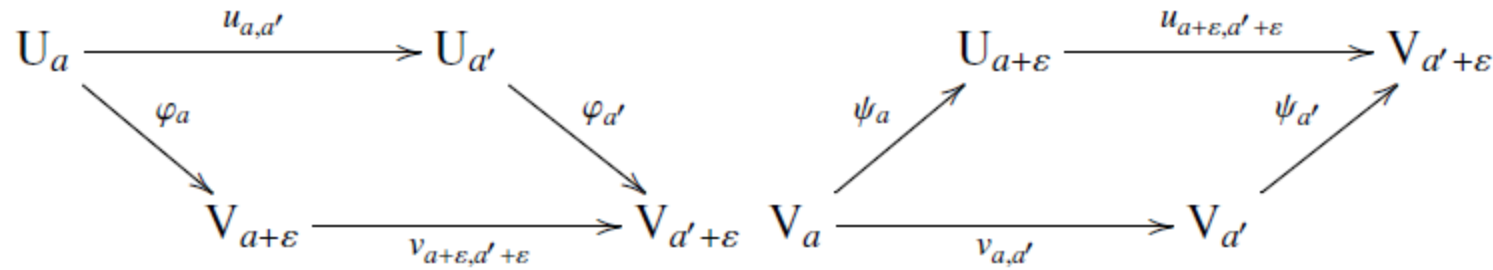
ϵ -Interleaving

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



ϵ -Interleaving

- ▶ U and V are ϵ -interleaving if there exists maps
 - ▶ $\varphi_a : U_a \rightarrow V_{a+\epsilon}$ and $\phi_a : V_a \rightarrow U_{a+\epsilon}$ 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)



ϵ -Interleaving

- ▶ U and V are ϵ -interleaving if there exists maps
 - ▶ $\varphi_a : U_a \rightarrow V_{a+\epsilon}$ and $\phi_a : V_a \rightarrow U_{a+\epsilon}$ 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 persistence 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 $I[10.1,10.2)$

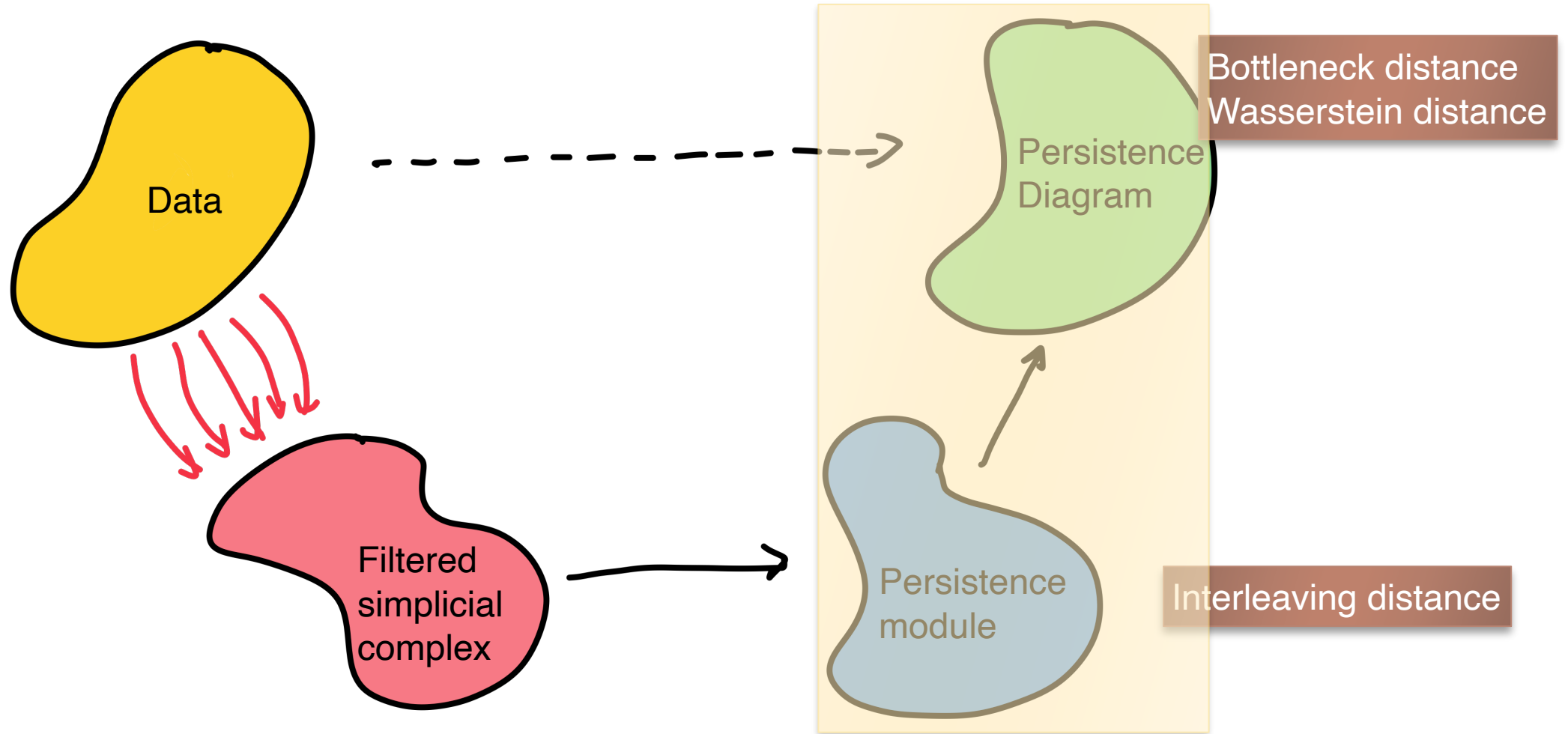
Interleaving distance between interval modules

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

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

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

Bottleneck distance vs interleaving distance



Recall: Finitely presented filtration

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

Finitely presented persistence modules

- ▶ A persistence module $V = \{V_t\}$ is called **finitely represented** if
 - ▶ There exist $0 = t_0 < t_1 < \dots < t_n$ such that
 - ▶ $\varphi : V_t \rightarrow V_{t'}$ is an isomorphism, $\forall t_i \leq t < t' < t_{i+1}$ and $i = 0, \dots, n$
($t_{n+1} := \infty$)

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

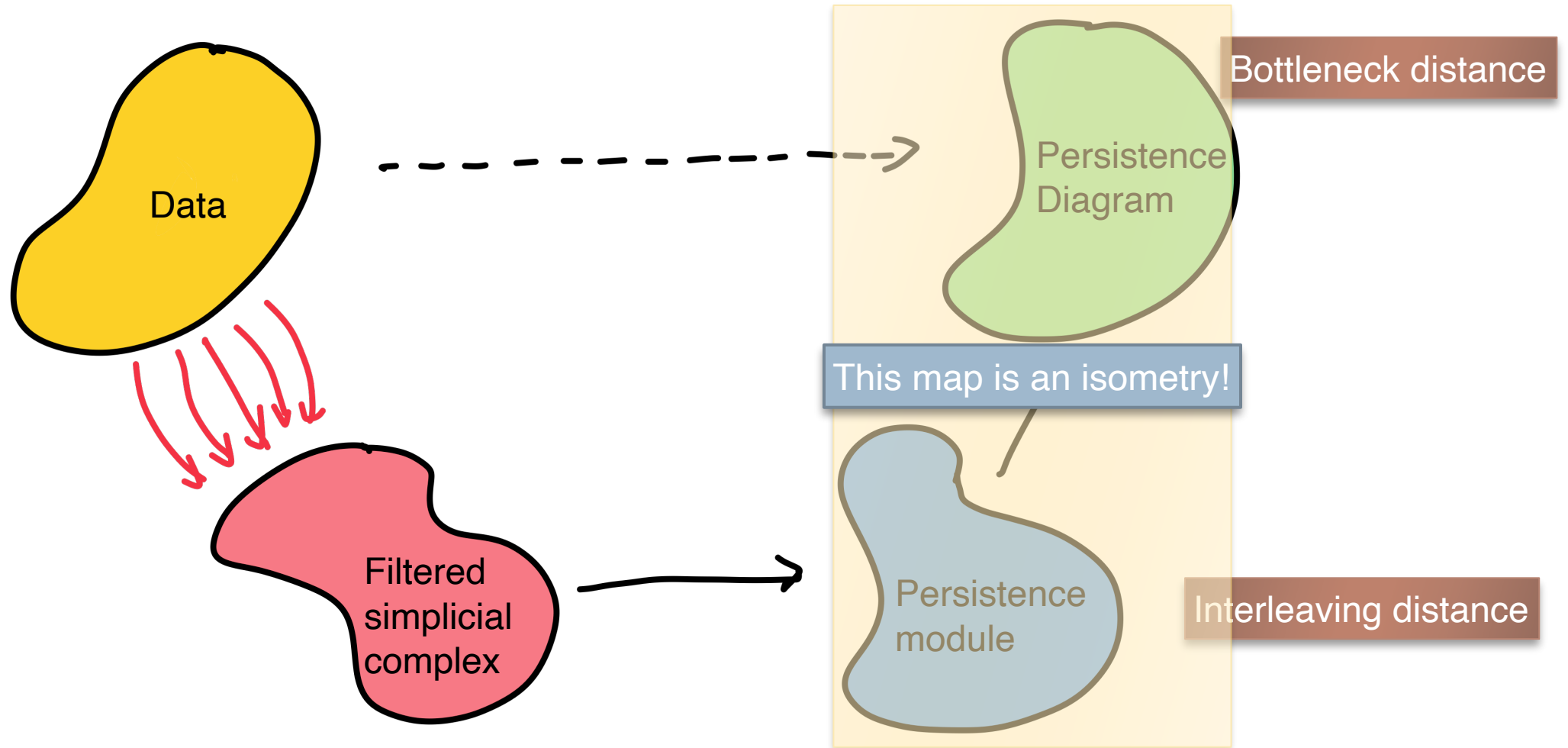
► $d_I(V, U) = \inf_{\epsilon \geq 0} \{V \text{ and } U \text{ are } \epsilon\text{-interleaved}\}$

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

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

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

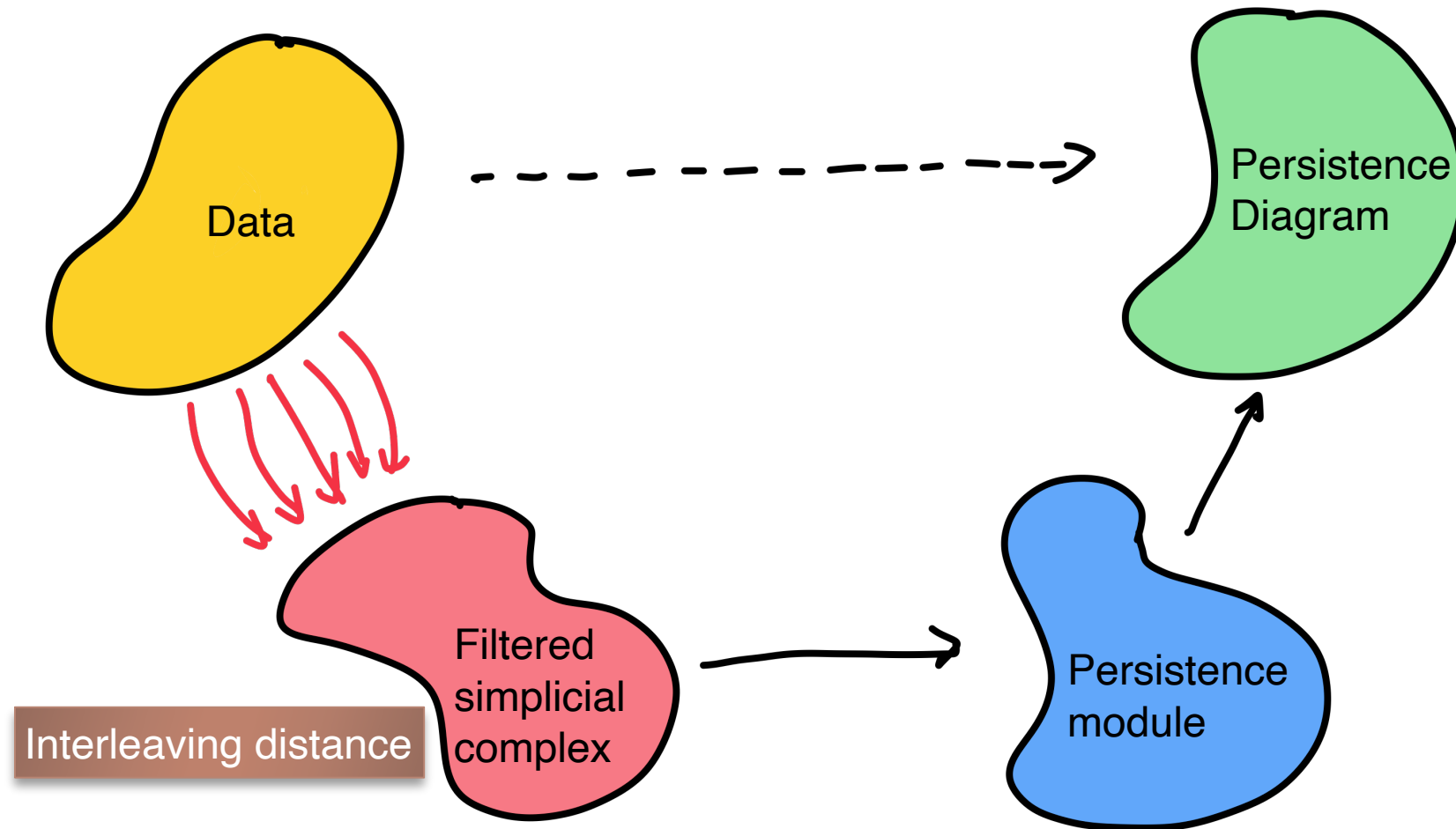
Bottleneck distance vs interleaving distance



Section 3:

Interleaving distance between filtrations

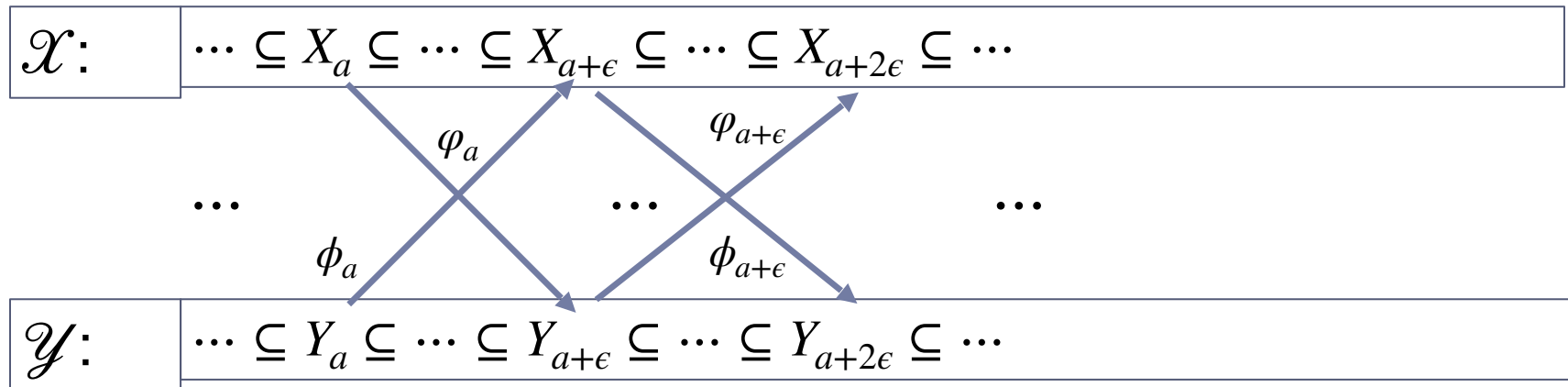
Bottleneck distance vs interleaving distance



Filtered simplicial complexes over the same vertex set

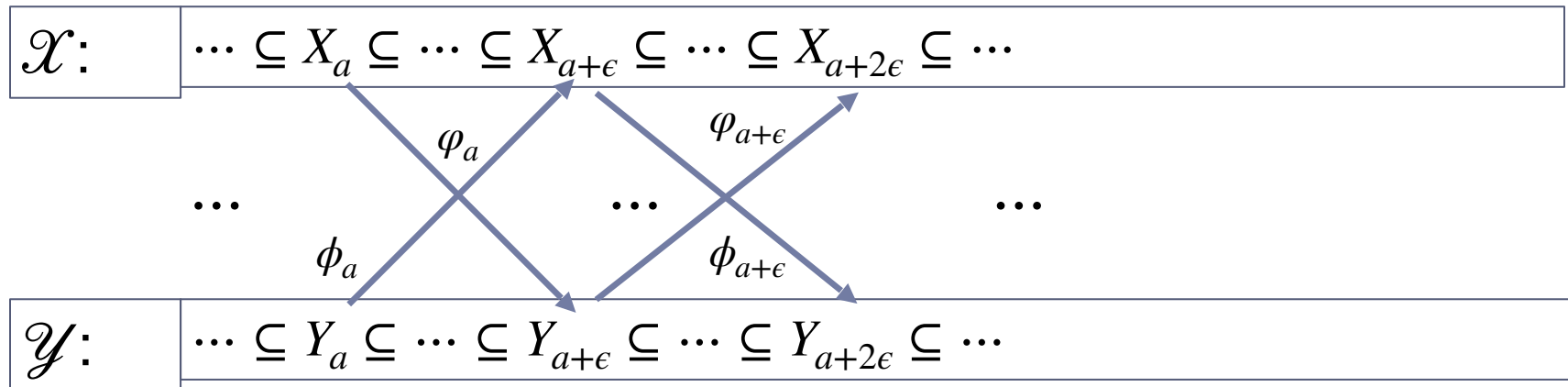
- ▶ Given two simplicial filtrations \mathcal{X} and \mathcal{Y} over the “same” vertex set V
- ▶ We say they are ϵ -interleaved if there exist **inclusion** maps

$\varphi_a : X_a \hookrightarrow Y_{a+\epsilon}$ and $\phi_{a'} : Y_{a'} \hookrightarrow X_{a'+\epsilon}$ such that the following diagram commutes



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

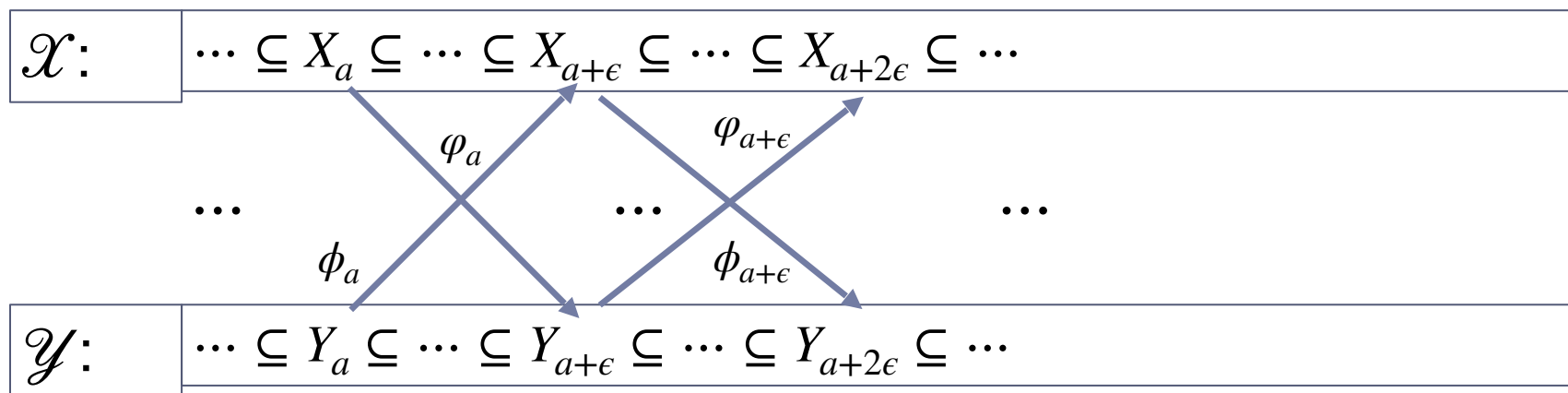


A first Interleaving distance

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



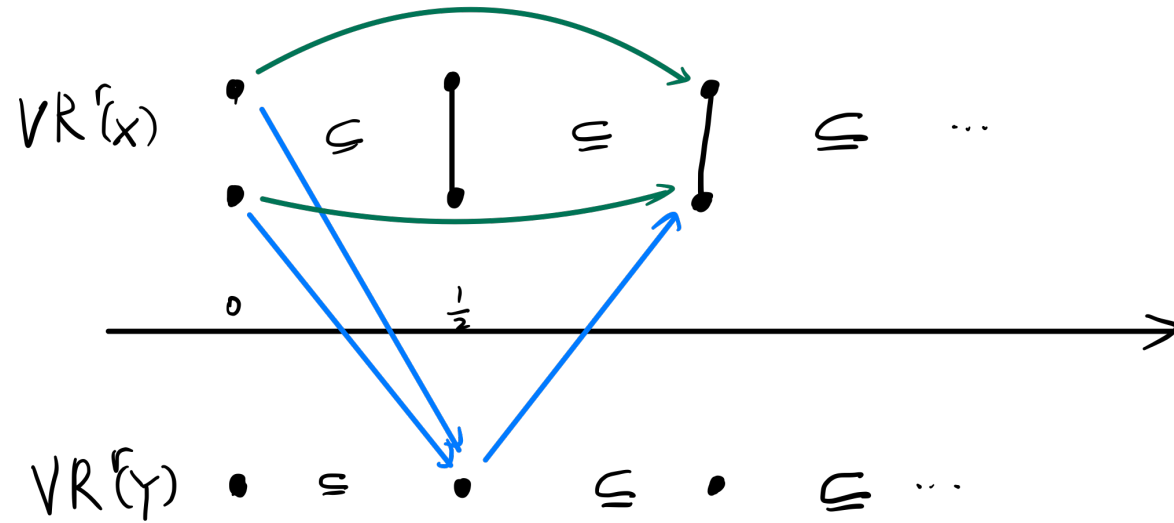
X



Y



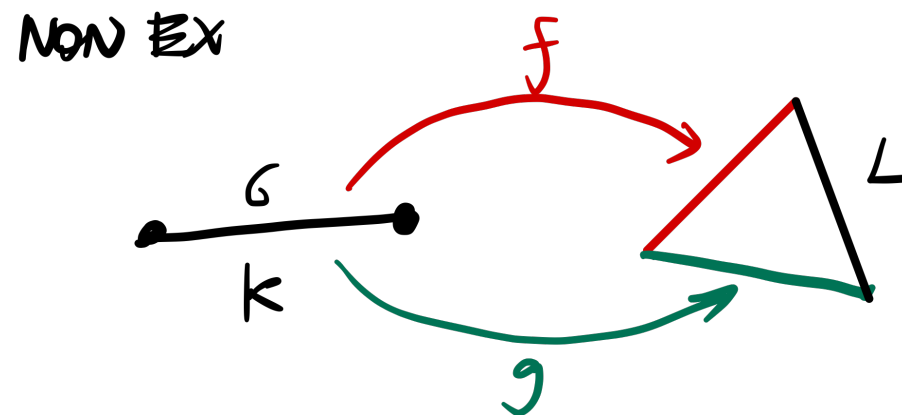
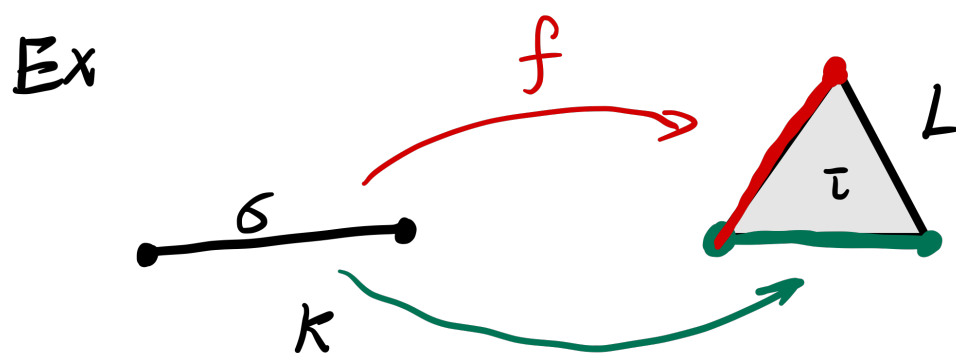
► Rips filtration



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

Contiguity

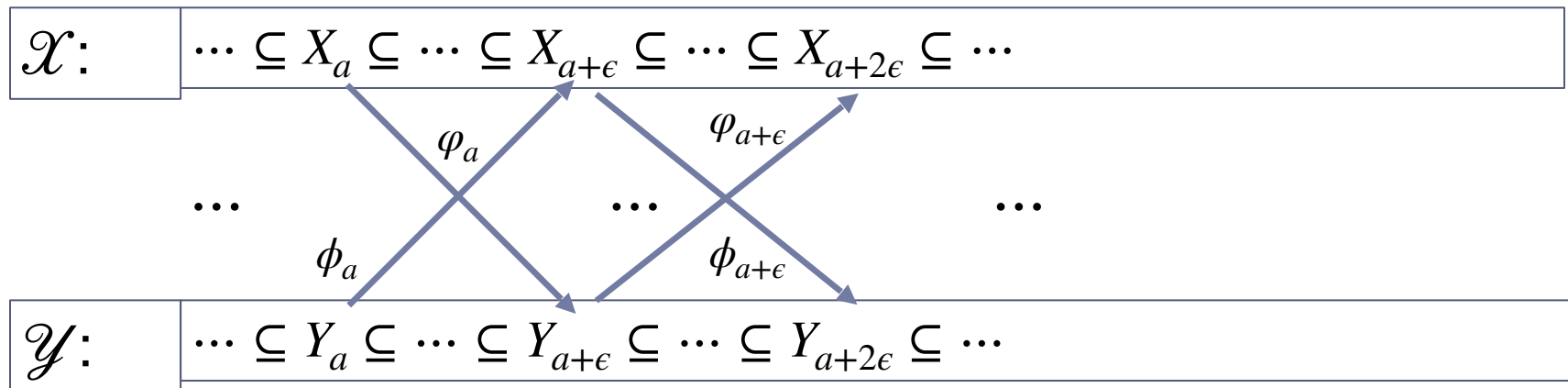
- Two simplicial maps $f, g : K \rightarrow L$ are contiguous if for any $\sigma \in \Sigma_K$ there exists a simplex $\tau \in \Sigma_L$ such that $f(\sigma) \cup g(\sigma) \subseteq \tau$



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

General filtered simplicial complexes

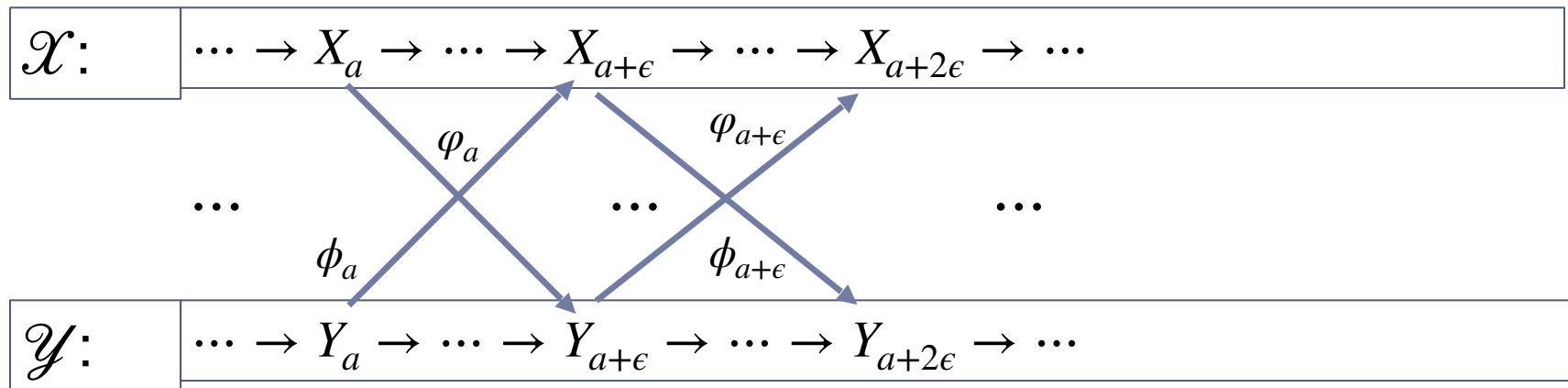
- ▶ Given two simplicial filtrations \mathcal{X} and \mathcal{Y}
- ▶ We say they are ϵ -interleaved if there exist **simplicial** maps $\varphi_a : X_a \rightarrow Y_{a+\epsilon}$ and $\phi_{a'} : Y_{a'} \rightarrow X_{a'+\epsilon}$ such that the following diagram commutes up to **contiguity**



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

A generalization to simplicial towers

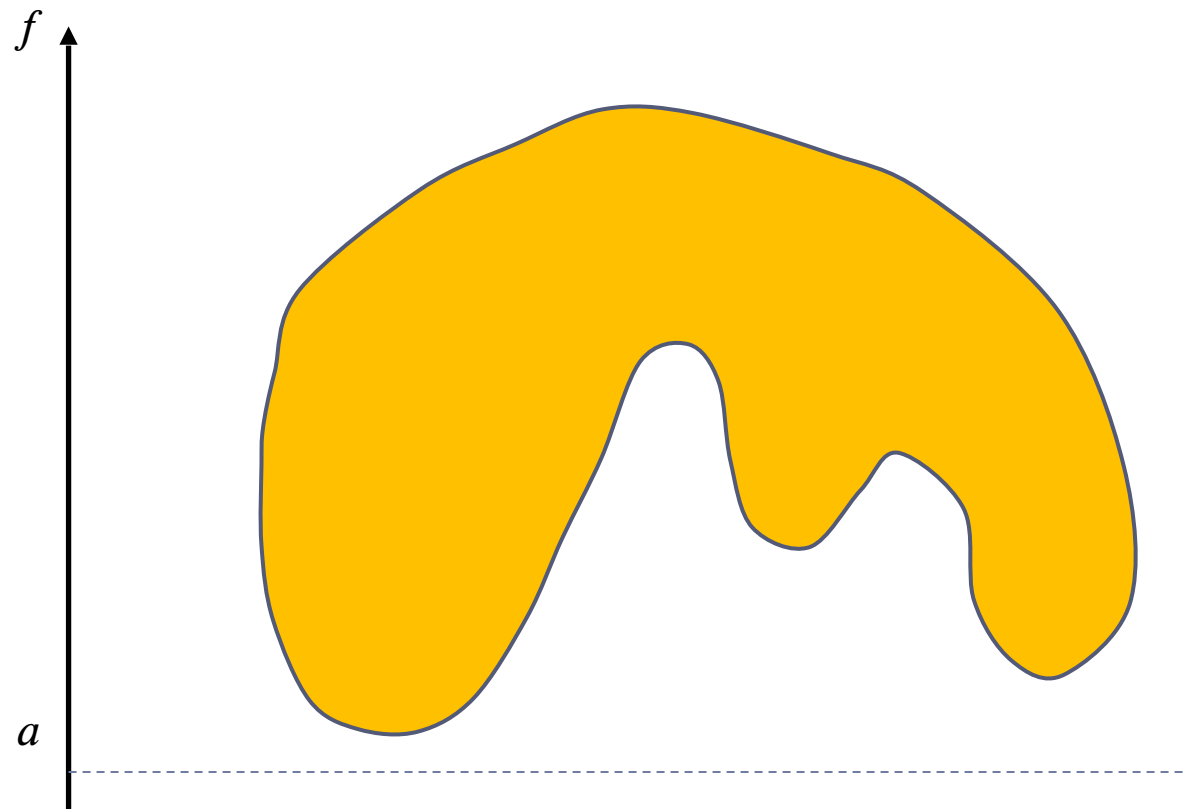
- A simplicial tower $\mathcal{X} : \dots \rightarrow X_a \rightarrow \dots \rightarrow X_{a+\epsilon} \xrightarrow{\text{Simplicial maps}} \dots \rightarrow X_{a+2\epsilon} \rightarrow \dots$
- We say two simplicial towers \mathcal{X} and \mathcal{Y} are ϵ -interleaved if there exist **simplicial**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



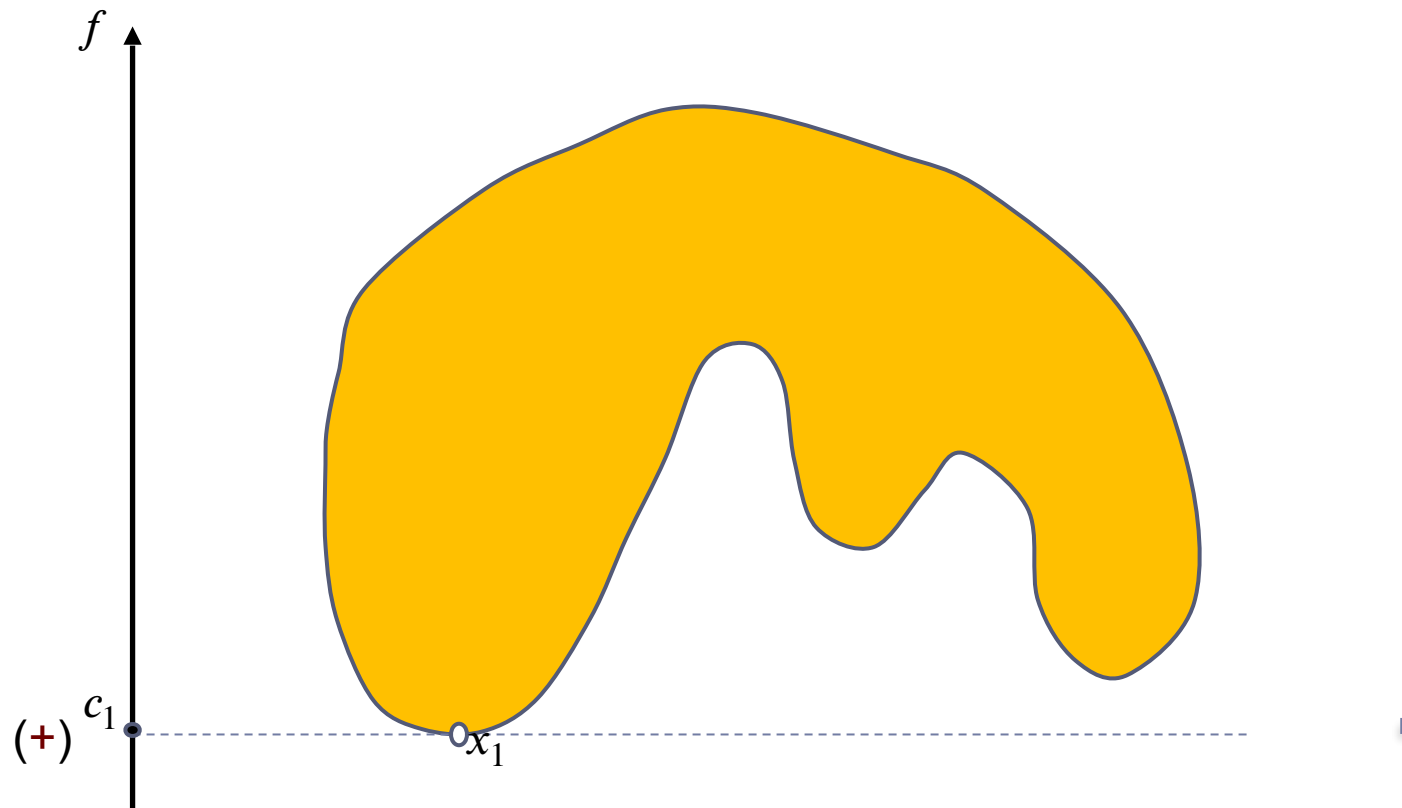
A special example - Merge tree

- ▶ What if each X_a in $\mathcal{X} = \{X_a\}_a$ is just a finite set (or $\dim(X_a) = 0$)?
- ▶ Merge tree: a 0-dim simplicial tower generated by tracking connected components of sub-level sets of some function

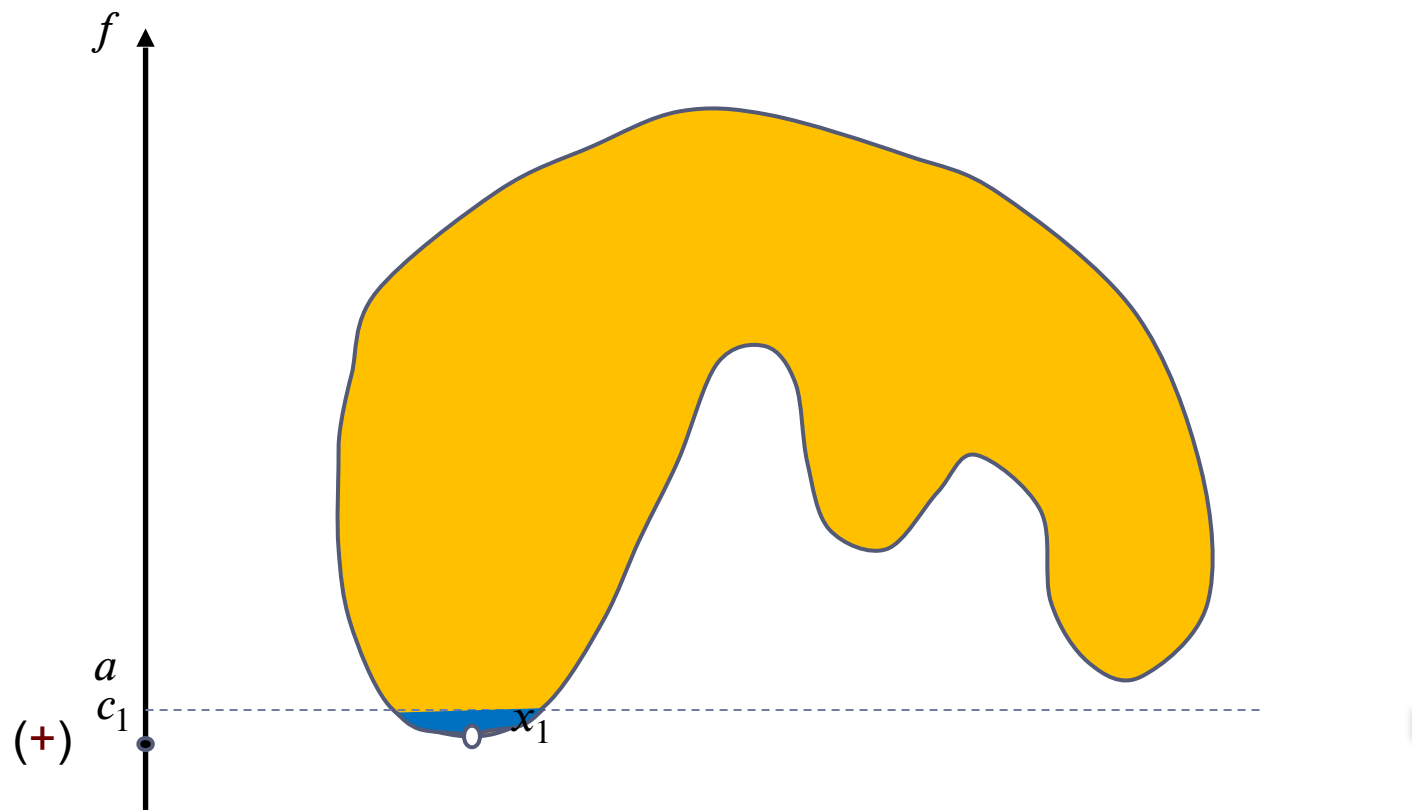
A Simple Example



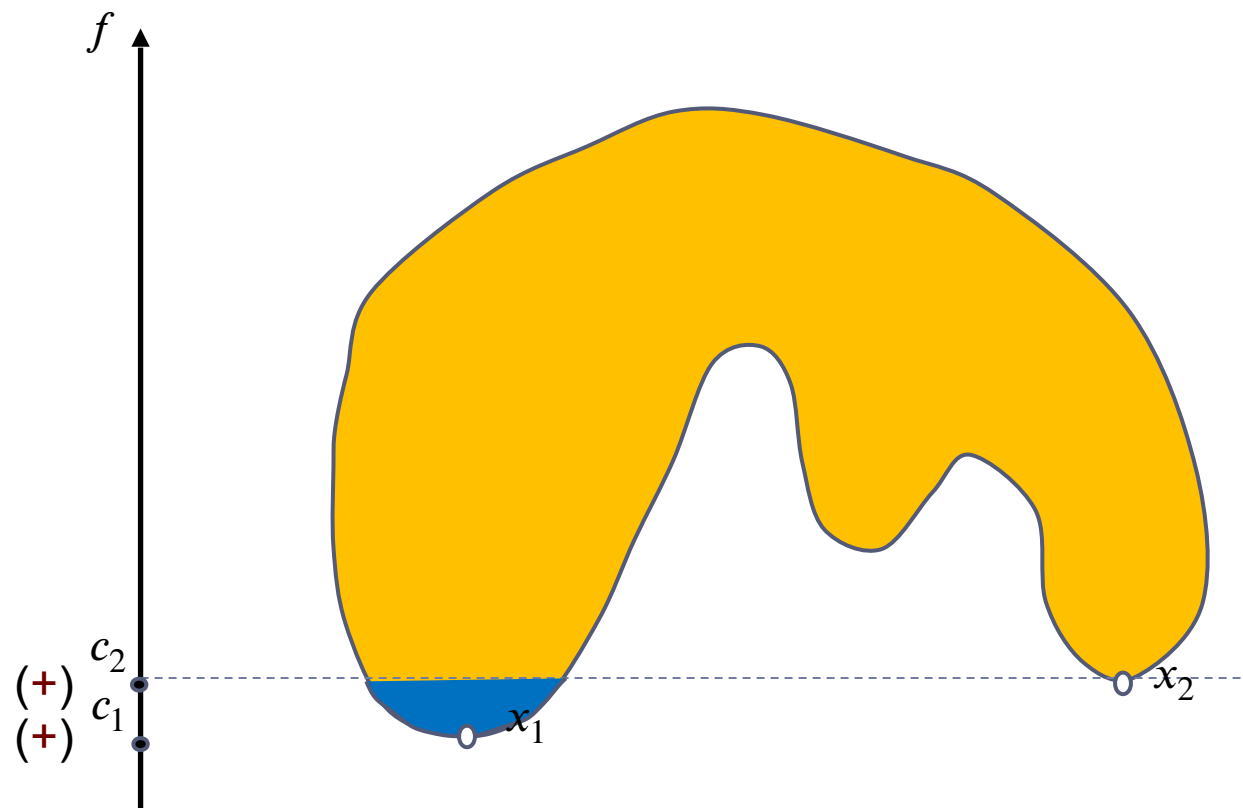
A Simple Example



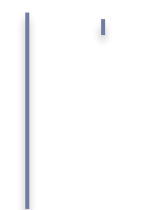
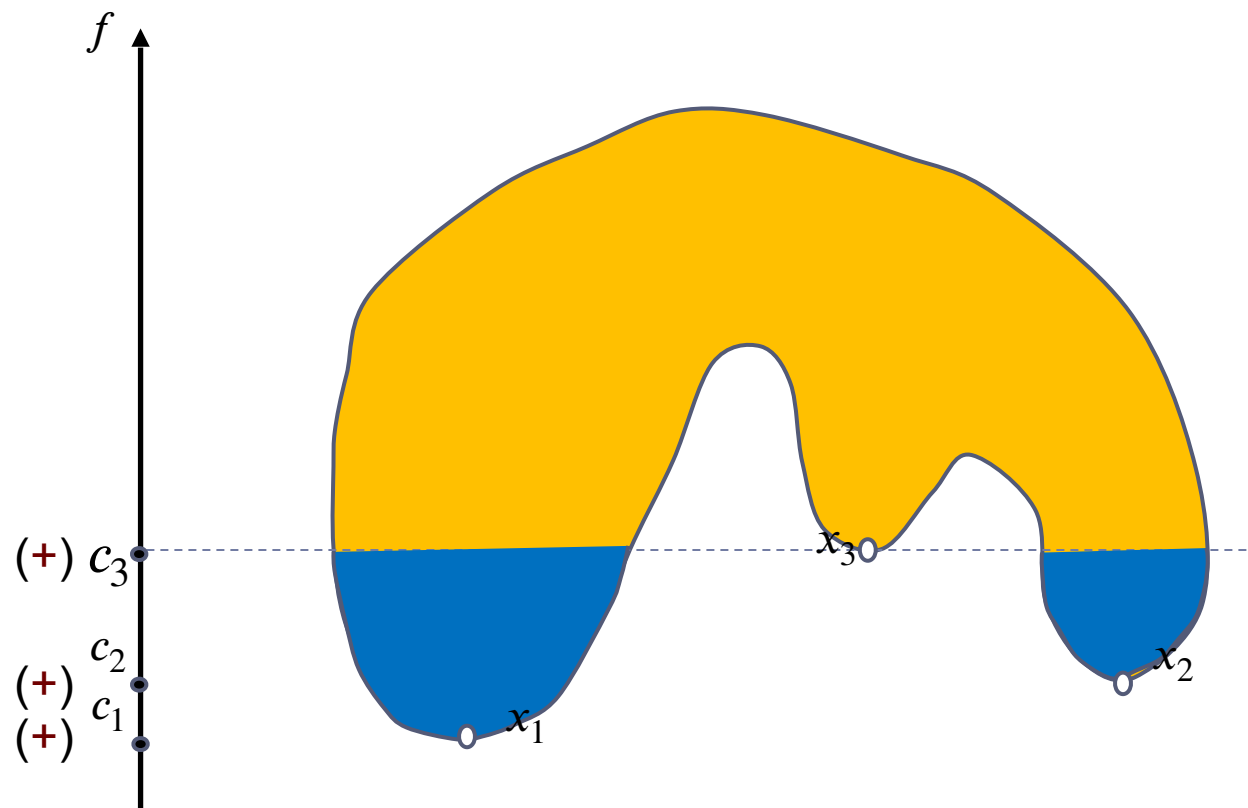
A Simple Example



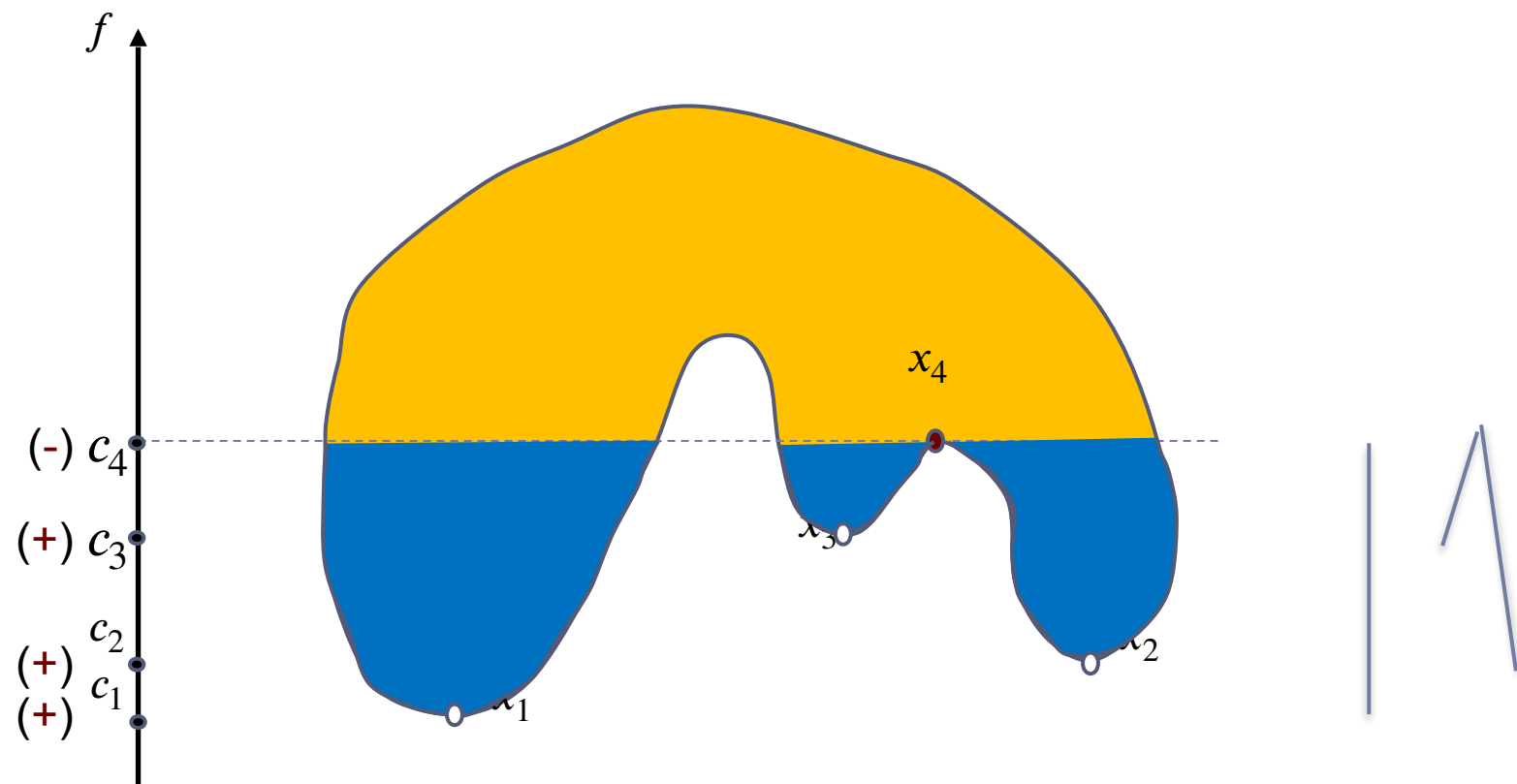
A Simple Example



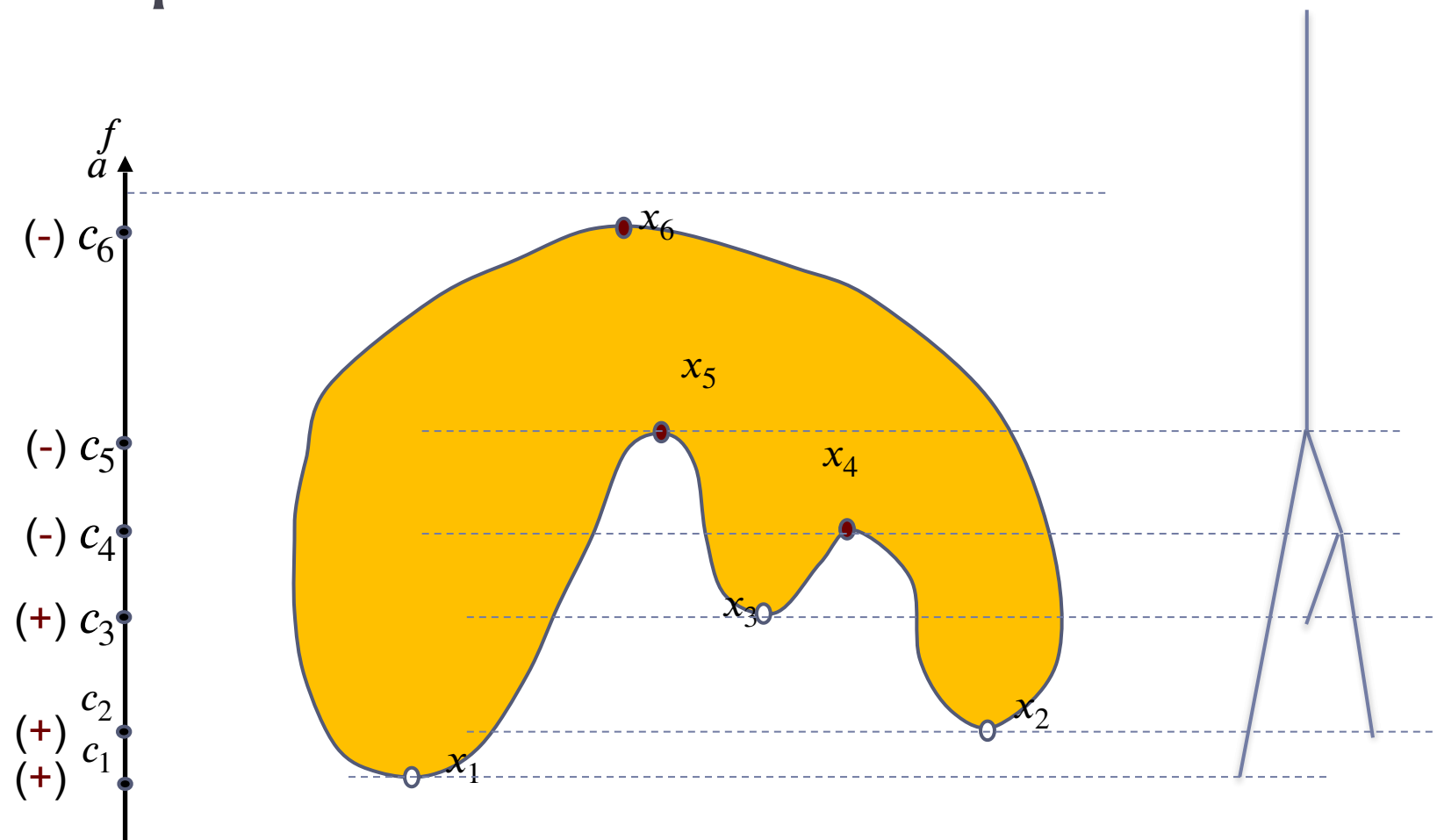
A Simple Example



A Simple Example



A Simple Example

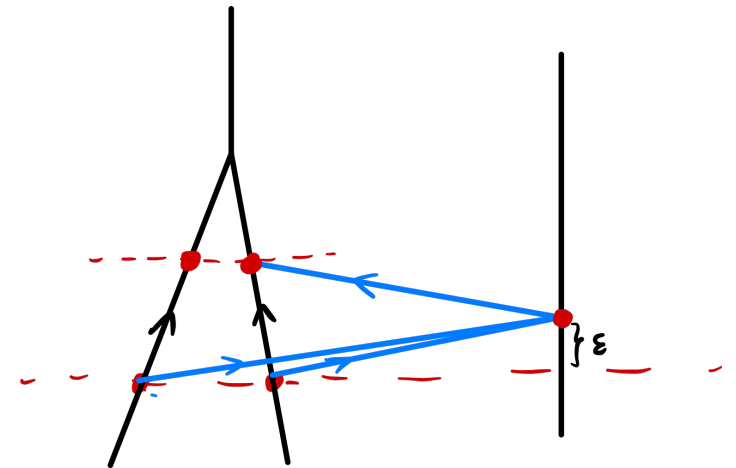


A special example - Merge tree

- ▶ What if each X_a in $\mathcal{X} = \{X_a\}_a$ is just a finite set (or $\dim(X_a) = 0$)?
- ▶ Merge tree: a 0-dim simplicial tower generated by tracking connected components of sub-level sets of some function
 - ▶ In general, it is a tree T endowed with a height function h whose value indicates the index in the filtration

A special example - Merge tree

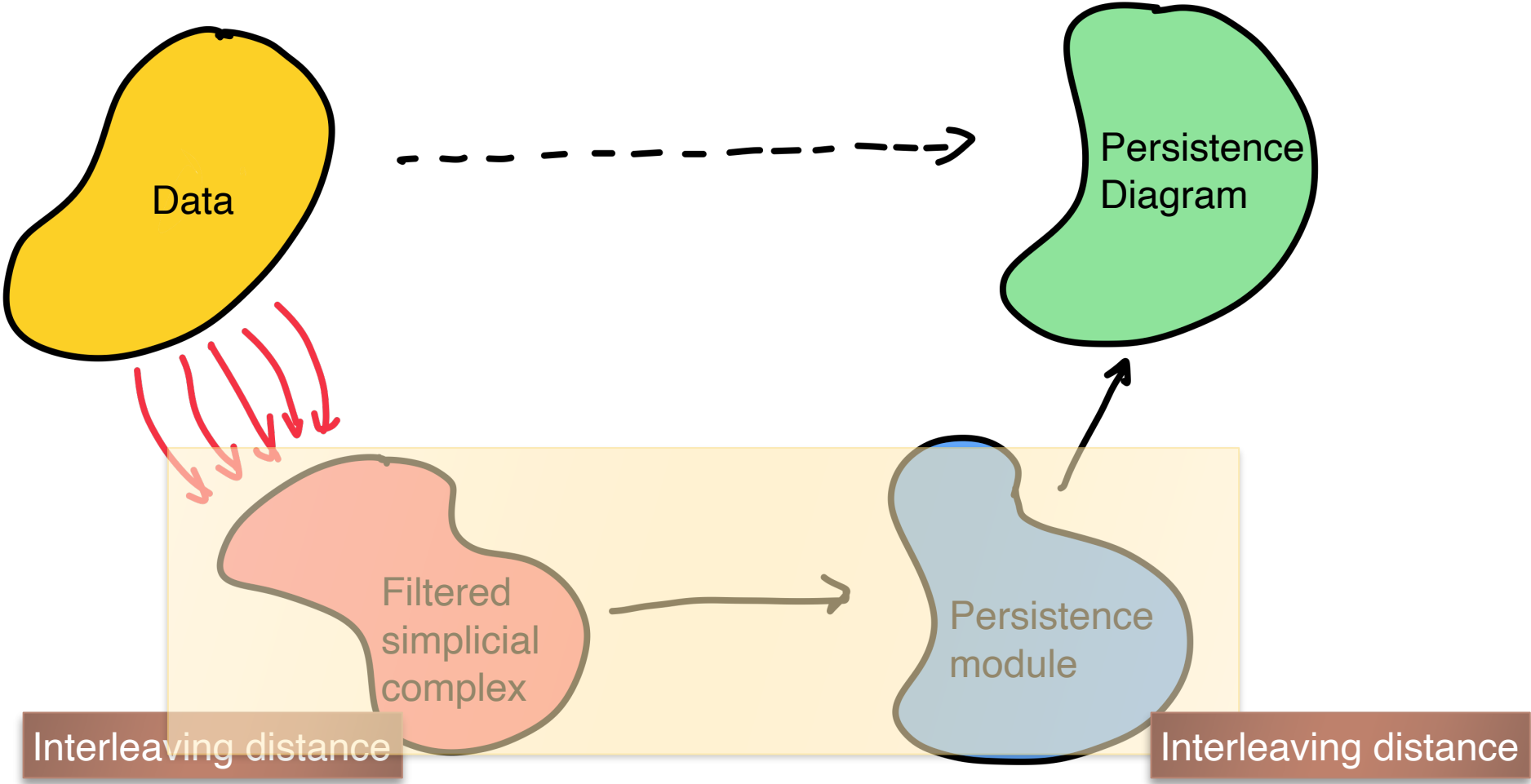
- ▶ What if each X_a in $\mathcal{X} = \{X_a\}_a$ is just a finite set (or $\dim(X_a) = 0$)?
- ▶ Merge tree: a simplicial tower generated by level sets
- ▶ One can use the interleaving distance between simplicial towers to define the interleaving distance between merge trees.
- ▶ The contiguity requirement can be replaced by the equality requirement
- ▶ This agrees with [Morozov et al. 2013]

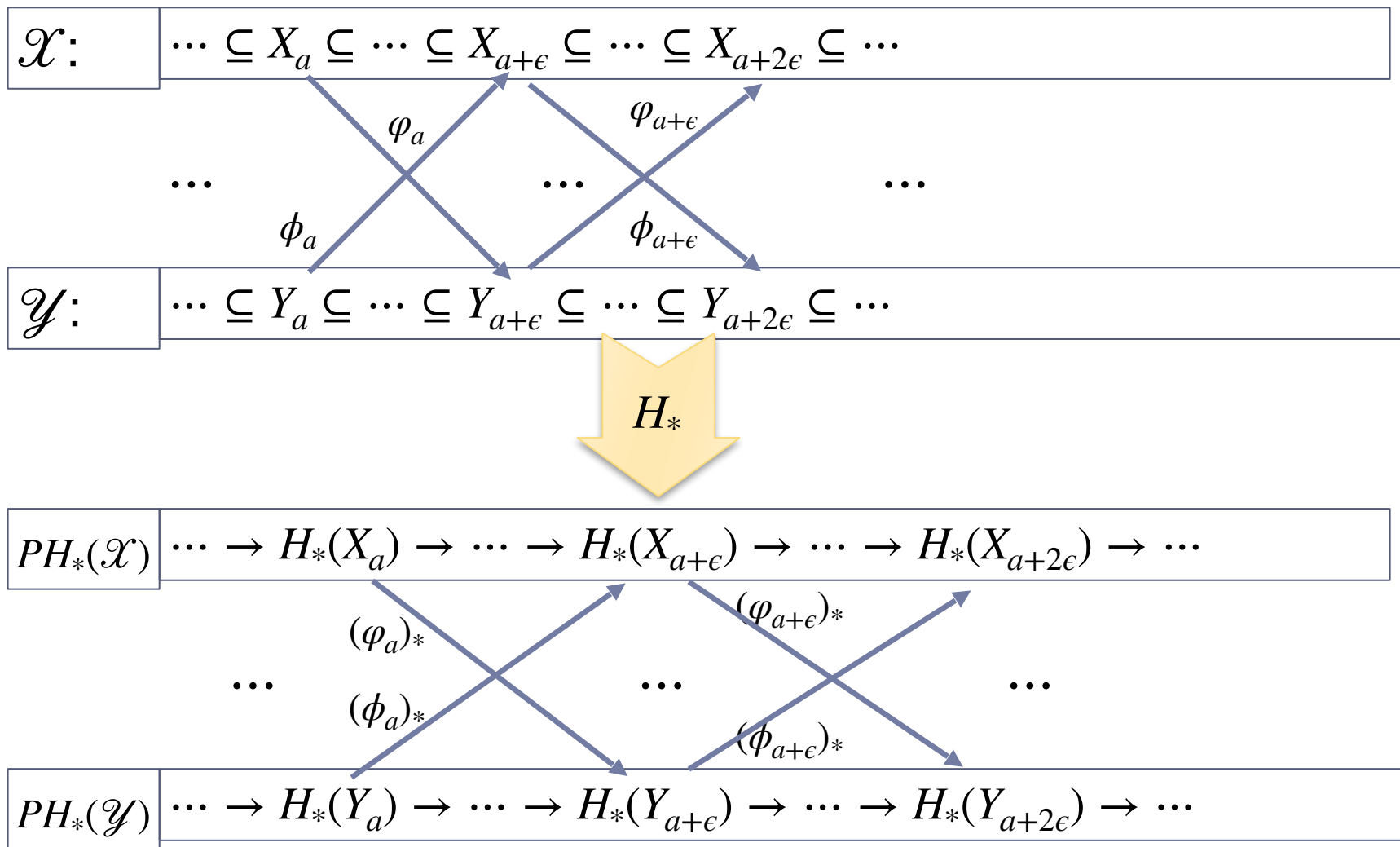


A special example - Merge tree

- ▶ What if each X_a in $\mathcal{X} = \{X_a\}_a$ is just a finite set (or $\dim(X_a) = 0$)?
- ▶ Merge tree: a simplicial tower generated by level sets
- ▶ One can use the interleaving distance between simplicial towers to define the interleaving distance between merge trees.
- ▶ It is NP-hard to compute the interleaving distance [Touli and Wang, 2019]!

Interleaving distance vs interleaving distance





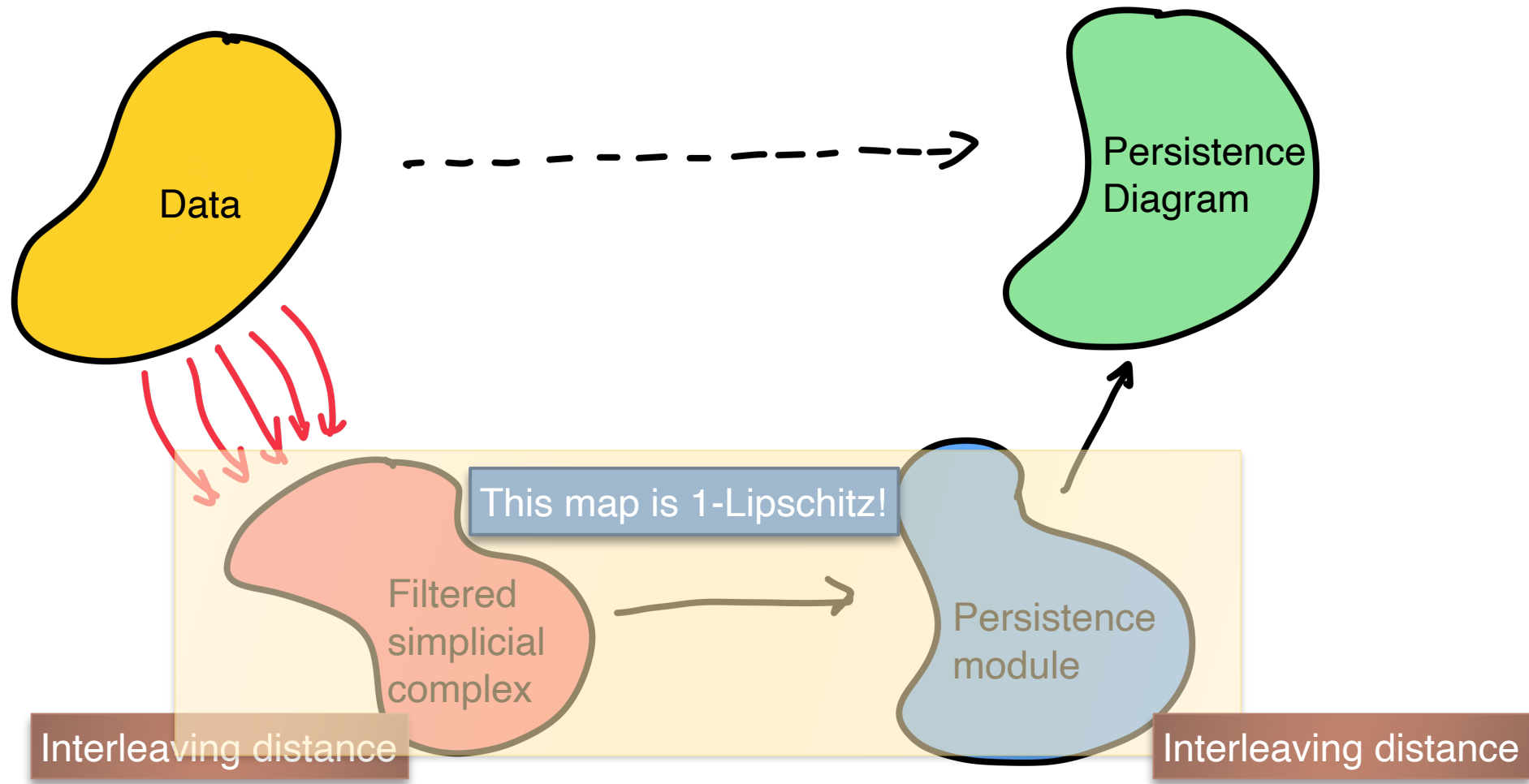
- ▶ An ϵ -interleaving between simplicial filtrations induces an ϵ -interleaving between persistence modules!

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:

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

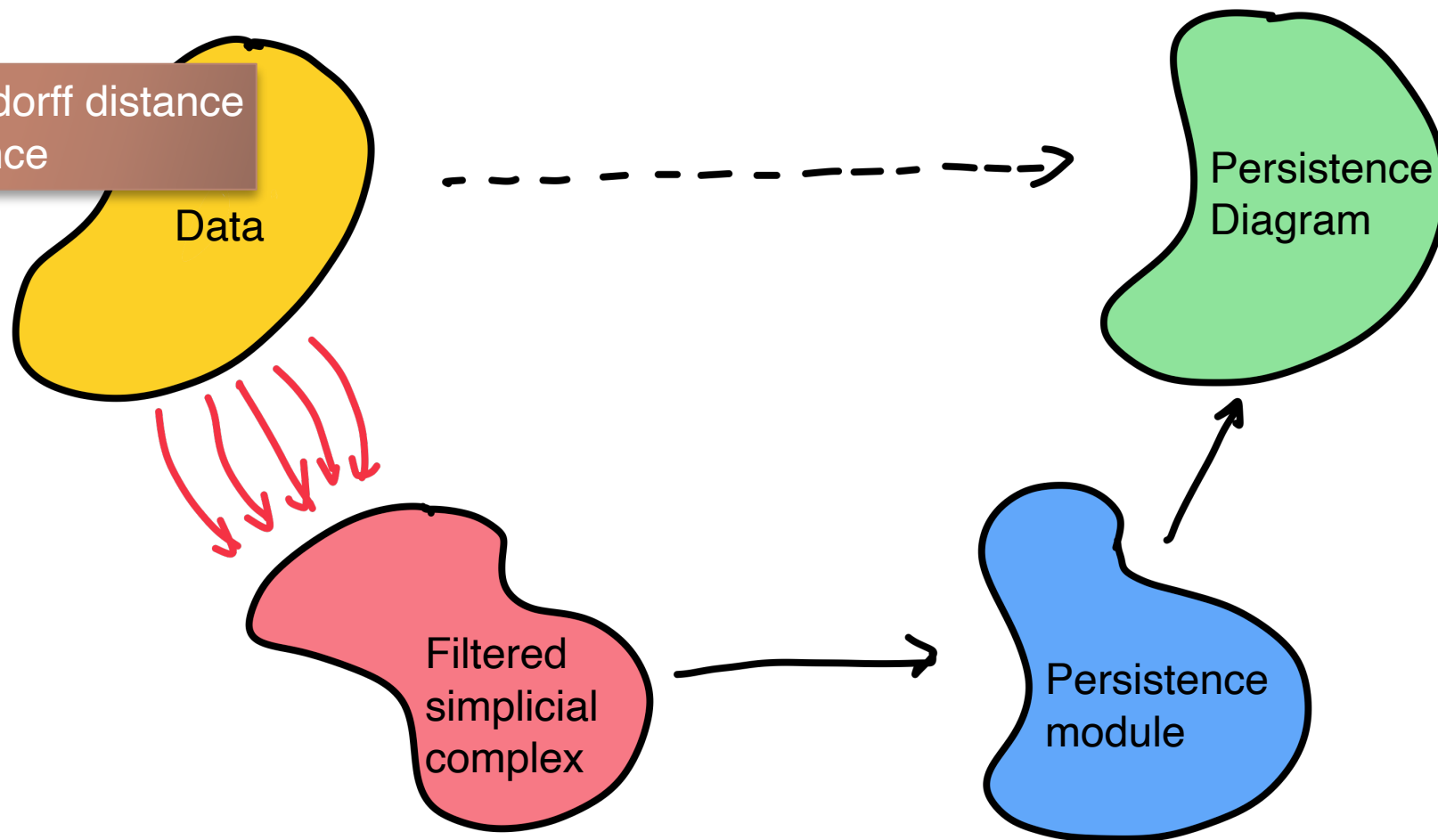
Interleaving distance vs interleaving distance



Section 4:

Distances for data and stability

Gromov-Hausdorff distance
And ℓ^∞ distance



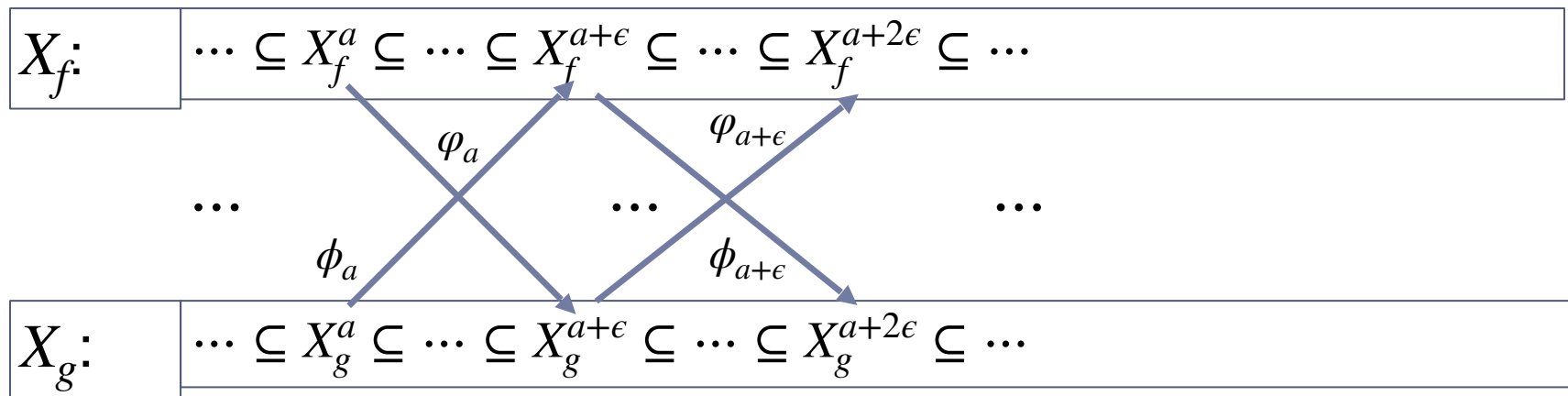
Functions

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

- ▶ Given a topological space X and two functions $f, g : X \rightarrow \mathbb{R}$
- ▶ 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 topological space X and two functions $f, g : X \rightarrow \mathbb{R}$
- ▶ 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



- ▶ Given a topological space X and two functions $f, g : X \rightarrow \mathbb{R}$
- ▶ 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 “nice” functions $f, g: X \rightarrow R$, 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:

$$d_B(D_f^*, D_g^*) = d_I(PH_*(X_f), PH_*(X_g)) \leq \|f - g\|_\infty$$

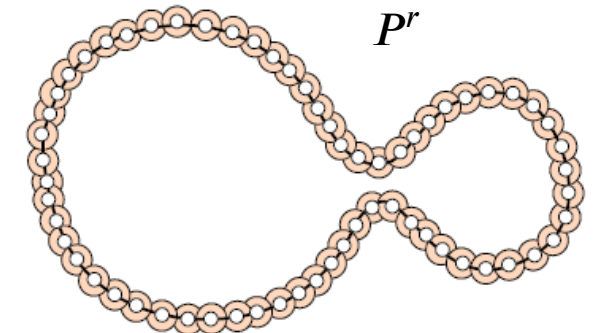
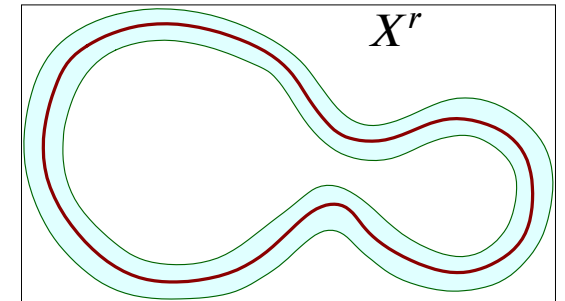
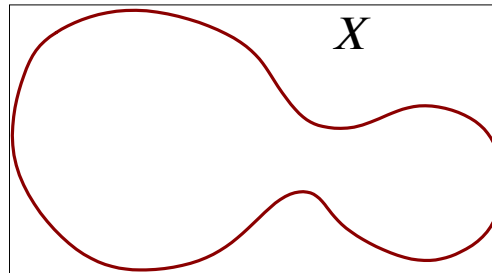
Metric spaces

Hausdorff distance between subsets

- ▶ Hausdorff distance between two sets $A, B \subset (Z, d_Z)$

- ▶ $d_H(A, B) = \max\{\max_{a \in A} \min_{b \in B} d_Z(a, b), \max_{b \in B} \min_{a \in A} d_Z(a, b)\}$

- ▶ $d_H(A, B) = \inf\{r : A \subseteq B^r, B \subseteq A^r\}$



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

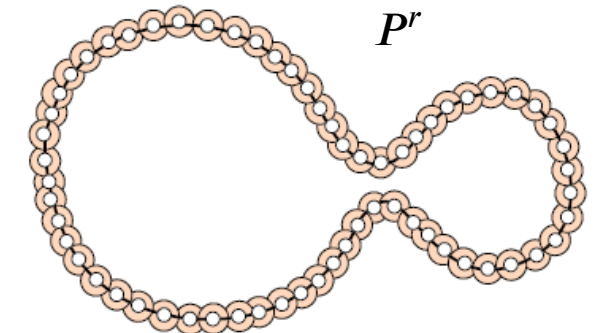
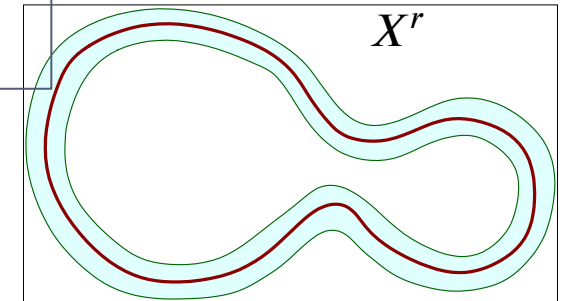
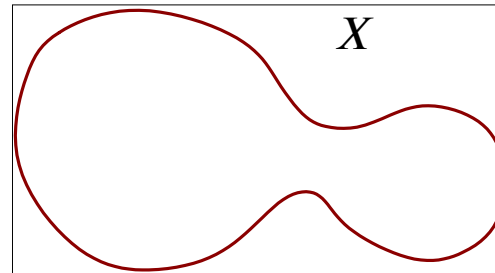
Hausdorff distance between subsets

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

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

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

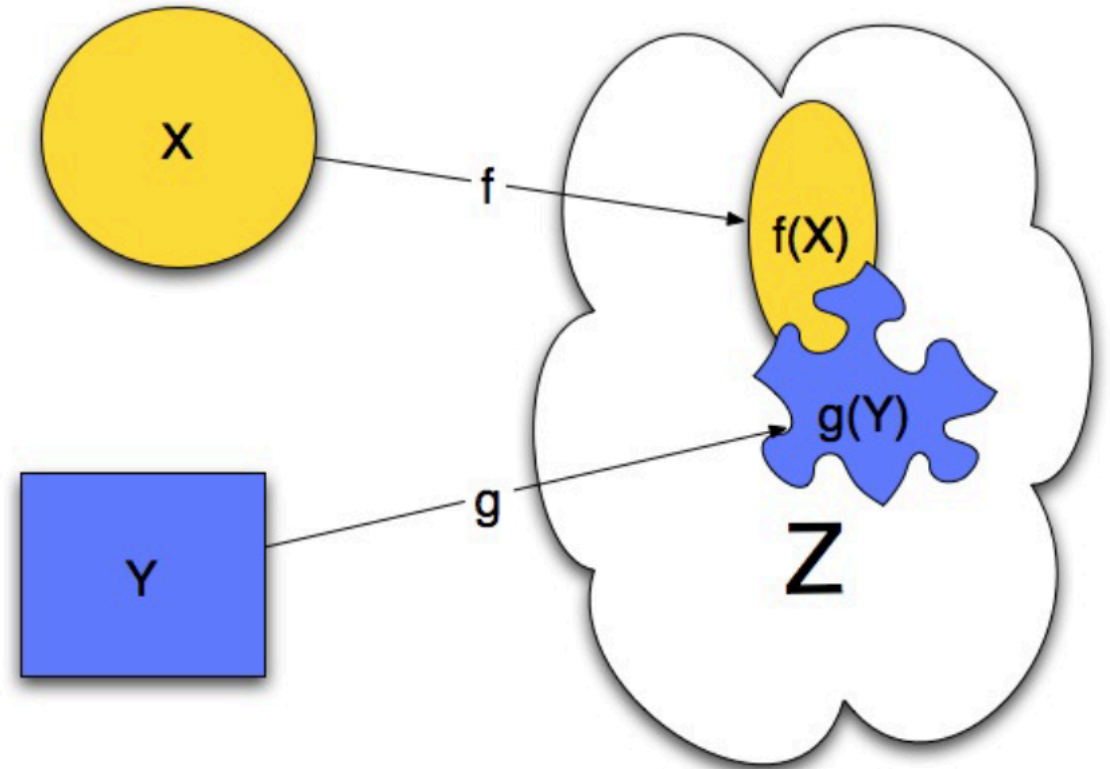
- ▶ Note that
 - ▶ $P^r \subset X^{r+\epsilon}$
 - ▶ $X^r \subset P^{r+\epsilon}$
- ▶ So $d_I(P, F_X) \leq \epsilon$



Gromov-Hausdorff distance between metric spaces

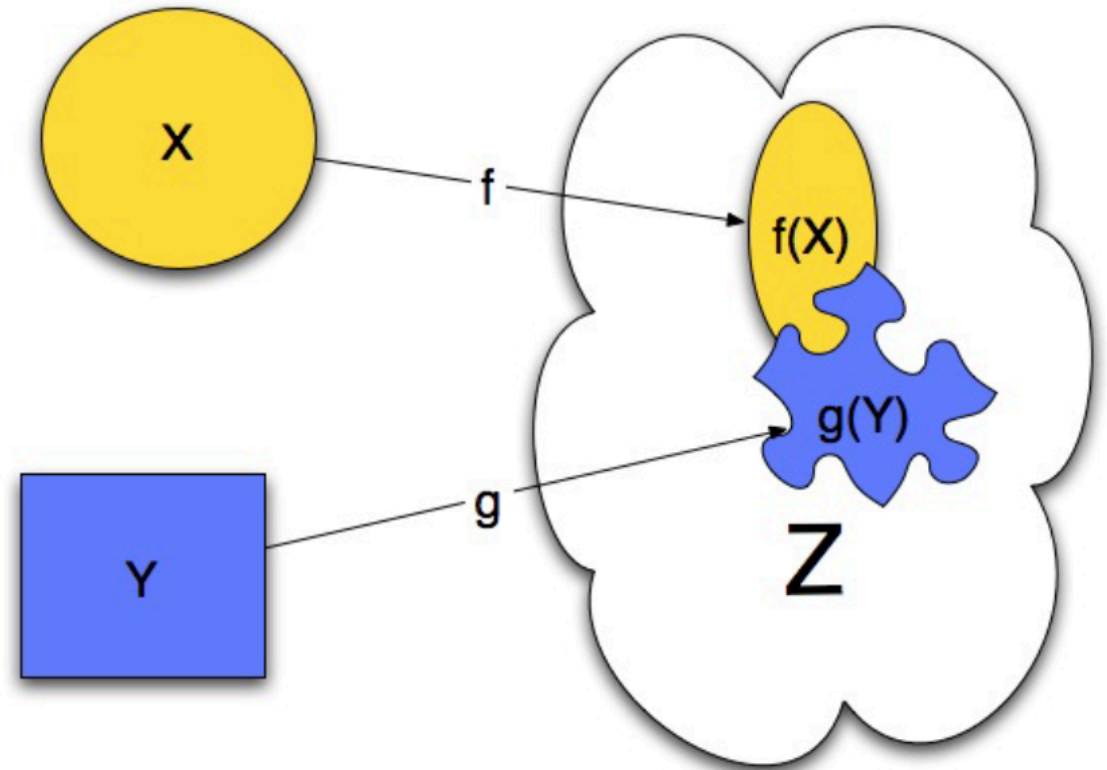
- Given two metric spaces X and Y , the **Gromov-Hausdorff distance** between them is defined as

$$d_{GH}(X, Y) := \inf_{X \hookrightarrow Z, Y \hookrightarrow Z} d_H^Z(X, Y)$$



Gromov-Hausdorff distance between metric spaces

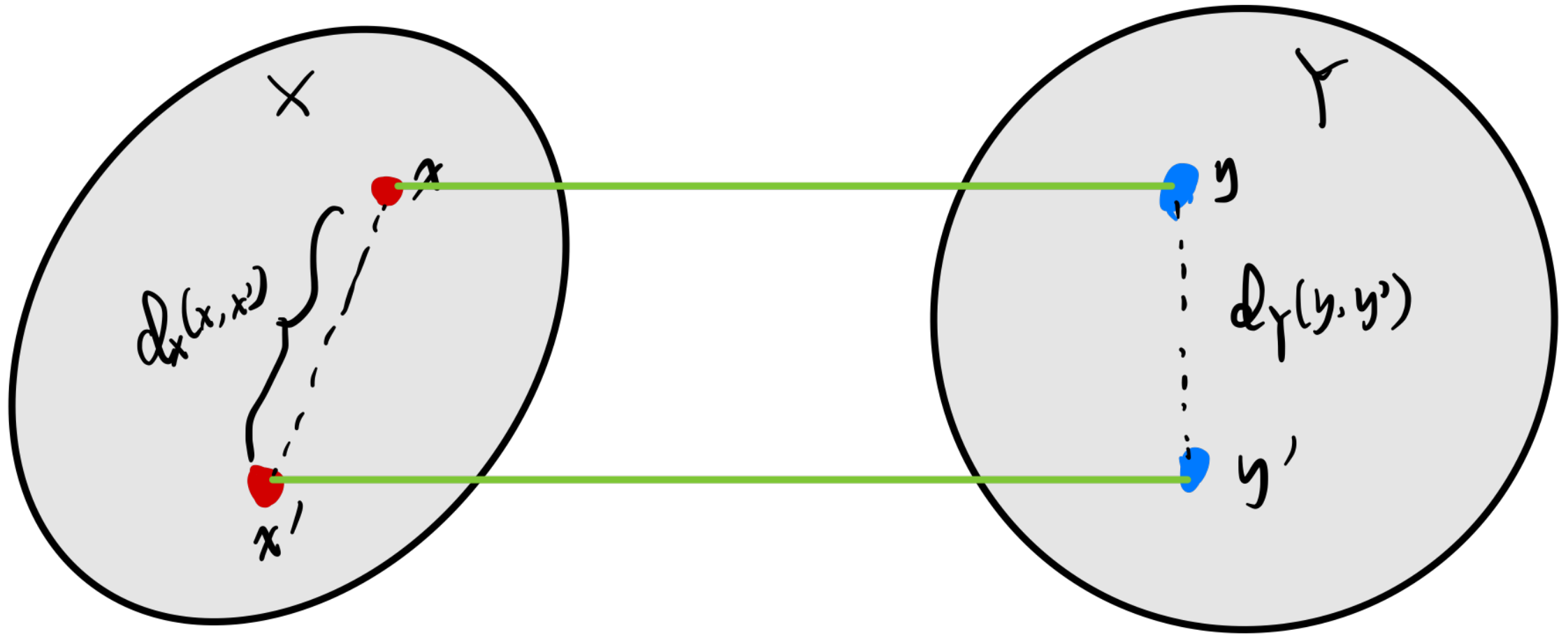
- ▶ The Gromov-Hausdorff distance is a metric on the class of compact metric spaces
- ▶ $d_{GH}(X, Y) = 0$ implies that X is isometric to Y



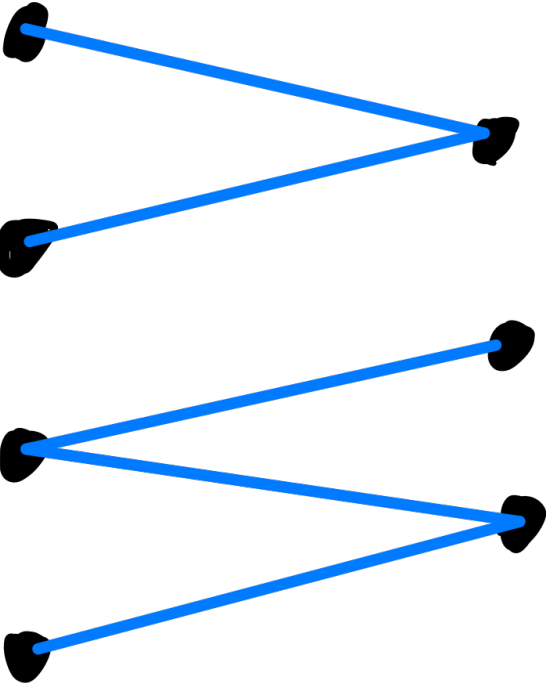
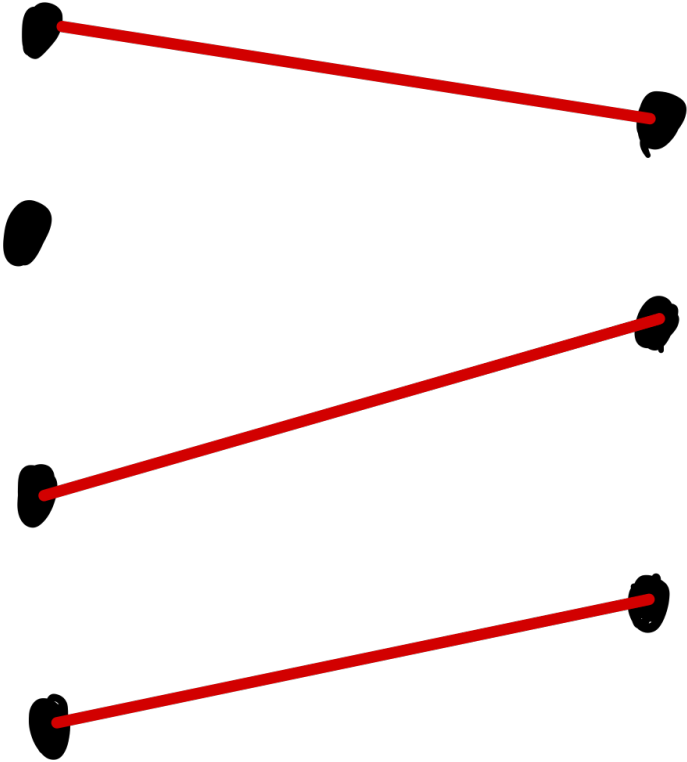
Application in mathematics

- ▶ [Wilhelm 1992] For any compact manifold X without boundary with sectional curvature ≥ 1 , if $d_{GH}(\mathbb{S}^n, X) \leq \tau(n)$ for some $\tau(n) > 0$ independent of X , then X is diffeomorphic to \mathbb{S}^n

Alternative formulation



Correspondence

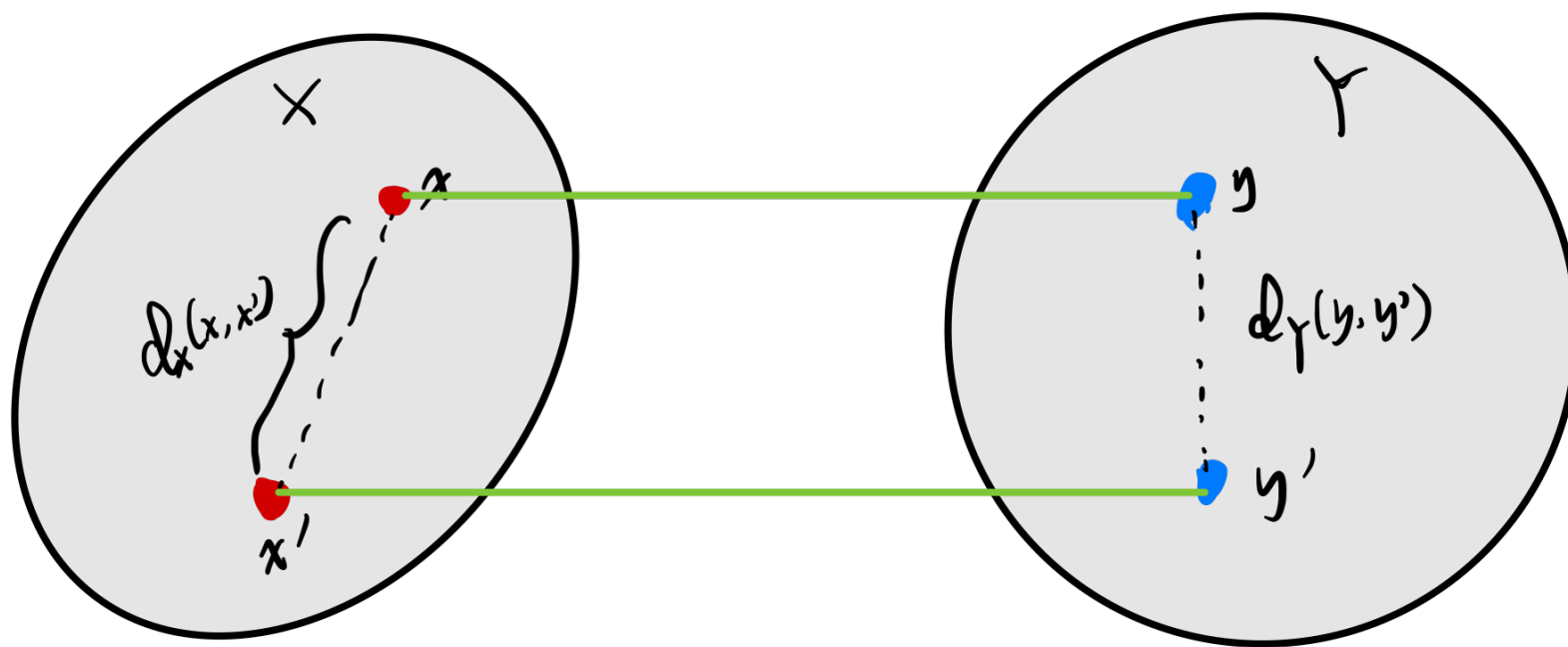


Correspondence

- ▶ A correspondence $R \subset X \times Y$ between two sets
 - ▶ For every $x \in X$, there exists $y \in Y$ such that $(x, y) \in R$
 - ▶ For every $y \in Y$, there exists $x \in X$ such that $(x, y) \in R$
- ▶ Cost (or distortion) of a correspondence R
 - ▶
$$dis(R) = \max_{(x,y),(x',y') \in R} |d_X(x, x') - d_Y(y, y')|$$

Alternative formulation

► $d_{GH}(X, Y) = \frac{1}{2} \inf_R \text{dis}(R)$



Example

▶ $d_{GH}(X, *) = ?$

Basic bounds

- ▶ $\text{diam}(X) = \sup_{x, x' \in X} d_X(x, x')$
- ▶ $d_{GH}(X, Y) \geq \frac{1}{2} |\text{diam}(X) - \text{diam}(Y)|$
- ▶ $d_{GH}(X, Y) \leq \frac{1}{2} \max(\text{diam}(X), \text{diam}(Y))$

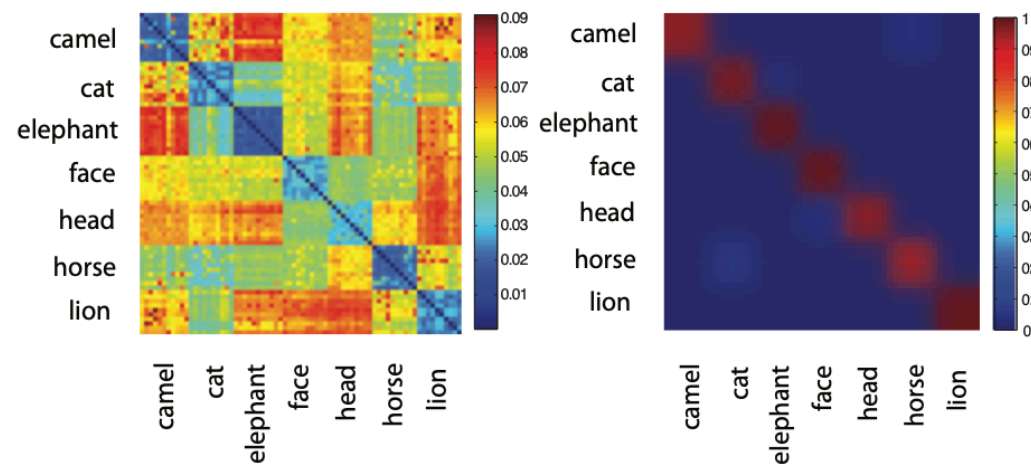
Gromov-Hausdorff distance vs Quadratic Assignment problem

- ▶ $d_{GH}(X, Y) = \frac{1}{2} \inf_R dis(R)$
- ▶ $d_{GH}(X, Y) = \frac{1}{2} \min \max_{i,k,j,l} \Gamma_{ikjl} \delta_{ij}^R \delta_{kl}^R$ where
- ▶ $\Gamma_{ikjl} = |d_X(x_i, x_k) - d_Y(y_j, y_l)|$
- ▶ $\delta_{ij}^R \in \{0,1\}$
- ▶ $\sum_i \delta_{ij}^R \geq 1$ and $\sum_j \delta_{ij}^R \geq 1$

Computational complexity

- ▶ Computing d_{GH} between finite metric spaces is NP-hard [Agarwal et al. 2018; Schmiedl 2017]
- ▶ The hardness holds even when restricted to ultrametric spaces
- ▶ FPT algorithm exists for computing d_{GH} between ultrametric spaces [Mémoli et al. 2021]

Shape comparison

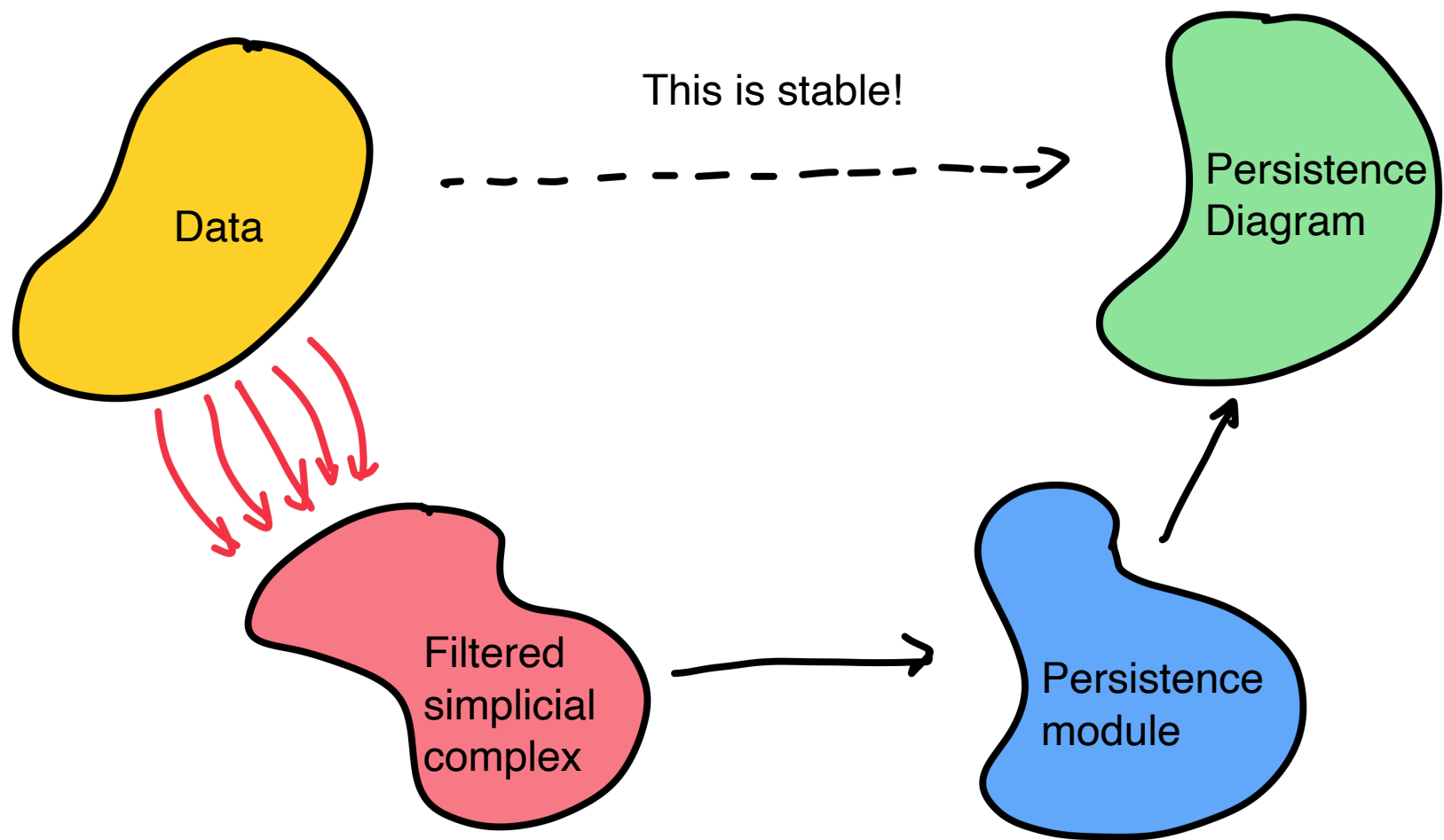


Courtesy of Mémoli 2007

Stability of persistence diagrams - metric spaces

- ▶ Given two metric spaces X and Y , one has that
 - ▶ $d_I(VR(X), VR(Y)) \leq d_{GH}(X, Y)$
 - ▶ $d_I(C(X), C(Y)) \leq 2d_{GH}(X, Y)$
- ▶ Therefore
 - ▶ $d_B(Dgm_*(VR(X)), Dgm_*(VR(Y))) \leq d_I(VR(X), VR(Y)) \leq d_{GH}(X, Y)$
 - ▶ $d_B(Dgm_*(C(X)), Dgm_*(C(Y))) \leq d_I(C(X), C(Y)) \leq 2d_{GH}(X, Y)$

Persistence-based Framework



FIN