MATH412/COMPSCI434/MATH713
Fall 2025
Topological Data Analysis

Topic 5: Stability

Instructor: Ling Zhou

Different choices of intervals

Definition 3.20 (Interval module)**.** Given an index set $A \subseteq \mathbb{R}$ and a pair of indices $b, d \in A$, $b \leq d$, four types of *interval modules* denoted $\mathbb{I}[b, d)$, $\mathbb{I}(b, d]$, $\mathbb{I}[b, d]$, $\mathbb{I}(b, d)$ respectively are special persistence modules defined as:

- (closed-open): $\mathbb{I}[b, d) : \{\mathrm{V_a} \xrightarrow{v_{a,a'}} \mathrm{V_{a'}}\}_{a,a' \in A}$ where (i) $\mathrm{V}_a = \mathbb{Z}_2$ for all $a \in [b, d)$ and $V_a = 0$ otherwise, (ii) $v_{a,a'}$ is identity map for $b \leq a \leq a' < d$ and zero map otherwise.

- (open-closed): $\mathbb{I}(b, d] : \{\mathrm{V_a} \xrightarrow{v_{a,a'}} \mathrm{V_{a'}}\}_{a,a' \in A}$ where (i) $\mathrm{V}_a = \mathbb{Z}_2$ for all $a \in (b, d]$ and $V_a = 0$ otherwise, (ii) $v_{a,a'}$ is identity map for $b < a \leq a' \leq d$ and zero map otherwise.

- (closed-closed): $\mathbb{I}[b, d] : \{\mathrm{V_a} \xrightarrow{v_{a,a'}} \mathrm{V_{a'}}\}_{a,a' \in A}$ where (i) $\mathrm{V}_a = \mathbb{Z}_2$ for all $a \in [b, d]$ and $V_a = 0$ otherwise, (ii) $v_{a,a'}$ is identity map for $b \leq a \leq a' \leq d$ and zero map otherwise.

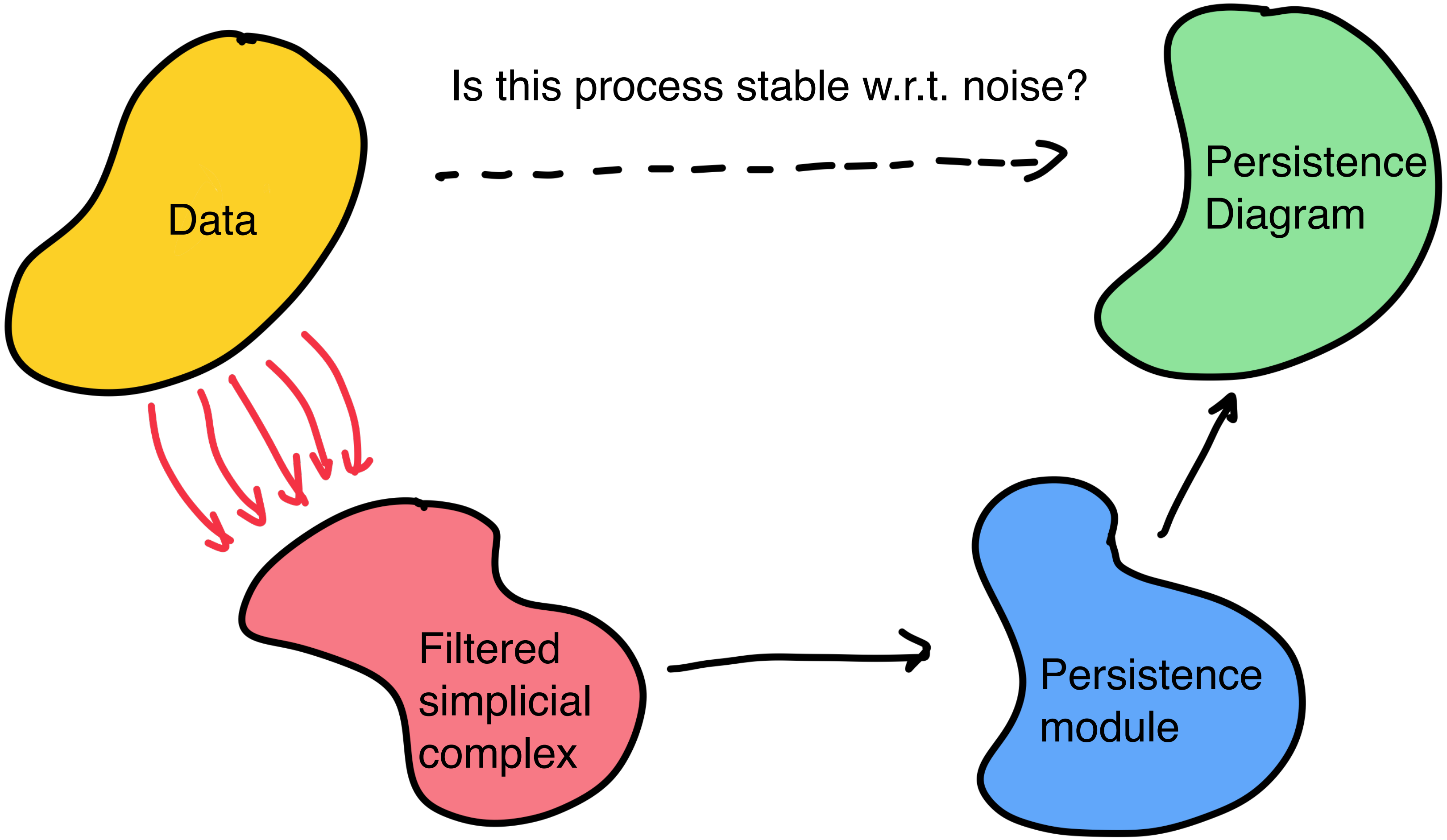
- (open-open): $\mathbb{I}(b, d) : \{\mathrm{V_a} \xrightarrow{v_{a,a'}} \mathrm{V_{a'}}\}_{a,a' \in A}$ where (i) $\mathrm{V}_a = \mathbb{Z}_2$ for all $a \in (b, d)$ and $V_a = 0$ otherwise, (ii) $v_{a,a'}$ is identity map for $b < a \leq a' < d$ and zero map otherwise.

Different choices of intervals

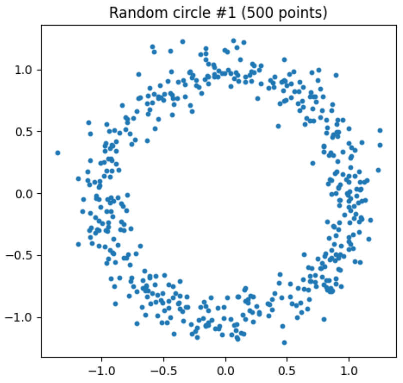
- (**closed-open**) $\beta_p^{i,j}$ = number of intervals **[k, l)** that contains [i, j]
- (**open-closed**) $\beta_p^{i,j}$ = number of intervals **(k, l]** that contains [i, j]
- (**closed-closed**) $\beta_p^{i,j}$ = number of intervals **[k, l]** that contains [i, j]
- (**open-open**) $\beta_p^{i,j}$ = number of intervals **(k, l)** that contains [i, j]

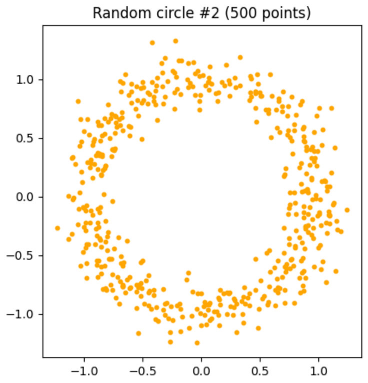
- $\mu_p^{i,j} = (\beta_p^{i,j-1} - \beta_p^{i,j}) - (\beta_p^{i-1,j-1} - \beta_p^{i-1,j})$

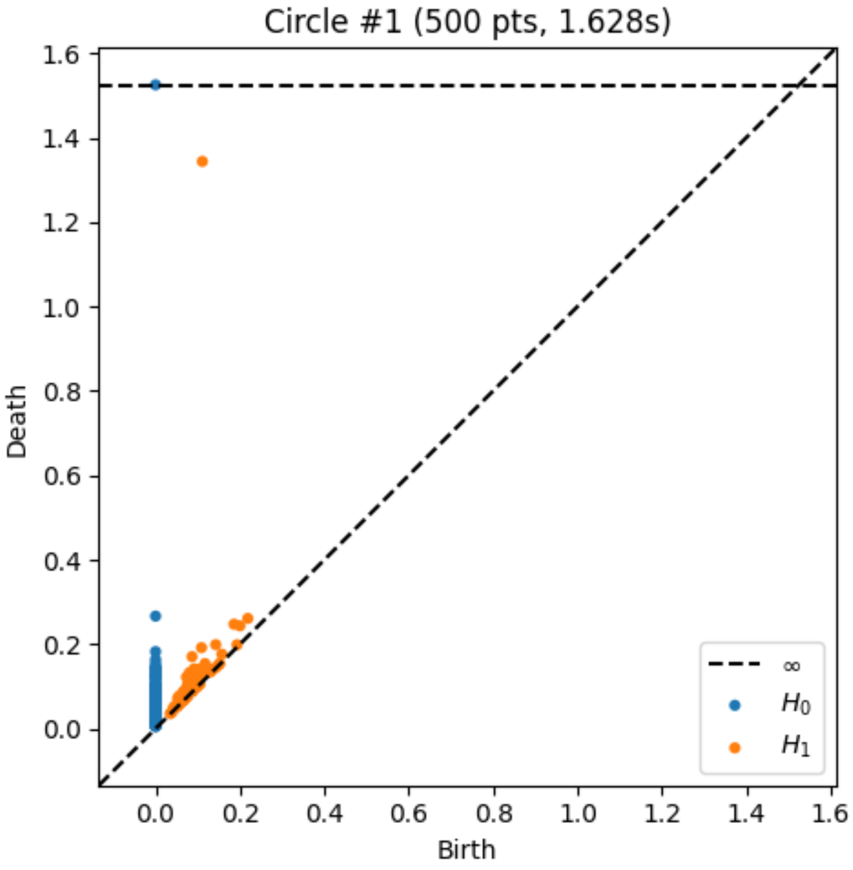
Persistence-based Framework

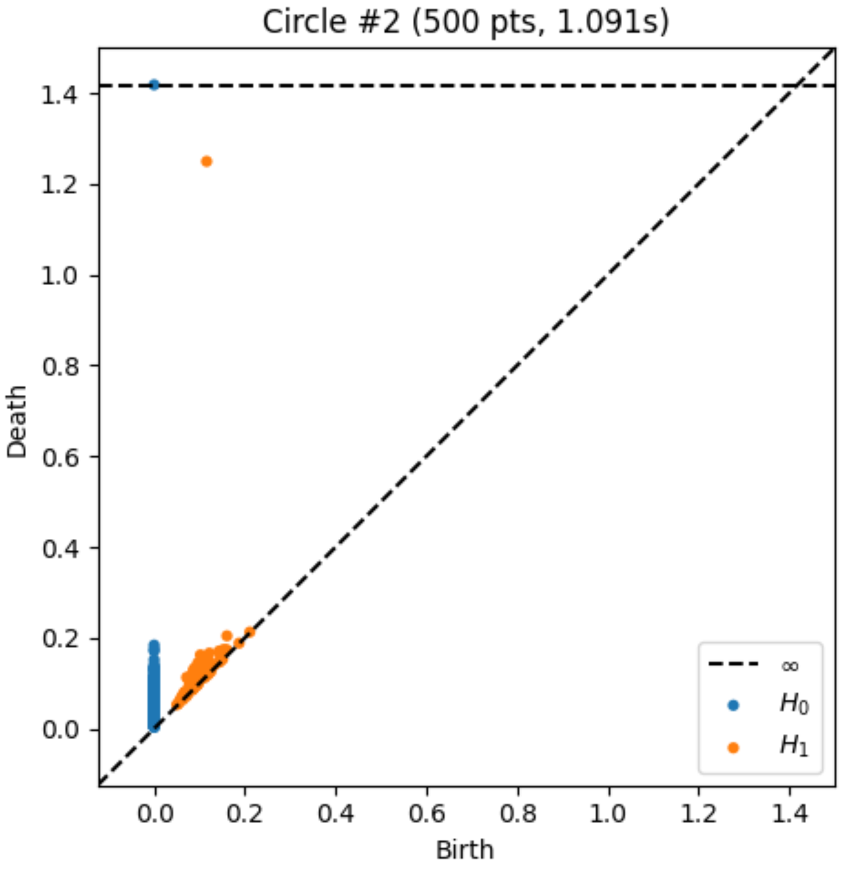


Similar Data => Similar PD



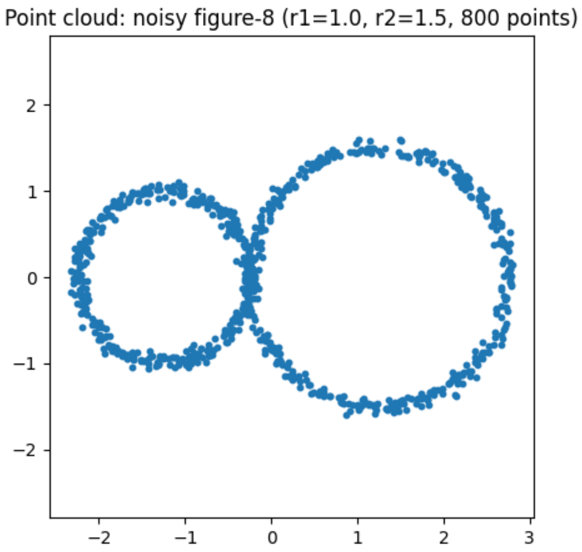


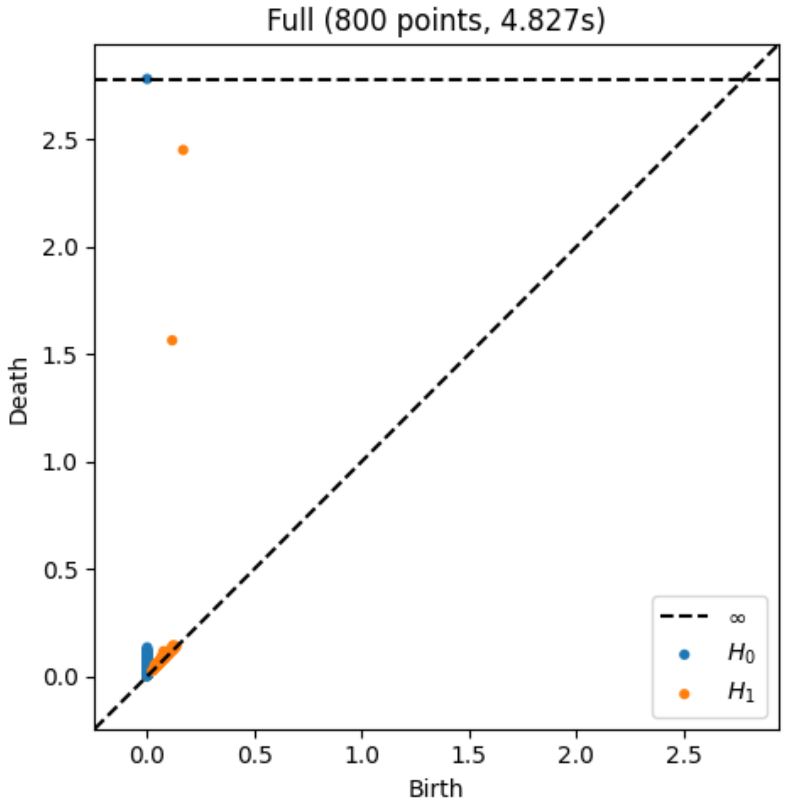


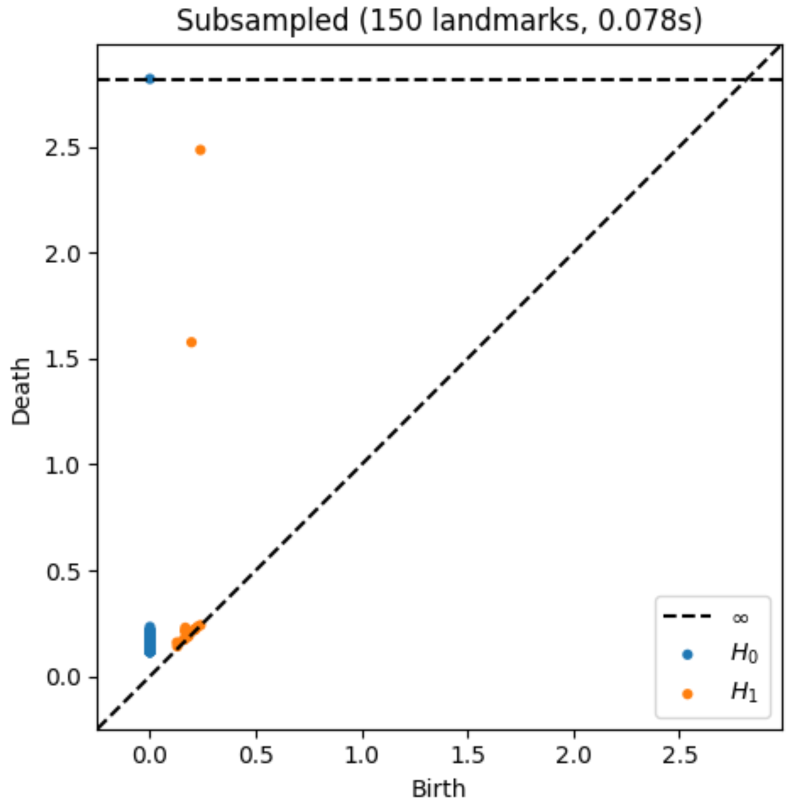


Similar Data => Similar PD

- Allows the use of subsampling to reduce runtime.





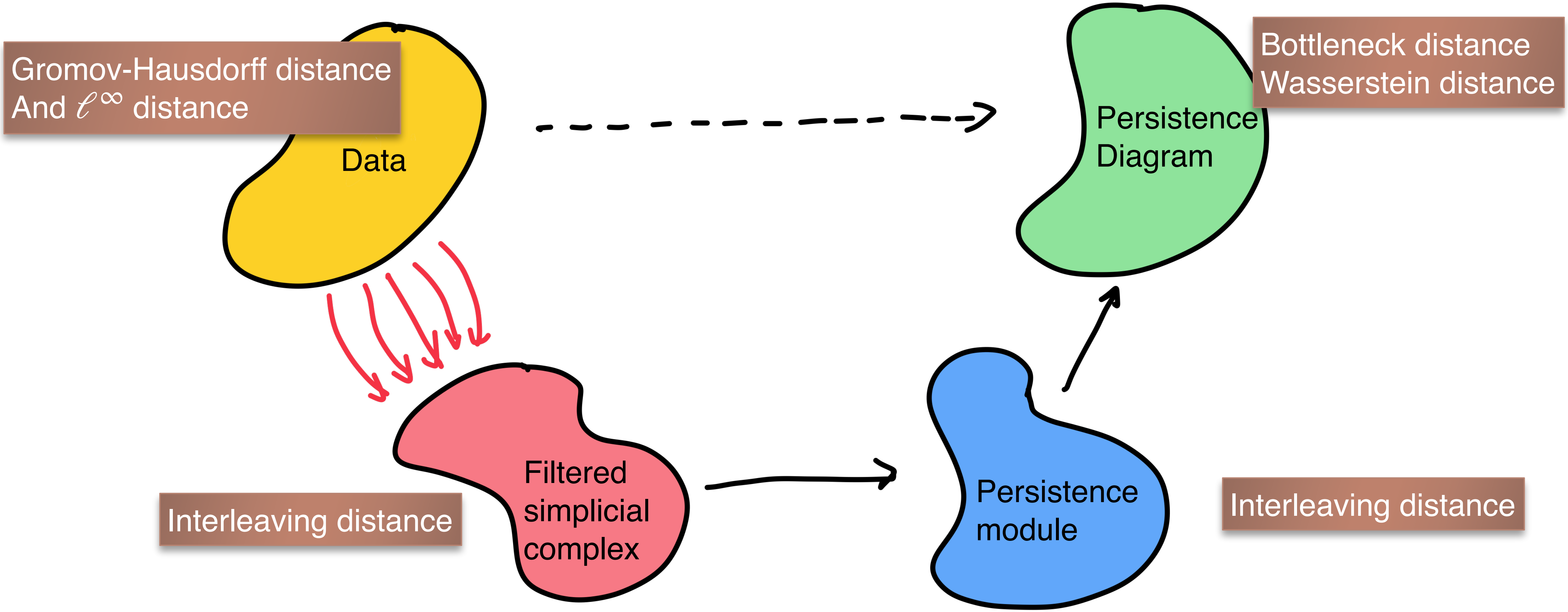


```
[uniform_square] n=  50, H0=50, H1=9, time=0.006s
[uniform_square] n= 100, H0=100, H1=19, time=0.015s
[uniform_square] n= 500, H0=500, H1=121, time=0.754s
[uniform_square] n=1000, H0=1000, H1=247, time=2.488s
[uniform_square] n=2000, H0=2000, H1=541, time=6.109s
[uniform_square] n=3000, H0=3000, H1=777, time=10.767s
-----
[gaussian_cloud] n=  50, H0=50, H1=7, time=0.004s
[gaussian_cloud] n= 100, H0=100, H1=18, time=0.006s
[gaussian_cloud] n= 500, H0=500, H1=120, time=0.248s
[gaussian_cloud] n=1000, H0=1000, H1=270, time=1.032s
[gaussian_cloud] n=2000, H0=2000, H1=527, time=5.476s
[gaussian_cloud] n=3000, H0=3000, H1=797, time=12.846s
-----
[circle_random] n=  50, H0=50, H1=1, time=0.005s
[circle_random] n= 100, H0=100, H1=1, time=0.020s
[circle_random] n= 500, H0=500, H1=1, time=3.363s
[circle_random] n=1000, H0=1000, H1=1, time=27.641s
```

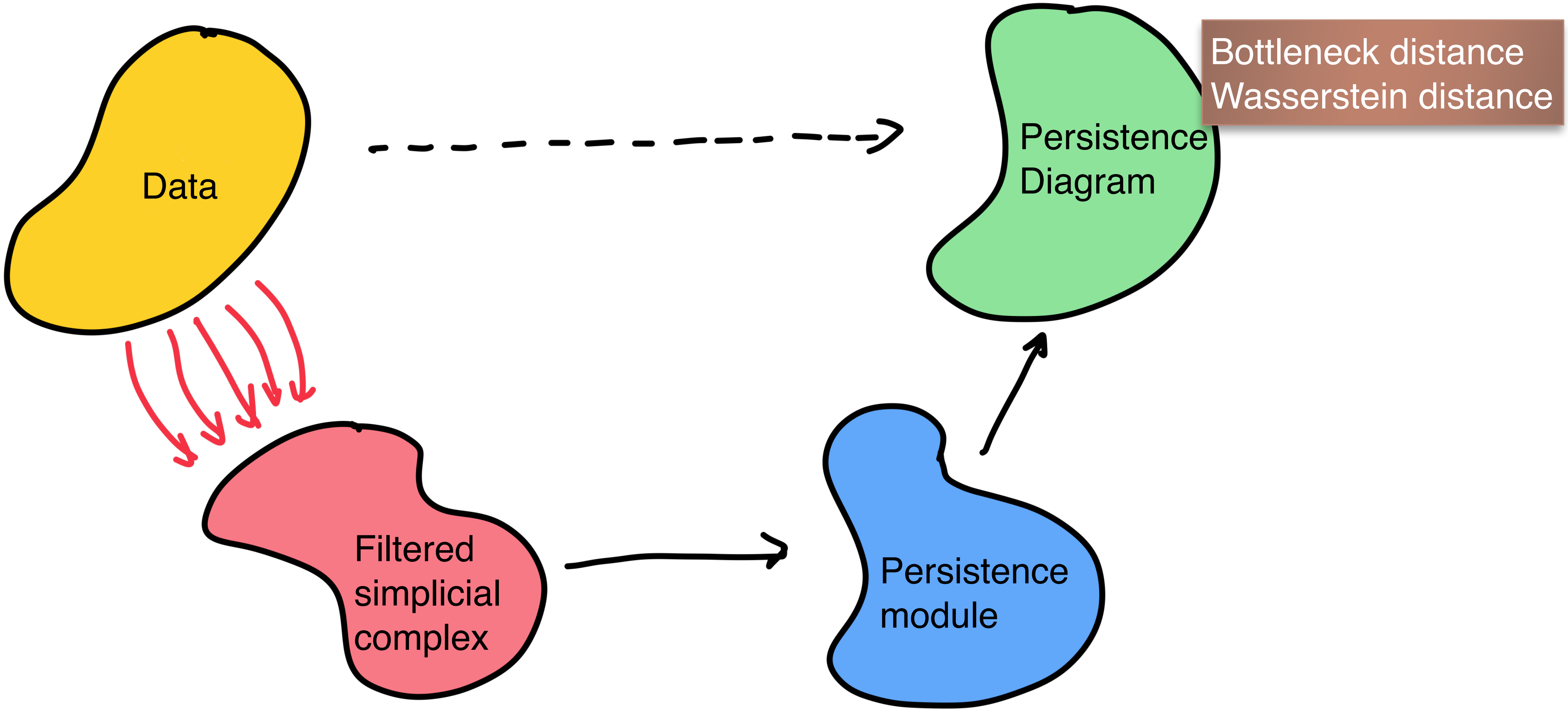
crashed…

```
[uniform_square] n=  50, H0=50, H1=9, H2=0, time=0.016s
[uniform_square] n= 100, H0=100, H1=24, H2=0, time=0.092s
[uniform_square] n= 200, H0=200, H1=55, H2=0, time=0.766s
[uniform_square] n= 400, H0=400, H1=96, H2=1, time=6.080s
[uniform_square] n= 600, H0=600, H1=155, H2=2, time=28.395s
-----
[gaussian_cloud] n=  50, H0=50, H1=8, H2=0, time=0.022s
[gaussian_cloud] n= 100, H0=100, H1=20, H2=1, time=0.186s
[gaussian_cloud] n= 200, H0=200, H1=51, H2=1, time=1.801s
[gaussian_cloud] n= 400, H0=400, H1=93, H2=1, time=13.391s
[gaussian_cloud] n= 600, H0=600, H1=167, H2=1, time=47.470s
-----
[circle_random] n=  50, H0=50, H1=1, H2=6, time=0.022s
[circle_random] n= 100, H0=100, H1=1, H2=11, time=0.246s
[circle_random] n= 200, H0=200, H1=1, H2=24, time=2.134s
[circle_random] n= 400, H0=400, H1=1, H2=56, time=16.814s
[circle_random] n= 600, H0=600, H1=1, H2=75, time=62.211s
-----
```

Using metrics to measure perturbations



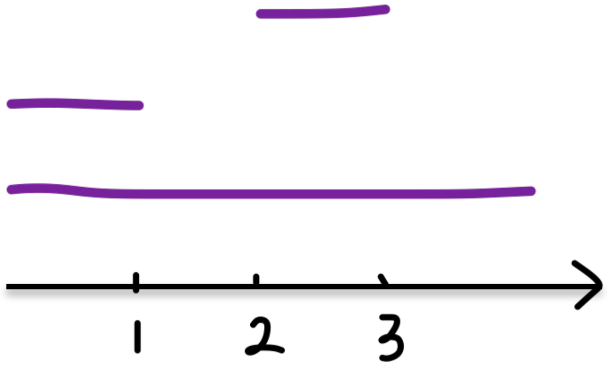
Bottleneck Distance for PD


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

Recall: Persistence Diagram and Barcodes

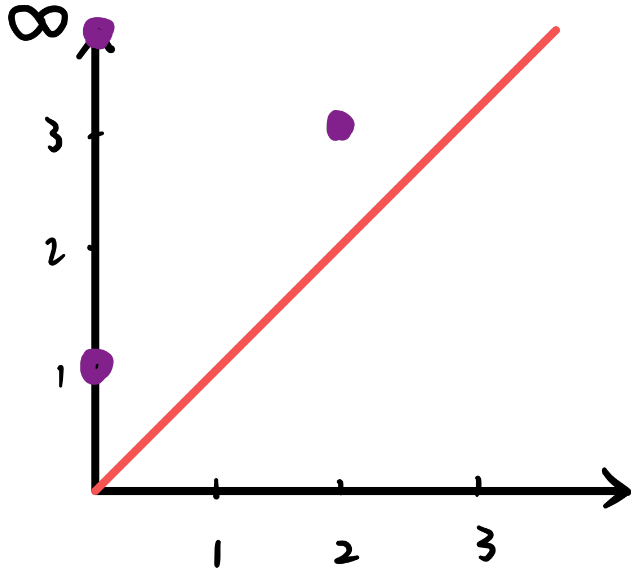
- Assume we have decomposed a persistence module as
  $V_\bullet \cong I[b_1, d_1) \oplus I[b_2, d_2) \oplus \cdots \oplus I[b_M, d_M)$
  - The collection of intervals $\{[b_j, d_j)\}_{j=1,\ldots,M}$ is called the **barcode** of $V_\bullet$
  - The multiset $D = \{(b_j, d_j)\}_{j=1,\ldots,M}$ is called the **persistence diagram** of $V_\bullet$

Barcode



$H_0(K_\bullet) \cong I[0,\infty) \oplus I[0,1) \oplus I[2,3)$

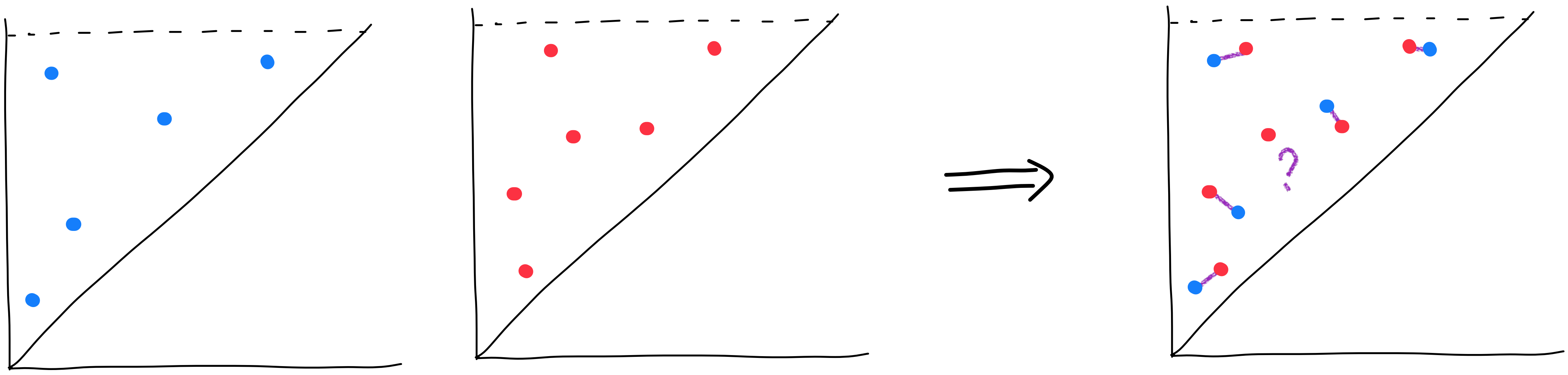
Persistence Diagram



(General) 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 (PD)**, where $0 \leq b_i < d_i \leq \infty$ for each $i = 1,\ldots,M$
- How to compare two different persistence diagrams?

Matching points

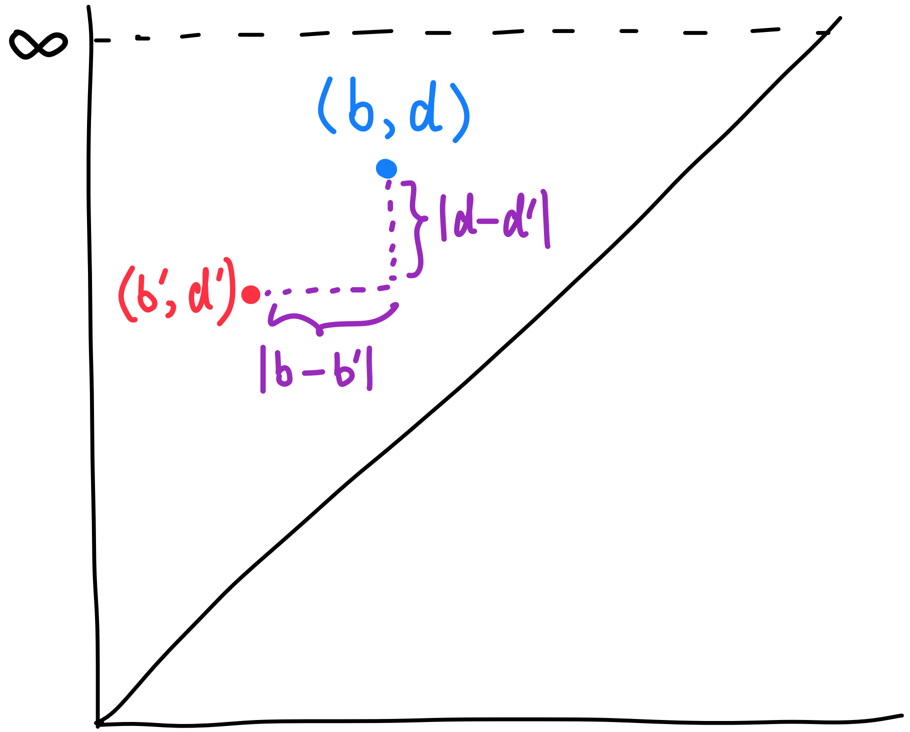


Distance Between Points

- Given two points $p = (b, d)$ and $q = (b', d') \in (\mathbb{R} \cup \infty)^2$
- Distance between two points is

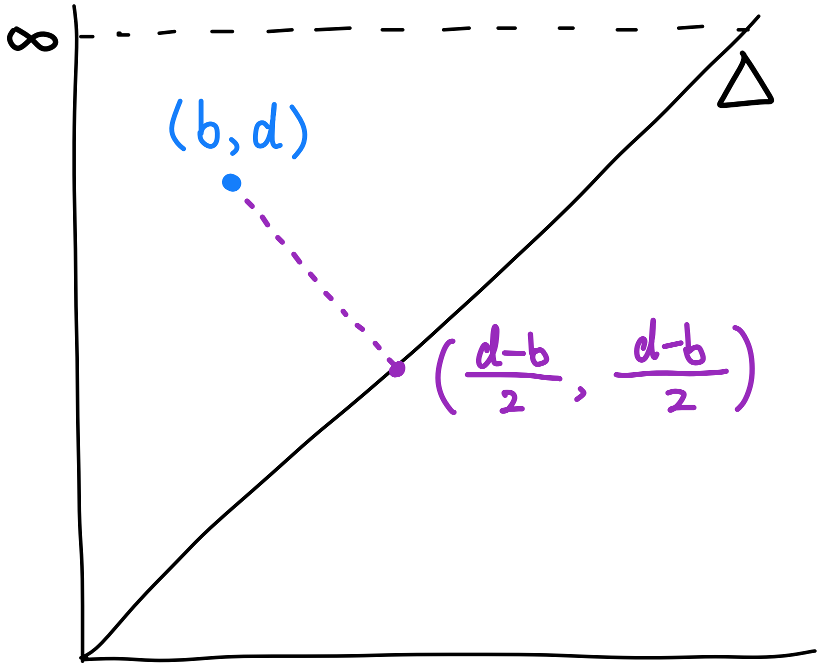
  $\|p - q\|_\infty = \max(|b - b'|, |d - d'|)$

  - Here, we assume:
    - $\infty - \infty = 0$
    - $\infty - \text{finite} = \infty$



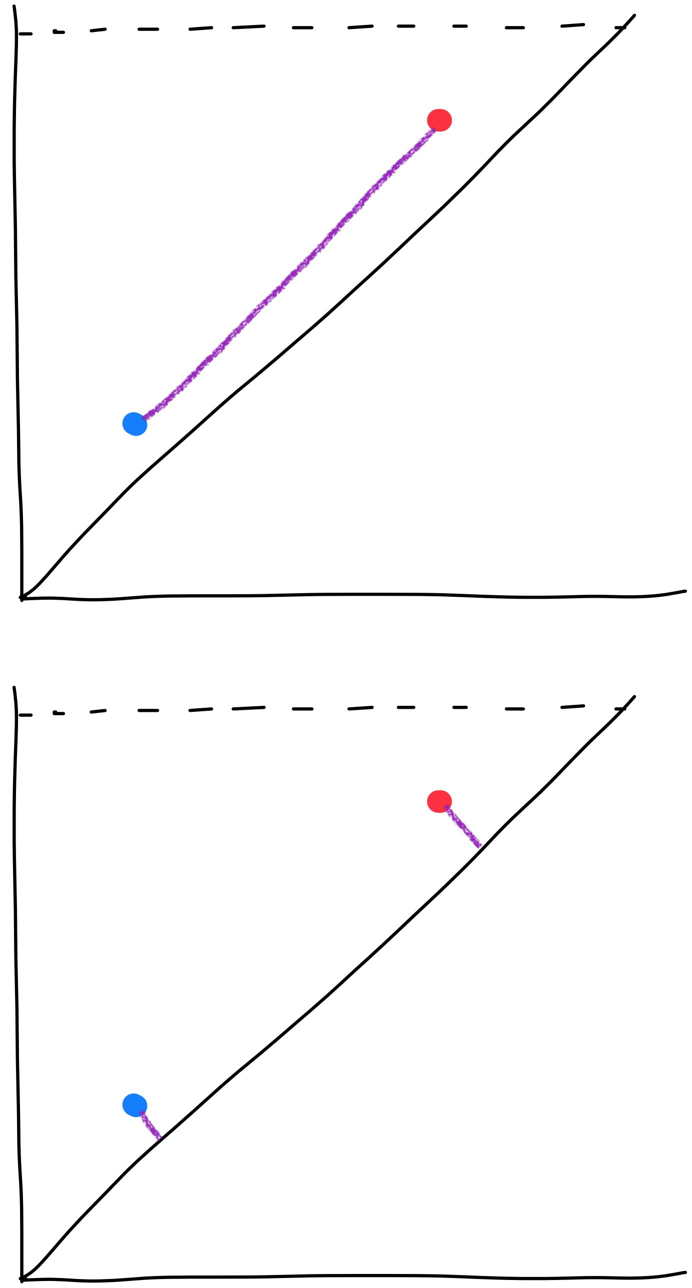
Distance Between Point and Diagonal

- Let $\Delta = \{(x, y) \mid x = y\}$ be the set of the diagonal points.
- Distance between a point and the diagonal is $\|p - \Delta\|_\infty = \dfrac{|b - d|}{2}$
  - $\|p - \Delta\|_\infty$ is the distance between p and its closest point on the diagonal



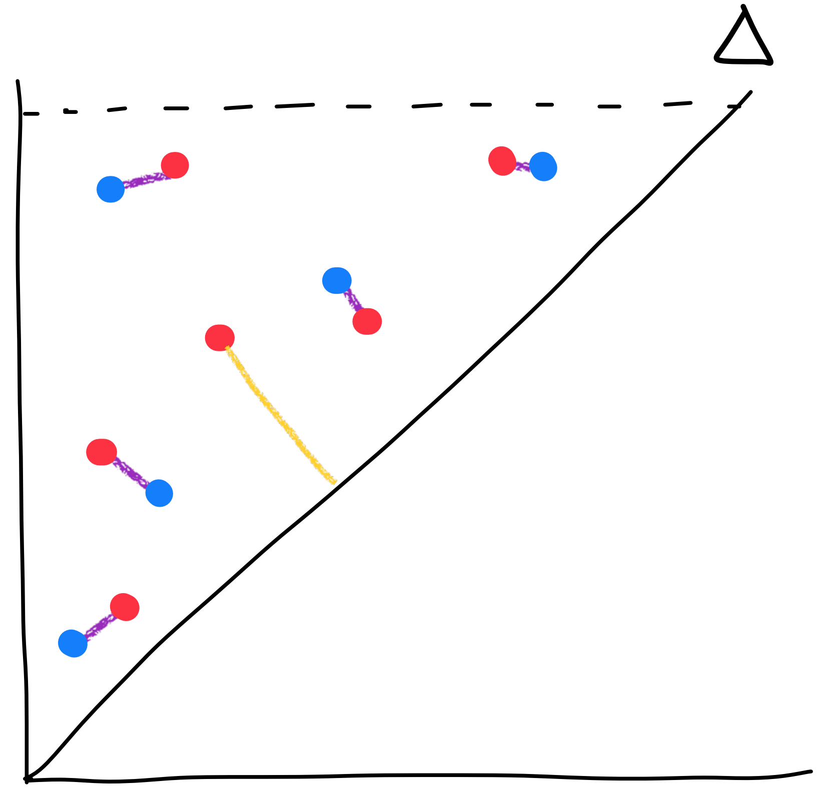
Motivation

- Points close to the diagonal
  - short bars, i.e. short lifetime
  - non-important features
- Non-important features should NOT be matched with large costs.



Motivating examples

- Two persistence diagrams D and D' may have different number of points
- There may be no bijection between D and D'
- Strategy:
  - Match part of D and part of D'
  - Compute ℓ^∞ between matched pairs
  - Record "importance" of unmatched points; (equivalently, these points are matched to Δ)



Partial-matching: how to match points

- 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** between D_1 and D_2 is $M \subseteq D_1 \times D_2 \quad s.t.$
  - $\forall p \in D_1, \ \exists \ at\ most\ one \ (p, x) \in M$
  - $\forall q \in D_2, \ \exists \ at\ most\ one \ (x, q) \in M$
- If $(p, q) \in M$, we say p is **matched** with q.

$M_1 = \{(p_1, q_1)\}$

<u>Example</u> $D_1 = \{p_1, p_2\}$

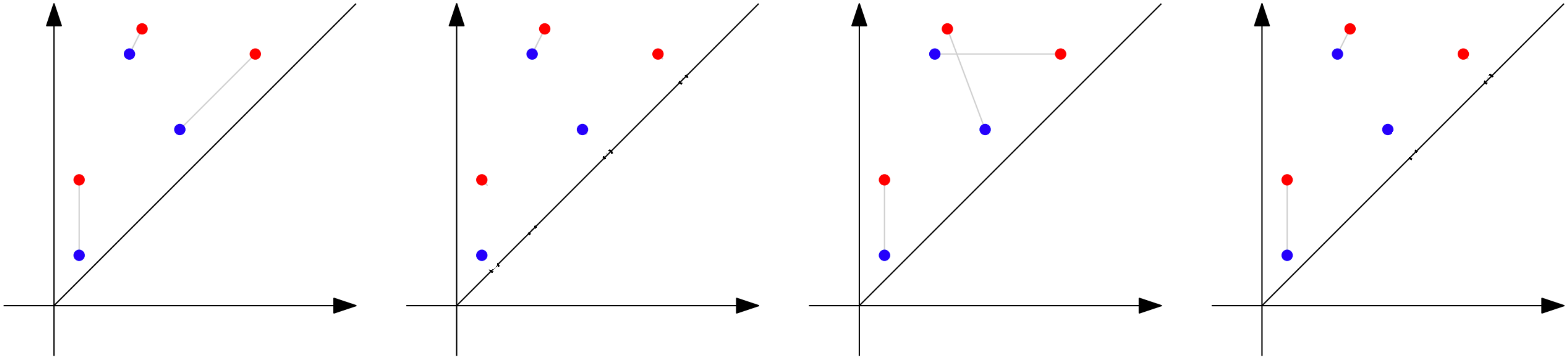
$M_2 = \{(p_1, q_1), (p_1, q_2)\}$

$D_2 = \{q_1, q_2\}$

$M_3 = \{(p_1, q_2), (p_2, q_1)\}$

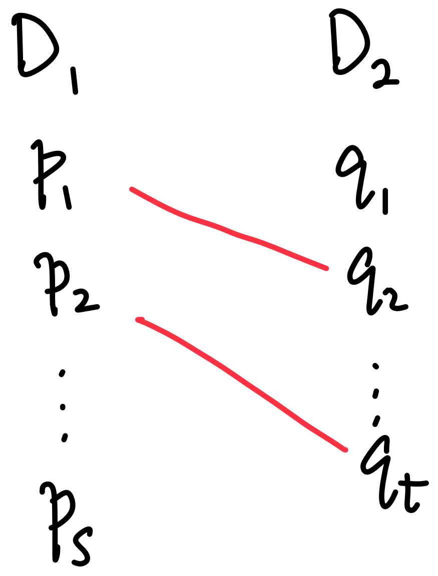
Partial-matching: how to match points

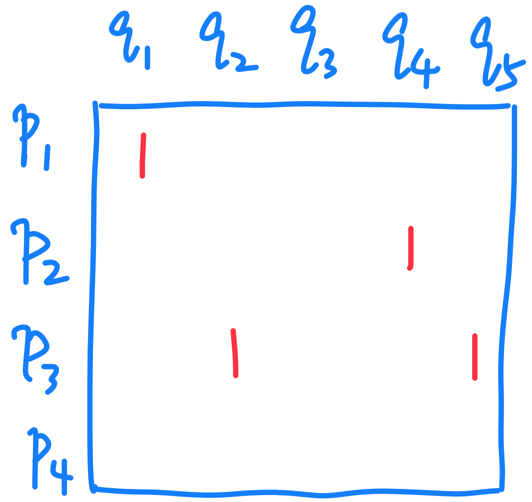
Examples of partial matchings:



Partial-matching: how to match points

- Different ways of understanding partial matching:
  - subset of $D_1 \times D_2$
  - bipartite graph
    - two lists of nodes
    - edges connect nodes in different lists
    - every node has at most one edge
  - binary matrix
    - at most one 1 in every row and every column

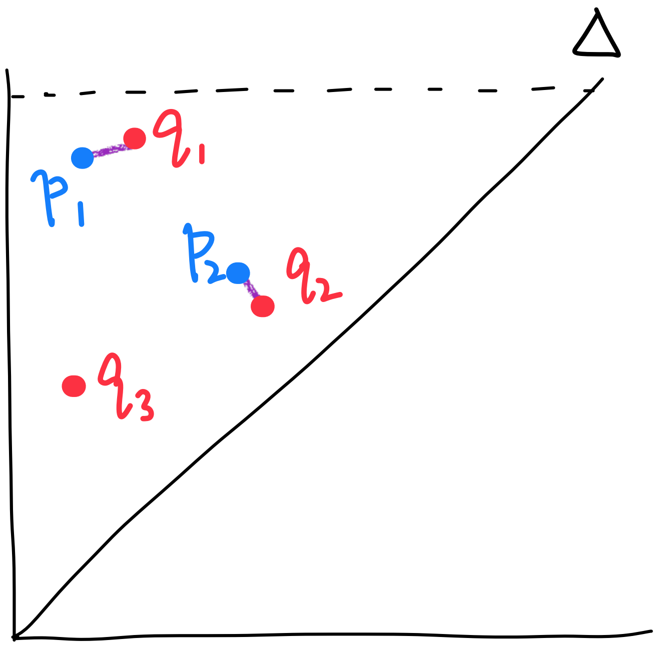




Cost of partial-matching

- The cost of a partial matching $M \subset D_1 \times D_2$ is

$$cost(M) = \max\left(\max_{(p,q)\in M} \|p - q\|_\infty, \max_{p \text{ unmatched}} \|p - \Delta\|_\infty \right)$$



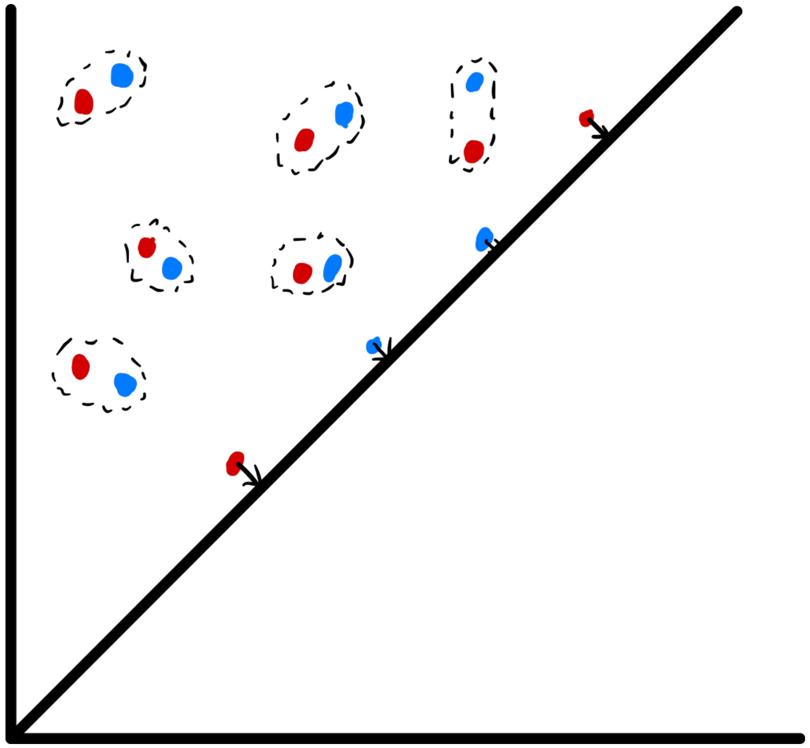
$D_1 = \{p_1, p_2\}$ $\quad D_2 = \{q_1, q_2, q_3\}$

$M = \{(p_1, q_1), (p_2, q_2)\}$

$$cost(M) = \max\{\|p_1 - q_1\|_\infty, \|p_2 - q_2\|_\infty, \|\Delta - q_3\|_\infty\}$$

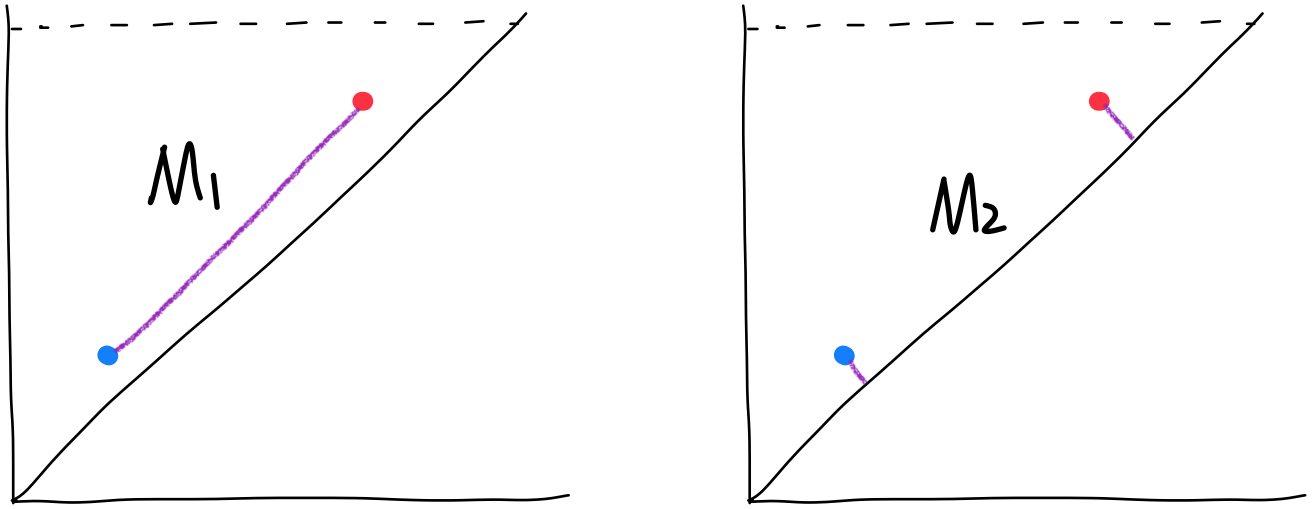
Bottleneck distance: optimal partial-matching

- *[Cohen-Steiner, Edelsbrunner, Harer, DCG 2007]*
- The bottleneck distance between D_1 and D_2 is
  - $d_B(D_1, D_2) := \min\{cost(M) \mid M \subset D_1 \times D_2 \text{ a partial matching}\}$



Bottleneck distance between 1-point PDs

- Assume that $D = \{p\}$ and $D' = \{q\}$ with p=(b,d) and q=(b', d').
- There are only two possible partial matchings:
  - $M_1 = \{(p, q)\}$ with $cost(M_1) = \|p - q\|_\infty$
  - $M_2 = \varnothing$ with $cost(M_2) = \max(\|p - \Delta\|_\infty, \|q - \Delta\|_\infty)$
- Thus, $d_B(D, D') = \min\left(\max(|b - b'|, |d - d'|), \max\left(\frac{|b - d|}{2}, \frac{|b' - d'|}{2}\right)\right)$



Bottleneck distance is an extended metric

- The bottleneck distance defines a metric on the set of PDs:
  - $d_B(D, D') = 0$ iff $D = D'$
  - $d_B(D, D') = d_B(D', D)$
  - $d_B(D, D') \leq d_B(D, D'') + d_B(D'', D')$
- We call d_B an *extended* metric, because it can take value ∞

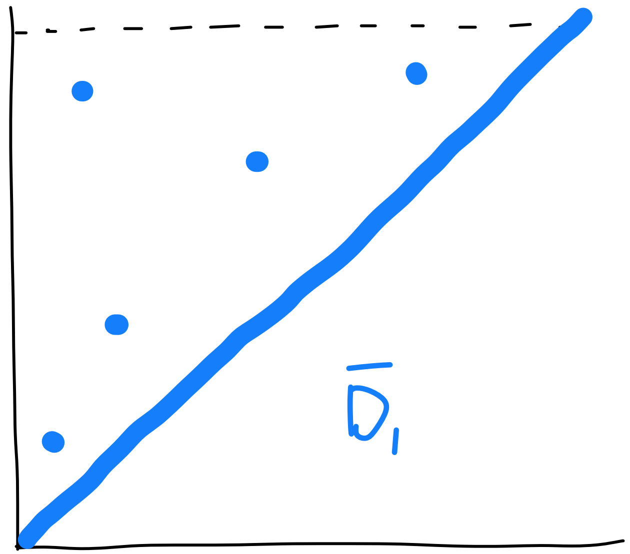
Example: $D = \{ \underbrace{(0, \infty)}_{=:p} \}$, $D' = \emptyset$.

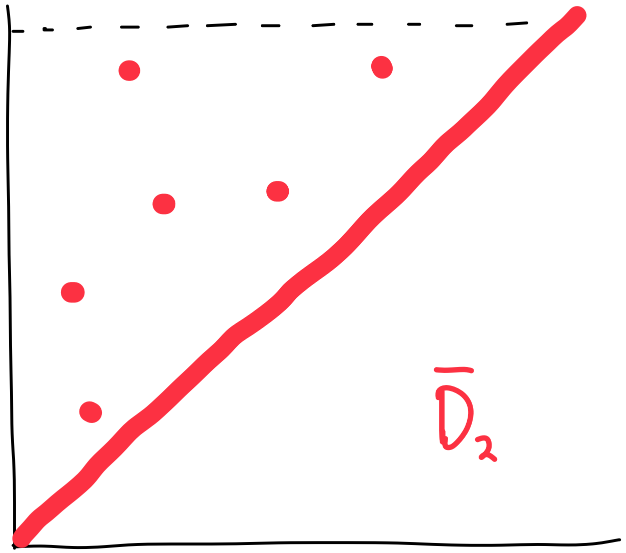
The only partial-matching is $M = \emptyset$

$$\Rightarrow d_B(D, D') = \text{cost}(M) = \|p - \Delta\|_\infty = \frac{\infty - 0}{2} = \infty$$

Alternative formulation of the 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 $D_2 = \{q_1, q_2, \ldots, q_t\}$
- Augment $\bar{D}_1 := D_1 \cup \Delta^\infty$ and $\bar{D}_2 := D_2 \cup \Delta^\infty$
  - where Δ^∞ is the diagonal points each with **infinite multiplicity**





- When PDs have different cardinalities, we cannot talk about bijections.
- Adding the extra copies of diagonal points forces the same cardinality.

Alternative formulation of the bottleneck distance

- A partial-matching between D_1 and D_2 is a **bijection** $\bar{M} \subset \bar{D}_1 \times \bar{D}_2$ with $cost(\bar{M}) := \max_{(p,q)\in\bar{M}} \|p - q\|_\infty$

For any p, let $\tilde{p} \in \Delta$ be p's closest point in Δ

$\forall M \subset D_1 \times D_2$, define $\bar{M} := M \cup \{(p, \tilde{p}) : p \in D_1 \text{ unmatched}\}$

$\cup \{(\tilde{q}, q) : q \in D_2 \text{ unmatched}\}$

$\cup \{(\tilde{p}, \tilde{p}) : \tilde{p} \in \Delta\}$

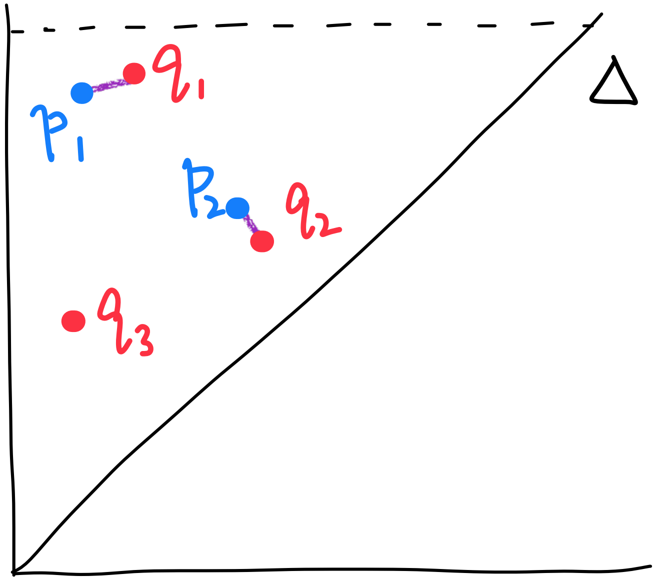
$\forall \bar{M} \subset (D_1 \cup \Delta^\infty) \times (D_2 \cup \Delta^\infty)$, define $M = \{(p, q) \in \bar{M} \mid p \notin \Delta, q \notin \Delta\}$

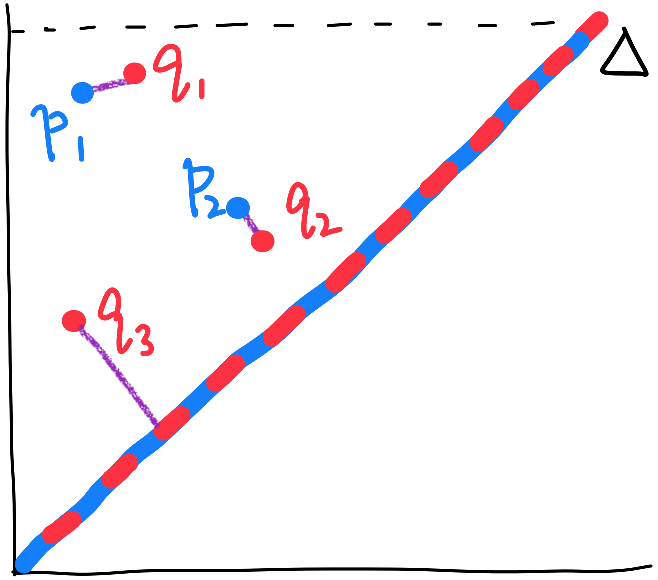
Alternative formulation of the bottleneck distance

- The bottleneck distance between D_1 and D_2 is

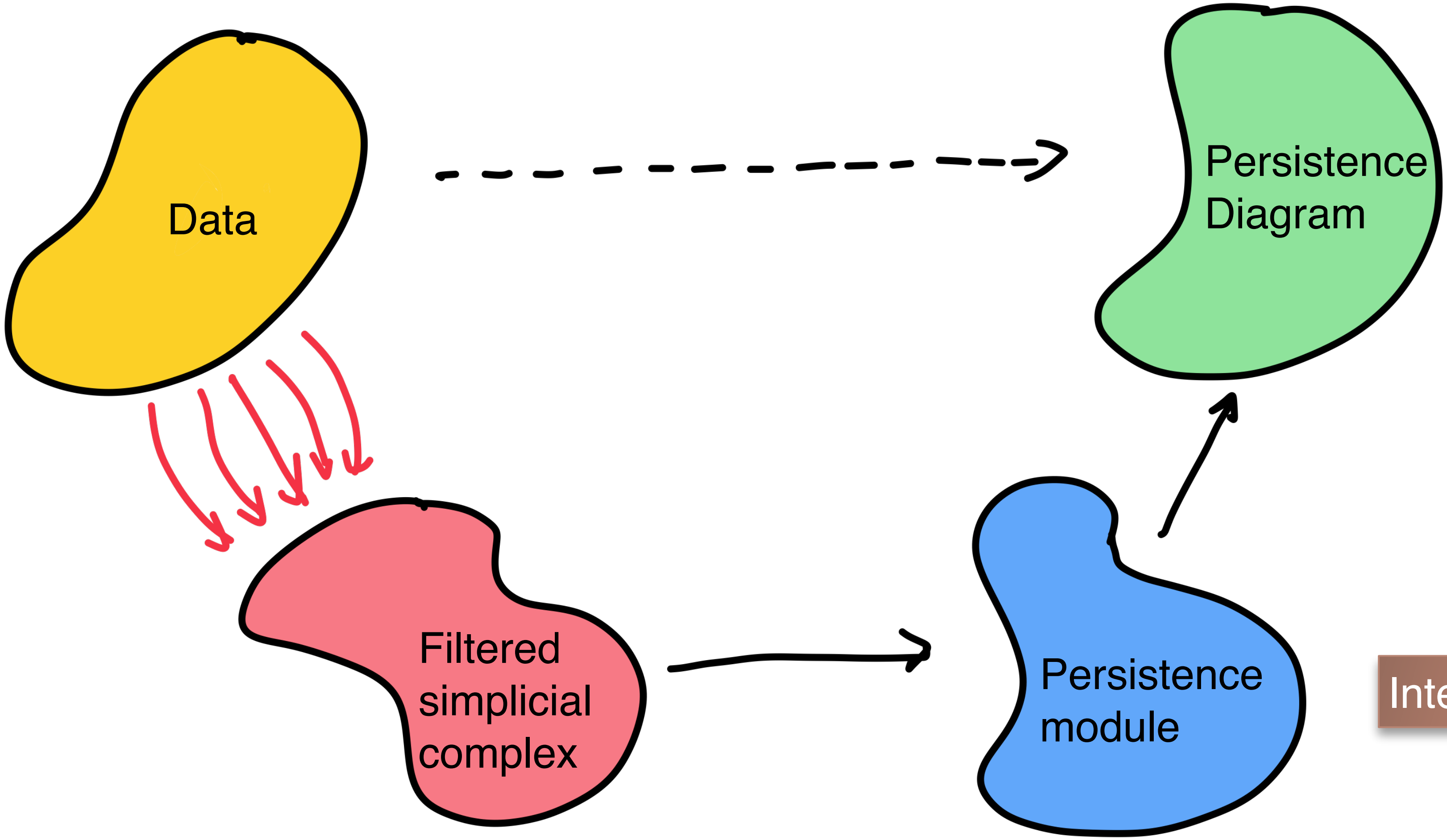
$$d_B(D_1, D_2) := \min\{cost(\bar{M}) \mid \bar{M} \subset \bar{D}_1 \times \bar{D}_2 \text{ a bijection}\} = \min_{\bar{M}} \max_{(p,q)\in\bar{M}} \|p - q\|_\infty$$

$$\left\{\begin{array}{c}\text{partial-matchings} \\ M \subset D_1 \times D_2\end{array}\right\} \overset{1-1}{\longleftrightarrow} \left\{\begin{array}{c}\text{bijections} \\ \bar{M} \subset \bar{D}_1 \times \bar{D}_2\end{array}\right\}$$





Interleaving distance between Persistence Modules



Interleaving distance

Intuition

- Isomorphic persistence modules

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

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

Intuition

- Isomorphic persistence modules

U: $\cdots \to U_a \to \cdots \to U_b \to \cdots \to \cdots$

V: $\cdots \to V_a \to \cdots \to V_b \to \cdots \to \cdots$

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

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

Intuition

- Isomorphic persistence modules

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

$\ldots \qquad \ldots \qquad \ldots$

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

ϵ-Interleaving

- U and V are ϵ-**interleaved** 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
- The maps φ_a and ϕ_a are called an ϵ-**interleaving**

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

ϵ-Interleaving

- U and V are **ϵ-interleaved** 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):

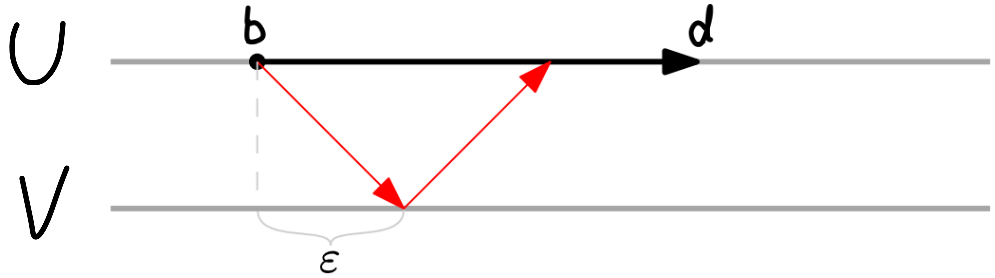
$$\begin{tikzcd}
U_a \arrow[rr, "u_{a,a'}"] \arrow[dr, "\varphi_a"] & & U_{a'} \arrow[dr, "\varphi_{a'}"] & \\
& V_{a+\varepsilon} \arrow[rr, "v_{a+\varepsilon,a'+\varepsilon}"'] & & V_{a'+\varepsilon}
\end{tikzcd}
\qquad
\begin{tikzcd}
& U_{a+\varepsilon} \arrow[rr, "u_{a+\varepsilon,a'+\varepsilon}"]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

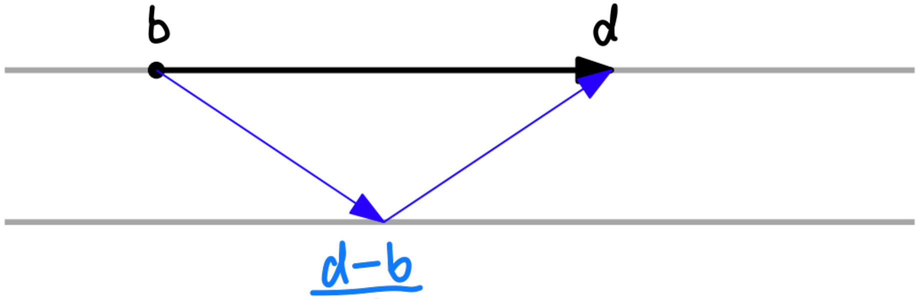
$$\begin{tikzcd}
U_a \arrow[rr, "u_{a,a+2\varepsilon}"] \arrow[dr, "\varphi_a"'] & & U_{a+2\varepsilon} \\
& V_{a+\varepsilon} \arrow[ur, "\psi_{a+\varepsilon}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
& U_{a+\varepsilon} \arrow[dr, "\varphi_{a+\varepsilon}"] & \\
V_a \arrow[ur, "\psi_a"] \arrow[rr, "v_{a,a+2\varepsilon}"]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

ϵ-Interleaving

- U and V are ϵ-**interleaved** 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

Example: $U = \mathbb{F}[b, d)$, $V = 0$ Let $\varphi_a = 0$ & $\phi_a = 0$





NOT ε-interleaved for $\varepsilon < \frac{d-b}{2}$

ε-interleaved for $\varepsilon \geq \frac{d-b}{2}$

ϵ-Interleaving

- U and V are **ϵ-interleaved** 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

ϕ_a & φ_a are linear iso.

$\Uparrow$

- Proposition: If U and V are 0-interleaved, then they are isomorphic

$\Downarrow$

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

$$\begin{tikzcd}
U_a \arrow[rr, "id_{\mathbb{F}}"] \arrow[dr, "\varphi_a"'] & & U_a \\
 & V_a \arrow[ur, "\phi_a"'] &
\end{tikzcd}$$

$$\begin{tikzcd}
 & U_a \arrow[dr, "\phi_a"] & \\
V_a \arrow[ur, "\varphi_a"] \arrow[rr, "id_{\mathbb{F}}"'] & & V_a
\end{tikzcd}$$

ϕ_a & φ_a are inverse to each other.

Interleaving Distance

- The **interleaving distance** between two persistence modules V and U is $d_I(V, U) := \inf\{\epsilon > 0 \mid V \text{ and } U \text{ is } \epsilon\text{-interleaved}\}$

- It is an extended pseudo-metric
  - It is symmetric and it satisfies triangle inequality
    - $d_I(U, W) \leq d_I(U, V) + d_I(V, W)$
  - (extended) Can take value ∞
  - (pseudo) Non isomorphic persistence modules can have zero distance

Interleaving Distance

- The **interleaving distance** between two persistence modules V and U is

$d_I(V, U) := \inf\{\epsilon > 0 \mid V \text{ and } U \text{ is } \epsilon\text{-interleaved}\}$

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:

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