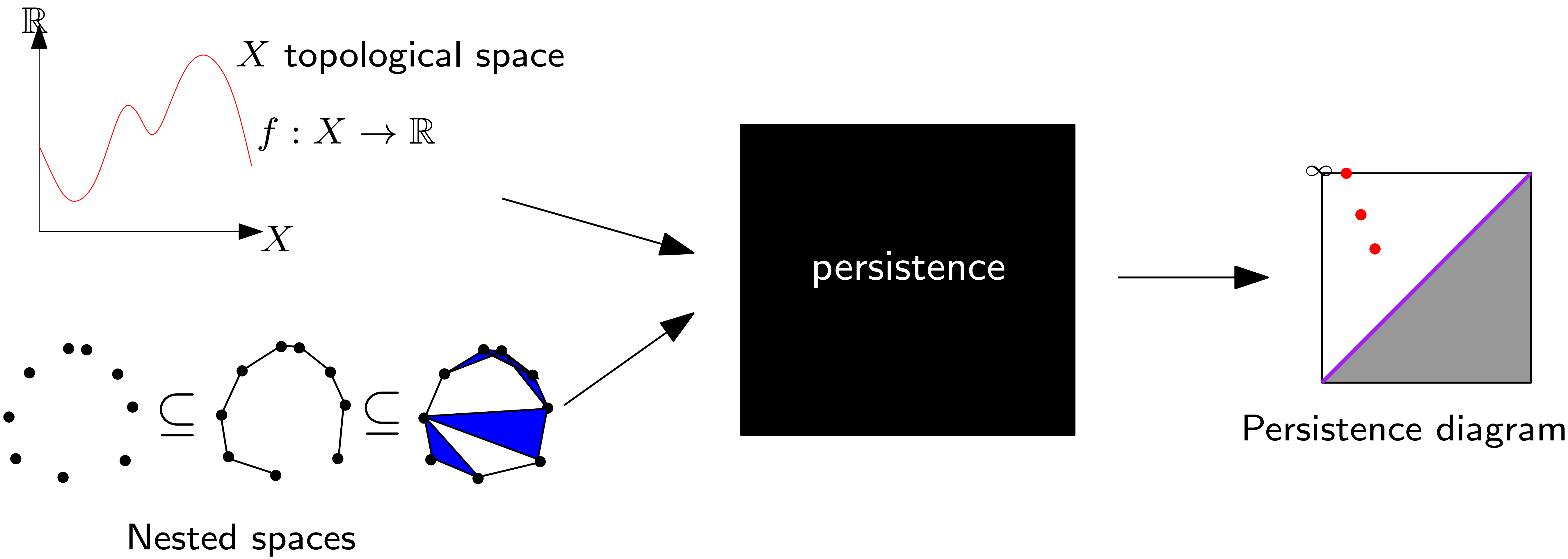
Sophia-Antipolis, January, 2017

Persistent homology in TDA

Frédéric Chazal
INRIA Saclay - Ile-de-France
frederic.chazal@inria.fr

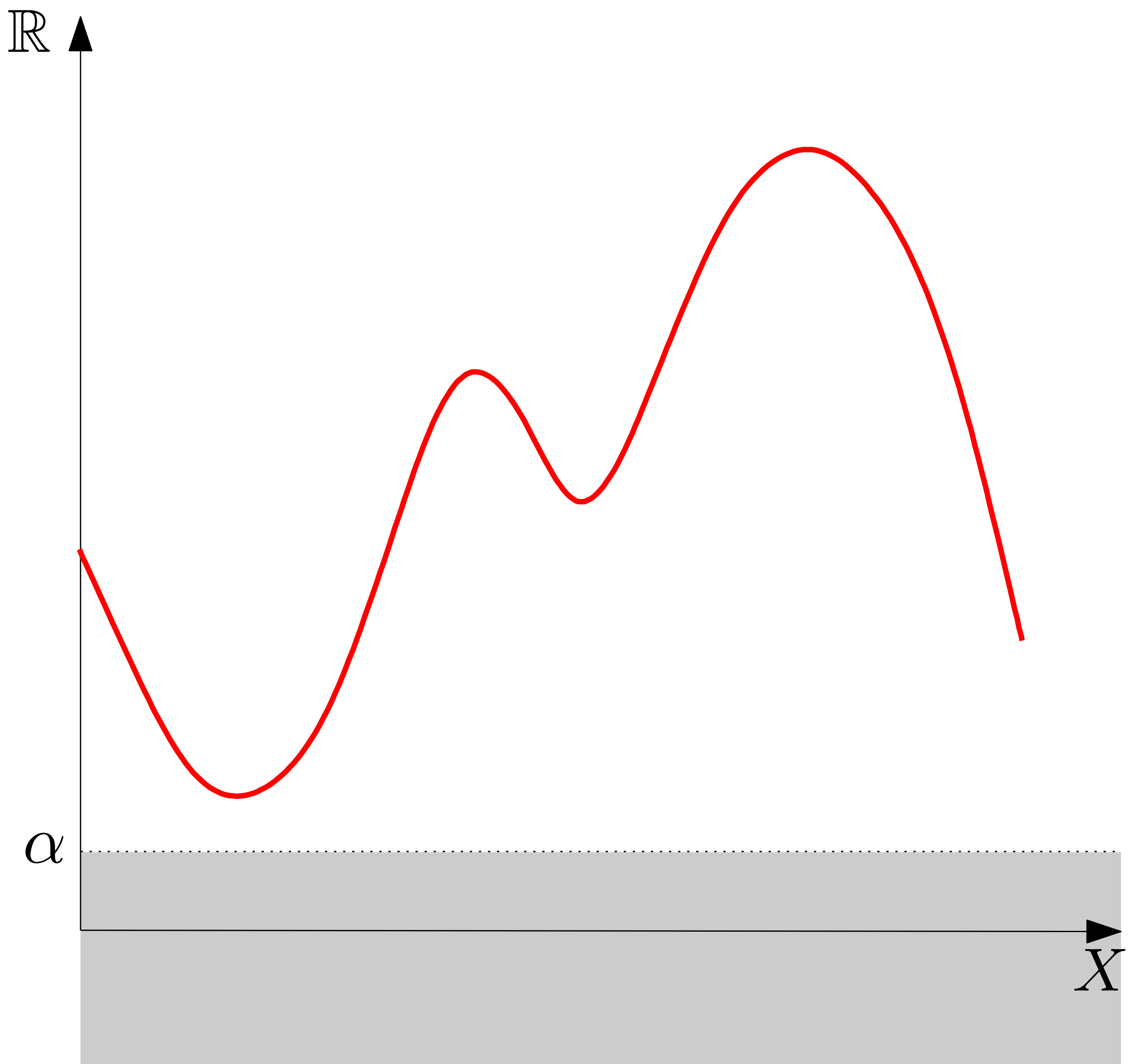
Persistent homology



- A general mathematical framework to encode the evolution of the topology (homology) of families of nested spaces (filtered complex, sublevel sets,...).
- Formalized by H. Edelsbrunner (2002) et al and G. Carlsson et al (2005) - wide development during the last decade.
- Multiscale topological information.
- Barcodes/persistence diagrams can be efficiently computed.
- Stability properties

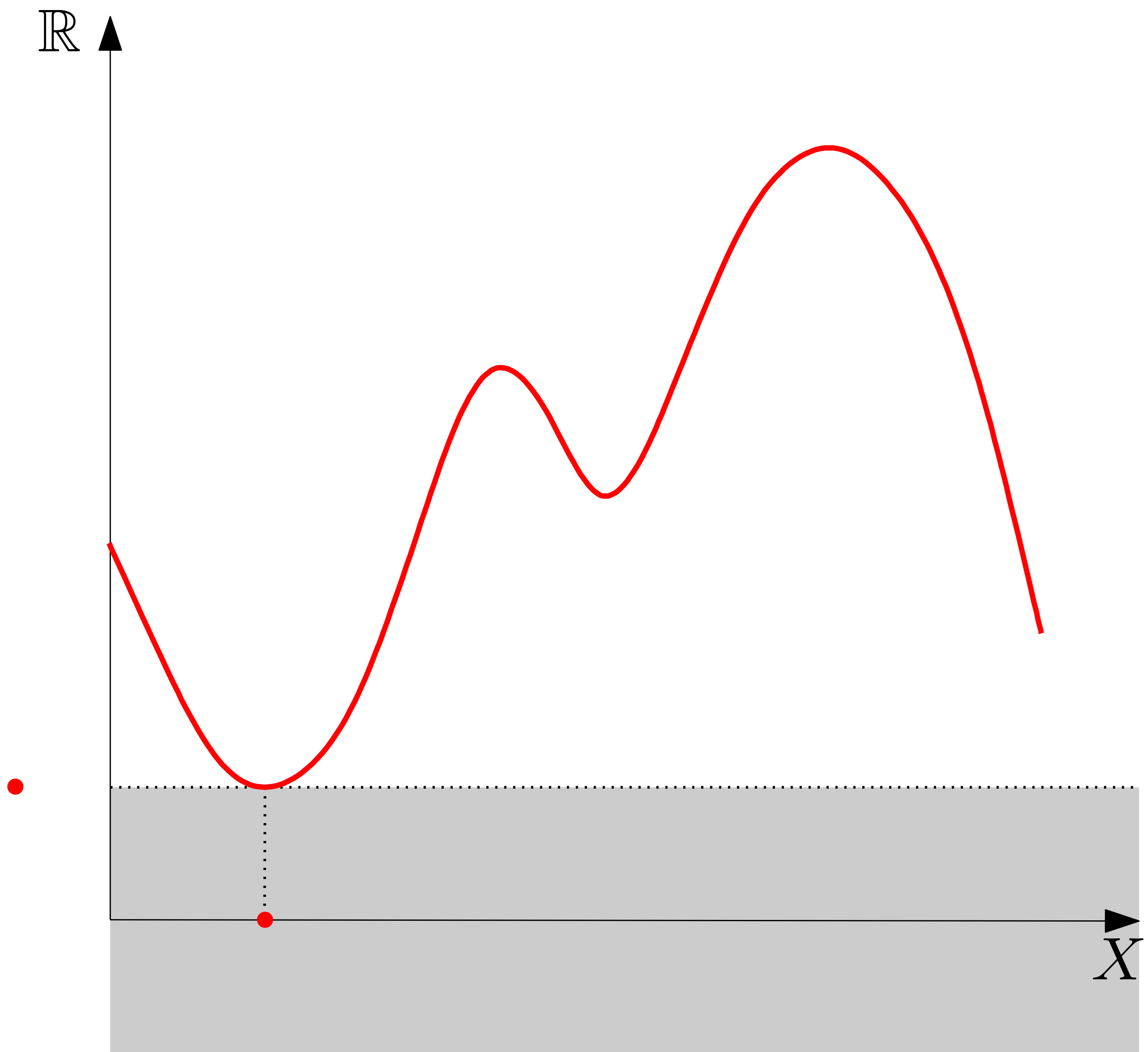
Persistent homology for functions

- Nested family (filtration) of sublevel-sets $f^{-1}((-\infty, \alpha])$ for $\alpha = -\infty$ to $+\infty$.
- Track evolution of homology throughout the family.



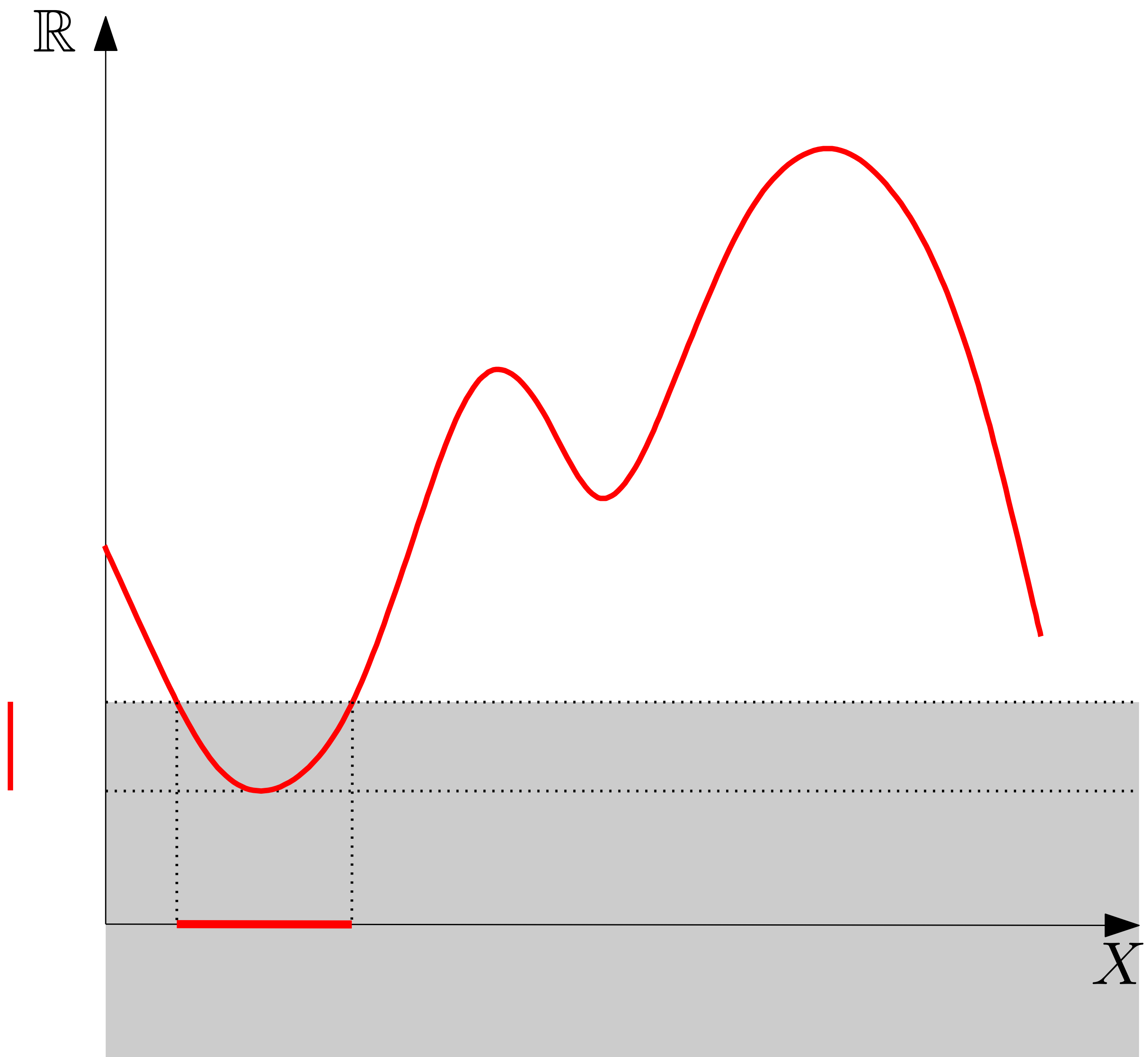
Persistent homology for functions

- Nested family (filtration) of sublevel-sets $f^{-1}((-\infty, \alpha])$ for $\alpha = -\infty$ to $+\infty$.
- Track evolution of homology throughout the family.



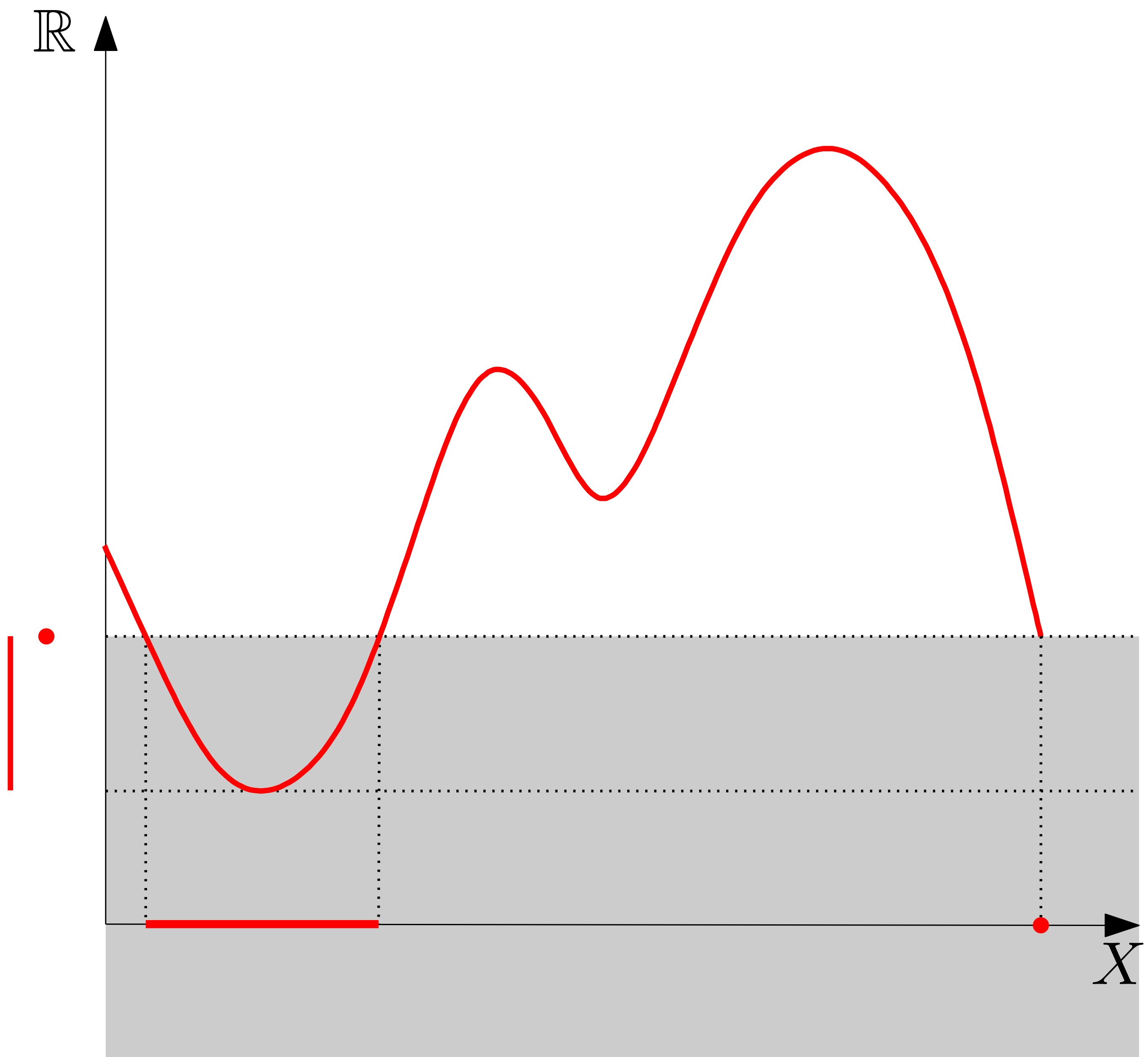
Persistent homology for functions

- Nested family (filtration) of sublevel-sets $f^{-1}((-\infty, \alpha])$ for $\alpha = -\infty$ to $+\infty$.
- Track evolution of homology throughout the family.



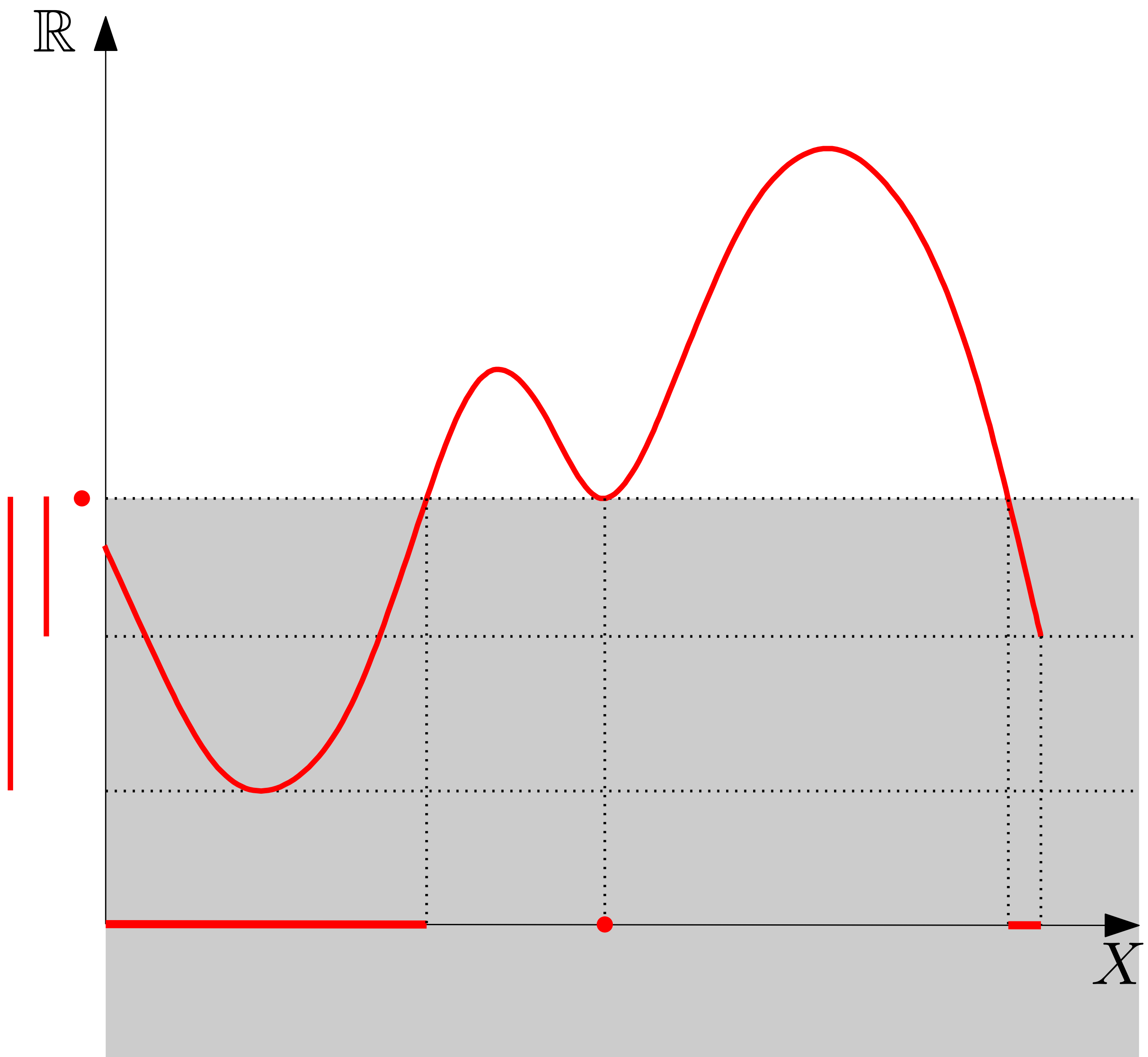
Persistent homology for functions

- Nested family (filtration) of sublevel-sets $f^{-1}((-\infty, \alpha])$ for $\alpha = -\infty$ to $+\infty$.
- Track evolution of homology throughout the family.



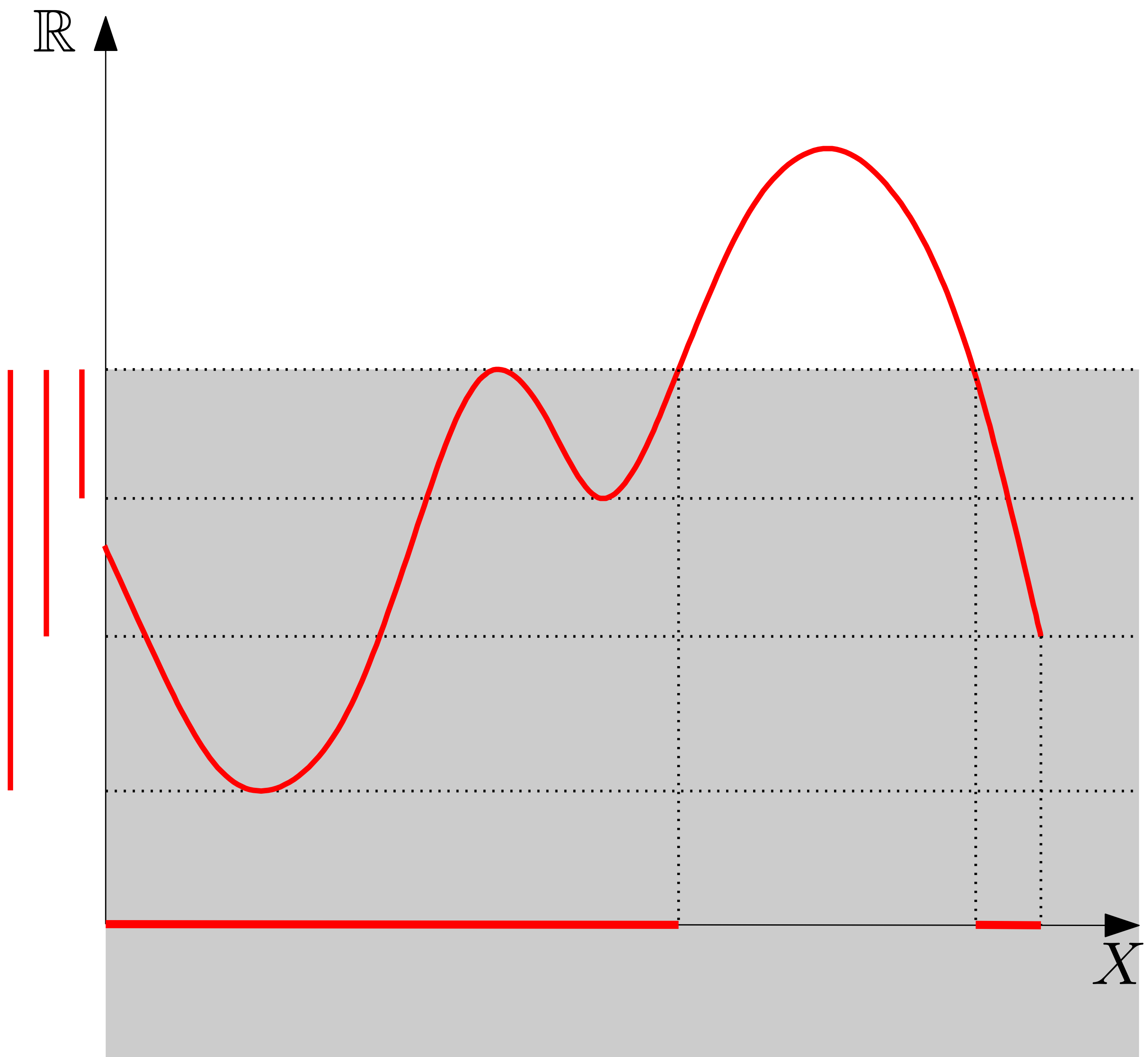
Persistent homology for functions

- Nested family (filtration) of sublevel-sets $f^{-1}((-\infty,\alpha])$ for $\alpha=-\infty$ to $+\infty$.
- Track evolution of homology throughout the family.



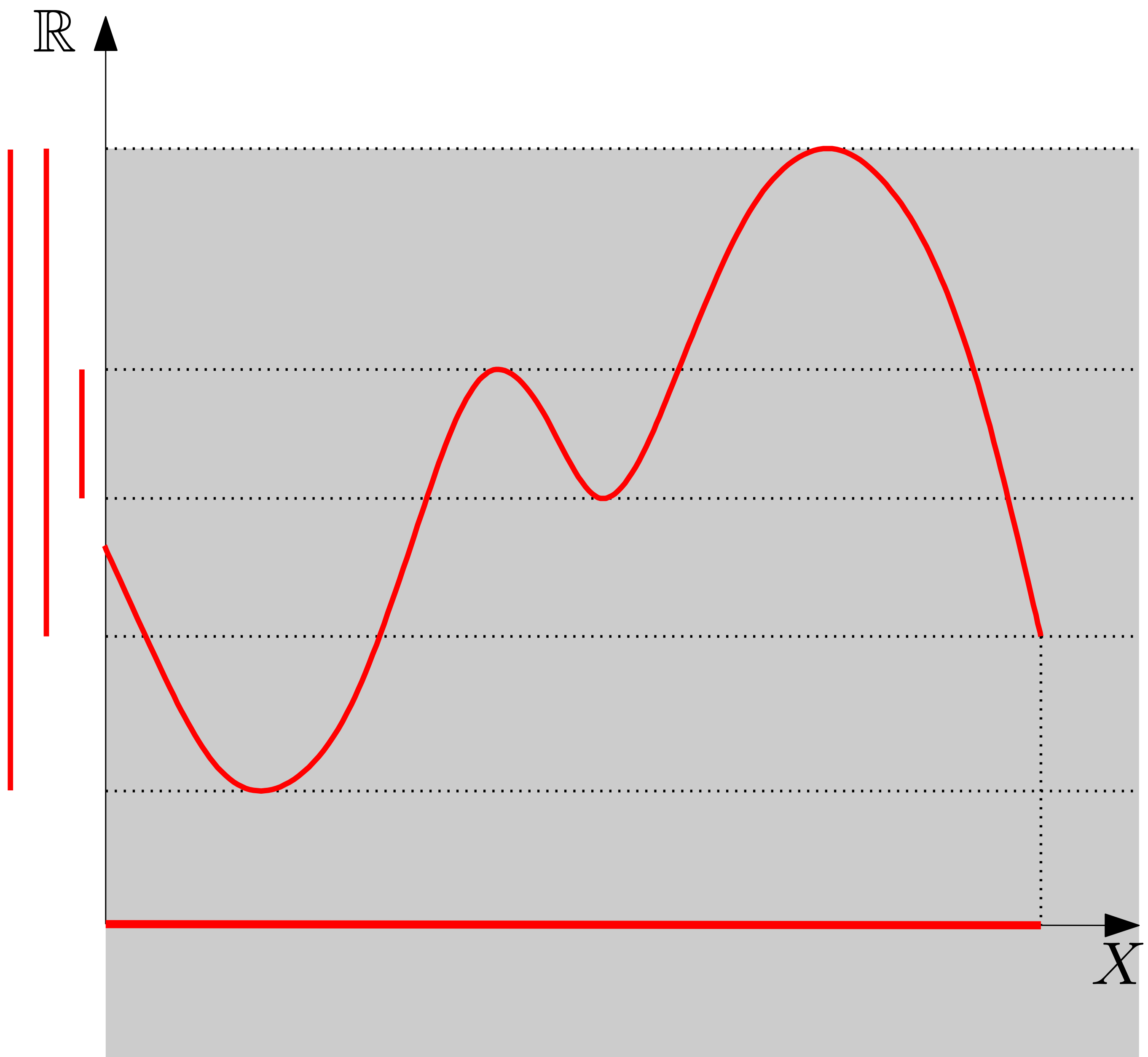
Persistent homology for functions

- Nested family (filtration) of sublevel-sets $f^{-1}((-\infty, \alpha])$ for $\alpha = -\infty$ to $+\infty$.
- Track evolution of homology throughout the family.



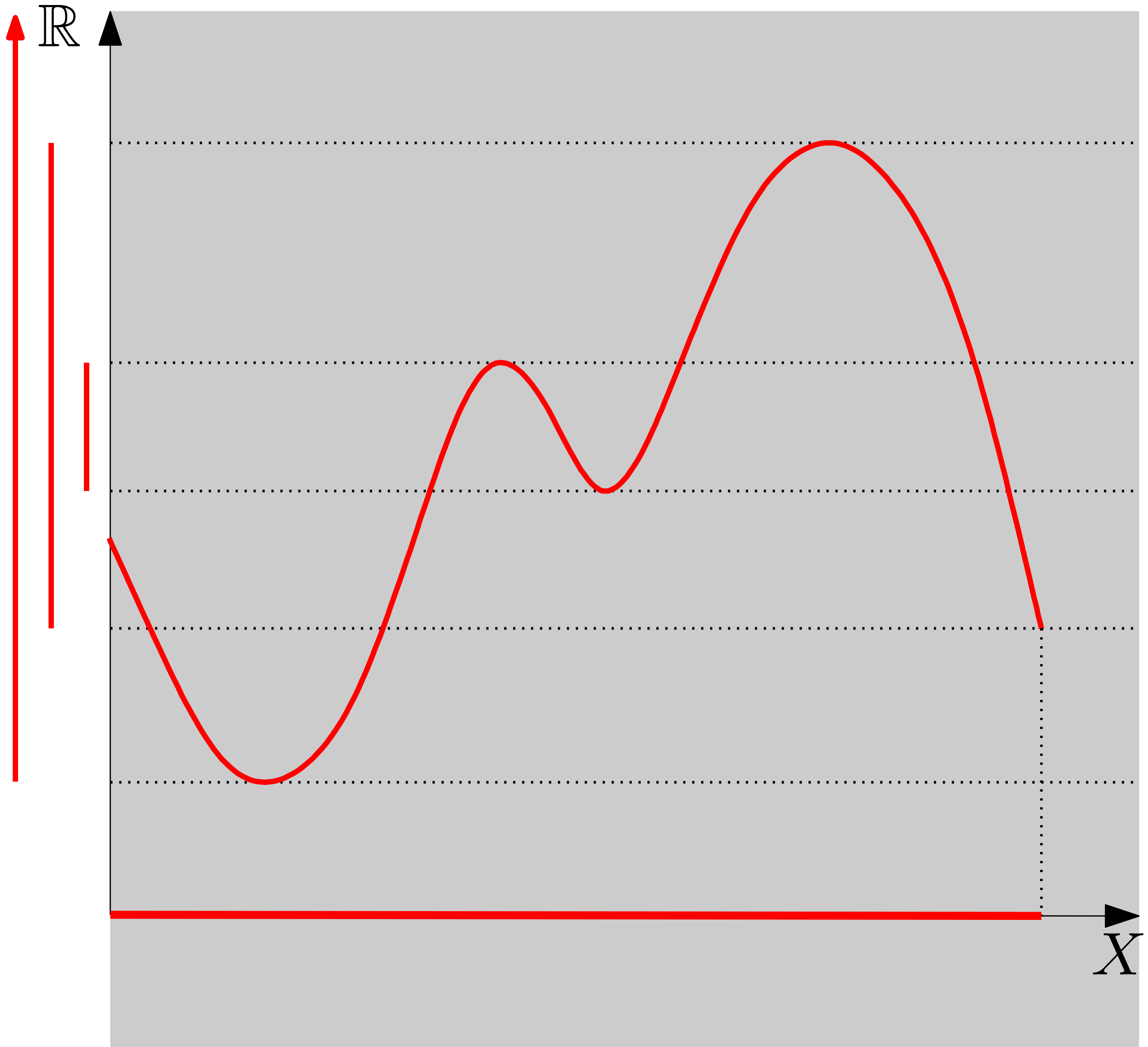
Persistent homology for functions

- Nested family (filtration) of sublevel-sets $f^{-1}((-\infty, \alpha])$ for $\alpha = -\infty$ to $+\infty$.
- Track evolution of homology throughout the family.



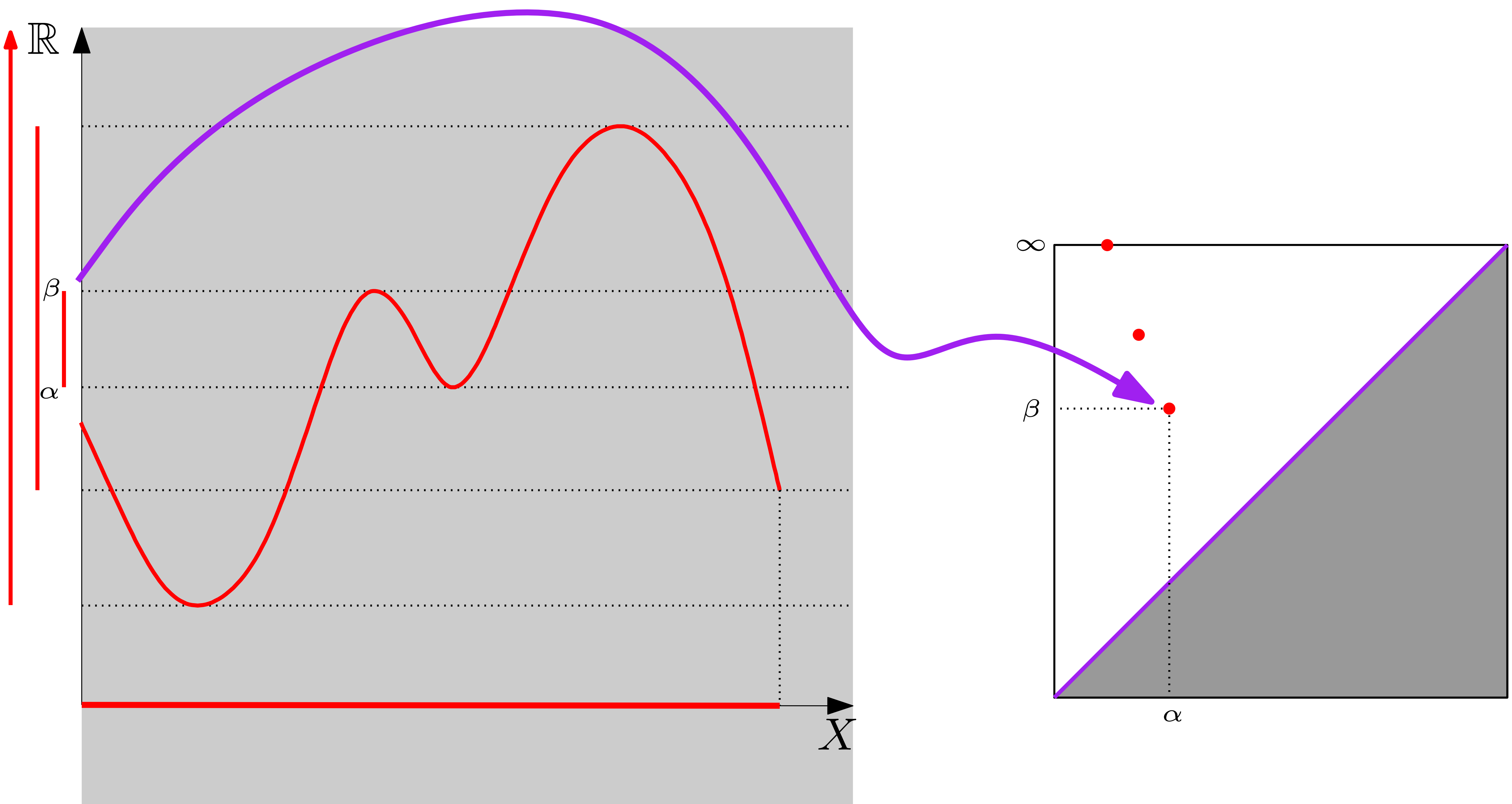
Persistent homology for functions

- Nested family (filtration) of sublevel-sets $f^{-1}((-\infty, \alpha])$ for $\alpha = -\infty$ to $+\infty$.
- Track evolution of homology throughout the family.
- Finite set of intervals (barcode) encodes births/deaths of topological features.



Persistent homology for functions

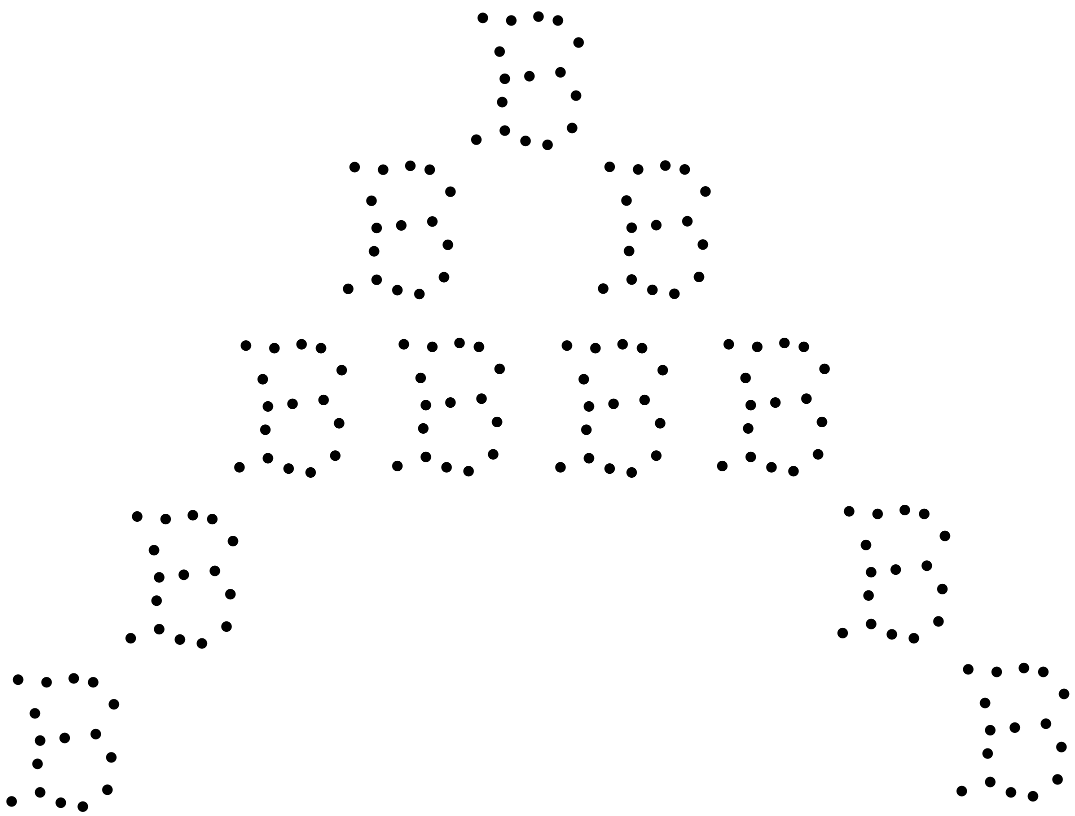
- Nested family (filtration) of sublevel-sets $f^{-1}((-\infty,\alpha])$ for $\alpha = -\infty$ to $+\infty$.
- Track evolution of homology throughout the family.
- Finite set of intervals (barcode) encodes births/deaths of topological features.



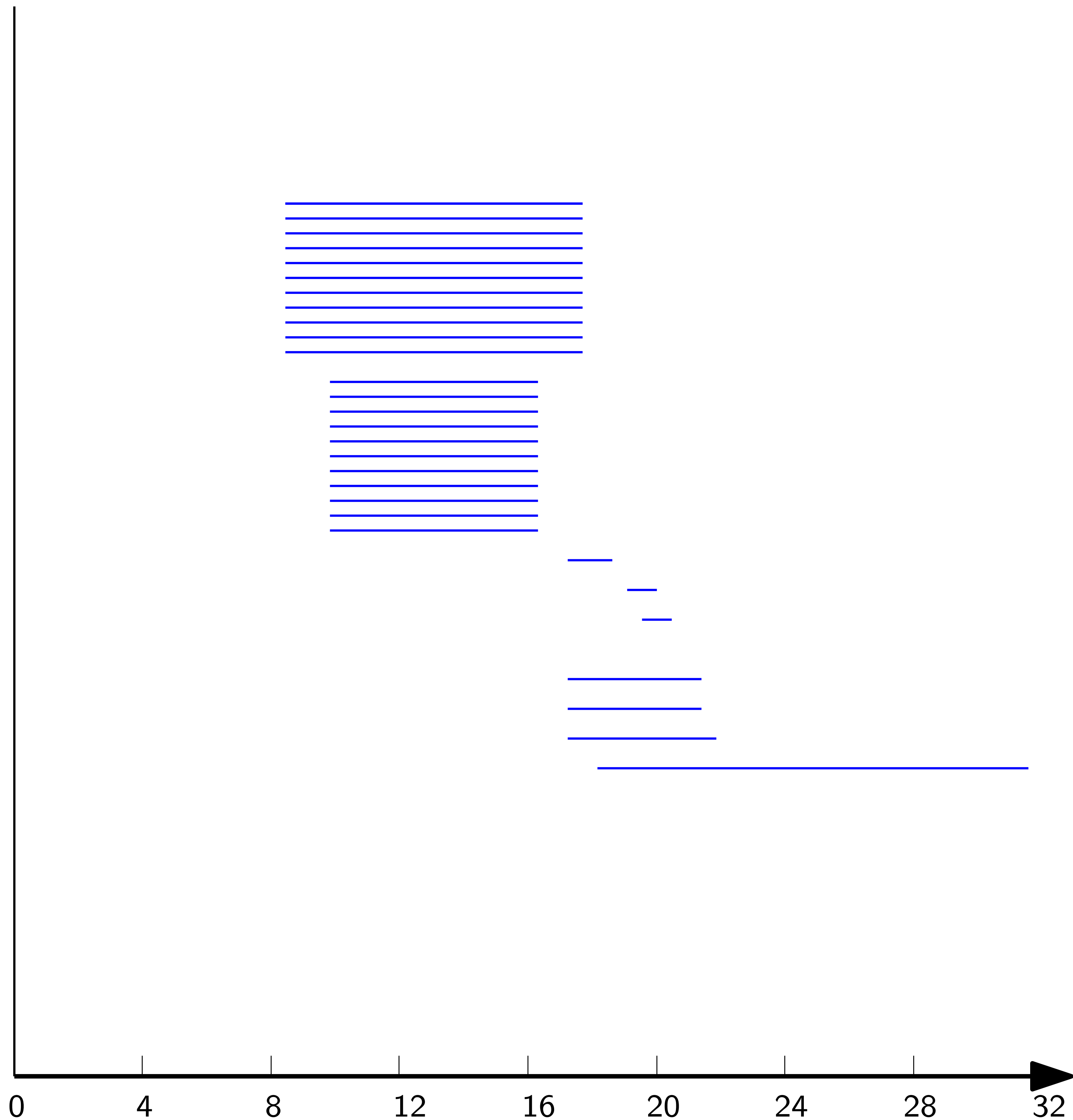
Persistent homology for functions

$$f_P : \mathbb{R}^2 \to \mathbb{R}$$

$$x \to \min_{p \in P} \|x - p\|_2$$



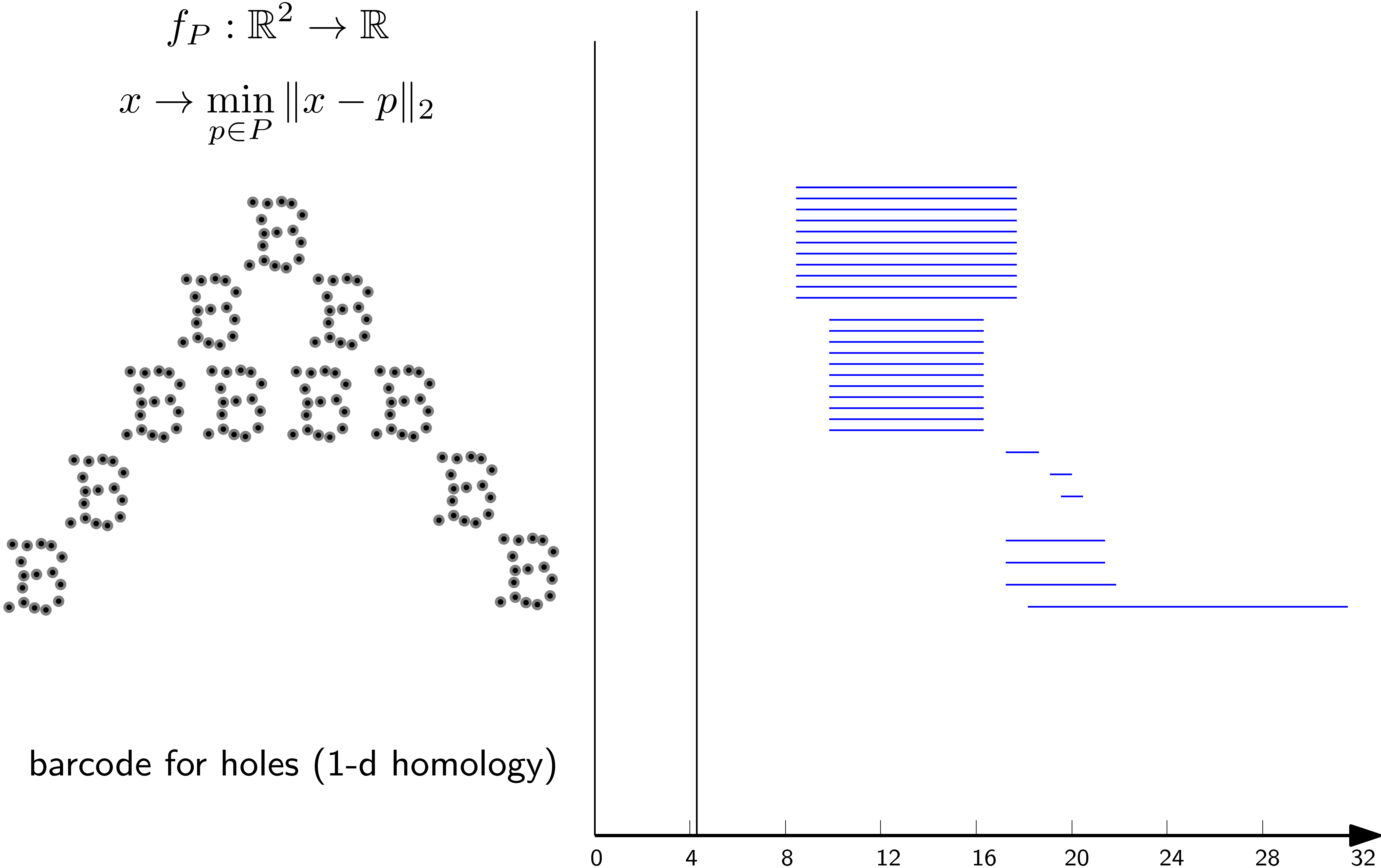
barcode for holes (1-d homology)



Persistent homology for functions

$$f_P : \mathbb{R}^2 \to \mathbb{R}$$

$$x \to \min_{p \in P} \|x - p\|_2$$

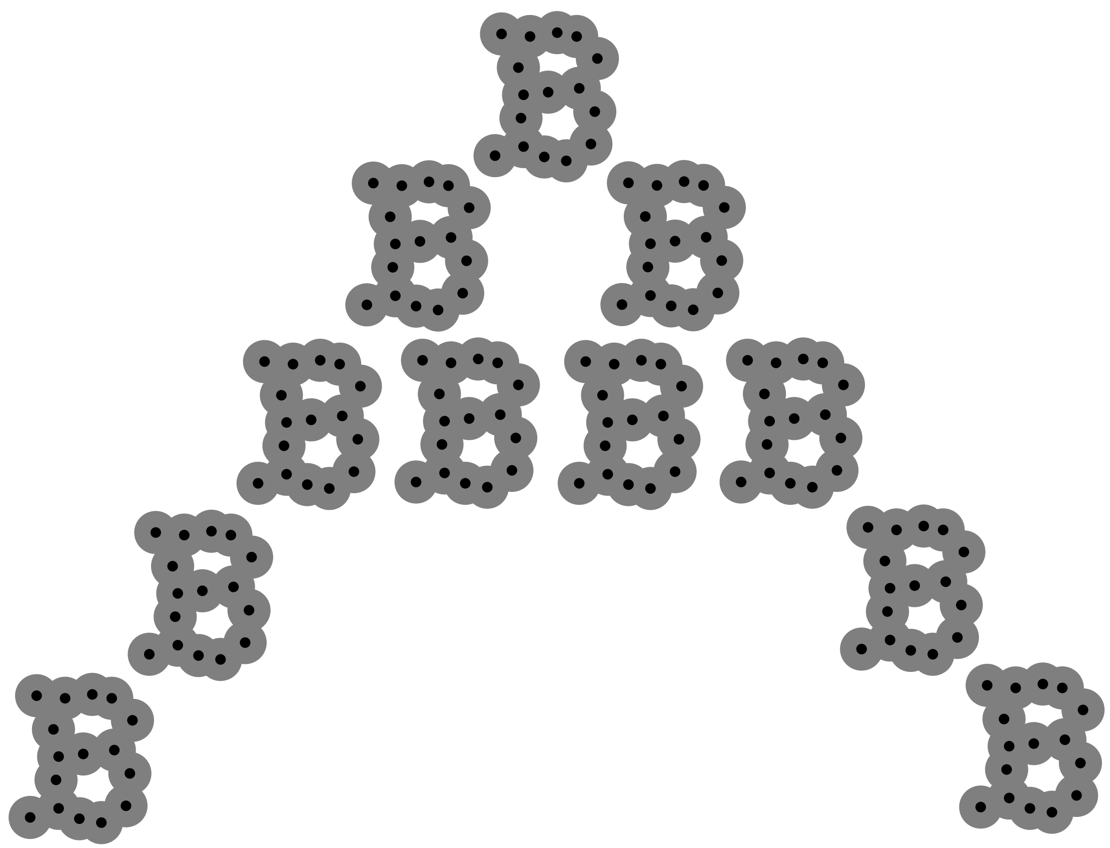


barcode for holes (1-d homology)

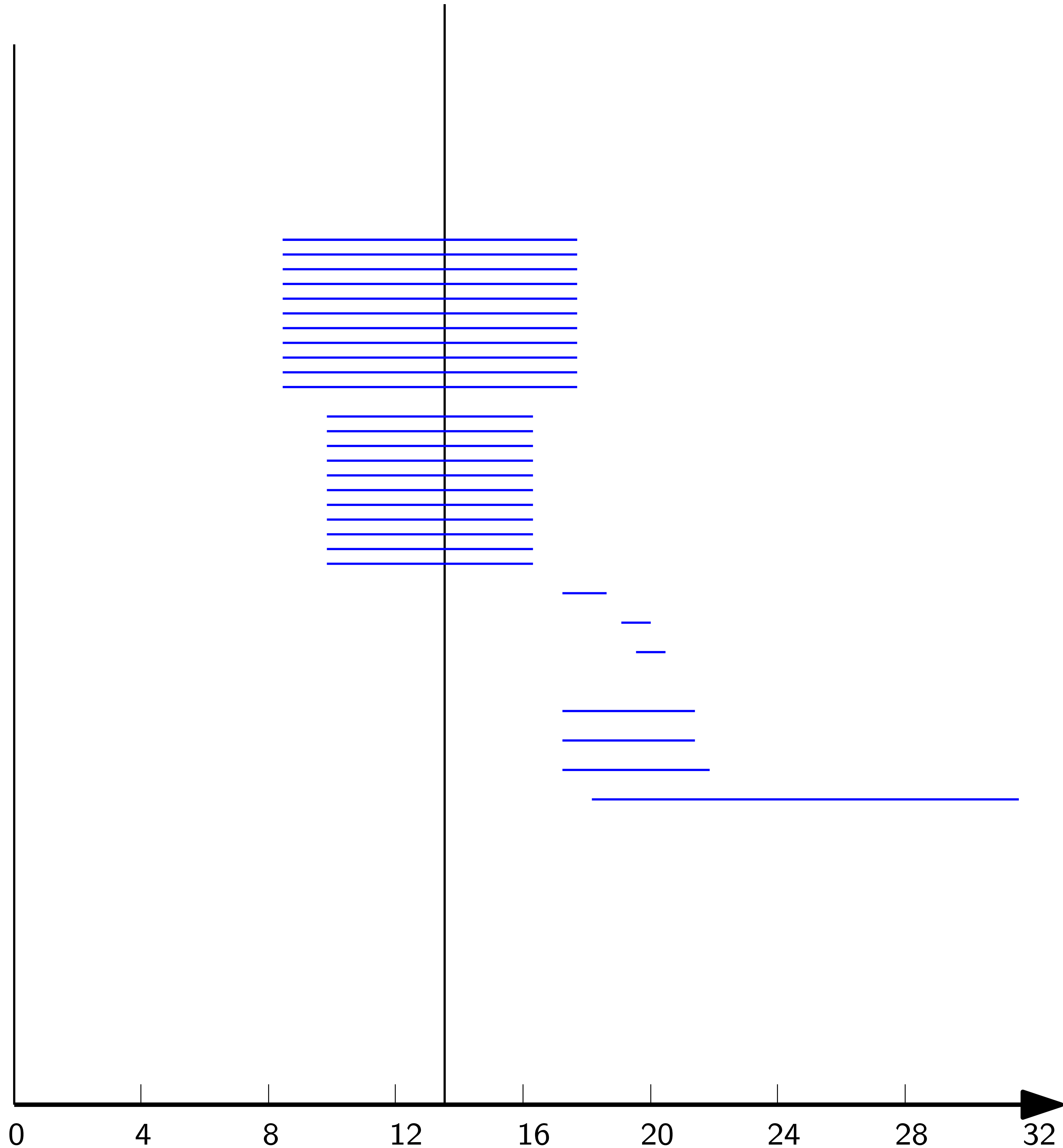
Persistent homology for functions

$$f_P : \mathbb{R}^2 \to \mathbb{R}$$

$$x \to \min_{p \in P} \|x - p\|_2$$



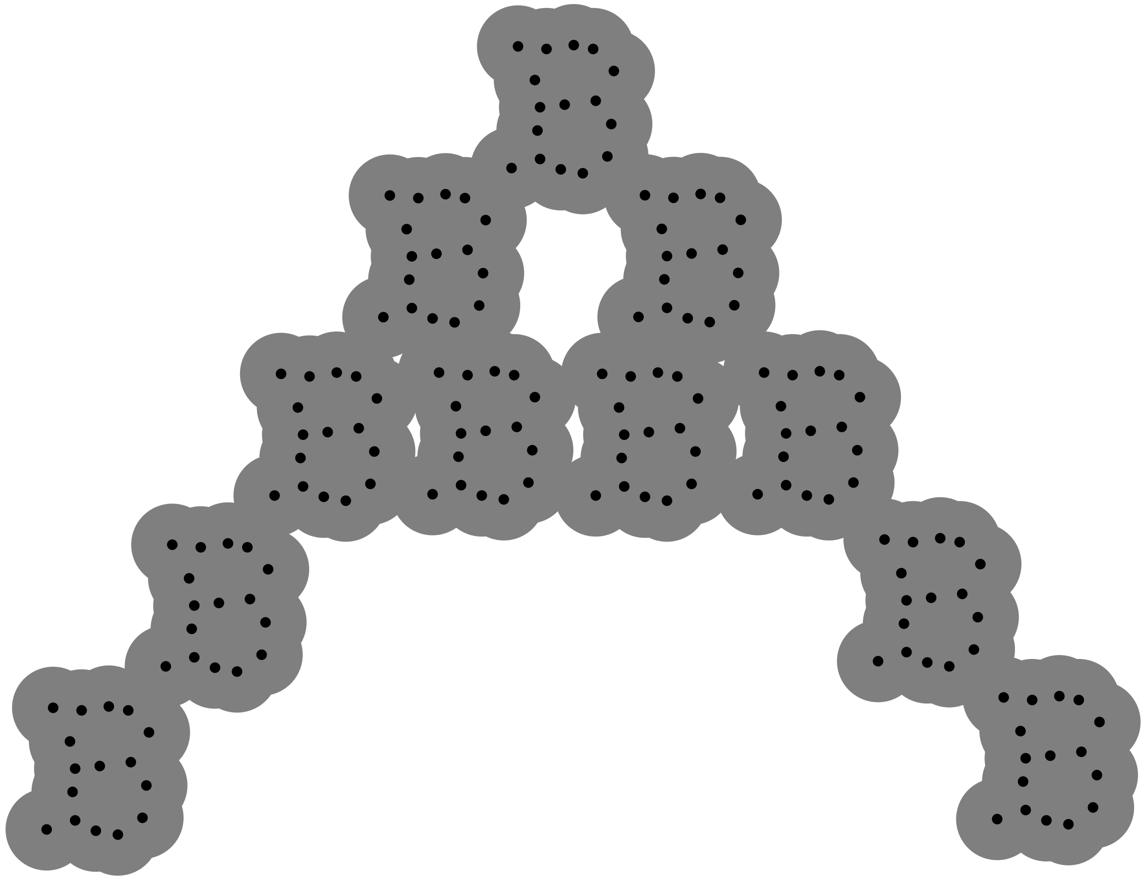
barcode for holes (1-d homology)



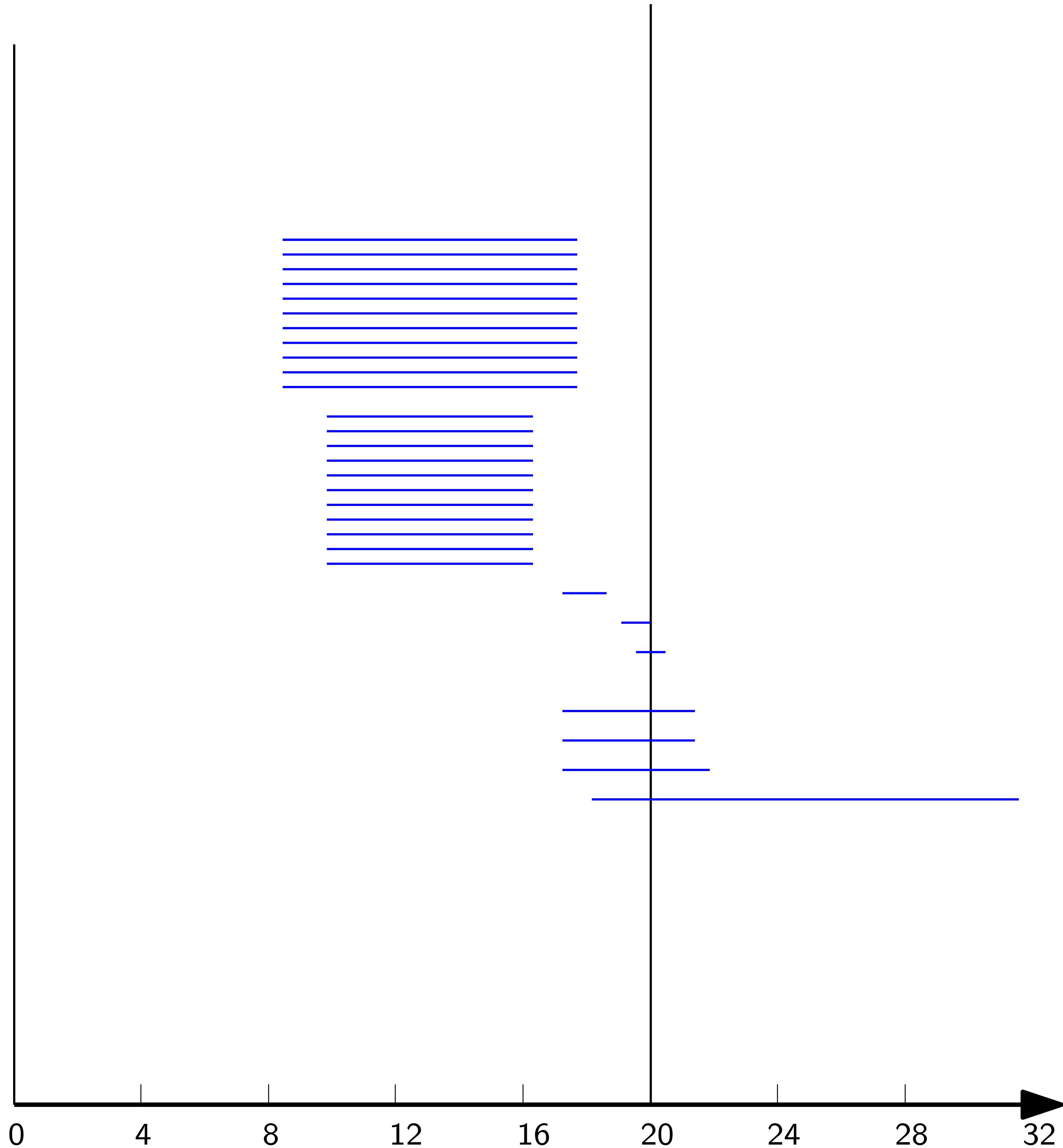
Persistent homology for functions

$$f_P : \mathbb{R}^2 \to \mathbb{R}$$

$$x \to \min_{p \in P} \|x - p\|_2$$



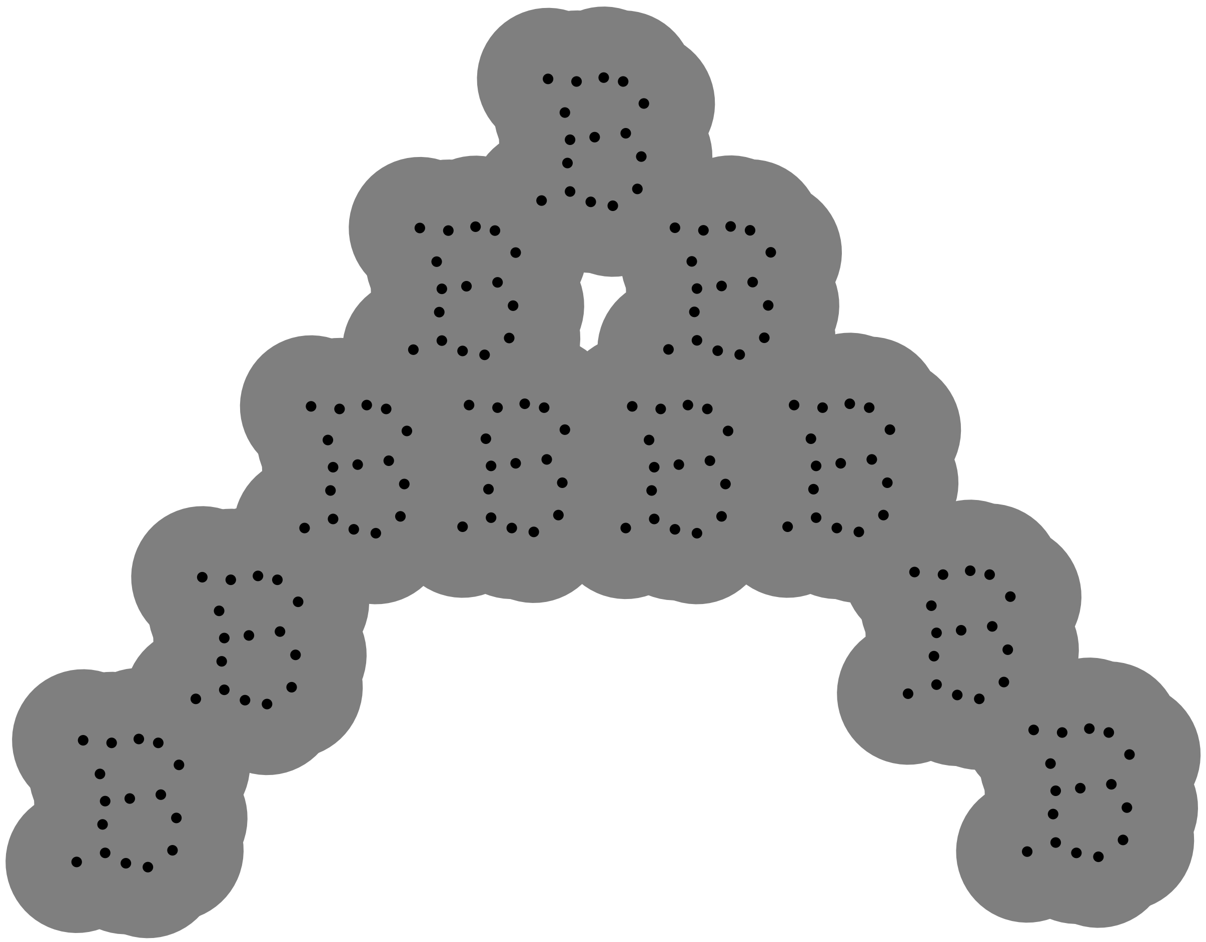
barcode for holes (1-d homology)



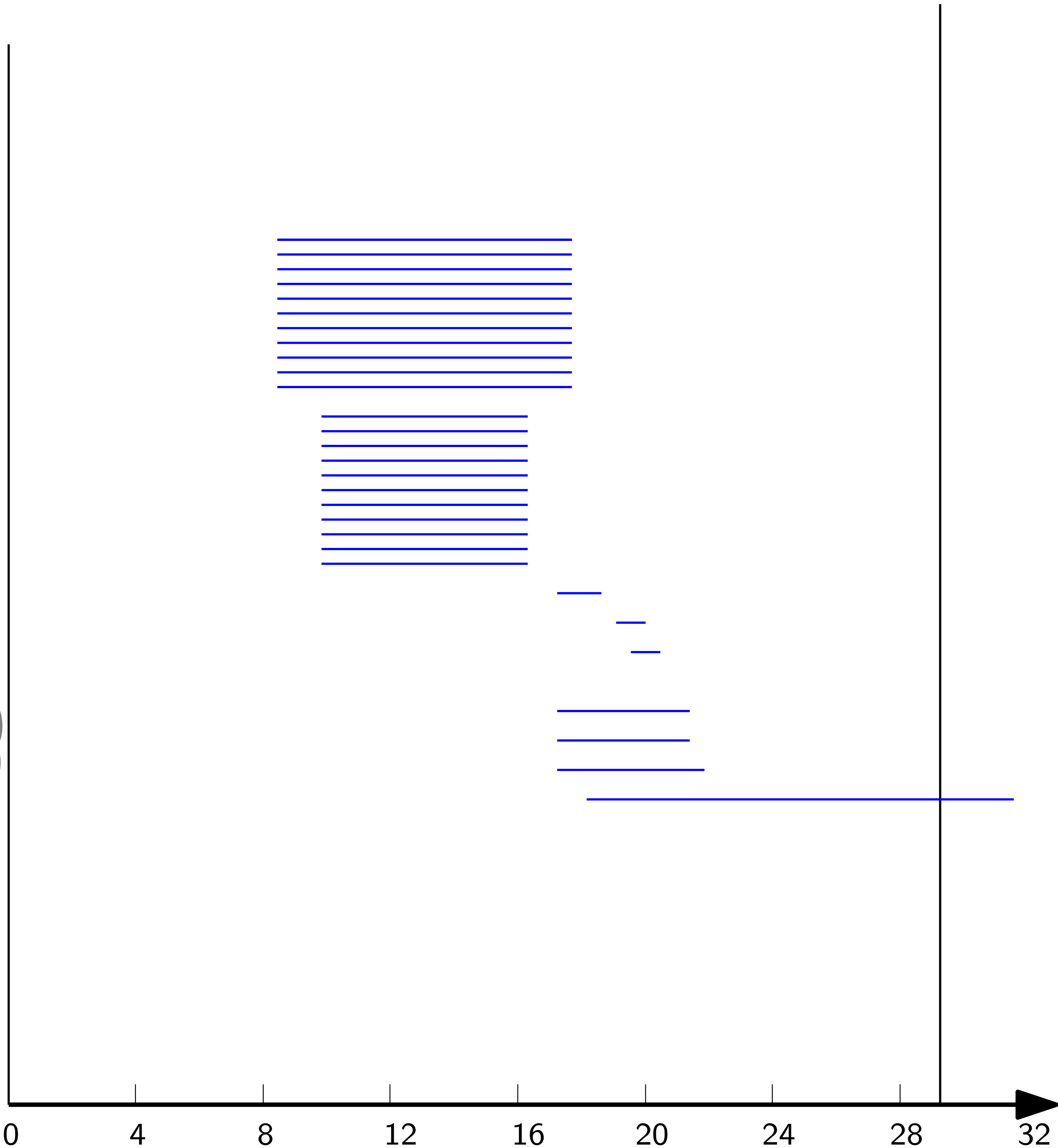
Persistent homology for functions

$$f_P : \mathbb{R}^2 \to \mathbb{R}$$

$$x \to \min_{p \in P} \|x - p\|_2$$



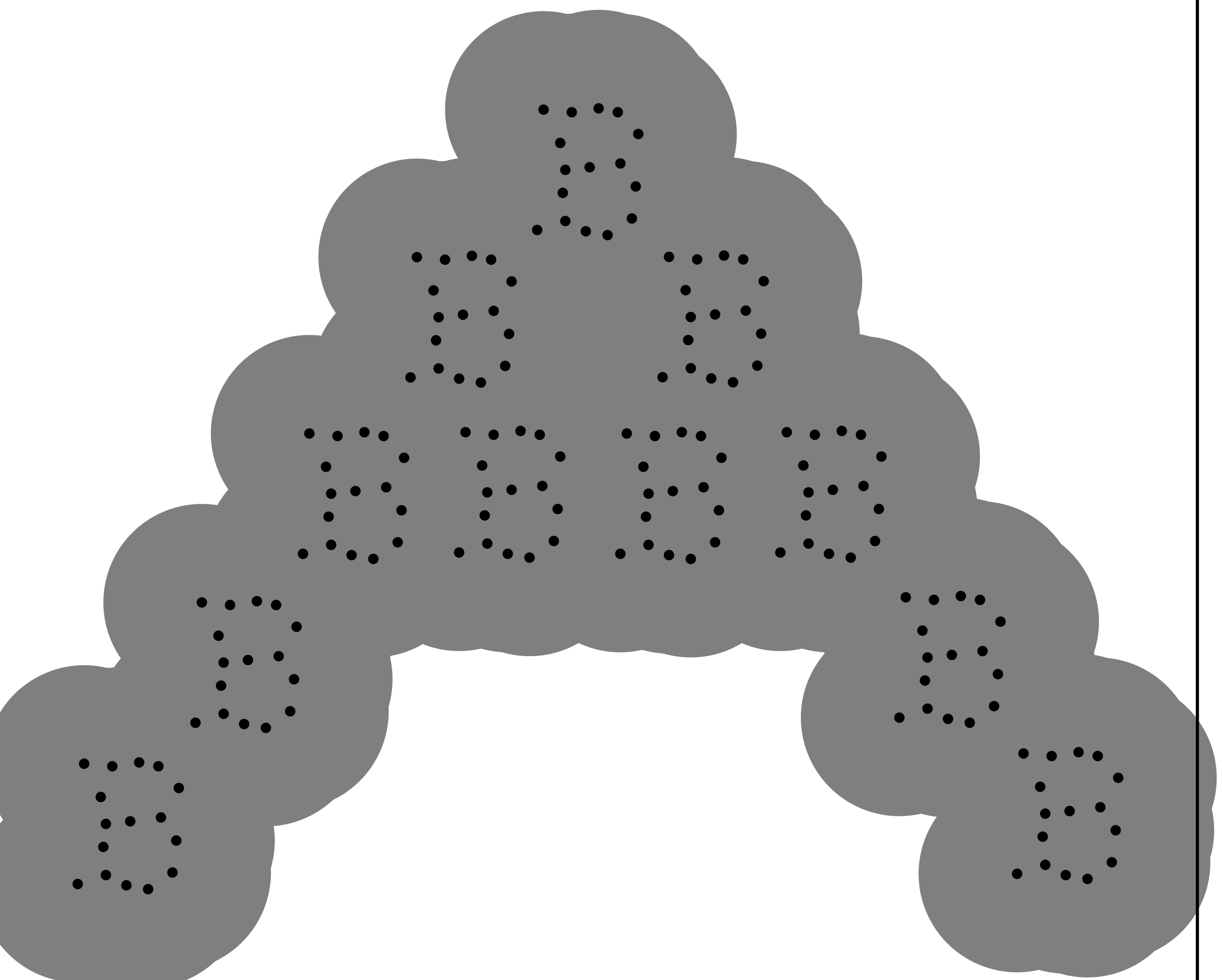
barcode for holes (1-d homology)



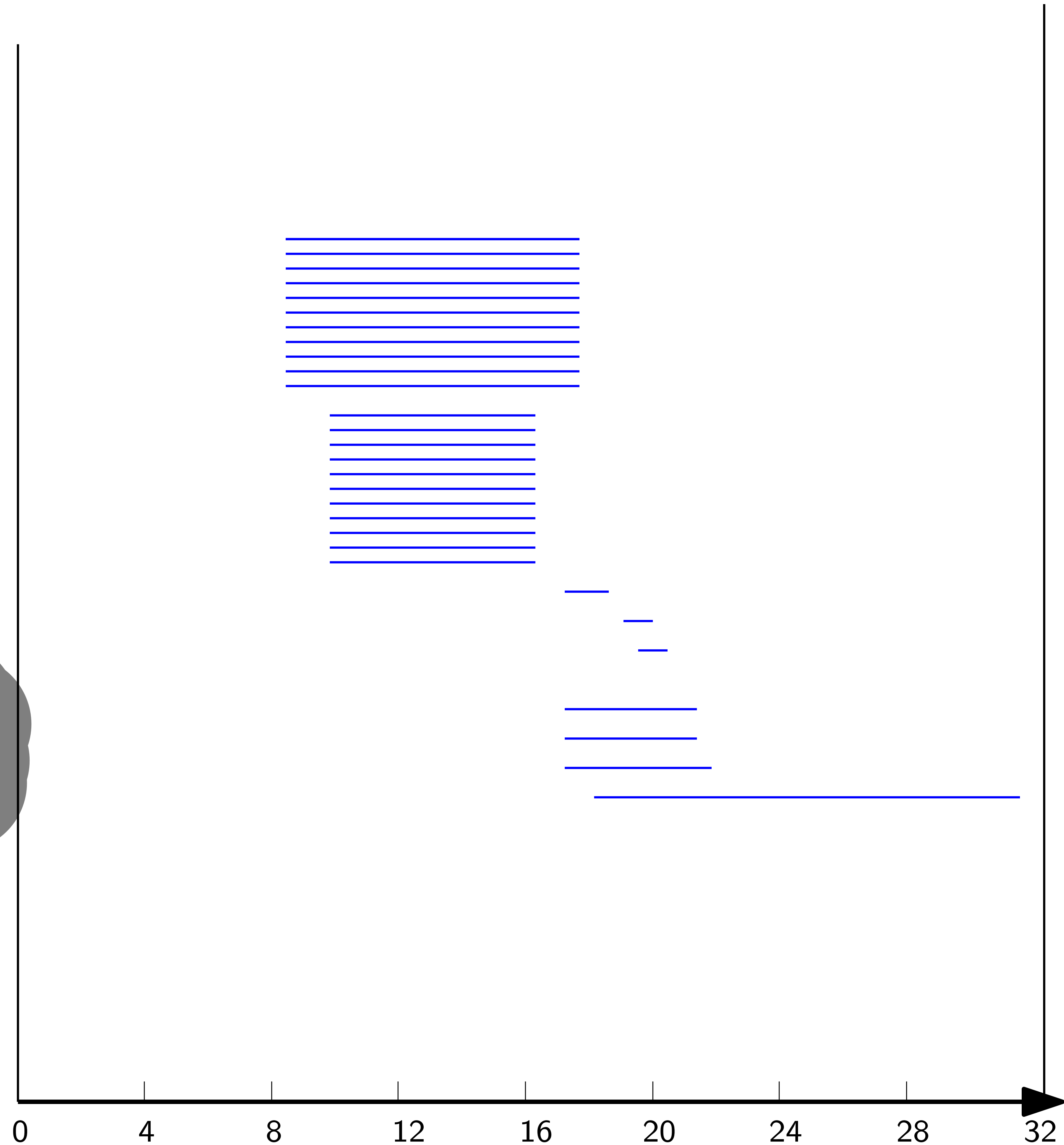
Persistent homology for functions

$$f_P : \mathbb{R}^2 \to \mathbb{R}$$

$$x \to \min_{p \in P} \|x - p\|_2$$



barcode for holes (1-d homology)



Persistent homology of filtered complexes

Let $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ be a filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

Persistent homology of filtered complexes

Let $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ be a filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

Relation between sublevel sets filtrations and filtered simplicial complexes:

- $\forall t \leq t' \in \mathbb{R}$, $f^{-1}((-\infty, t]) \subseteq f^{-1}((-\infty, t']) \rightarrow$ filtration of X by the sublevel sets of f.
- If f is defined at the vertices of a simplicial complex K , the sublevel sets filtration is a filtration of the simplicial complex K.

- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

Persistent homology of filtered complexes

Let $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ be a filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

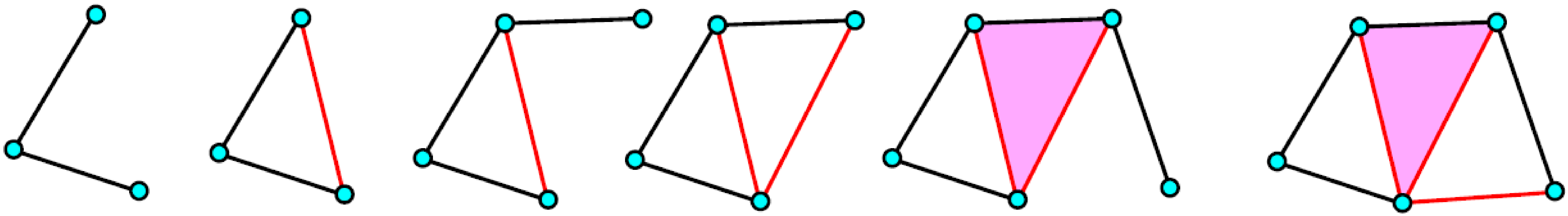
Algorithm to compute the Betti numbers $\beta_0, \beta_1, \cdots, \beta_d$ of K:

$\beta_0 = \beta_1 = \cdots = \beta_d = 0$;
for $i = 1$ to m
  $k = \dim \sigma^i - 1$;
  if σ^i is contained in a $(k+1)$-cycle in K^i
    then $\beta_{k+1} = \beta_{k+1} + 1$;
    else $\beta_k = \beta_k - 1$;
  end if;
end for;
output $(\beta_0, \beta_1, \cdots, \beta_d)$;

The algorithm can be easily adapted to keep track of an homology basis and pairs positive simplices (birth of a new homological class) to negative simplices (death of an existing homology class).

Notation: $H_k^i = H_k(K^i)$

Cycle associated to a positive simplex

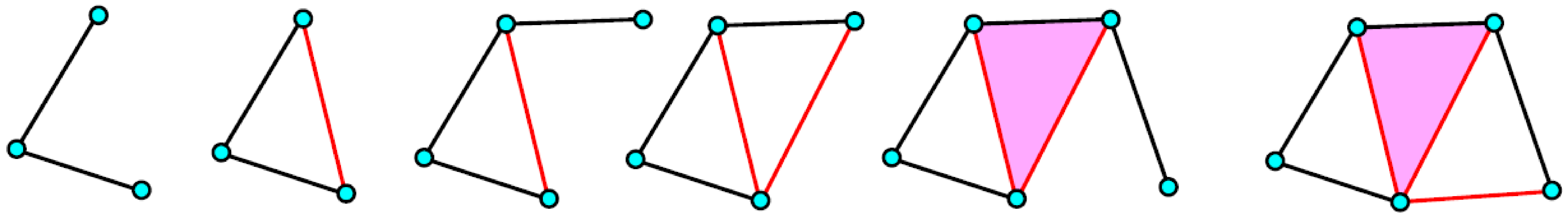


Lemma: If σ^i is a positive k-cycle, then there exists a k-cycle c_σ s.t.:
- c_σ is not a boundary in K^i,
- c_σ contains σ^i but no other positive k-simplex.

The cycle c^σ is unique.

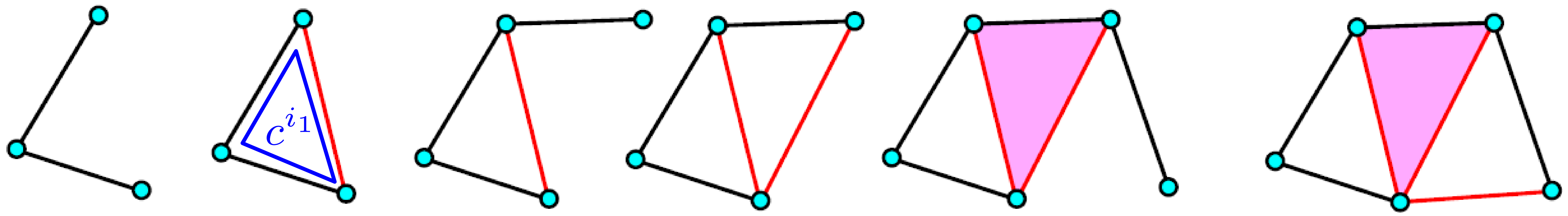
Proof:
By induction on the order of appearence of the simplices in the filtration.

Homology basis



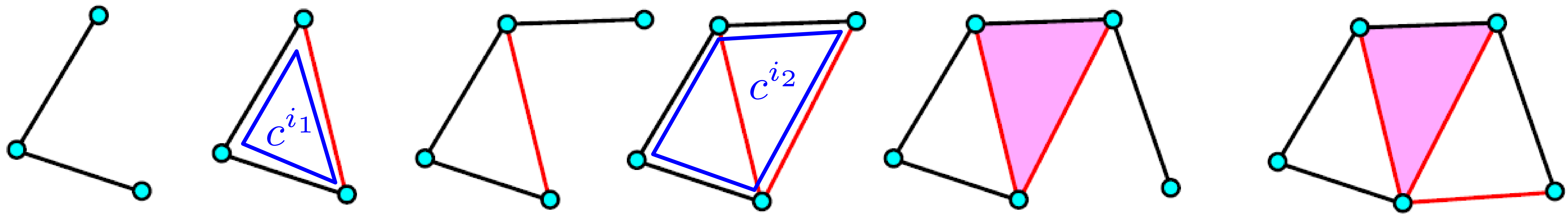
- At the beginning: the basis of H_k^0 is empty.
- If a basis of H_k^{i-1} has been built and σ^i is a positive k-simplex then one adds the homology class of the cycle c^i associated to σ^i to the basis of H_k^{i-1} $\Rightarrow$ basis of H_k^i.
- If a basis of H_k^{j-1} has been built and σ^j is a negative $(k+1)$-simplex:
  - let $c^{i_1}, \cdots, c^{i_p}$ be the cycles associated to the positive simplices $\sigma^{i_1}, \cdots, \sigma^{i_p}$ that form a basis of H_k^{j-1}
  - $d = \partial\sigma^j = \sum_{k=1}^{p} \varepsilon_k c^{i_k} + b$
  - $l(j) = \max\{i_k : \varepsilon_k = 1\}$
  - Remove the homology class of $c^{l(j)}$ from the basis of H_k^{j-1} $\Rightarrow$ basis of H_k^j.

Homology basis



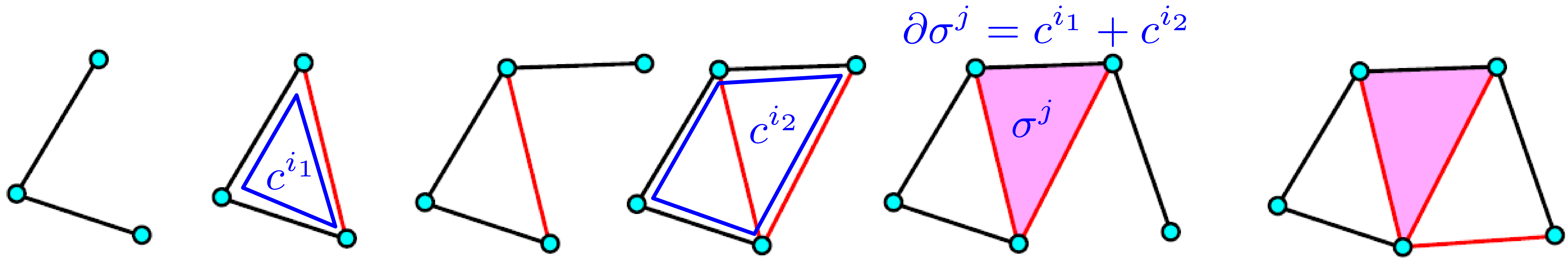
- At the beginning: the basis of H_k^0 is empty.
- If a basis of H_k^{i-1} has been built and σ^i is a positive k-simplex then one adds the homology class of the cycle c^i associated to σ^i to the basis of H_k^{i-1} $\Rightarrow$ basis of H_k^i.
- If a basis of H_k^{j-1} has been built and σ^j is a negative $(k+1)$-simplex:
  - let $c^{i_1}, \cdots, c^{i_p}$ be the cycles associated to the positive simplices $\sigma^{i_1}, \cdots, \sigma^{i_p}$ that form a basis of H_k^{j-1}
  - $d = \partial\sigma^j = \sum_{k=1}^{p} \varepsilon_k c^{i_k} + b$
  - $l(j) = \max\{i_k : \varepsilon_k = 1\}$
  - Remove the homology class of $c^{l(j)}$ from the basis of H_k^{j-1} $\Rightarrow$ basis of H_k^j.

Homology basis



- At the beginning: the basis of H_k^0 is empty.
- If a basis of H_k^{i-1} has been built and σ^i is a positive k-simplex then one adds the homology class of the cycle c^i associated to σ^i to the basis of H_k^{i-1} $\Rightarrow$ basis of H_k^i.
- If a basis of H_k^{j-1} has been built and σ^j is a negative $(k+1)$-simplex:
  - let $c^{i_1}, \cdots, c^{i_p}$ be the cycles associated to the positive simplices $\sigma^{i_1}, \cdots, \sigma^{i_p}$ that form a basis of H_k^{j-1}
  - $d = \partial\sigma^j = \sum_{k=1}^{p} \varepsilon_k c^{i_k} + b$
  - $l(j) = \max\{i_k : \varepsilon_k = 1\}$
  - Remove the homology class of $c^{l(j)}$ from the basis of H_k^{j-1} $\Rightarrow$ basis of H_k^j.

Homology basis



- At the beginning: the basis of H_k^0 is empty.
- If a basis of H_k^{i-1} has been built and σ^i is a positive k-simplex then one adds the homology class of the cycle c^i associated to σ^i to the basis of $H_k^{i-1} \Rightarrow$ basis of H_k^i.
- If a basis of H_k^{j-1} has been built and σ^j is a negative $(k+1)$-simplex:
  - let $c^{i_1}, \cdots, c^{i_p}$ be the cycles associated to the positive simplices $\sigma^{i_1}, \cdots, \sigma^{i_p}$ that form a basis of H_k^{j-1}
  - $d = \partial\sigma^j = \sum_{k=1}^{p} \varepsilon_k c^{i_k} + b$
  - $l(j) = \max\{i_k : \varepsilon_k = 1\}$
  - Remove the homology class of $c^{l(j)}$ from the basis of $H_k^{j-1} \Rightarrow$ basis of H_k^j.

Pairing simplices

- If a basis of H_k^{j-1} has been built and σ^j is a negative $(k+1)$-simplex:
  - let $c^{i_1}, \cdots, c^{i_p}$ be the cycles associated to the positive simplices $\sigma^{i_1}, \cdots, \sigma^{i_p}$ that form a basis of H_k^{j-1}
  - $d = \partial\sigma^j = \sum_{k=1}^{p} \varepsilon_k c^{i_k} + b$
  - $l(j) = \max\{i_k : \varepsilon_k = 1\}$
  - Remove the homology class of $c^{l(j)}$ from the basis of $H_k^{j-1} \Rightarrow$ basis of H_k^j.

The simplices $\sigma^{l(j)}$ and σ^j are paired to form a persistent pair $(\sigma^{l(j)}, \sigma^j)$.
$\rightarrow$ The homology class created by $\sigma^{l(j)}$ in $K^{l(j)}$ is killed by σ^j in K^j. The persistence (or life-time) of this cycle is : $j - l(j) - 1$.

Remark: filtrations of K can be indexed by increasing sequences α_i of real numbers (useful when working with a function defined on the vertices of a simplicial complex).

Persistence algorithm: first version

Input: $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ a d-dimensional filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

$L_0 = L_1 = \cdots = L_{d-1} = \emptyset$
For $j = 0$ to m
    $k = \dim \sigma^j - 1$;
    if σ^j is a negative simplex
        $l(j) =$ highest index of the positive simplices associated to $\partial\sigma^j$;
        $L_k = L_k \cup \{(\sigma^{l(j)}, \sigma^j)\}$;
    end if
end for
output $L_0, L_1, \cdots, L_{d-1}$;

Persistence algorithm: first version

Input: $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ a d-dimensional filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

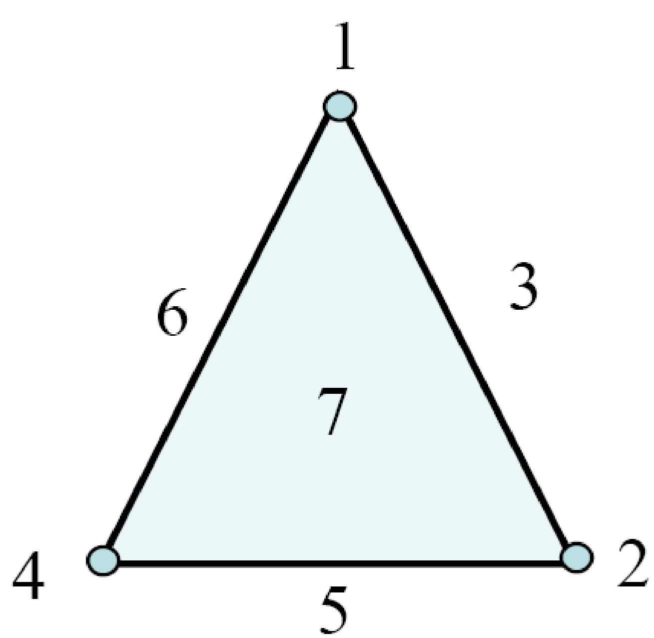
$L_0 = L_1 = \cdots = L_{d-1} = \emptyset$
For $j = 0$ to m
    $k = \dim \sigma^j - 1$;
    if σ^j is a negative simplex
        $l(j) =$ highest index of the positive simplices associated to $\partial \sigma^j$;
        $L_k = L_k \cup \{(\sigma^{l(j)}, \sigma^j)\}$;
    end if
end for
output $L_0, L_1, \cdots, L_{d-1}$;

How to test this condition?

The persistence algorithm: matrix version

Input: $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ a d-dimensional filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

The matrix of the boundary operator:



$$\begin{array}{cc} \begin{array}{ccccccc} & & & & j & & \end{array} & \\ \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} & l(j) \end{array}$$

- $M = (m_{ij})_{i,j=1,\cdots,m}$ with coefficient in $\mathbb{Z}/2$ defined by

$$m_{ij} = 1 \text{ if } \sigma^i \text{ is a face of } \sigma^j \text{ and } m_{ij} = 0 \text{ otherwise}$$

- For any column C_j, $l(j)$ is defined by

$$(i = l(j)) \Leftrightarrow (m_{ij} = 1 \text{ and } m_{i'j} = 0 \ \ \forall i' > i)$$

The persistence algorithm: matrix version

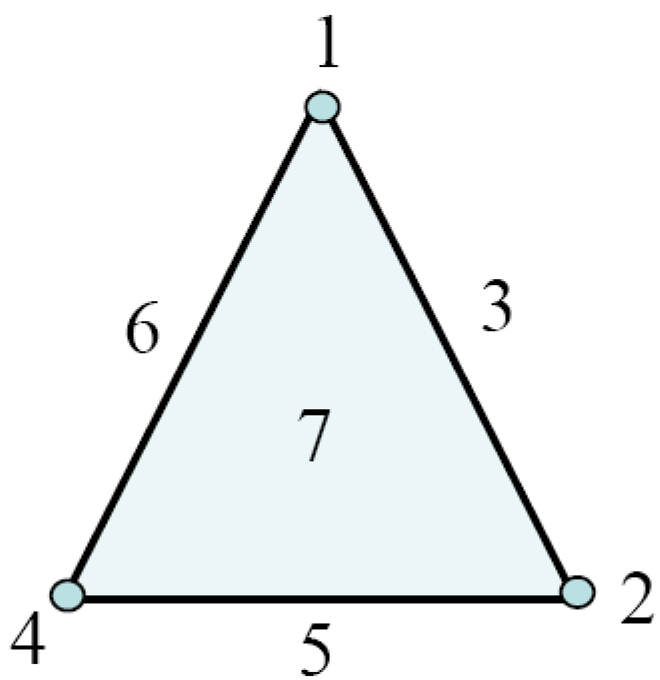
Input: $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ a d-dimensional filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

Compute the matrix of the boundary operator M
For $j = 0$ to m
    While (there exists $j' < j$ such that $l(j') == l(j)$)
        $C_j = C_j + C_{j'}$ mod(2);
    End while
End for
Output the pairs $(l(j), j)$;

Remark: The worst case complexity of the algorithm is $O(m^3)$ but much lower in most practical cases.

The persistence algorithm: matrix version

A simple example:



$$\begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & \boxed{1} & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{array}{c} \hspace{6em} C_5 + C_6 \\ \hspace{6em}\downarrow \\ \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & \boxed{1} & 0 & 1 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$\begin{array}{c} \hspace{6em} C_3 + C_6 \\ \hspace{6em}\downarrow \\ \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Paires : (2,3) (4,5) (6,7)

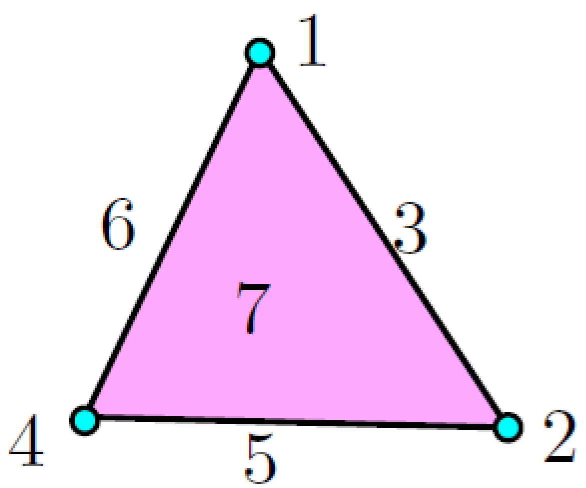
Correctness of the algorithm

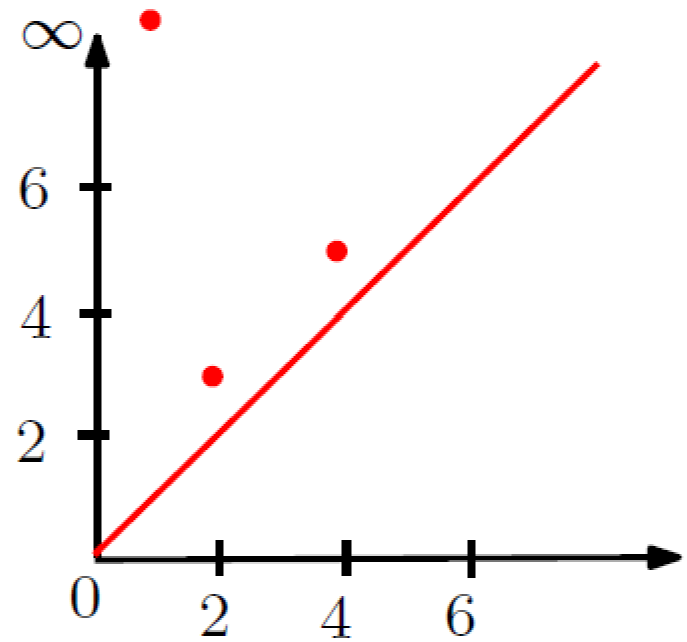
Proposition: the second algorithm (matric version) outputs the persistence pairs.

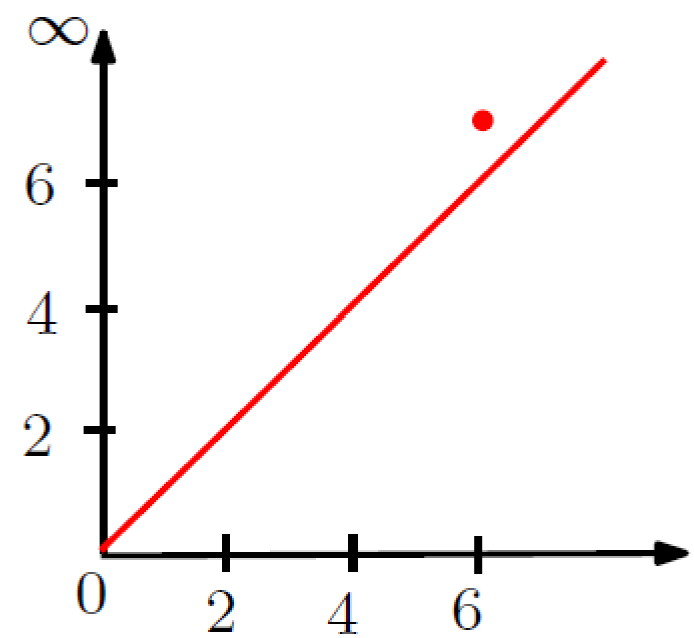
Proof: follows from the four remarks below.

1. At each step of the algorithm, the column C_j represents a chain of the form
$$\partial\left(\sigma^j + \sum_{i<j} \varepsilon_i \sigma^i\right) \text{ with } \varepsilon_i \in \{0,1\}$$
2. At this end of the algorithm, if j is s.t. $l(j)$ is defined then $\sigma^{l(j)}$ is a positive simplex.
3. If at the end of the algorithm if the column C_j is zero then σ^j is positive.
4. If at the end of the algorithm the column C_j is not zero then $(\sigma^{l(j)}, \sigma^j)$ is a persistence pair.

Persistence diagram

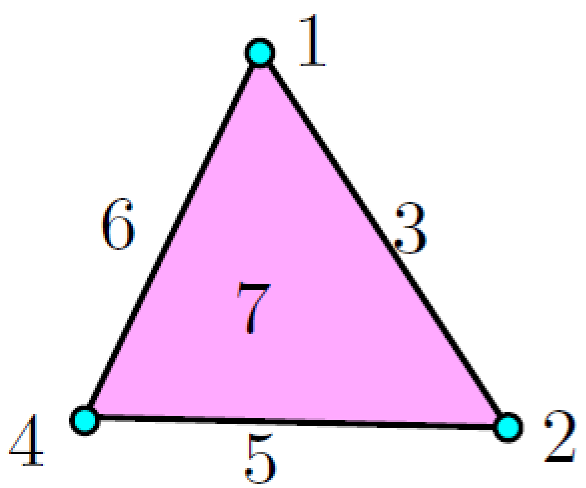


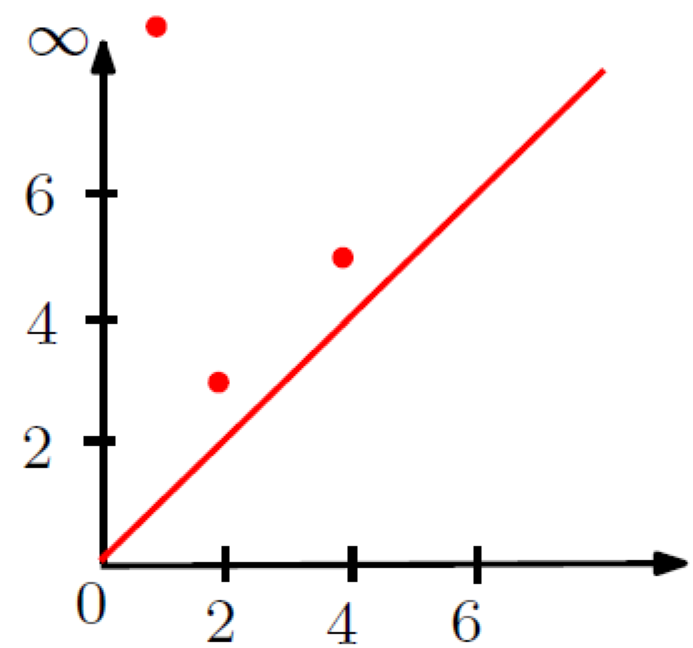


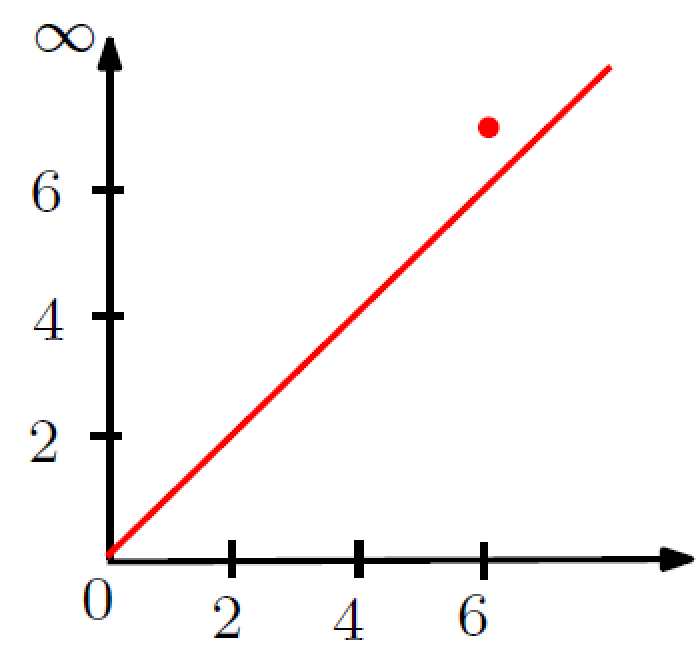


- each pair $(\sigma^{l(j)}, \sigma^j)$ is represented by $(l(j), j)$ or $(f(\sigma^{l(j)}), f(\sigma^j)) \in \mathbb{R}^2$ when considering filtrations induced by functions, or $(\alpha_{l(j)}, \alpha_j)$ if the filtration is indexed by a real valued sequence $(\alpha_i)_{i \in I}$.
- The diagonal $\{y = x\}$ is added to the persistence diagram.
- Unpaired positive simplex $\sigma^i \to (i, +\infty)$.

Persistence diagram



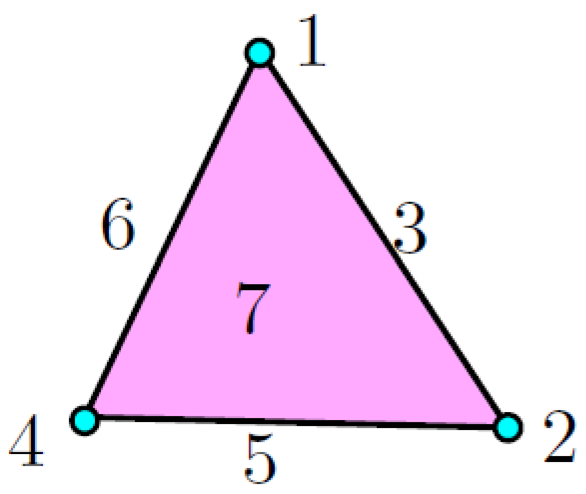


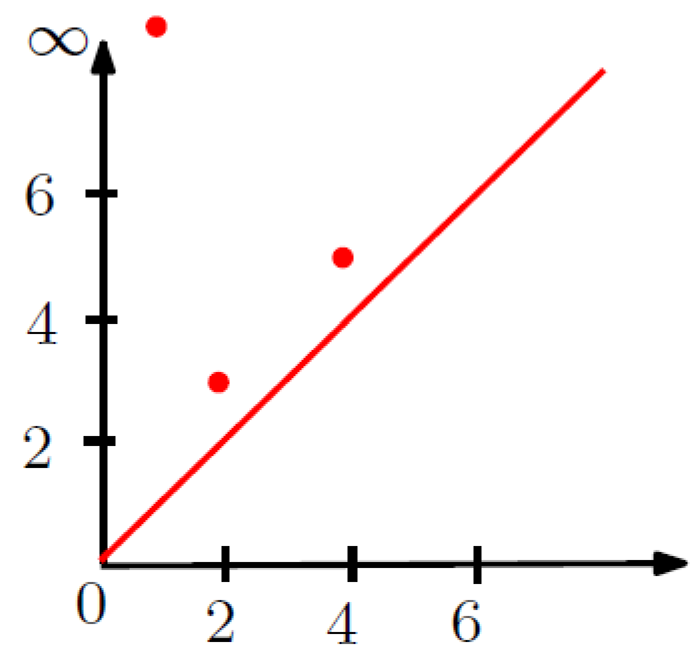


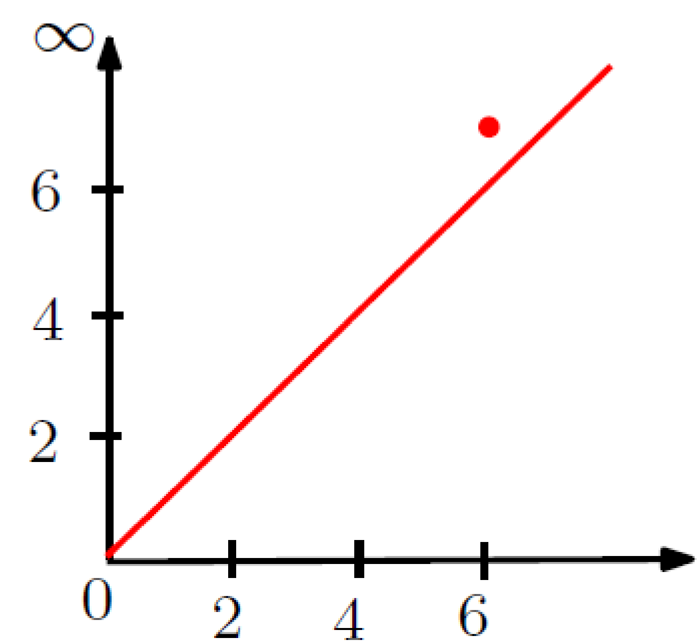
- each pair $(\sigma^{l(j)}, \sigma^j)$ is represented by $(l(j), j)$ or $(f(\sigma^{l(j)}), f(\sigma^j)) \in \mathbb{R}^2$ when considering filtrations induced by functions, or $(\alpha_{l(j)}, \alpha_j)$ if the filtration is indexed by a real valued sequence $(\alpha_i)_{i \in I}$.
- The diagonal $\{y = x\}$ is added to the persistence diagram.
- Unpaired positive simplex $\sigma^i \to (i, +\infty)$.

Points may have multiplicity

Persistence diagram

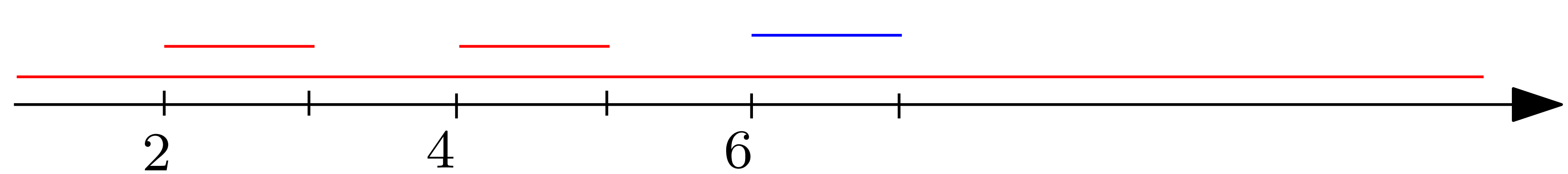




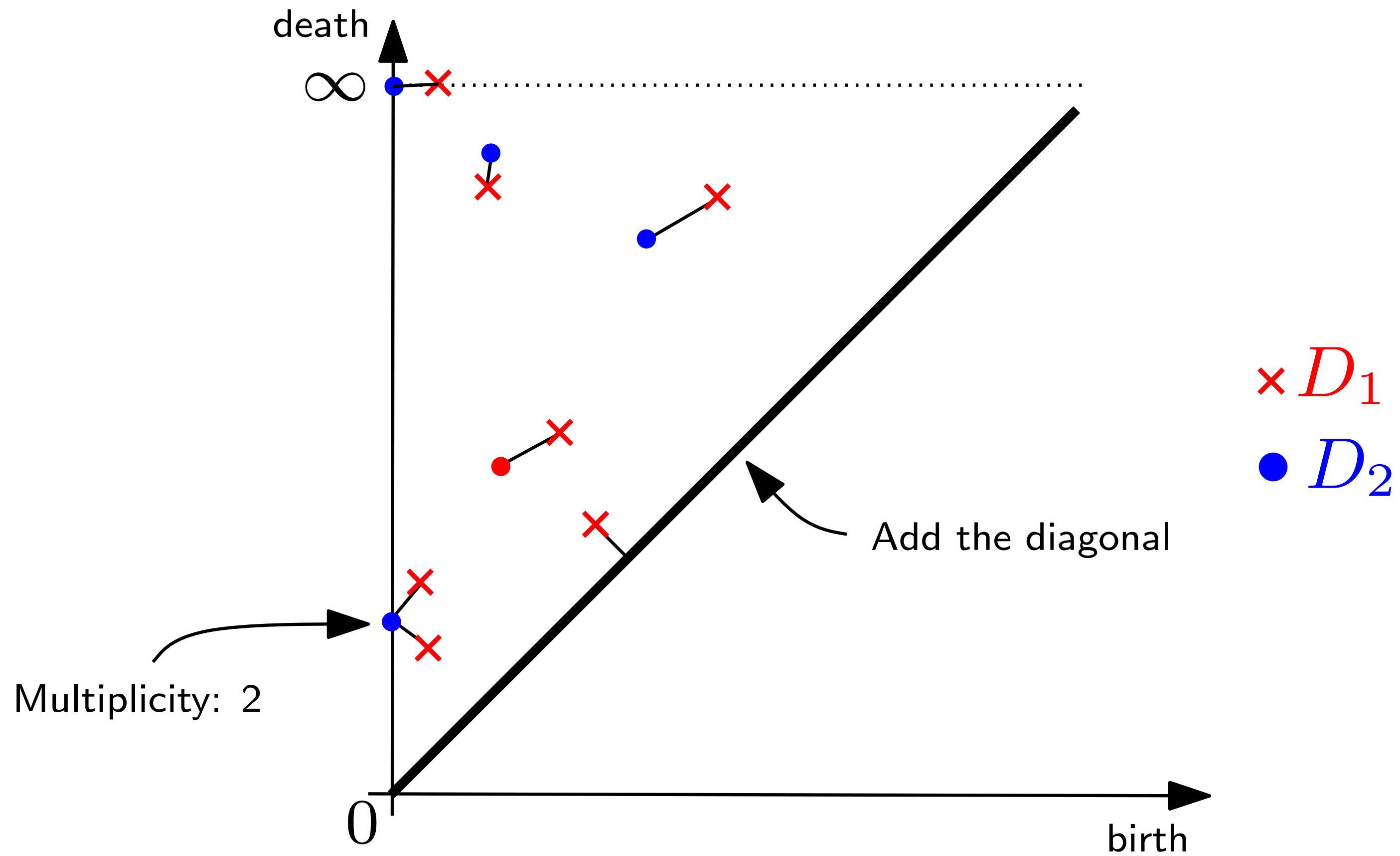


- each pair $(\sigma^{l(j)}, \sigma^j)$ is represented by $(l(j), j)$ or $(f(\sigma^{l(j)}), f(\sigma^j)) \in \mathbb{R}^2$ when considering filtrations induced by functions, or $(\alpha_{l(j)}, \alpha_j)$ if the filtration is indexed by a real valued sequence $(\alpha_i)_{i \in I}$.
- The diagonal $\{y = x\}$ is added to the persistence diagram.
- Unpaired positive simplex $\sigma^i \to (i, +\infty)$.

Barcodes: an alternative (equivalent) representation where each pair (i, j) is represented by the interval $[i, j]$



Distance between persistence diagrams

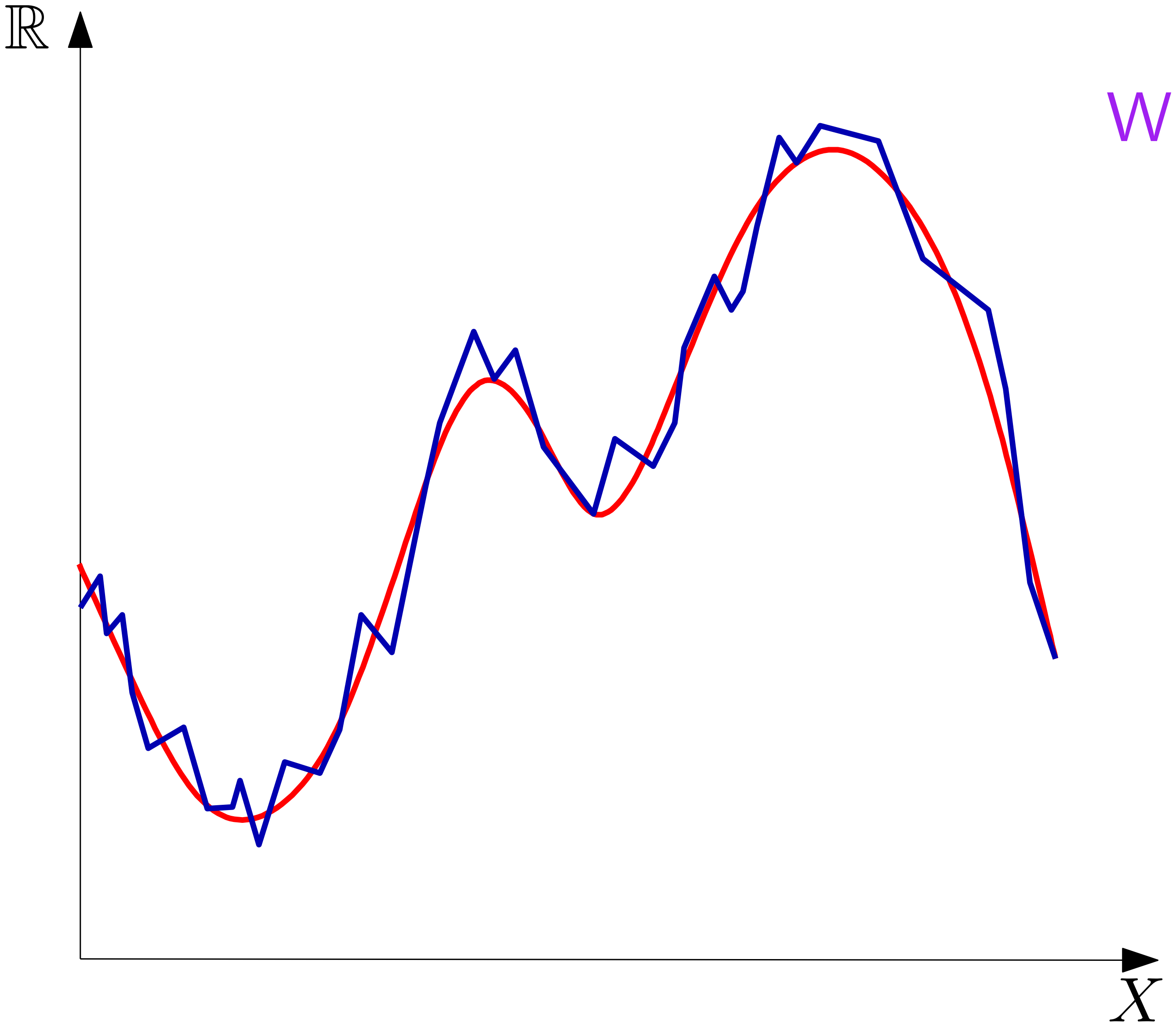


The bottleneck distance between two diagrams D_1 and D_2 is

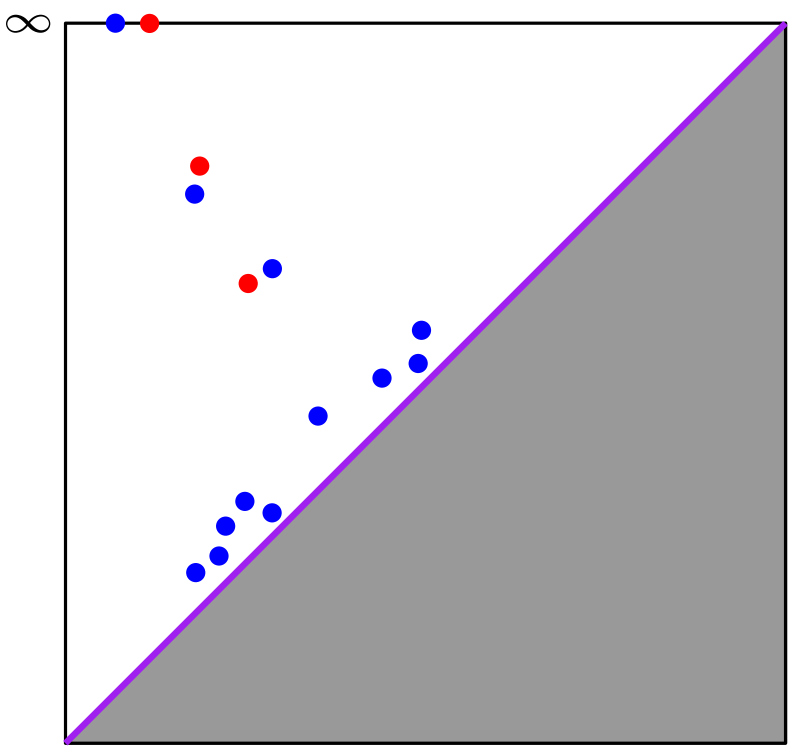
$$d_B(D_1, D_2) = \inf_{\gamma \in \Gamma} \sup_{p \in D_1} \|p - \gamma(p)\|_\infty$$

where Γ is the set of all the bijections between D_1 and D_2 and $\|p - q\|_\infty = \max(|x_p - x_q|, |y_p - y_q|)$.

Stability properties



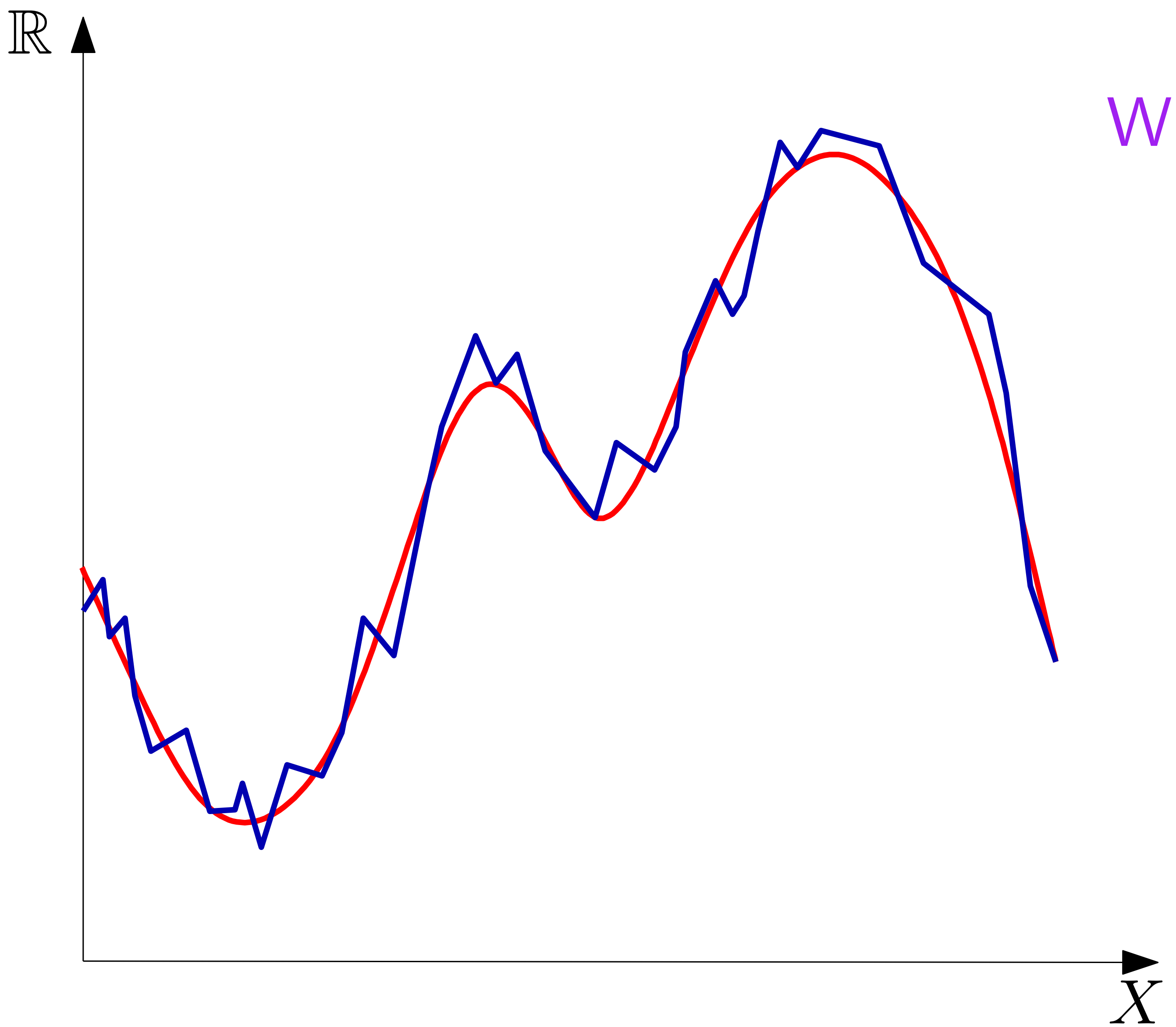
What if f is slightly perturbed?



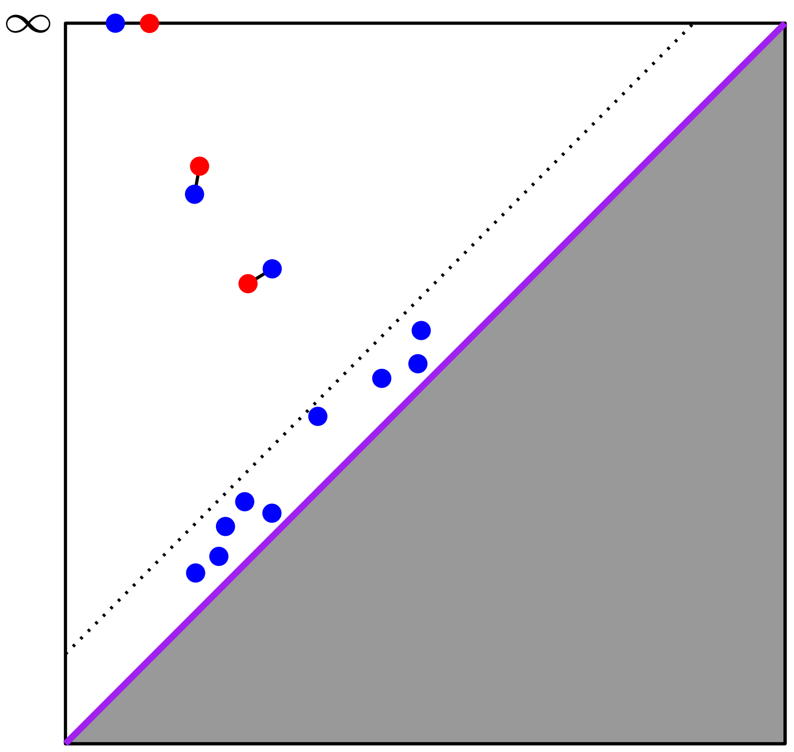
Stability properties

Theorem (Stability):
For any *tame* functions $f, g : \mathbb{X} \to \mathbb{R}$, $\mathrm{d}_\mathrm{B}^\infty(\mathrm{D}_f, \mathrm{D}_g) \leq \|f - g\|_\infty$.

[Cohen-Steiner, Edelsbrunner, Harer 05], [C., Cohen-Steiner, Glisse, Guibas, Oudot - SoCG 09], [C., de Silva, Glisse, Oudot 12]



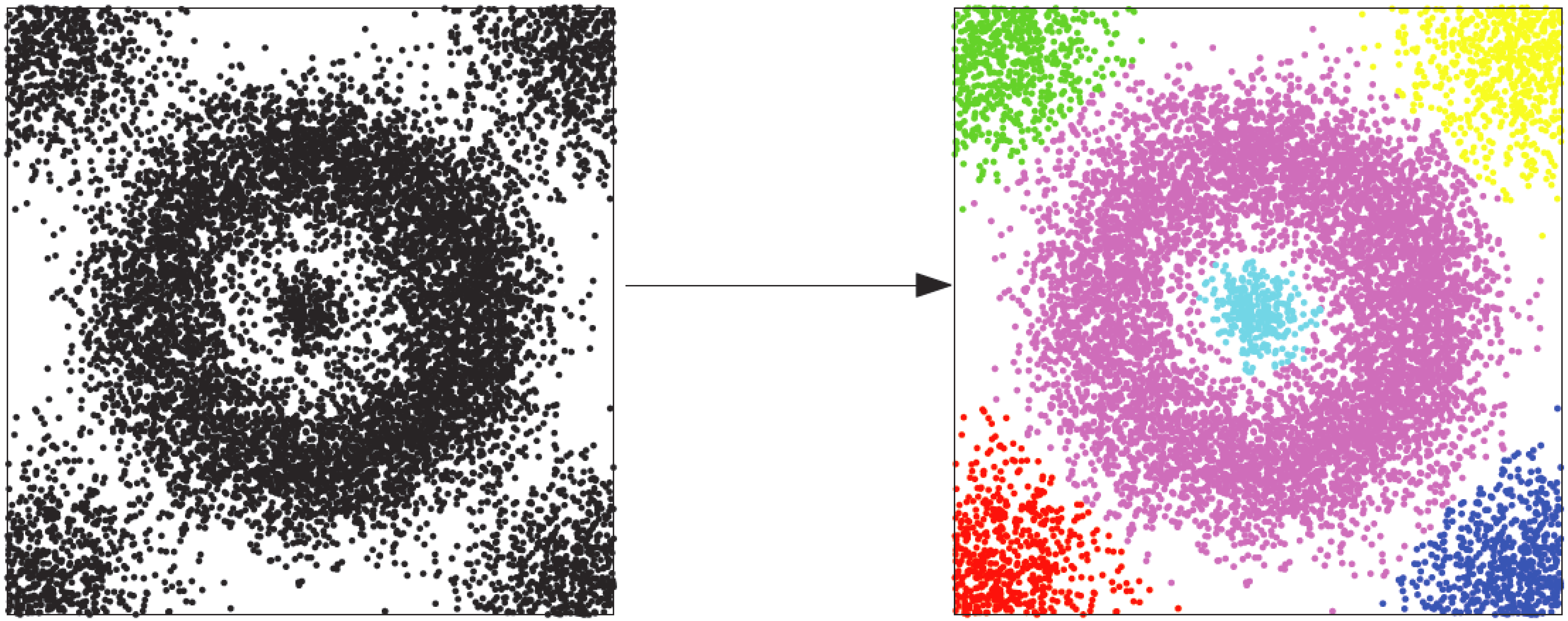
What if f is slightly perturbed?



Persistence-based clustering

Combine a mode seeking approach with (0-dim) persistence computation.

[C.,Guibas,Oudot,Skraba - J. ACM 2013]



Input:

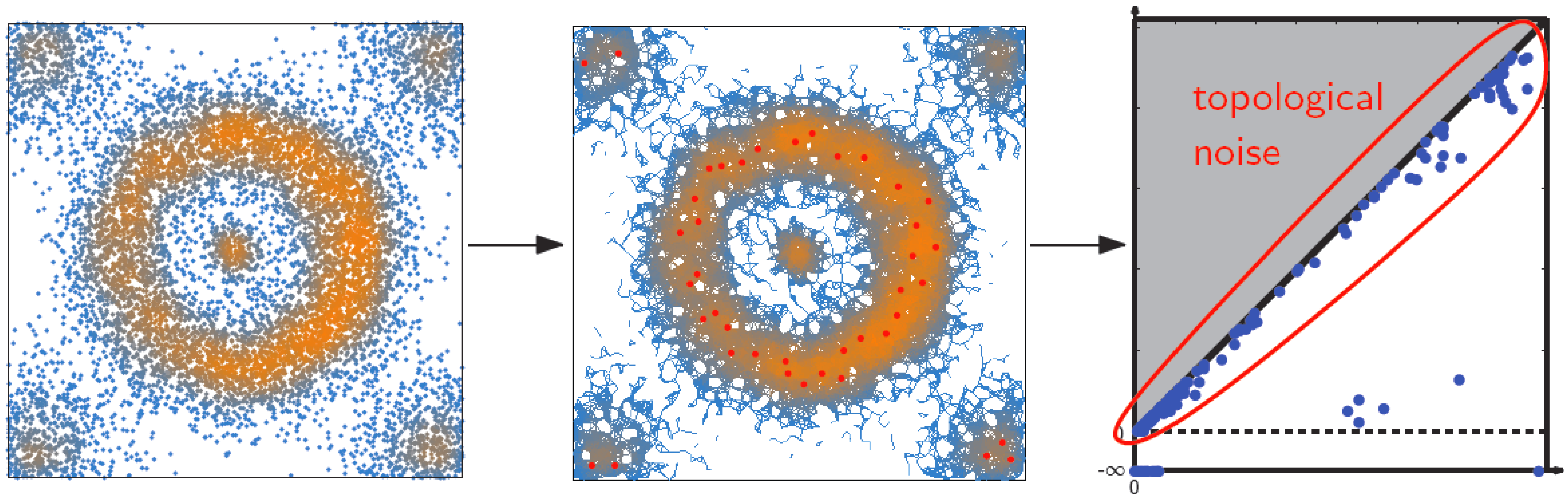
1. A finite set X of observations (point cloud with coordinates or pairwise distance matrix),
2. A real valued function f defined on the observations (e.g. density estimate).

Goal: Partition the data according to the basins of attraction of the peaks of f

Persistence-based clustering

Combine a mode seeking approach with (0-dim) persistence computation.

[C.,Guibas,Oudot,Skraba - J. ACM 2013]



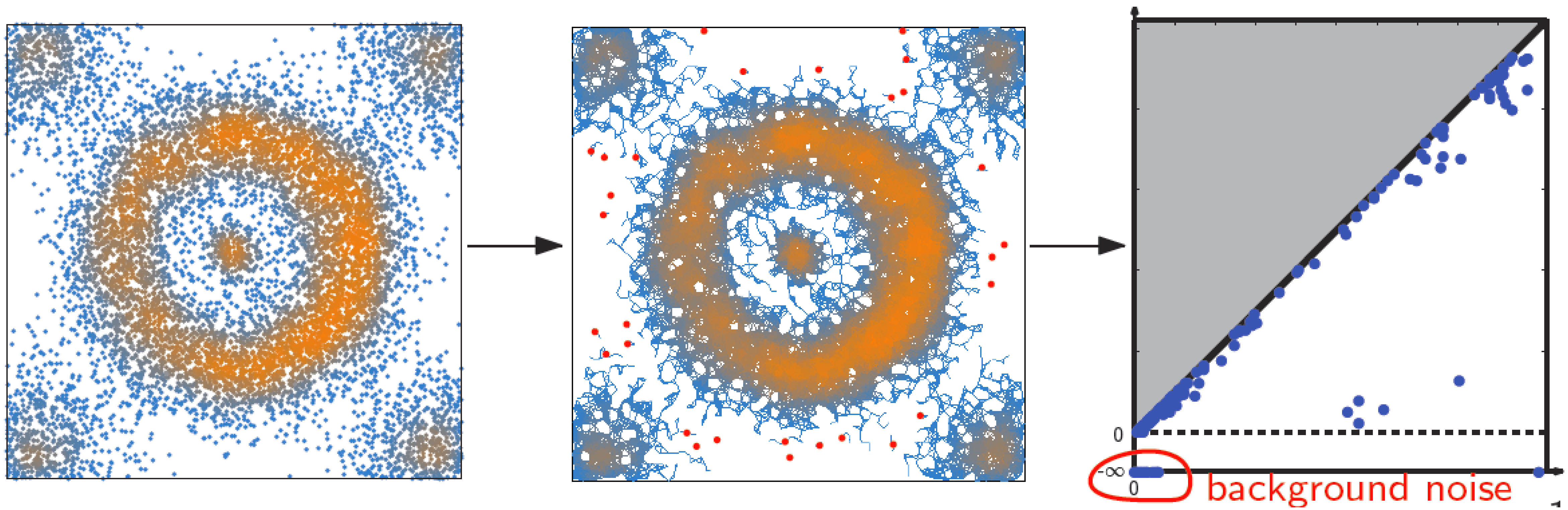
1. Build a neighborhing graph G on top of X.

2. Compute the (0-dim) persistence of f to identify prominent peaks $\rightarrow$ number of clusters (union-find algorithm).

Persistence-based clustering

Combine a mode seeking approach with (0-dim) persistence computation.

[C.,Guibas,Oudot,Skraba - J. ACM 2013]



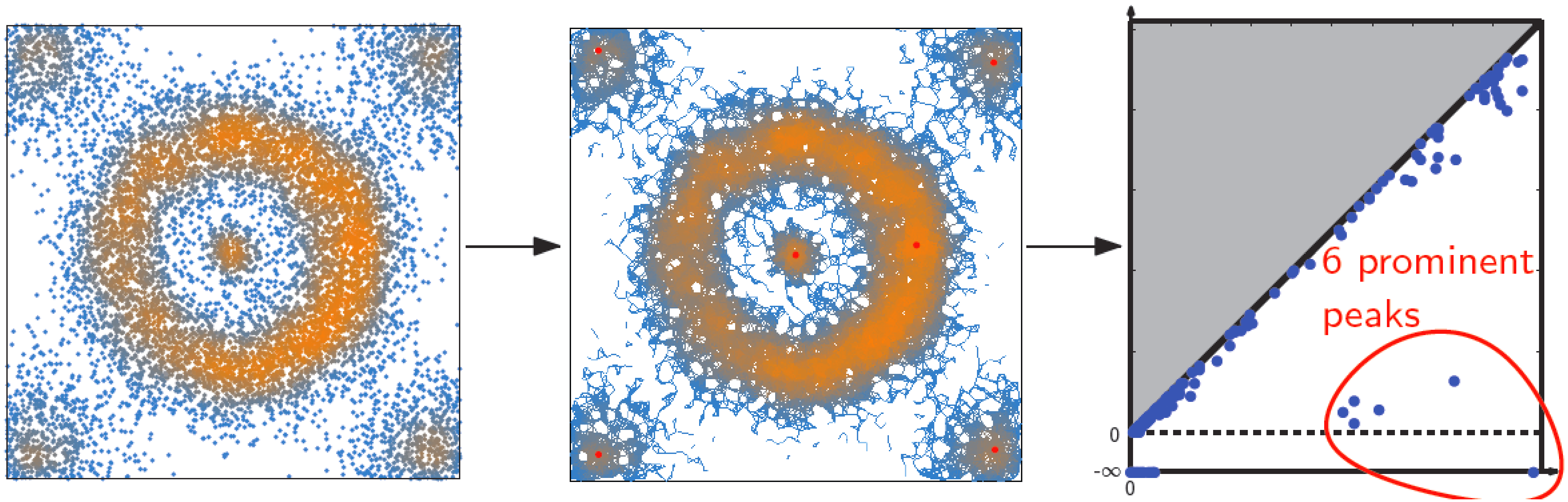
1. Build a neighborhing graph G on top of X.

2. Compute the (0-dim) persistence of f to identify prominent peaks $\to$ number of clusters (union-find algorithm).

Persistence-based clustering

Combine a mode seeking approach with (0-dim) persistence computation.

[C.,Guibas,Oudot,Skraba - J. ACM 2013]



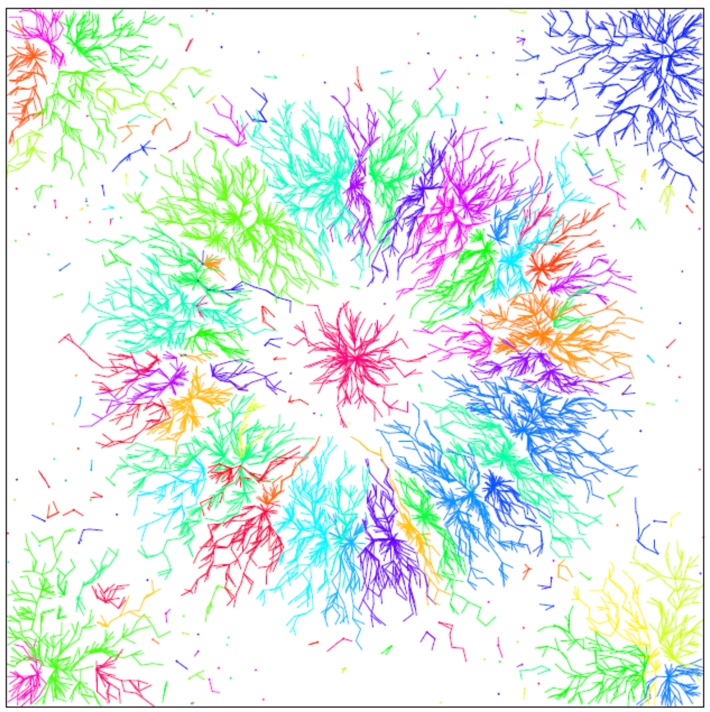
1. Build a neighborhing graph G on top of X.

2. Compute the (0-dim) persistence of f to identify prominent peaks $\rightarrow$ number of clusters (union-find algorithm).

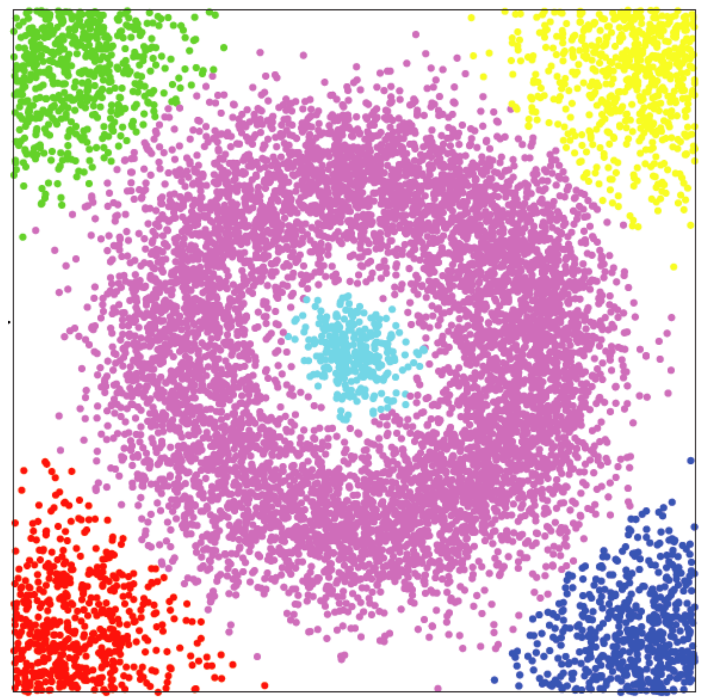
Persistence-based clustering

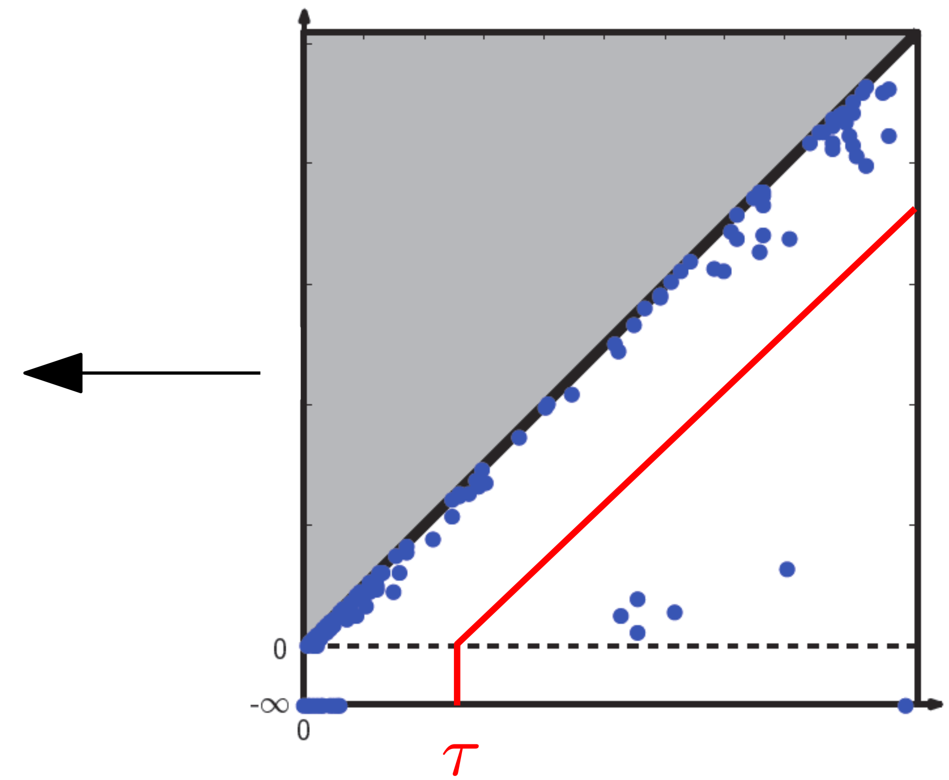
Combine a mode seeking approach with (0-dim) persistence computation.

[C.,Guibas,Oudot,Skraba - J. ACM 2013]



$\tau = 0$





1. Build a neighborhing graph G on top of X.

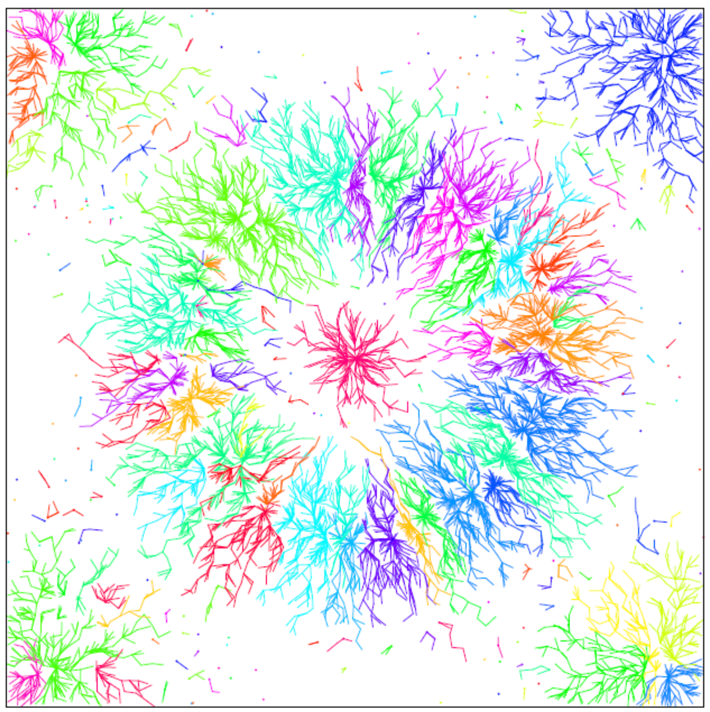
2. Compute the (0-dim) persistence of f to identify prominent peaks $\rightarrow$ number of clusters (union-find algorithm).

3. Chose a threshold $\tau > 0$ and use the persistence algorithm to merge components with prominence less than τ.

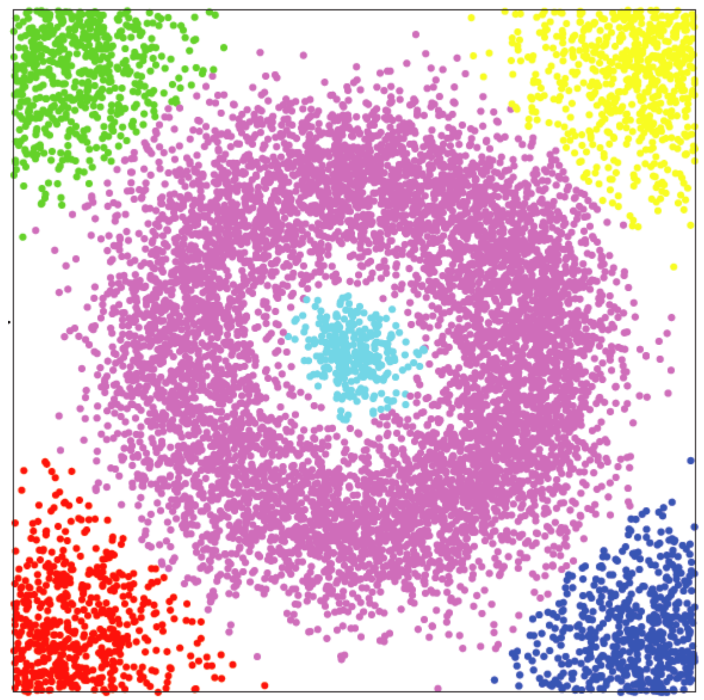
Persistence-based clustering

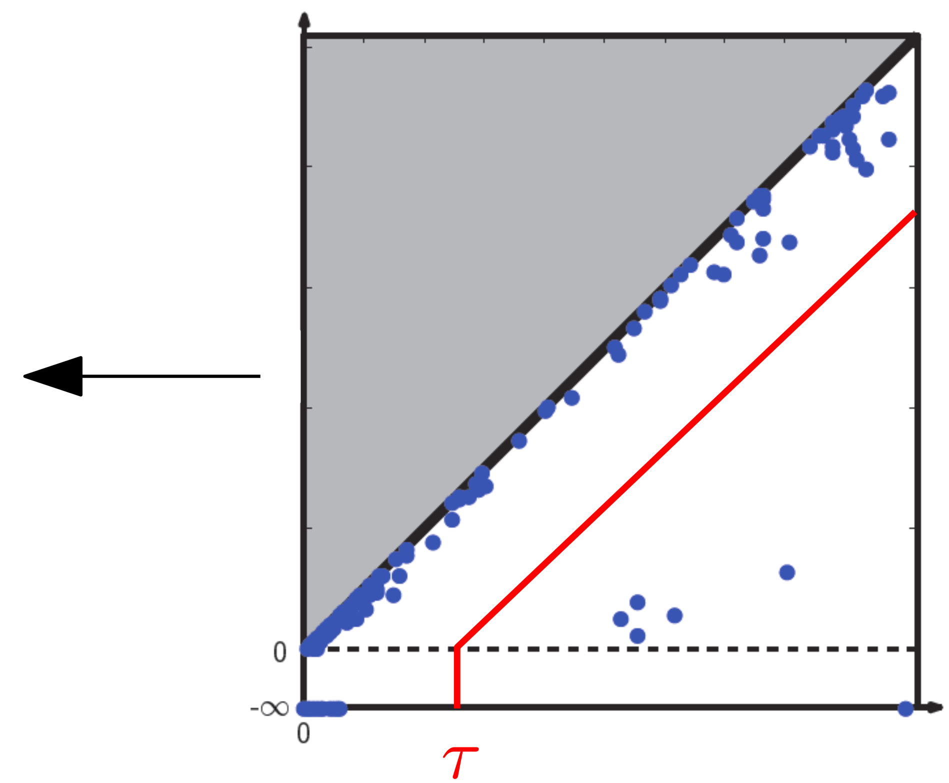
Combine a mode seeking approach with (0-dim) persistence computation.

[C.,Guibas,Oudot,Skraba - J. ACM 2013]



$\tau = 0$





Complexity of the algorithm: $\mathrm{O}(n \log n)$

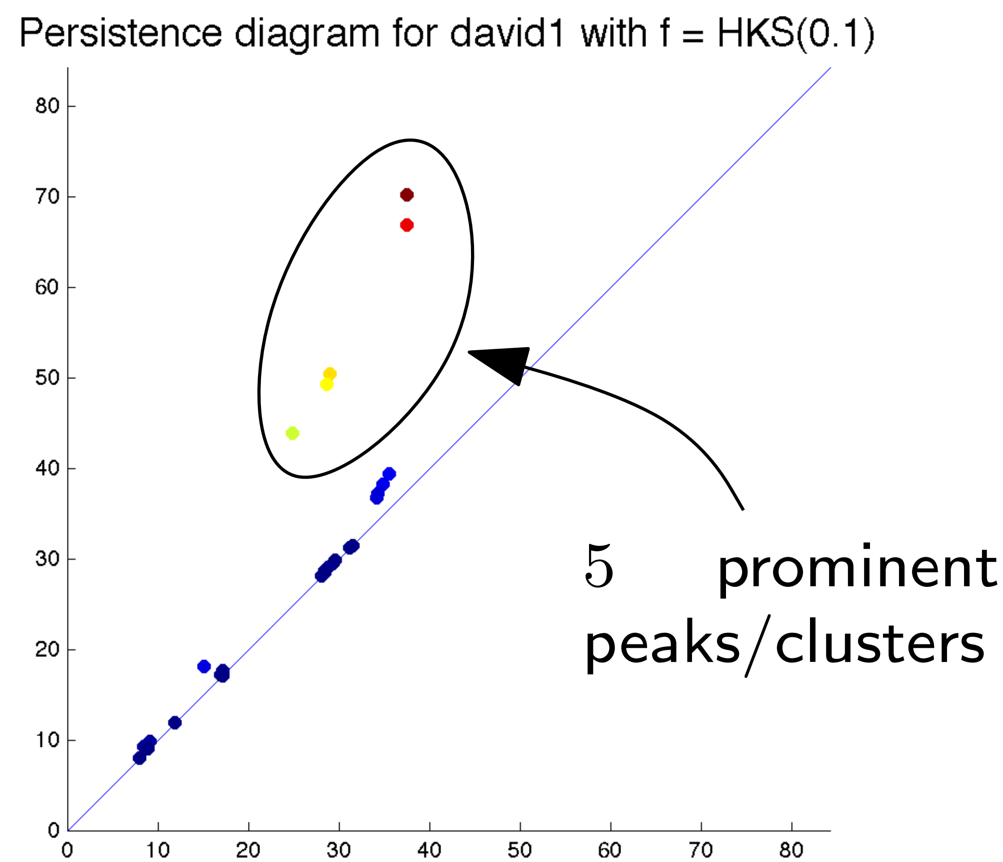
Theoretical guarantees:

- Stability of the number of clusters (w.r.t. perturbations of X and f).
- Partial stability of clusters: well identified stable parts in each cluster.

"soft " clustering

Application to non-rigid shape segmentation

[Skraba, Ovsjanikov, C.,Guibas, NORDIA 10]



X : a 3D shape
f = HKS function on X

Problem: some part of clusters are unstable $\rightarrow$ dirty segments

Application to non-rigid shape segmentation

[Skraba, Ovsjanikov, C.,Guibas, NORDIA 10]

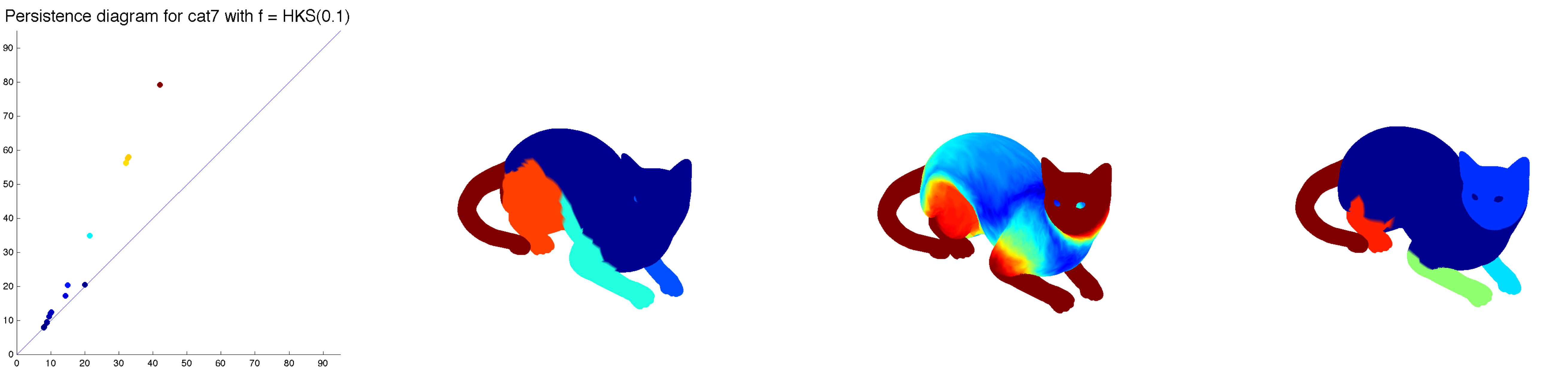
Problem: some part of clusters are unstable $\rightarrow$ dirty segments

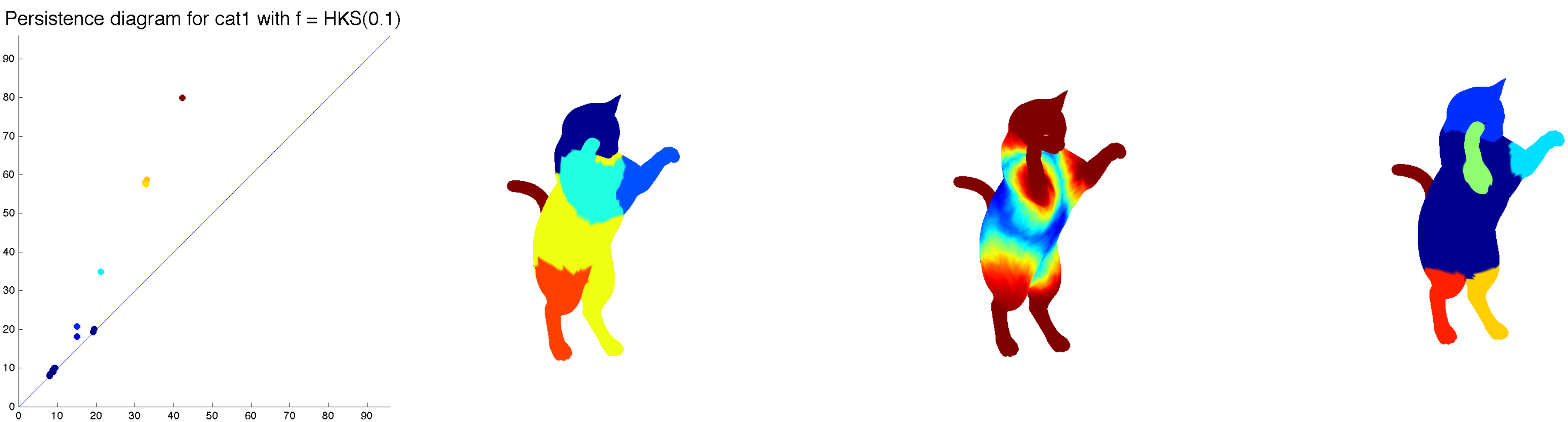
Idea:

- Run the persistence based algorithm several times on random perturbations of f (size bounded by the "persistence" gap).
- Partial stability of clusters allows to establish correspondences between clusters across the different runs $\rightarrow$ for any $x \in X$, a vector giving the probability for x to belong to each cluster.

Application to non-rigid shape segmentation

[Skraba, Ovsjanikov, C.,Guibas, NORDIA 10]

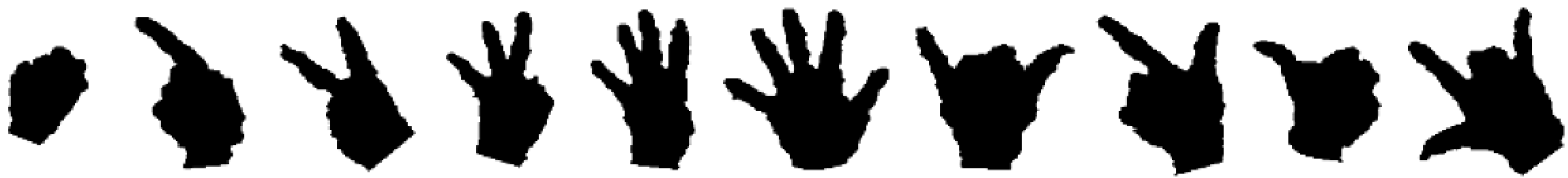


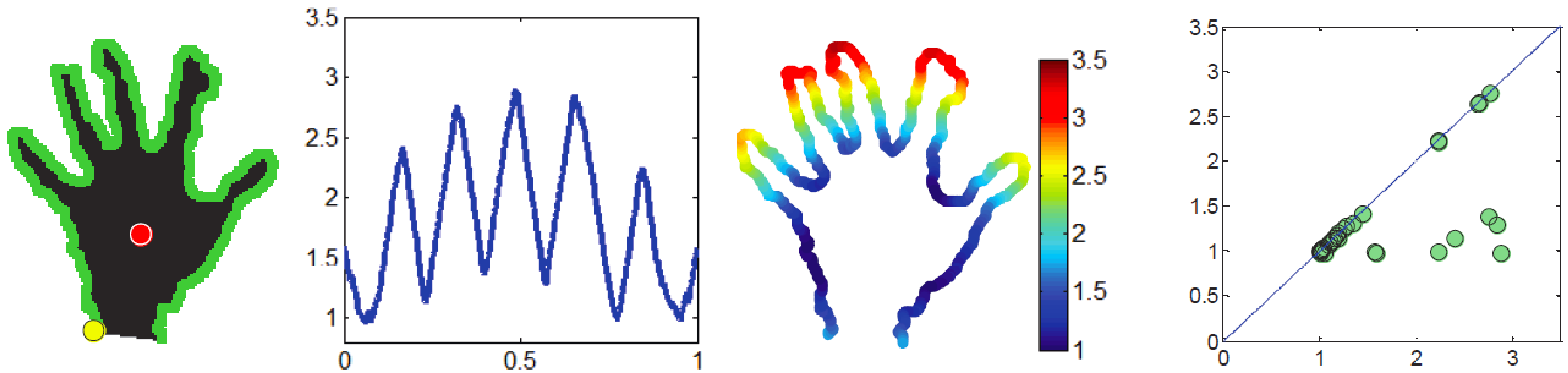


Other applications: classification, object recognition

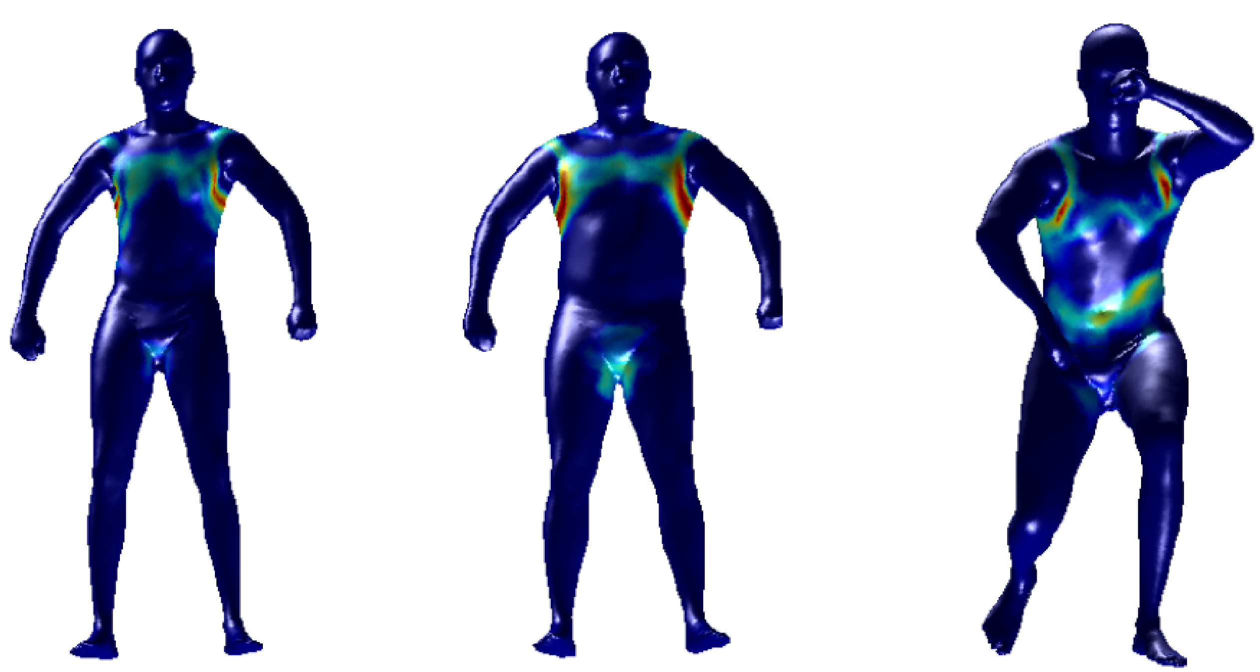
Examples:

- Hand gesture recognition [Li, Ovsjanikov, C. - CVPR'14]

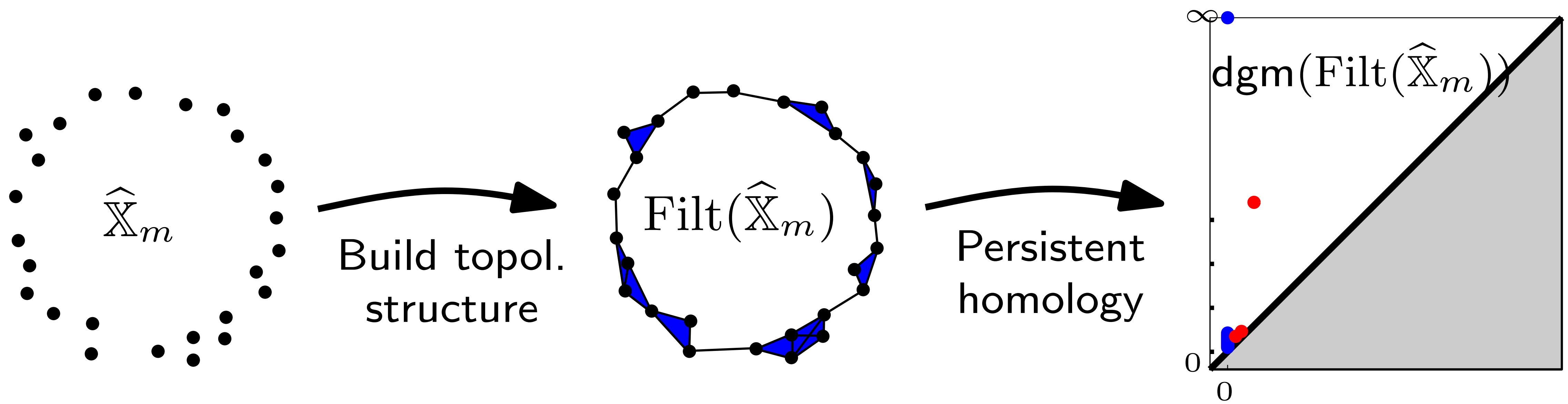




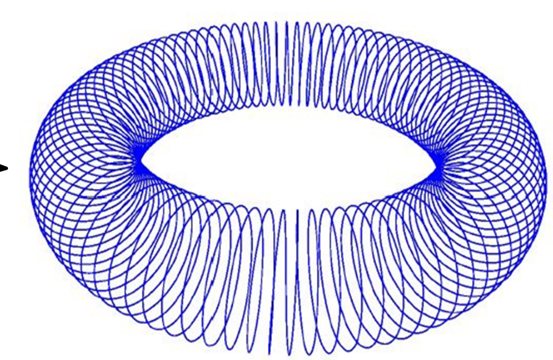
- Persistence-based pooling for shape recognition [Bonis, Ovsjanikov, Oudot, C. 2016]



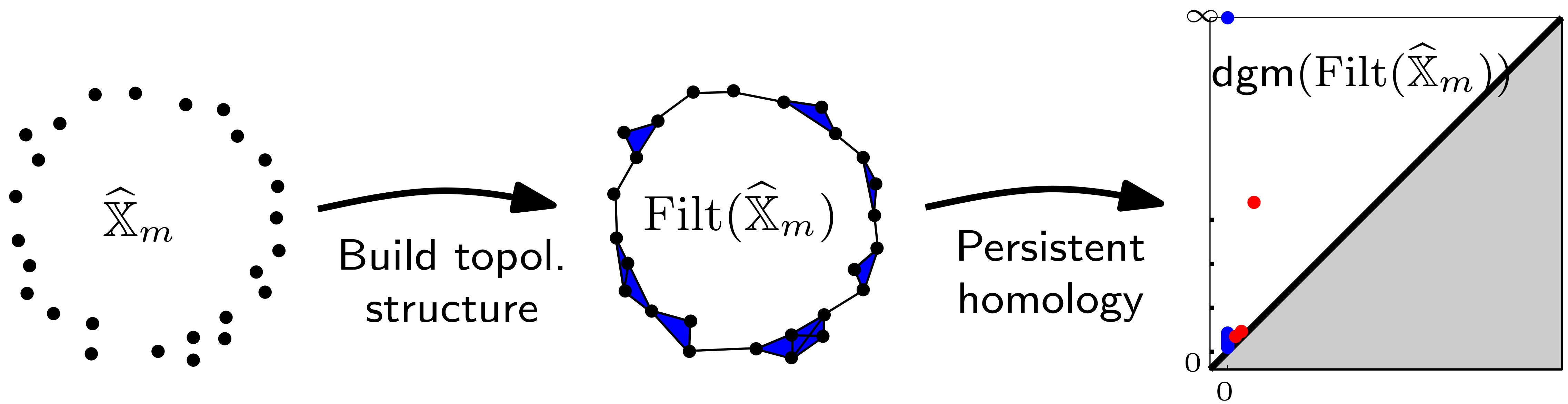
Persistent homology for (point cloud) data



- Challenges and goals:
  $\rightarrow$ no direct access to topological/geometric information: need of intermediate constructions (simplicial complexes);
  $\rightarrow$ distinguish topological "signal" from noise;
  $\rightarrow$ topological information may be multiscale;
  $\rightarrow$ statistical analysis of topological information.

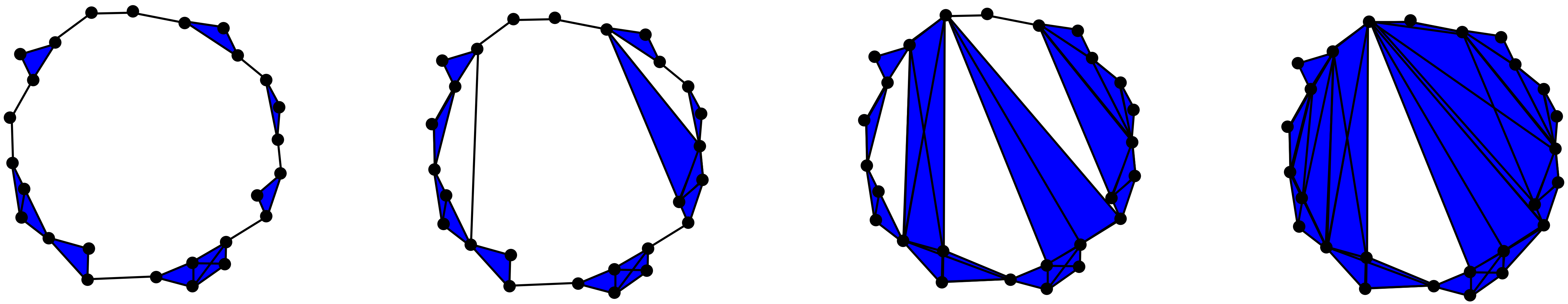


Persistent homology for (point cloud) data



- Build a geometric filtered simplicial complex on top of $\widehat{\mathbb{X}}_m$ $\rightarrow$ multiscale topol. structure.
- Compute the persistent homology of the complex $\rightarrow$ multiscale topol. signature.
- Compare the signatures of "close" data sets $\rightarrow$ robustness and stability results.
- Statistical properties of signatures

Filtered complexes and filtrations



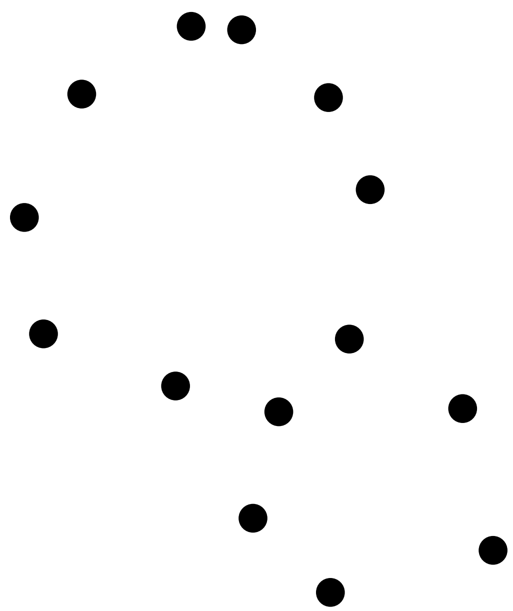
A filtered simplicial complex $\mathbb{S}$ built on top of a set X is a family $(\mathbb{S}_a \mid a \in \mathbf{R})$ of subcomplexes of some fixed simplicial complex $\overline{\mathbb{S}}$ with vertex set X s. t. $\mathbb{S}_a \subseteq \mathbb{S}_b$ for any $a \leq b$.

A filtration $\mathbb{F}$ of a space $\mathbb{X}$ is a nested family $(\mathbb{F}_a \mid a \in \mathbf{R})$ of subspaces of $\mathbb{X}$ such that $\mathbb{F}_a \subseteq \mathbb{F}_b$ for any $a \leq b$.

Example: If $f : \mathbb{X} \to \mathbf{R}$ is a function, then the sublevelsets of f, $\mathbb{F}_a = f^{-1}((-\infty, a])$ define the sublevel set filtration associated to f.

Example: Rips and Cech filtrations

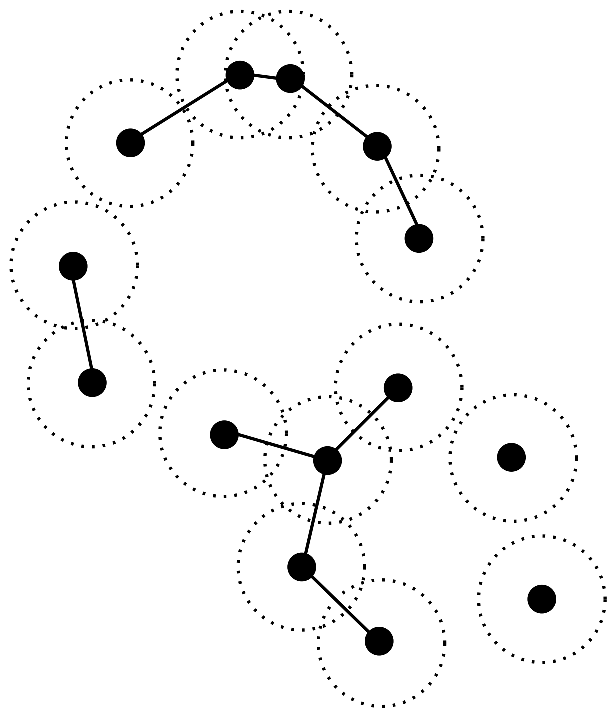
Persistent homology for (point cloud) data



$\widehat{\mathbb{X}}_m$: metric data set

- Build a geometric filtered simplicial complex on top of $\widehat{\mathbb{X}}_m \to$ multiscale topol. structure.
- Compute the persistent homology of the complex $\to$ multiscale topol. signature.
- Compare the signatures of "close" data sets $\to$ robustness and stability results.
- Statistical properties of signatures

Persistent homology for (point cloud) data

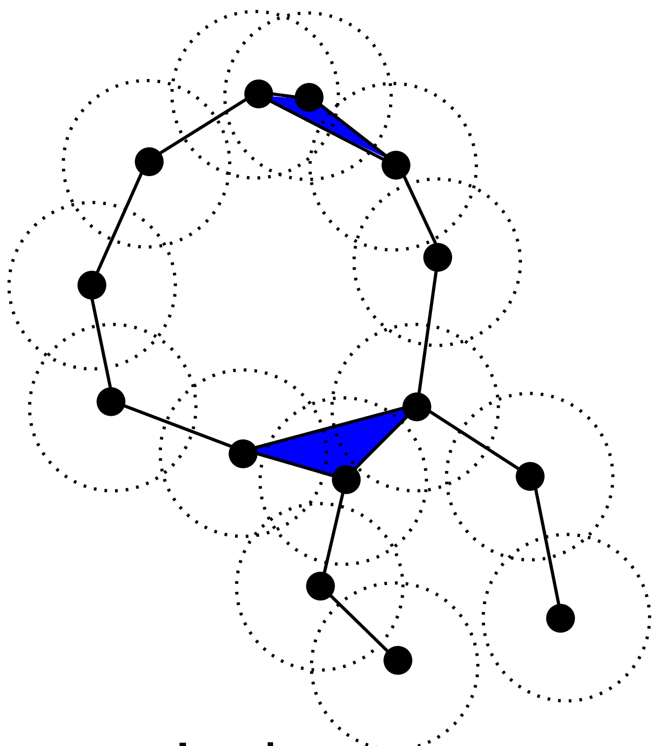


$\widehat{\mathbb{X}}_m$: metric data set

$\mathrm{Filt}(\widehat{\mathbb{X}}_m)$: filtered simplicial complex

- Build a geometric filtered simplicial complex on top of $\widehat{\mathbb{X}}_m$ $\to$ multiscale topol. structure.
- Compute the persistent homology of the complex $\to$ multiscale topol. signature.
- Compare the signatures of "close" data sets $\to$ robustness and stability results.
- Statistical properties of signatures

Persistent homology for (point cloud) data

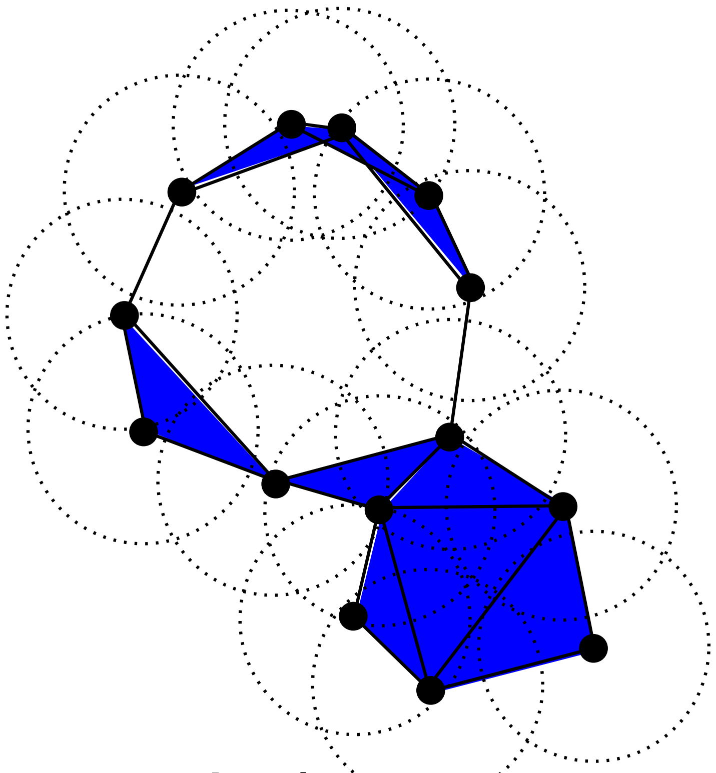


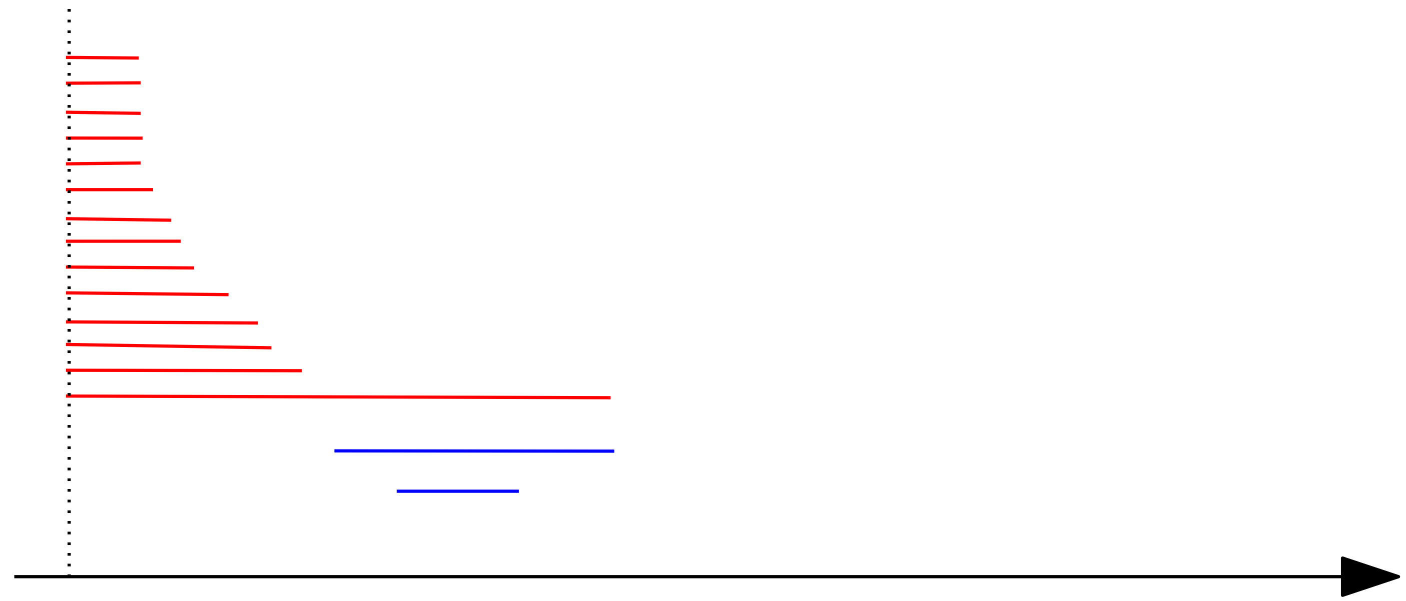
$\widehat{\mathbb{X}}_m$: metric data set

$\mathrm{Filt}(\widehat{\mathbb{X}}_m)$: filtered simplicial complex

- Build a geometric filtered simplicial complex on top of $\widehat{\mathbb{X}}_m \rightarrow$ multiscale topol. structure.
- Compute the persistent homology of the complex $\rightarrow$ multiscale topol. signature.
- Compare the signatures of "close" data sets $\rightarrow$ robustness and stability results.
- Statistical properties of signatures

Persistent homology for (point cloud) data



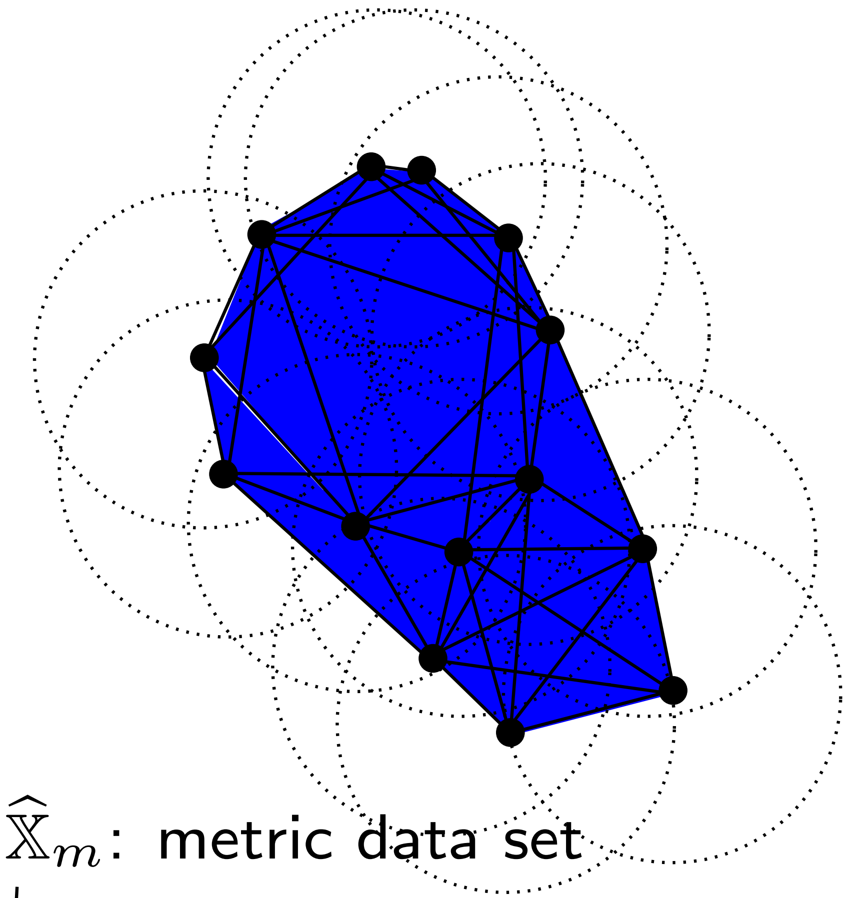


$\widehat{\mathbb{X}}_m$: metric data set

$\mathrm{Filt}(\widehat{\mathbb{X}}_m)$: filtered simplicial complex

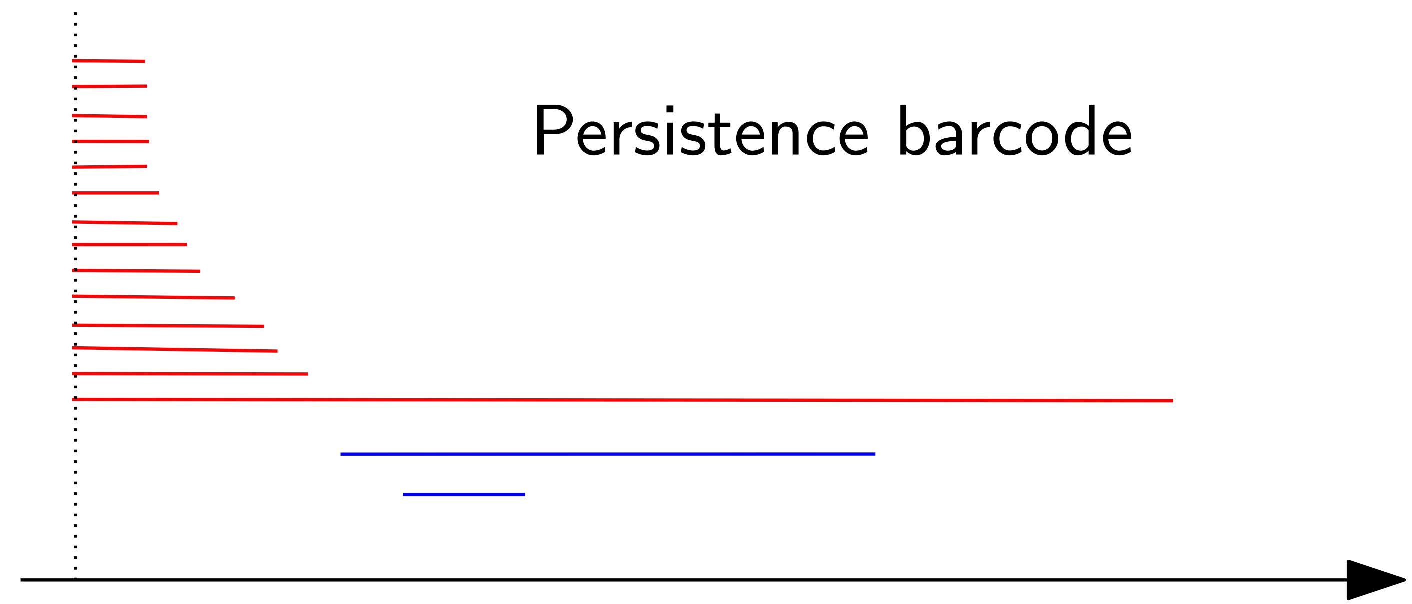
- Build a geometric filtered simplicial complex on top of $\widehat{\mathbb{X}}_m$ $\rightarrow$ multiscale topol. structure.
- Compute the persistent homology of the complex $\rightarrow$ multiscale topol. signature.
- Compare the signatures of "close" data sets $\rightarrow$ robustness and stability results.
- Statistical properties of signatures

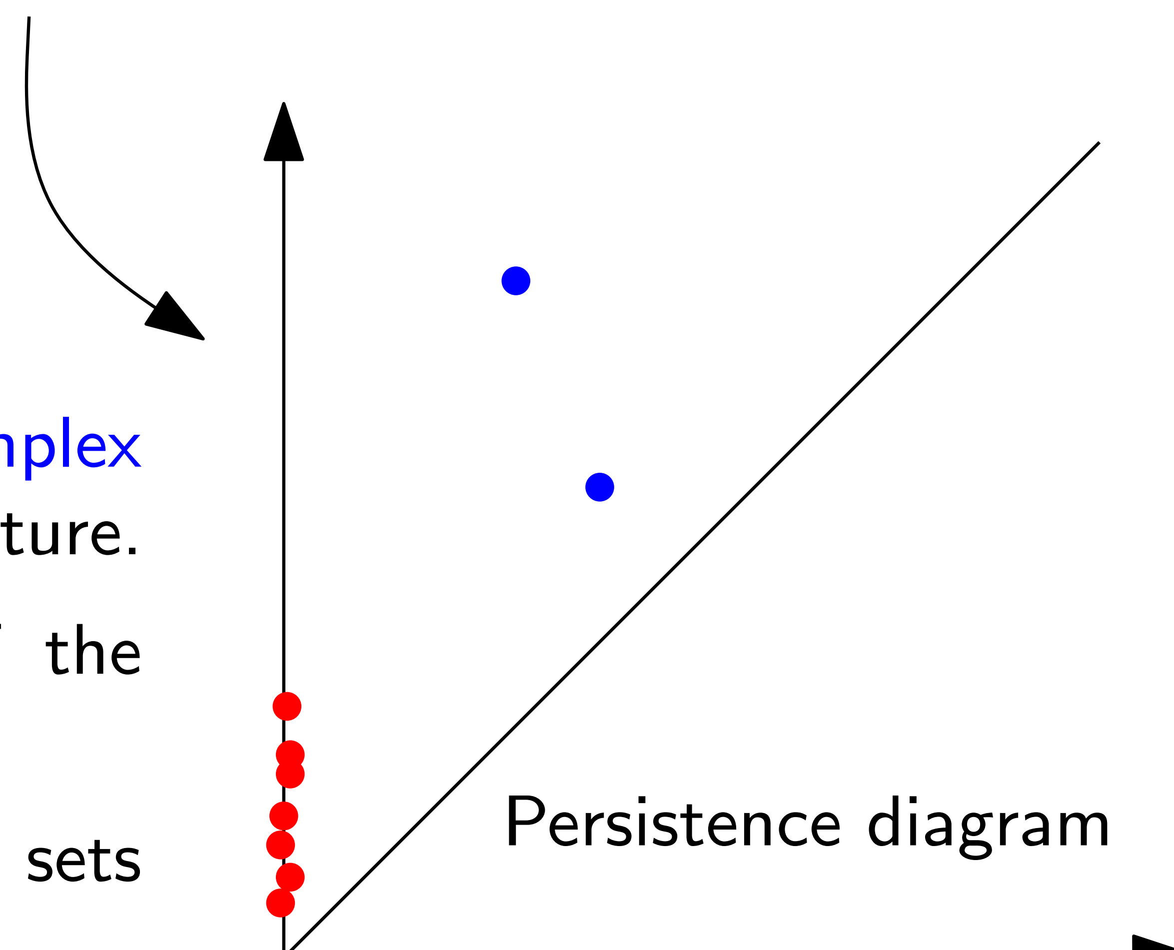
Persistent homology for (point cloud) data



$\widehat{\mathbb{X}}_m$: metric data set

$\text{Filt}(\widehat{\mathbb{X}}_m)$: filtered simplicial complex





- Build a geometric filtered simplicial complex on top of $\widehat{\mathbb{X}}_m$ $\rightarrow$ multiscale topol. structure.
- Compute the persistent homology of the complex $\rightarrow$ multiscale topol. signature.
- Compare the signatures of "close" data sets $\rightarrow$ robustness and stability results.
- Statistical properties of signatures

Stability properties

"Stability theorem": Close spaces/data sets have close persistence diagrams!

[C., de Silva, Oudot - Geom. Dedicata 2013].

If $\mathbb{X}$ and $\mathbb{Y}$ are pre-compact metric spaces, then

$$\boxed{\mathrm{d_b}(\mathsf{dgm}(\mathrm{Rips}(\mathbb{X})), \mathsf{dgm}(\mathrm{Rips}(\mathbb{Y}))) \leq d_{GH}(\mathbb{X}, \mathbb{Y}).}$$

Bottleneck distance

Gromov-Hausdorff distance

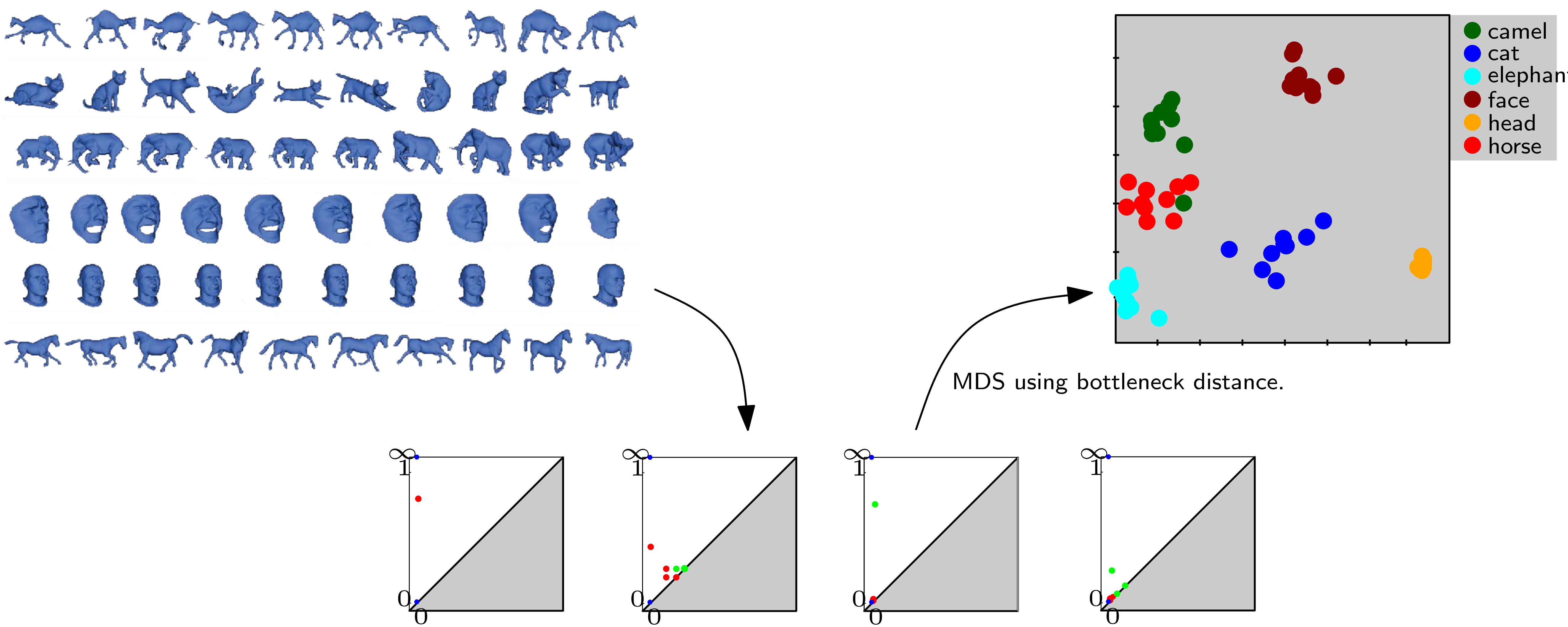
$$d_{GH}(\mathbb{X}, \mathbb{Y}) := \inf_{\mathbb{Z}, \gamma_1, \gamma_2} d_H(\gamma_1(\mathbb{X}), \gamma_2(\mathbb{X}))$$

$\mathbb{Z}$ metric space, $\gamma_1 : \mathbb{X} \to \mathbb{Z}$ and $\gamma_2 : \mathbb{Y} \to \mathbb{Z}$ isometric embeddings.

Rem: This result also holds for other families of filtrations (particular case of a more general theorem).

Application: non rigid shape classification

[C., Cohen-Steiner, Guibas, Mémoli, Oudot - SGP '09]



- Non rigid shapes in a same class are almost isometric, but computing Gromov-Hausdorff distance between shapes is extremely expensive.
- Compare diagrams of sampled shapes instead of shapes themselves.

Where do stability results come from?

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Examples:

- Let $\mathbb{S}$ be a filtered simplicial complex. If $V_a = \mathrm{H}(\mathbb{S}_a)$ and $v_a^b : \mathrm{H}(\mathbb{S}_a) \to \mathrm{H}(\mathbb{S}_b)$ is the linear map induced by the inclusion $\mathbb{S}_a \hookrightarrow \mathbb{S}_b$ then $(\mathrm{H}(\mathbb{S}_a) \mid a \in \mathbf{R})$ is a persistence module.
- Given a metric space $(\mathbb{X}, d_{\mathbb{X}})$, $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ is a persistence module.
- If $f : X \to \mathbf{R}$ is a function, then the filtration defined by the sublevel sets of f, $\mathbb{F}_a = f^{-1}((-\infty, a])$, induces a persistence module at homology level.

Where do stability results come from?

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Definition: A persistence module $\mathbb{V}$ is q-tame if for any $a < b$, v_a^b has a finite rank.

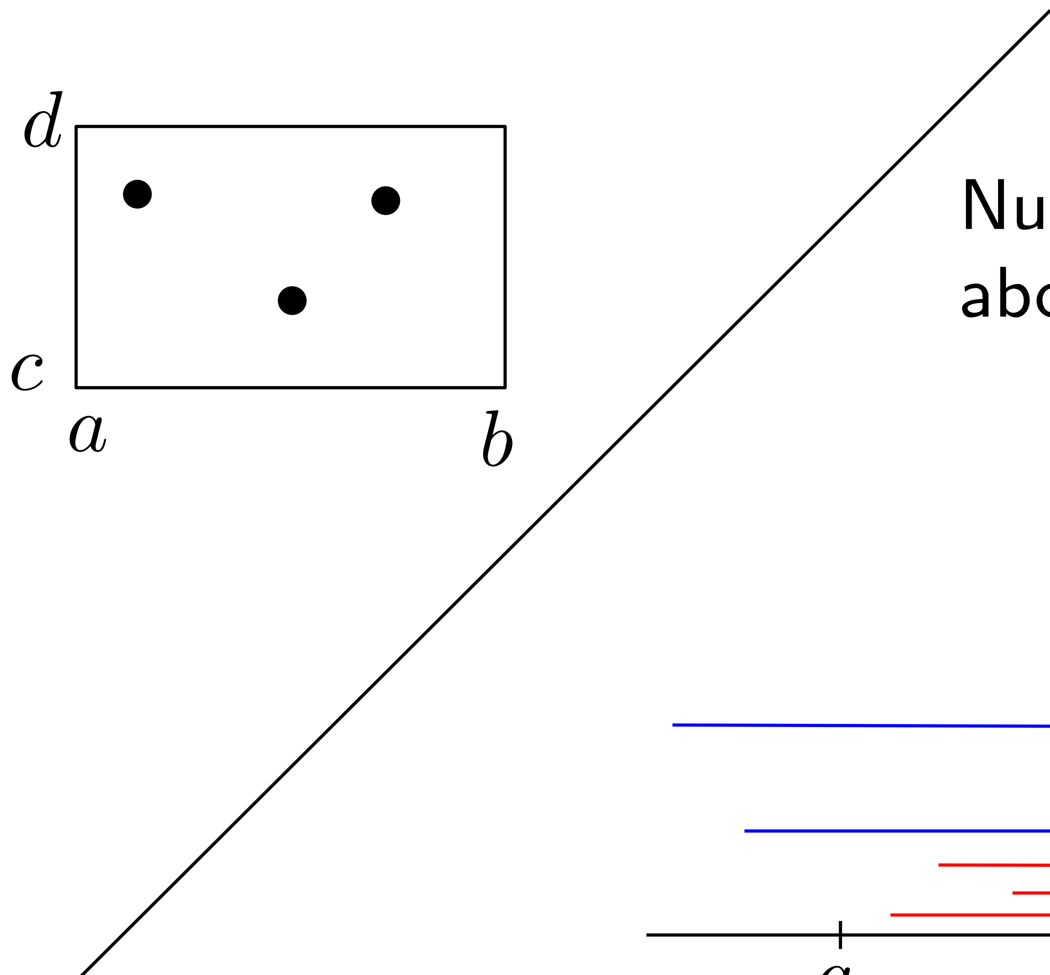
Theorem: [C., Cohen-Steiner, Glisse, Guibas, Oudot - SoCG'09], [C., de Silva, Glisse, Oudot 12]

q-tame persistence modules have well-defined persistence diagrams.

Where do stability results come from?

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

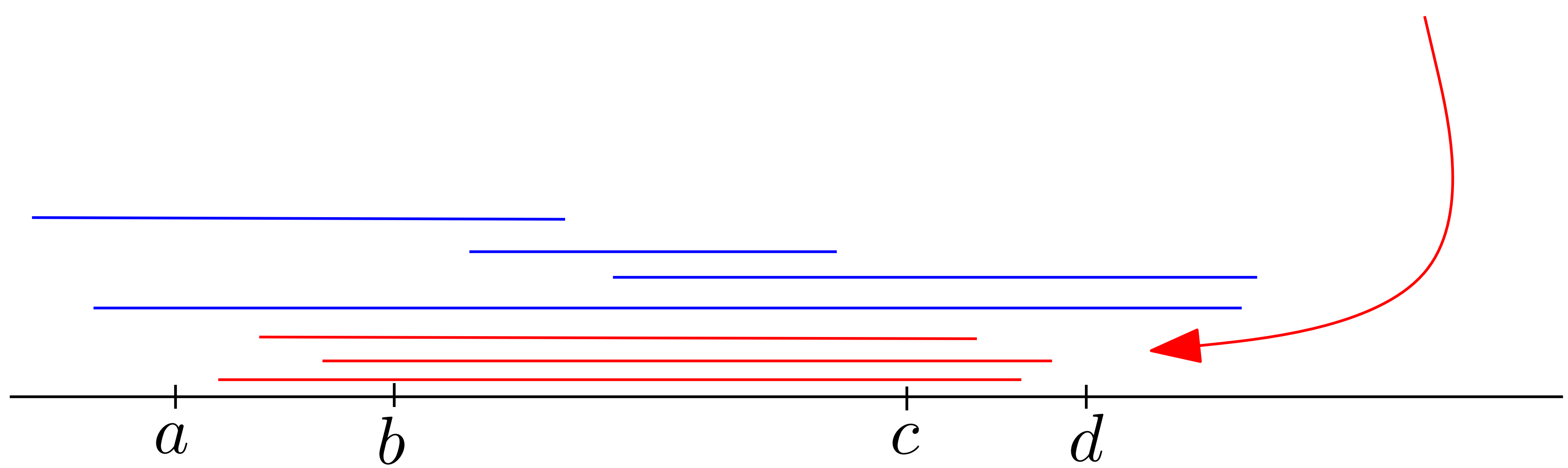
An idea about the definition of persistence diagrams:



Measures on rectangles:

Number of points in any rectangle $[a, b] \times [c, d]$ above the diagonal:

$$rk(v_b^c) - rk(v_b^d) + rk(v_a^d) - rk(v_a^c)$$



Where do stability results come from?

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Definition: A persistence module $\mathbb{V}$ is q-tame if for any $a < b$, v_a^b has a finite rank.

Theorem: [C., Cohen-Steiner, Glisse, Guibas, Oudot - SoCG'09], [C., de Silva, Glisse, Oudot 12]

q-tame persistence modules have well-defined persistence diagrams.

Exercise: Let $\mathbb{X}$ be a precompact metric space. Then $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ are q-tame.

Recall that a metric space $(\mathbb{X}, \rho)$ is precompact if for any $\epsilon > 0$ there exists a finite subset $F_\epsilon \subset \mathbb{X}$ such that $d_H(\mathbb{X}, F_\epsilon) < \epsilon$ (i.e. $\forall x \in X, \exists p \in F_\epsilon$ s.t. $\rho(x, p) < \epsilon$).

Where do stability results come from?

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

A homomorphism of degree ϵ between two persistence modules $\mathbb{U}$ and $\mathbb{V}$ is a collection Φ of linear maps

$$(\phi_a : U_a \to V_{a+\epsilon} \mid a \in \mathbf{R})$$

such that $v_{a+\epsilon}^{b+\epsilon} \circ \phi_a = \phi_b \circ u_a^b$ for all $a \leq b$.

$$\begin{tikzcd}
U^a \arrow[r] \arrow[d] \arrow[dr, phantom, "\circlearrowright"] & U^b \arrow[d] \\
V^{a+\epsilon} \arrow[r] & V^{b+\epsilon}
\end{tikzcd}$$

An ε-interleaving between $\mathbb{U}$ and $\mathbb{V}$ is specified by two homomorphisms of degree ϵ $\Phi : \mathbb{U} \to \mathbb{V}$ and $\Psi : \mathbb{V} \to \mathbb{U}$ s.t. $\Phi \circ \Psi$ and $\Psi \circ \Phi$ are the "shifts" of degree 2ϵ between $\mathbb{U}$ and $\mathbb{V}$.

$$\begin{tikzcd}
\cdots \arrow[rr] & & U^a \arrow[rr, "u_a^{a+2\epsilon}"] \arrow[dr, "\phi_a"] & & U^{a+2\epsilon} \arrow[rr] \arrow[dr] & & \cdots & \\
& \cdots \arrow[ur] \arrow[rr] & & V^{a+\epsilon} \arrow[ur, "\psi_{a+\epsilon}"'] \arrow[rr, "v_{a+\epsilon}^{a+3\epsilon}"'] & & V^{a+3\epsilon} \arrow[rr] & & \cdots
\end{tikzcd}$$

Where do stability results come from?

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Stability Thm: [C., Cohen-Steiner, Glisse, Guibas, Oudot - SoCG '09], [C., de Silva, Glisse Oudot 12]

If $\mathbb{U}$ and $\mathbb{V}$ are q-tame and ϵ-interleaved for some $\epsilon \geq 0$ then

$$d_B(\mathsf{dgm}(\mathbb{U}), \mathsf{dgm}(\mathbb{V})) \leq \epsilon$$

Where do stability results come from?

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

Stability Thm: [C., Cohen-Steiner, Glisse, Guibas, Oudot - SoCG '09], [C., de Silva, Glisse Oudot 12]

If $\mathbb{U}$ and $\mathbb{V}$ are q-tame and ϵ-interleaved for some $\epsilon \geq 0$ then

$$d_B(\mathrm{dgm}(\mathbb{U}), \mathrm{dgm}(\mathbb{V})) \leq \epsilon$$

Exercise: Show the stability theorem for (tame) functions :
let $\mathbb{X}$ be a topological space and let $f, g : \mathbb{X} \to \mathbb{R}$ be two *tame* functions. Then

$$\mathrm{d_B}(\mathrm{D}_f, \mathrm{D}_g) \leq \|f - g\|_\infty .$$

Where do stability results come from?

Definition: A persistence module $\mathbb{V}$ is an indexed family of vector spaces $(V_a \mid a \in \mathbf{R})$ and a doubly-indexed family of linear maps $(v_a^b : V_a \to V_b \mid a \leq b)$ which satisfy the composition law $v_b^c \circ v_a^b = v_a^c$ whenever $a \leq b \leq c$, and where v_a^a is the identity map on V_a.

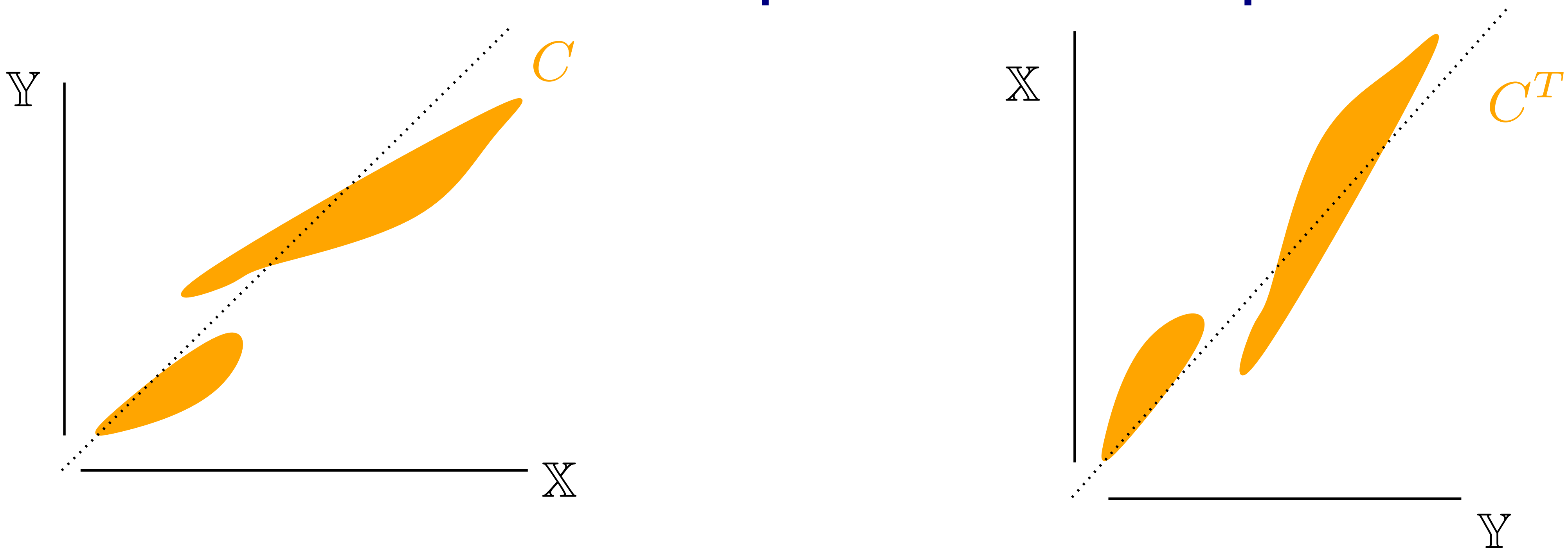
Stability Thm: [C., Cohen-Steiner, Glisse, Guibas, Oudot - SoCG '09], [C., de Silva, Glisse Oudot 12]

If $\mathbb{U}$ and $\mathbb{V}$ are q-tame and ϵ-interleaved for some $\epsilon \geq 0$ then

$$d_B(\mathrm{dgm}(\mathbb{U}), \mathrm{dgm}(\mathbb{V})) \leq \epsilon$$

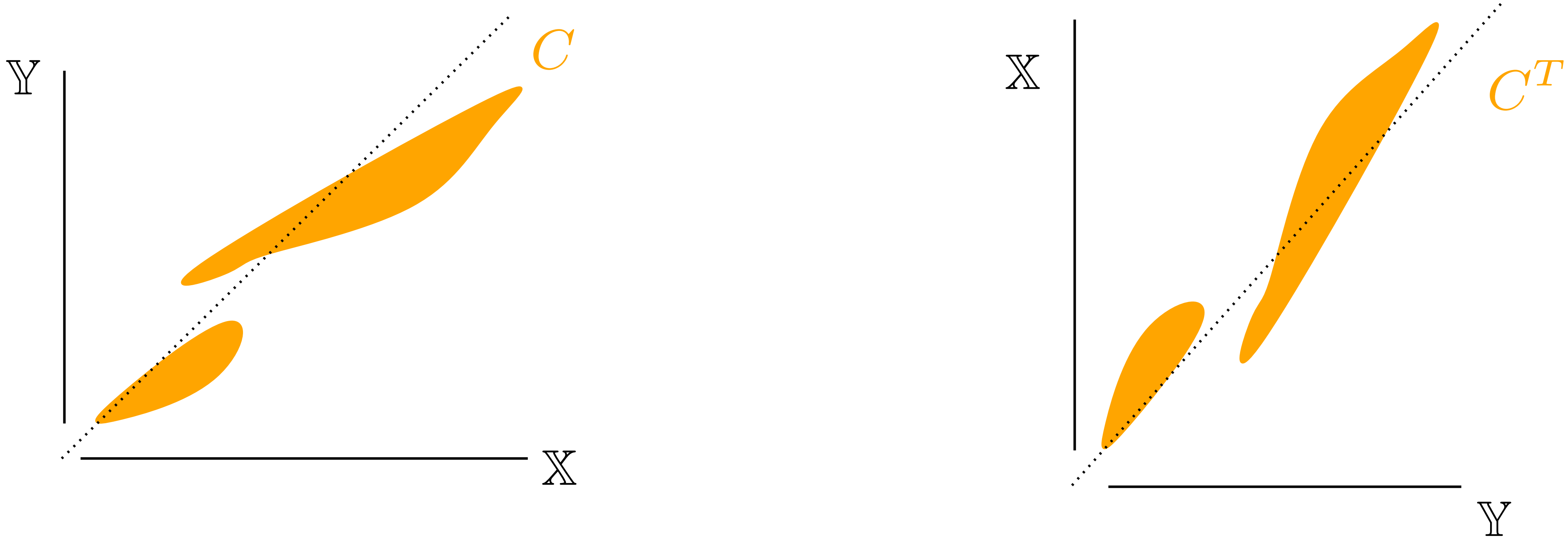
Strategy: build filtrations that induce **q-tame** homology persistence modules and that turn out to be **ϵ-interleaved** when the considered spaces/functions are *$O(\epsilon)$-close*.

Multivalued maps and correspondences



A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ from a set $\mathbb{X}$ to a set $\mathbb{Y}$ is a subset of $\mathbb{X} \times \mathbb{Y}$, also denoted C, that projects surjectively onto $\mathbb{X}$ through the canonical projection $\pi_{\mathbb{X}} : \mathbb{X} \times \mathbb{Y} \to \mathbb{X}$. The image $C(\sigma)$ of a subset σ of $\mathbb{X}$ is the canonical projection onto $\mathbb{Y}$ of the preimage of σ through $\pi_{\mathbb{X}}$.

Multivalued maps and correspondences

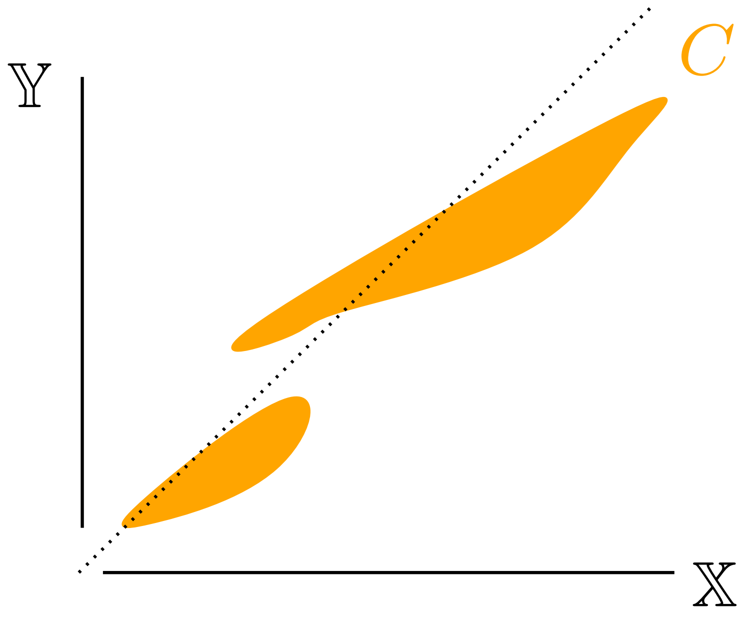


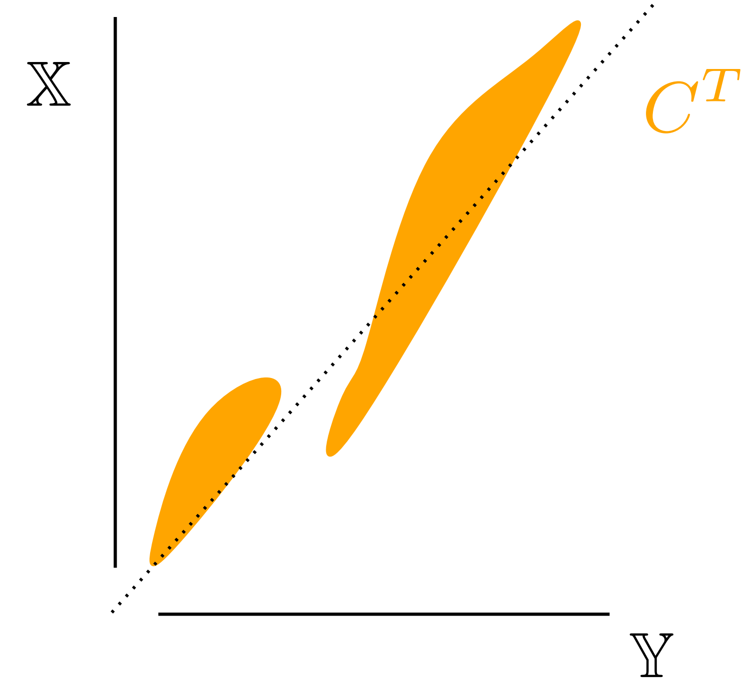
A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ from a set $\mathbb{X}$ to a set $\mathbb{Y}$ is a subset of $\mathbb{X} \times \mathbb{Y}$, also denoted C, that projects surjectively onto $\mathbb{X}$ through the canonical projection $\pi_{\mathbb{X}} : \mathbb{X} \times \mathbb{Y} \to \mathbb{X}$. The image $C(\sigma)$ of a subset σ of $\mathbb{X}$ is the canonical projection onto $\mathbb{Y}$ of the preimage of σ through $\pi_{\mathbb{X}}$.

The transpose of C, denoted C^T, is the image of C through the symmetry map $(x, y) \mapsto (y, x)$.

A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is a correspondence if C^T is also a multivalued map.

Multivalued maps and correspondences

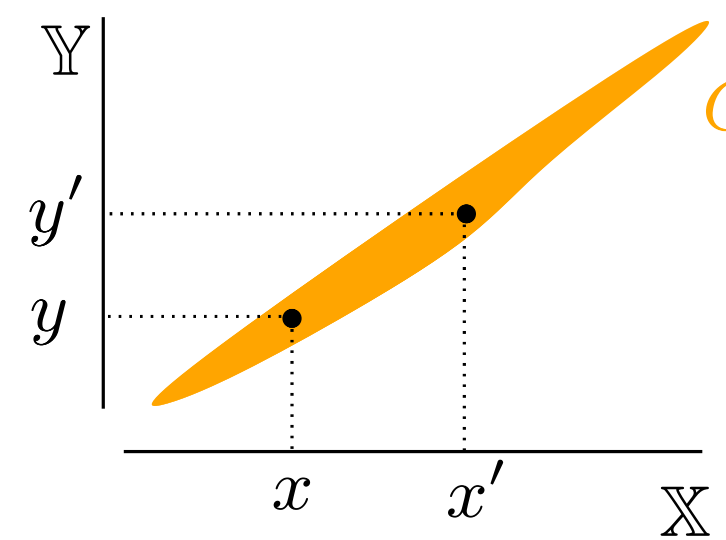




A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ from a set $\mathbb{X}$ to a set $\mathbb{Y}$ is a subset of $\mathbb{X} \times \mathbb{Y}$, also denoted C, that projects surjectively onto $\mathbb{X}$ through the canonical projection $\pi_{\mathbb{X}} : \mathbb{X} \times \mathbb{Y} \to \mathbb{X}$. The image $C(\sigma)$ of a subset σ of $\mathbb{X}$ is the canonical projection onto $\mathbb{Y}$ of the preimage of σ through $\pi_{\mathbb{X}}$.

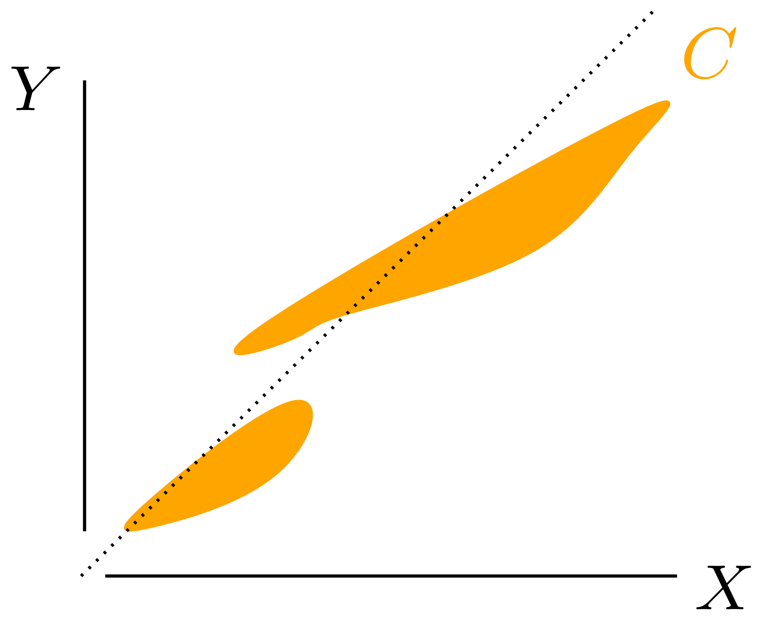
Example: ϵ-correspondence and Gromov-Hausdorff distance.

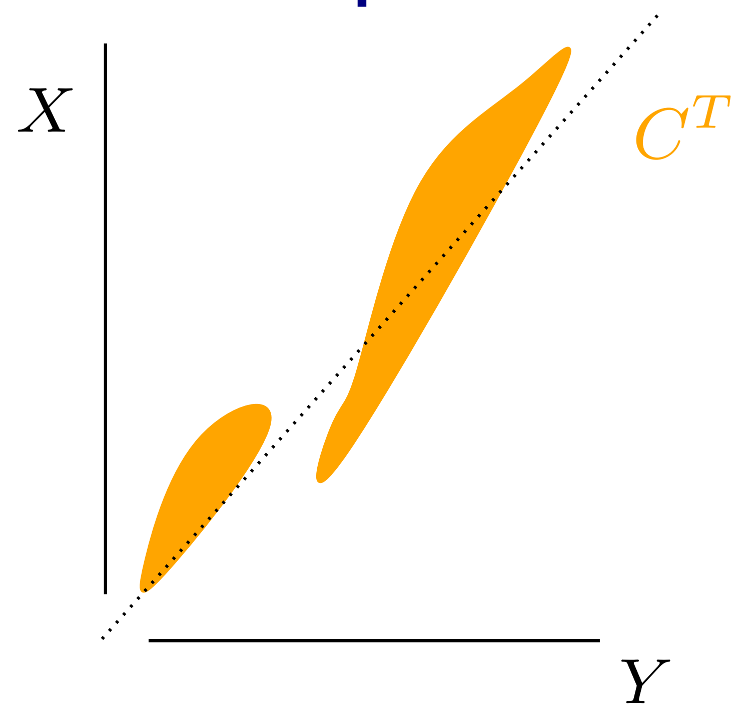
Let $(\mathbb{X}, \rho_{\mathbb{X}})$ and $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be compact metric spaces.
A correspondence $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is an ϵ-correspondence if $\forall (x, y), (x', y') \in C$, $|\rho_{\mathbb{X}}(x, x') - \rho_{\mathbb{Y}}(y, y')| \leq \varepsilon$.



$$d_{GH}(\mathbb{X}, \mathbb{Y}) = \frac{1}{2} \inf\{\varepsilon \geq 0 : \text{there exists an } \varepsilon\text{-correspondence between } \mathbb{X} \text{and } \mathbb{Y}\}$$

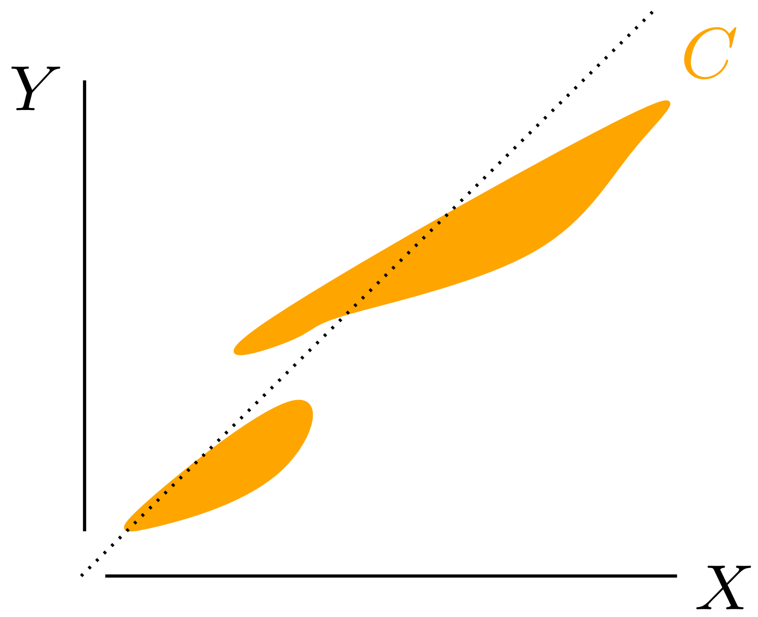
Multivalued simplicial maps

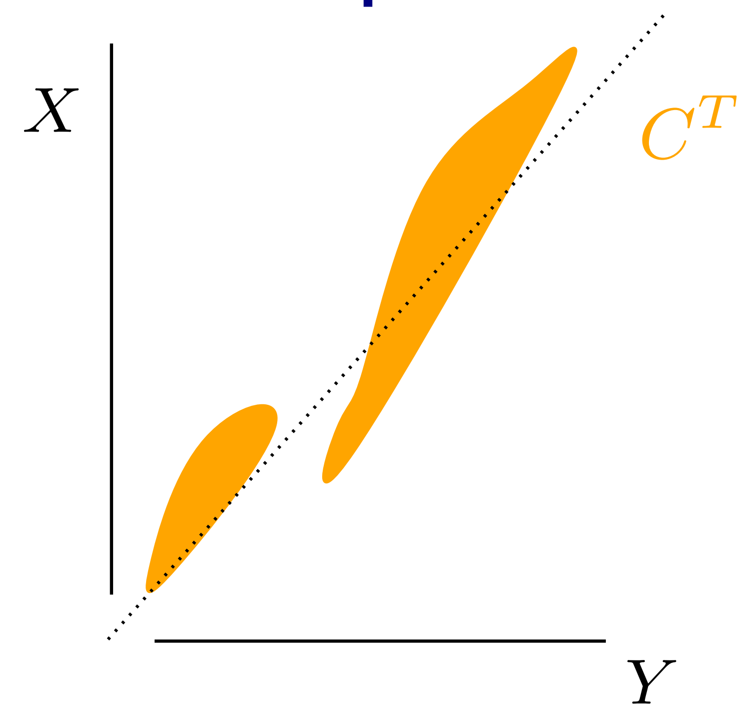




Let $\mathbb{S}$ and $\mathbb{T}$ be two filtered simplicial complexes with vertex sets $\mathbb{X}$ and $\mathbb{Y}$ respectively. A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is ε-simplicial from $\mathbb{S}$ to $\mathbb{T}$ if for any $a \in \mathbf{R}$ and any simplex $\sigma \in \mathbb{S}_a$, every finite subset of $C(\sigma)$ is a simplex of $\mathbb{T}_{a+\varepsilon}$.

Multivalued simplicial maps





Let $\mathbb{S}$ and $\mathbb{T}$ be two filtered simplicial complexes with vertex sets $\mathbb{X}$ and $\mathbb{Y}$ respectively. A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is ε-simplicial from $\mathbb{S}$ to $\mathbb{T}$ if for any $a \in \mathbf{R}$ and any simplex $\sigma \in \mathbb{S}_a$, every finite subset of $C(\sigma)$ is a simplex of $\mathbb{T}_{a+\varepsilon}$.

Proposition: Let $\mathbb{S}$, $\mathbb{T}$ be filtered complexes with vertex sets $\mathbb{X}$, $\mathbb{Y}$ respectively. If $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is a correspondence such that C and C^T are both ε-simplicial, then together they induce a canonical ε-interleaving between $\mathrm{H}(\mathbb{S})$ and $\mathrm{H}(\mathbb{T})$.

The example of the Rips and Čech filtrations

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}})$, $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\mathrm{Rips}(\mathbb{Y}))$ are ϵ-interleaved.

The example of the Rips and Čech filtrations

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}})$, $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\mathrm{Rips}(\mathbb{Y}))$ are ϵ-interleaved.

Proof: Let $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ be a correspondence with distortion at most ϵ.
If $\sigma \in \mathrm{Rips}(\mathbb{X}, a)$ then $\rho_{\mathbb{X}}(x, x') \leq a$ for all $x, x' \in \sigma$.
Let $\tau \subseteq C(\sigma)$ be any finite subset.
For any $y, y' \in \tau$ there exist $x, x' \in \sigma$ s. t. $y \in C(x)$, $y' \in C(x')$ so

$$\rho_{\mathbb{Y}}(y, y') \leq \rho_{\mathbb{X}}(x, x') + \epsilon \leq a + \epsilon \text{ and } \tau \in \mathrm{Rips}(\mathbb{Y}, a + \epsilon)$$

$\Rightarrow$ C is ϵ-simplicial from $\mathrm{Rips}(\mathbb{X})$ to $\mathrm{Rips}(\mathbb{Y})$.
Symetrically, C^T is ϵ-simplicial from $\mathrm{Rips}(\mathbb{Y})$ to $\mathrm{Rips}(\mathbb{X})$.

The example of the Rips and Čech filtrations

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}})$, $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d_{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\mathrm{Rips}(\mathbb{Y}))$ are ϵ-interleaved.

Proof: Let $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ be a correspondence with distortion at most ϵ.
If $\sigma \in \mathrm{Rips}(\mathbb{X}, a)$ then $\rho_{\mathbb{X}}(x, x') \leq a$ for all $x, x' \in \sigma$.
Let $\tau \subseteq C(\sigma)$ be any finite subset.
For any $y, y' \in \tau$ there exist $x, x' \in \sigma$ s. t. $y \in C(x)$, $y' \in C(x')$ so

$$\rho_{\mathbb{Y}}(y, y') \leq \rho_{\mathbb{X}}(x, x') + \epsilon \leq a + \epsilon \text{ and } \tau \in \mathrm{Rips}(\mathbb{Y}, a + \epsilon)$$

$\Rightarrow C$ is ϵ-simplicial from $\mathrm{Rips}(\mathbb{X})$ to $\mathrm{Rips}(\mathbb{Y})$.
Symetrically, C^T is ϵ-simplicial from $\mathrm{Rips}(\mathbb{Y})$ to $\mathrm{Rips}(\mathbb{X})$.

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}}), (\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d_{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{Y}))$ are ϵ-interleaved.

The example of the Rips and Čech filtrations

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}})$, $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\mathrm{Rips}(\mathbb{Y}))$ are ϵ-interleaved.

Proof: Let $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ be a correspondence with distortion at most ϵ.
If $\sigma \in \mathrm{Rips}(\mathbb{X}, a)$ then $\rho_{\mathbb{X}}(x, x') \leq a$ for all $x, x' \in \sigma$.
Let $\tau \subseteq C(\sigma)$ be any finite subset.
For any $y, y' \in \tau$ there exist $x, x' \in \sigma$ s. t. $y \in C(x)$, $y' \in C(x')$ so

$$\rho_{\mathbb{Y}}(y, y') \leq \rho_{\mathbb{X}}(x, x') + \epsilon \leq a + \epsilon \text{ and } \tau \in \mathrm{Rips}(\mathbb{Y}, a + \epsilon)$$

$\Rightarrow C$ is ϵ-simplicial from $\mathrm{Rips}(\mathbb{X})$ to $\mathrm{Rips}(\mathbb{Y})$.
Symetrically, C^T is ϵ-simplicial from $\mathrm{Rips}(\mathbb{Y})$ to $\mathrm{Rips}(\mathbb{X})$.

Proposition: Let $(\mathbb{X}, \rho_{\mathbb{X}}), (\mathbb{Y}, \rho_{\mathbb{Y}})$ be metric spaces. For any $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ the persistence modules $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{Y}))$ are ϵ-interleaved.

Remark: Similar results for witness complexes (fixed landmarks)

Tameness of the Rips and Čech filtrations

Theorem: Let $\mathbb{X}$ be a compact metric space. Then $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ are q-tame.

As a consequence dgm$(\mathrm{H}(\mathrm{Rips}(\mathbb{X})))$ and dgm$(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X})))$ are well-defined!

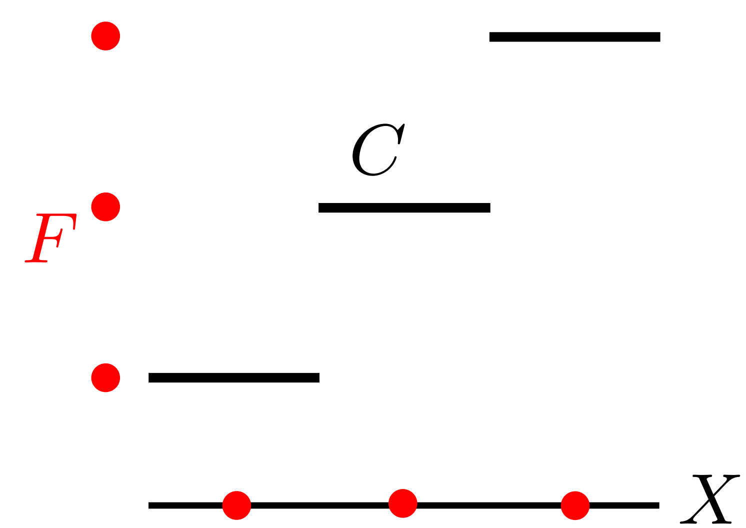
Tameness of the Rips and Čech filtrations

Theorem: Let $\mathbb{X}$ be a compact metric space. Then $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ are q-tame.
As a consequence $\mathrm{dgm}(\mathrm{H}(\mathrm{Rips}(\mathbb{X})))$ and $\mathrm{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X})))$ are well-defined!

Proof: show that $I_a^b : \mathrm{H}(\mathrm{Rips}(X, a)) \to \mathrm{H}(\mathrm{Rips}(X, b))$ has finite rank whenever $a < b$.

Let $\epsilon = (b - a)/2$ and let $F \subset X$ be finite s. t. $d_H(X, F) \leq \epsilon/2$.

Then $C = \{(x, f) \in X \times F | d(x, f) \leq \epsilon/2\}$ is an ϵ-correspondence.



Using the interleaving map, I_a^b factorizes as

$$\mathbf{H}\mathrm{Rips}(X, a) \to \mathbf{H}\mathrm{Rips}(F, a + \varepsilon) \to \mathbf{H}\mathrm{Rips}(X, a + 2\varepsilon) = \mathbf{H}\mathrm{Rips}(X, b)$$

finite dimensional

Tameness of the Rips and Čech filtrations

Theorem: Let $\mathbb{X}$ be a compact metric space. Then $\mathrm{H}(\mathrm{Rips}(\mathbb{X}))$ and $\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))$ are q-tame.
As a consequence $\mathsf{dgm}(\mathrm{H}(\mathrm{Rips}(\mathbb{X})))$ and $\mathsf{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X})))$ are well-defined!

Theorem: Let $\mathbb{X}, \mathbb{Y}$ be compact metric spaces. Then

$$\mathrm{d_b}(\mathsf{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))), \mathsf{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{Y})))) \leq 2\mathrm{d_{GH}}(\mathbb{X}, \mathbb{Y}),$$

$$\mathrm{d_b}(\mathsf{dgm}(\mathrm{H}(\mathrm{Rips}(\mathbb{X}))), \mathsf{dgm}(\mathrm{H}(\mathrm{Rips}(\mathbb{Y})))) \leq 2\mathrm{d_{GH}}(\mathbb{X}, \mathbb{Y}).$$

Remark: The proofs never use the triangle inequality! The previous approch and results easily extend to other settings like, e.g. spaces endowed with a similarity measure.

Why persistence

- Even when X is compact, $\mathrm{H}_p(\mathrm{Rips}(X, a))$, $p \geq 1$, might be infinite dimensional for some value of a:

It is also possible to build such an example with the open Rips complex:

$[x_0, x_1, \cdots, x_k] \in \mathrm{Rips}(X, a^-) \Leftrightarrow d_X(x_i, x_j) < a$, for all i, j

Why persistence

- Even when X is compact, $\mathrm{H}_p(\mathrm{Rips}(X,a))$, $p \geq 1$, might be infinite dimensional for some value of a:

It is also possible to build such an example with the open Rips complex:

$[x_0, x_1, \cdots, x_k] \in \mathrm{Rips}(X, a^-) \Leftrightarrow d_X(x_i, x_j) < a$, for all i, j

- For any $\alpha, \beta \in \mathbf{R}$ such that $0 < \alpha \leq \beta$ and any integer k there exists a compact metric space X such that for any $a \in [\alpha, \beta]$, $\mathrm{H}_k(\mathrm{Rips}(X,a))$ has a non countable infinite dimension (can be embedded in $\mathbf{R}^4$ [Droz 2013]).

Why persistence

- Even when X is compact, $\mathrm{H}_p(\mathrm{Rips}(X,a))$, $p \geq 1$, might be infinite dimensional for some value of a:

It is also possible to build such an example with the open Rips complex:

$$[x_0, x_1, \cdots, x_k] \in \mathrm{Rips}(X, a^-) \Leftrightarrow d_X(x_i, x_j) < a, \text{ for all } i, j$$

- For any $\alpha, \beta \in \mathbf{R}$ such that $0 < \alpha \leq \beta$ and any integer k there exists a compact metric space X such that for any $a \in [\alpha, \beta]$, $\mathrm{H}_k(\mathrm{Rips}(X,a))$ has a non countable infinite dimension (can be embedded in $\mathbf{R}^4$ [Droz 2013]).

- If X is compact, then $\dim \mathrm{H}_1(\check{\mathrm{C}}\mathrm{ech}(X,a)) < +\infty$ for all a ([Smale-Smale, C.-de Silva]).

- If X is geodesic, then $\dim \mathrm{H}_1(\mathrm{Rips}(X,a)) < +\infty$ for all $a > 0$ and $\mathrm{Dgm}(\mathrm{H}_1(\mathbb{R}\mathrm{ips}(X)))$ is contained in the vertical line $x = 0$.

- If X is a geodesic δ-hyperbolic space then $\mathrm{Dgm}(\mathrm{H}_2(\mathbb{R}\mathrm{ips}(X)))$ is contained in a vertical band of width $O(\delta)$.

Some weaknesses

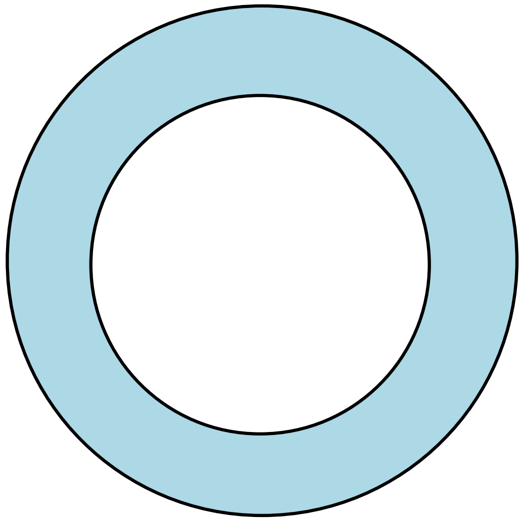
If $\mathbb{X}$ and $\mathbb{Y}$ are pre-compact metric spaces, then

$$\boxed{\mathrm{d_b}(\mathsf{dgm}(\mathrm{Rips}(\mathbb{X})), \mathsf{dgm}(\mathrm{Rips}(\mathbb{Y}))) \leq d_{GH}(\mathbb{X}, \mathbb{Y}).}$$

$\rightarrow$ Vietoris-Rips (or Cech, witness) filtrations quickly become prohibitively large as the size of the data increases ($\mathrm{O}(|\mathbb{X}|^d)$), making the computation of persistence practically almost impossible.

$\rightarrow$ Persistence diagrams of Rips-Vietoris (and Cěch, witness,..) filtrations and Gromov-Hausdorff distance are very sensitive to noise and outliers.

Statistical setting



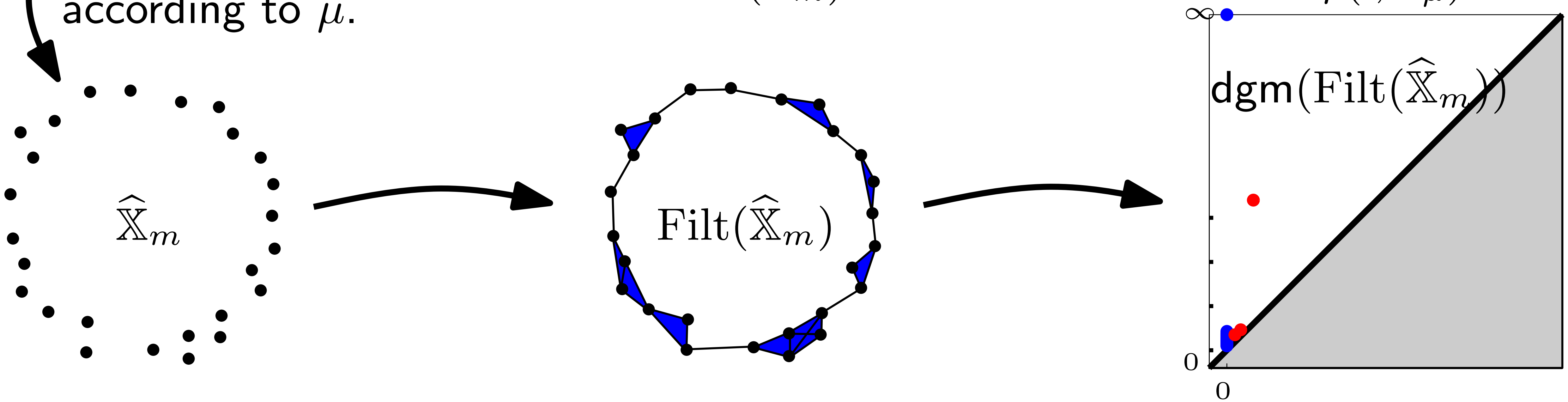
$(\mathbb{M}, \rho)$ metric space

μ a probability measure with compact support $\mathbb{X}_\mu$.

Examples:

- $\mathrm{Filt}(\widehat{\mathbb{X}}_m) = \mathrm{Rips}_\alpha(\widehat{\mathbb{X}}_m)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_m) = \check{\mathrm{C}}\mathrm{ech}_\alpha(\widehat{\mathbb{X}}_m)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_m) =$ sublevelset filtration of $\rho(., \mathbb{X}_\mu)$.

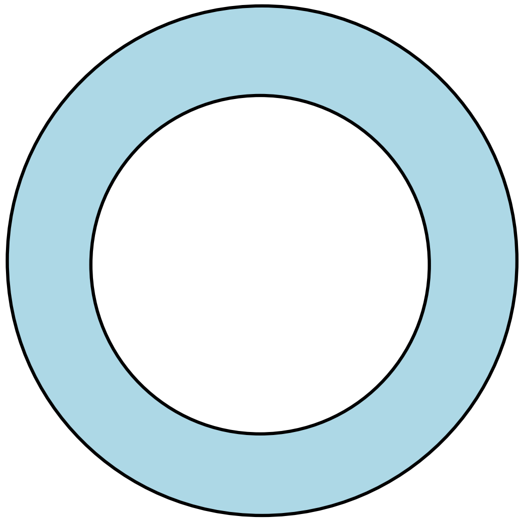
Sample m points according to μ.



Questions:

- Statistical properties of $\mathrm{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_m))$? $\mathrm{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_m)) \to ?$ as $m \to +\infty$?

Statistical setting



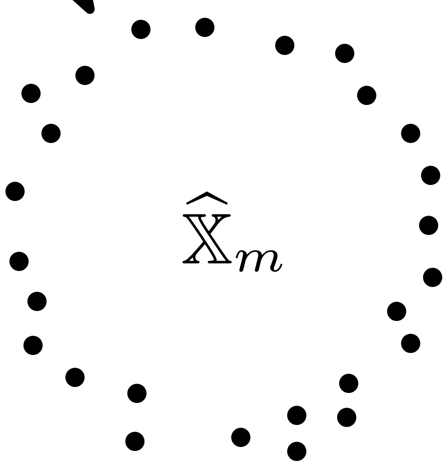
$(\mathbb{M}, \rho)$ metric space

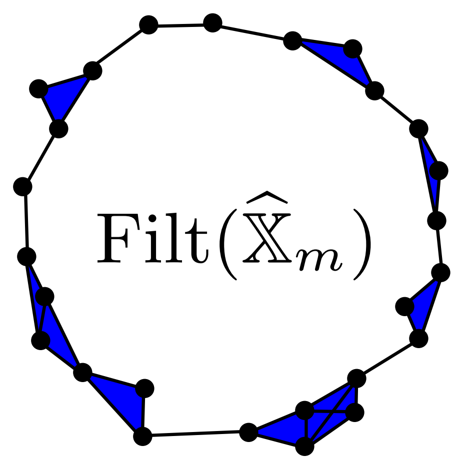
μ a probability measure with compact support $\mathbb{X}_\mu$.

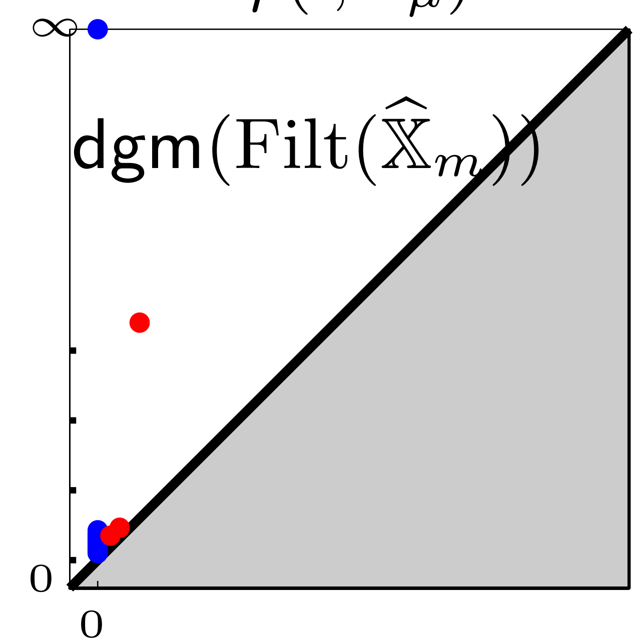
Examples:

- $\mathrm{Filt}(\widehat{\mathbb{X}}_m) = \mathrm{Rips}_\alpha(\widehat{\mathbb{X}}_m)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_m) = \check{\mathrm{C}}\mathrm{ech}_\alpha(\widehat{\mathbb{X}}_m)$
- $\mathrm{Filt}(\widehat{\mathbb{X}}_m) =$ sublevelset filtration of $\rho(., \mathbb{X}_\mu)$.

Sample m points according to μ.



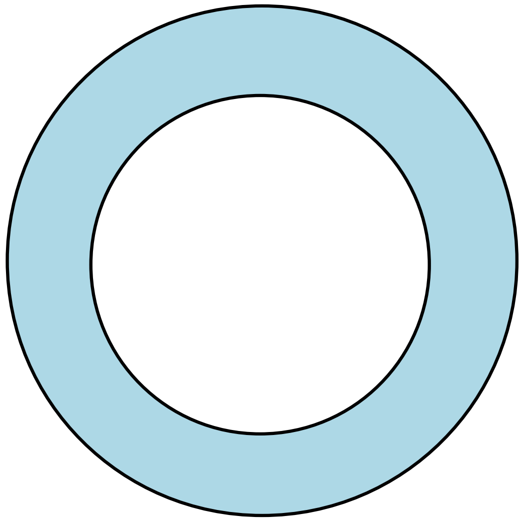




Questions:

- Statistical properties of dgm$(\mathrm{Filt}(\widehat{\mathbb{X}}_m))$? dgm$(\mathrm{Filt}(\widehat{\mathbb{X}}_m)) \to ?$ as $m \to +\infty$?
- Can we do more statistics with persistence diagrams?

Statistical setting

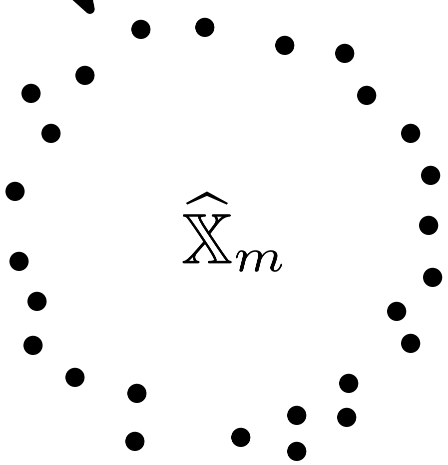


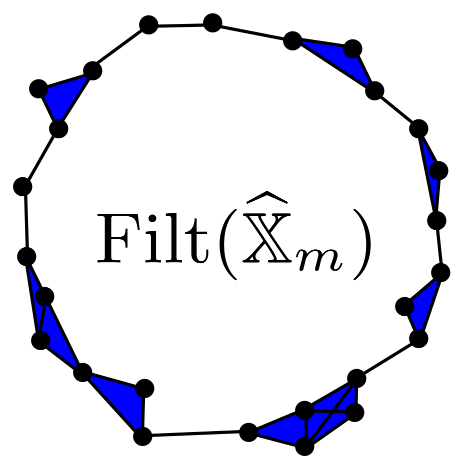
$(\mathbb{M}, \rho)$ metric space

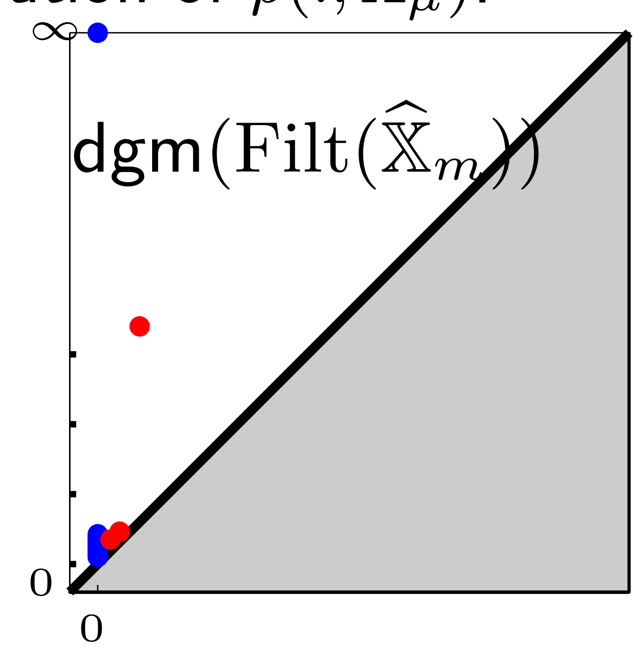
μ a probability measure with compact support $\mathbb{X}_\mu$.

Examples:
- $\text{Filt}(\widehat{\mathbb{X}}_m) = \text{Rips}_\alpha(\widehat{\mathbb{X}}_m)$
- $\text{Filt}(\widehat{\mathbb{X}}_m) = \check{\text{C}}\text{ech}_\alpha(\widehat{\mathbb{X}}_m)$
- $\text{Filt}(\widehat{\mathbb{X}}_m) =$ sublevelset filtration of $\rho(., \mathbb{X}_\mu)$.

Sample m points according to μ.







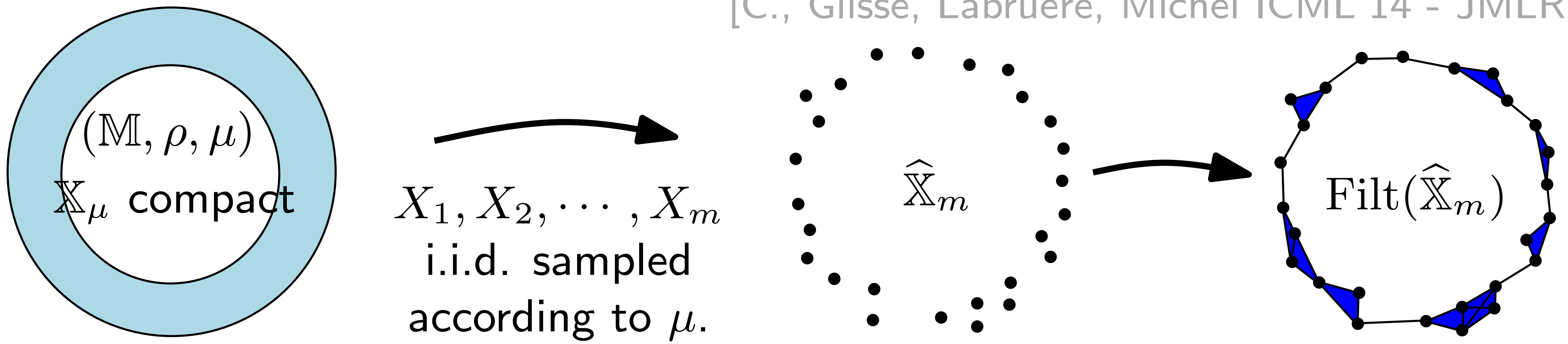
Stability thm: $\mathrm{d_b}(\text{dgm}(\text{Filt}(\mathbb{X}_\mu)), \text{dgm}(\text{Filt}(\widehat{\mathbb{X}}_m))) \leq 2 d_{GH}(\mathbb{X}_\mu, \widehat{\mathbb{X}}_m)$

So, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\mathrm{d_b}\left(\text{dgm}(\text{Filt}(\mathbb{X}_\mu)), \text{dgm}(\text{Filt}(\widehat{\mathbb{X}}_m))\right) > \varepsilon\right) \leq \mathbb{P}\left(d_{GH}(\mathbb{X}_\mu, \widehat{\mathbb{X}}_m) > \frac{\varepsilon}{2}\right)$$

Deviation inequality

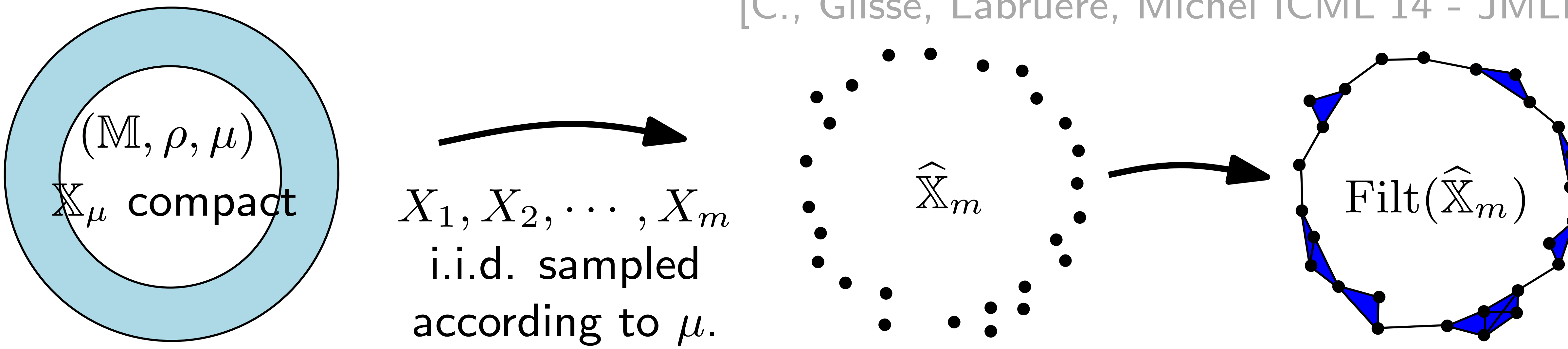
[C., Glisse, Labruère, Michel ICML'14 - JMLR'15]



For $a, b > 0$, μ satisfies the (a, b)-standard assumption if for any $x \in \mathbb{X}_\mu$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

Deviation inequality

[C., Glisse, Labruère, Michel ICML'14 - JMLR'15]



For $a, b > 0$, μ satisfies the (a, b)-standard assumption if for any $x \in \mathbb{X}_\mu$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

Theorem: If μ satisfies the (a, b)-standard assumption, then for any $\varepsilon > 0$:

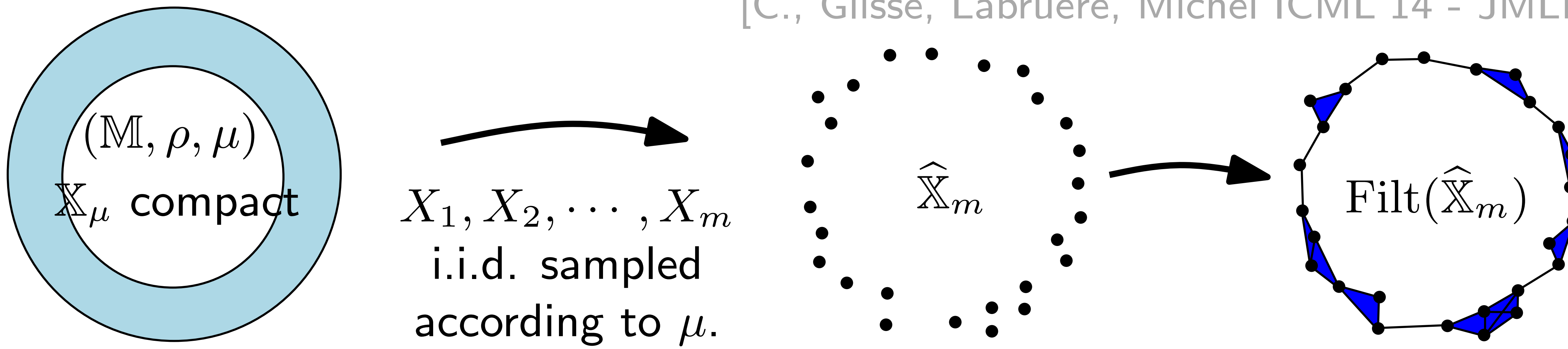
$$\mathbb{P}\left(\mathrm{d_b}\left(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathsf{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_m))\right) > \varepsilon\right) \leq \min(\frac{8^b}{a\varepsilon^b}\exp(-ma\varepsilon^b), 1).$$

Moreover $$\lim_{n\to\infty} \mathbb{P}\left(\mathrm{d_b}\left(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathsf{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_m))\right) \leq C_1 \left(\frac{\log m}{m}\right)^{1/b}\right) = 1.$$

where C_1 is a constant only depending on a and b.

Deviation inequality

[C., Glisse, Labruère, Michel ICML'14 - JMLR'15]



For $a, b > 0$, μ satisfies the (a,b)-standard assumption if for any $x \in \mathbb{X}_\mu$ and any $r > 0$, we have $\mu(B(x,r)) \geq \min(ar^b, 1)$.

Sketch of proof:

1. Upperbound $\mathbb{P}\left(d_H(\mathbb{X}_\mu, \widehat{\mathbb{X}}_m) > \frac{\varepsilon}{2}\right)$.
2. (a,b) standard assumption $\Rightarrow$ an explicit upperbound for the covering number of $\mathbb{X}_\mu$ (by balls of radius $\varepsilon/2$).
3. Apply "union bound" argument.

$C(\varepsilon) \leq P(\varepsilon/2)$

$+ \ \mu(B(x, \varepsilon/2)) \geq a(\varepsilon/2)^b$

Minimax rate of convergence

[C., Glisse, Labruère, Michel ICML'14 - JMLR'15]

Let $\mathcal{P}(a, b, \mathbb{M})$ be the set of all the probability measures on the metric space $(\mathbb{M}, \rho)$ satisfying the (a, b)-standard assumption on $\mathbb{M}$:

Minimax rate of convergence

[C., Glisse, Labruère, Michel ICML'14 - JMLR'15]

Let $\mathcal{P}(a,b,\mathbb{M})$ be the set of all the probability measures on the metric space $(\mathbb{M},\rho)$ satisfying the (a,b)-standard assumption on $\mathbb{M}$:

Theorem: Let $\mathcal{P}(a,b,\mathbb{M})$ be the set of (a,b)-standard proba measures on $\mathbb{M}$. Then:

$$\sup_{\mu \in \mathcal{P}(a,b,\mathbb{M})} \mathbb{E}\left[\mathrm{d_b}(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathsf{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_m)))\right] \le C\left(\frac{\ln m}{m}\right)^{1/b}$$

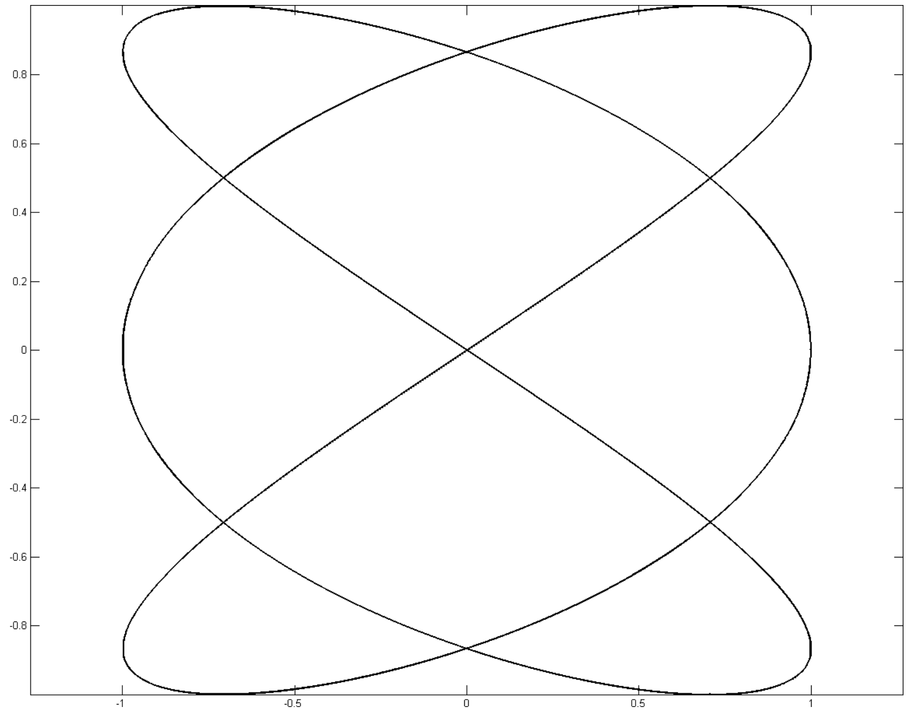
where the constant C only depends on a and b (not on $\mathbb{M}$!). Assume moreover that there exists a non isolated point x in $\mathbb{M}$ and let x_m be a sequence in $\mathbb{M}\setminus\{x\}$ such that $\rho(x,x_m) \le (am)^{-1/b}$. Then for any estimator $\widehat{\mathsf{dgm}}_m$ of $\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu))$:

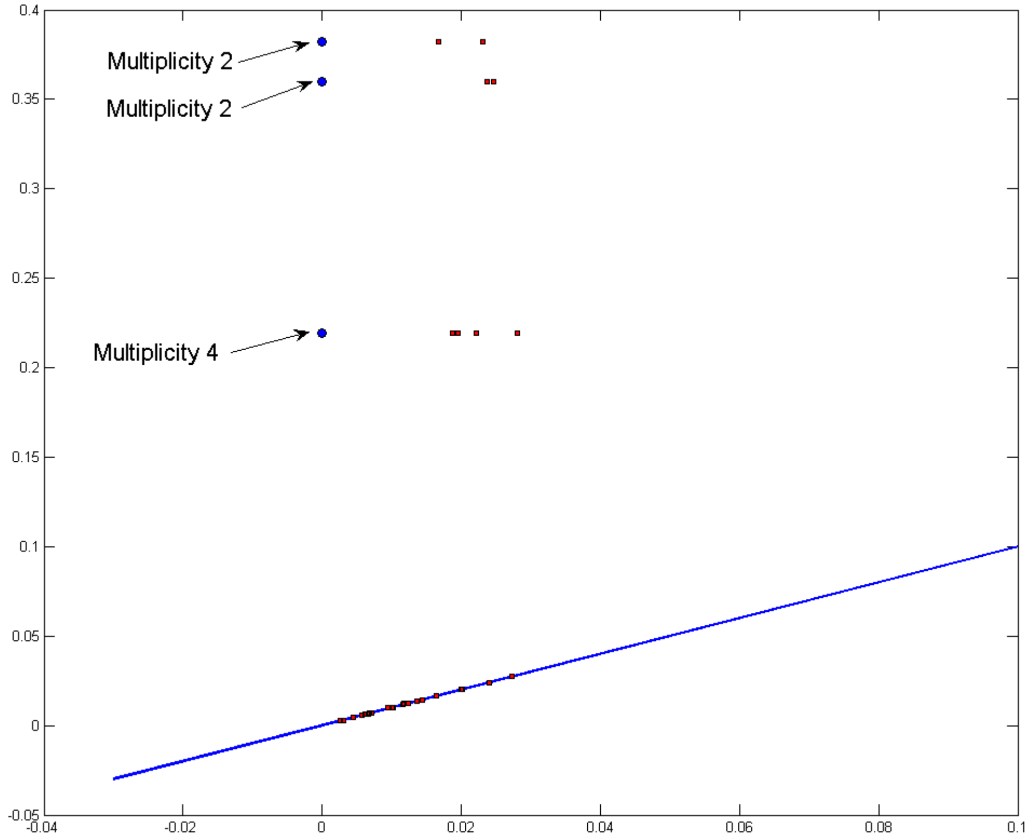
$$\liminf_{m\to\infty} \rho(x,x_m)^{-1} \sup_{\mu \in \mathcal{P}(a,b,\mathbb{M})} \mathbb{E}\left[\mathrm{d_b}(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \widehat{\mathsf{dgm}}_m)\right] \ge C'$$

where C' is an absolute constant.

Remark: we can obtain slightly better bounds if $\mathbb{X}_\mu$ is a submanifold of $\mathbb{R}^D$ - see [Genovese, Perone-Pacifico,Verdinelli, Wasserman 2011, 2012]

Numerical illustrations

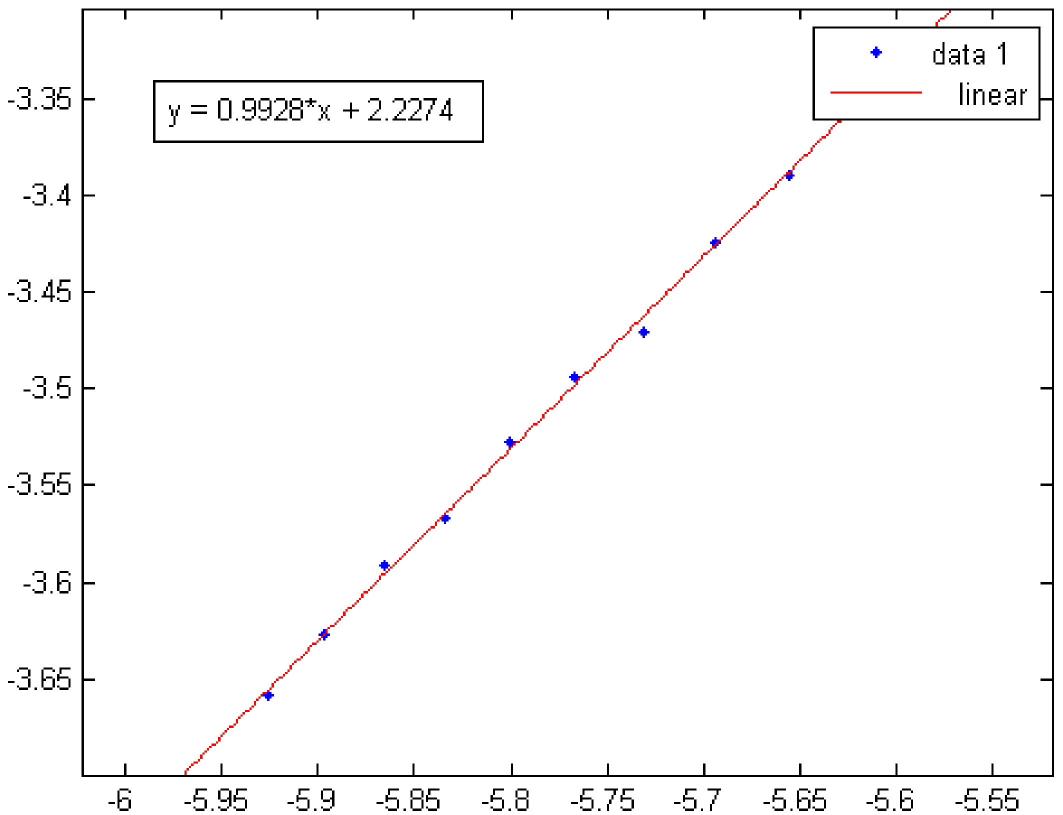




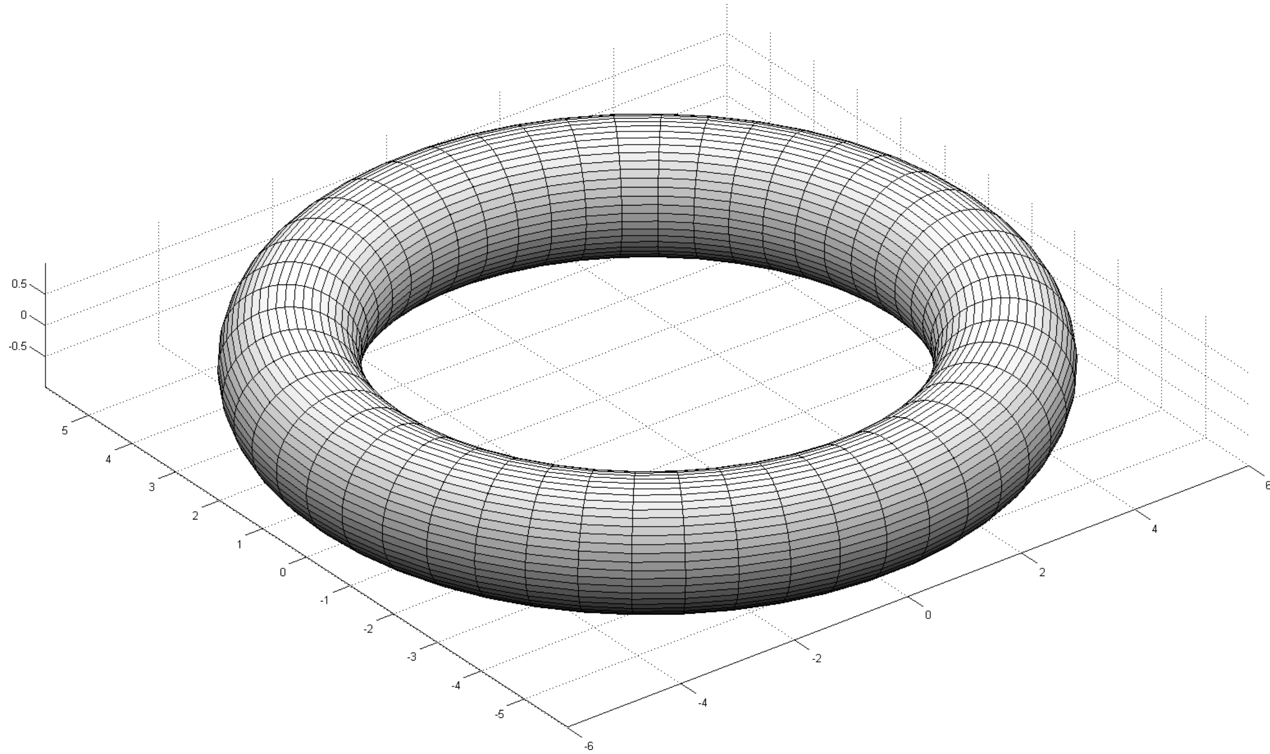
- μ: unif. measure on Lissajous curve $\mathbb{X}_\mu$.
- Filt: distance to $\mathbb{X}_\mu$ in $\mathbb{R}^2$.
- sample $k = 300$ sets of m points for $m = [2100 : 100 : 3000]$.
- compute

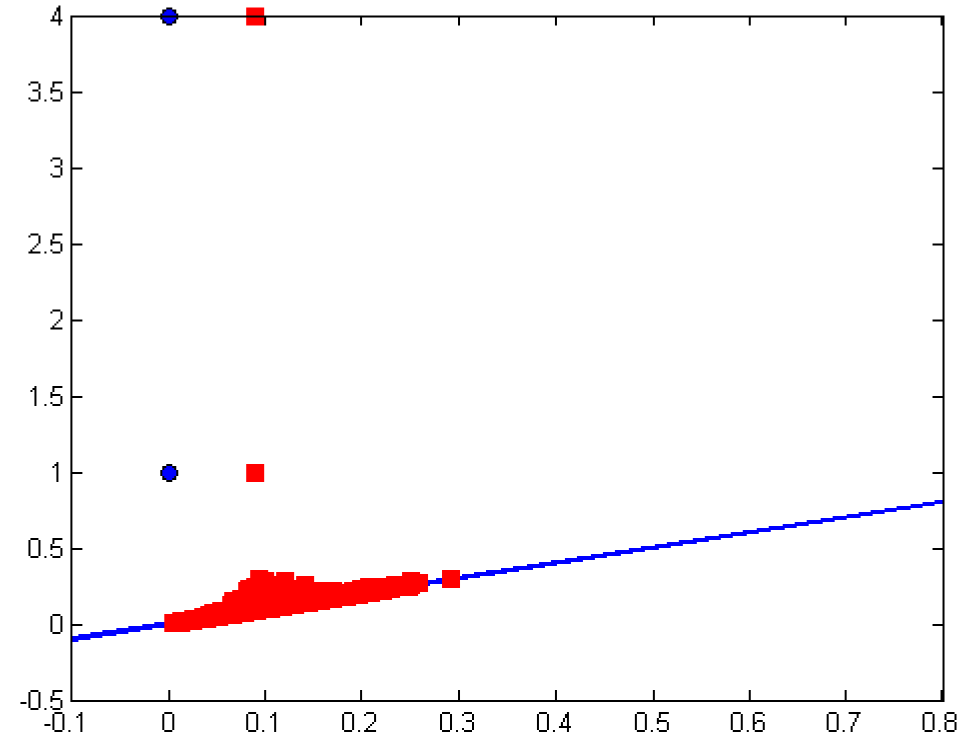
$$\widehat{\mathbb{E}}_m = \widehat{\mathbb{E}}[d_B(\mathrm{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathrm{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}_n})))].$$

- plot $\log(\widehat{\mathbb{E}}_m)$ as a function of $\log(\log(m)/m)$.



Numerical illustrations

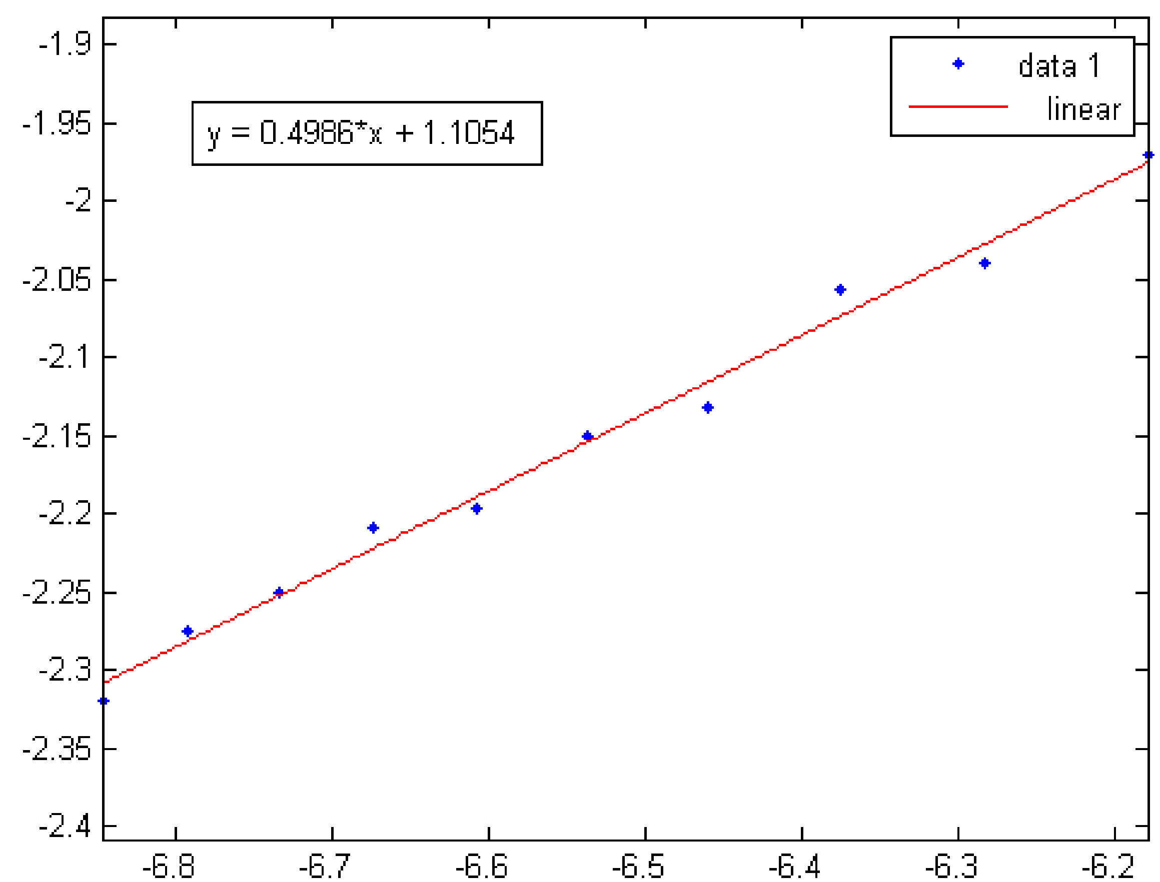




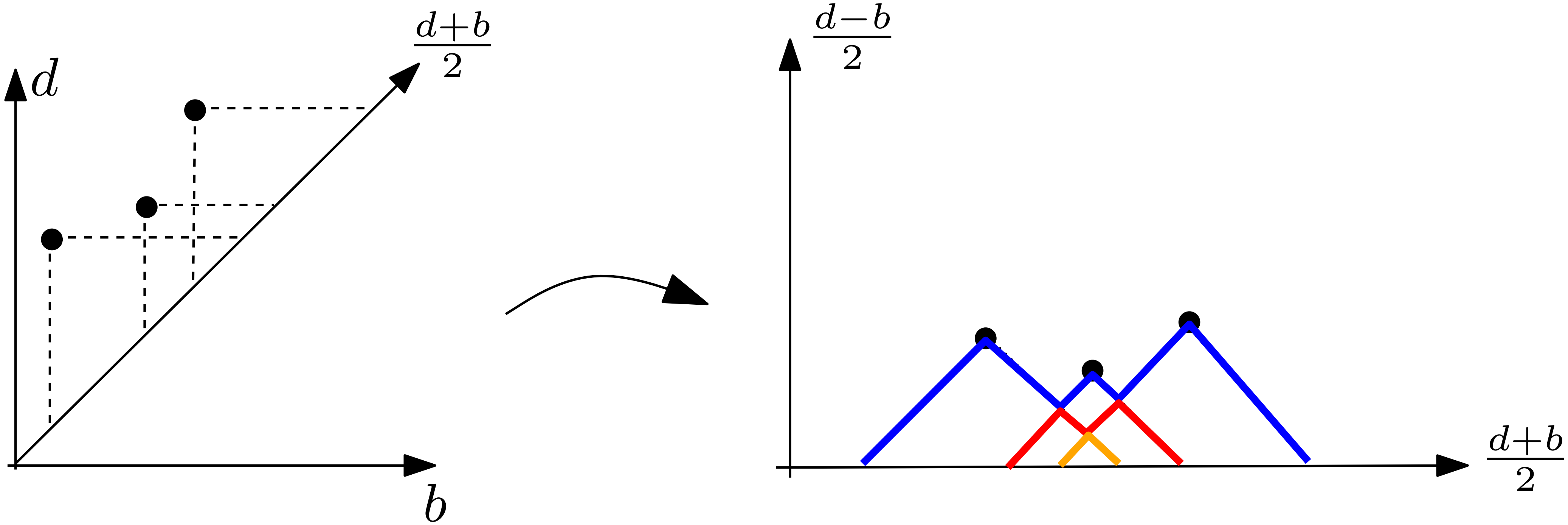
- μ: unif. measure on a torus $\mathbb{X}_\mu$.
- Filt: distance to $\mathbb{X}_\mu$ in $\mathbb{R}^3$.
- sample $k = 300$ sets of n points for $m = [12000 : 1000 : 21000]$.
- compute

$$\widehat{\mathbb{E}}_m = \widehat{\mathbb{E}}[d_B(\mathrm{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathrm{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}_m})))].$$

- plot $\log(\widehat{\mathbb{E}}_m)$ as a function of $\log(\log(m)/m)$.



Persistence landscapes



$D = \{(\frac{d_i+b_i}{2}, \frac{d_i+b_i}{2})\}i \in I$

For $p = (\frac{b+d}{2}, \frac{d-b}{2}) \in D$,

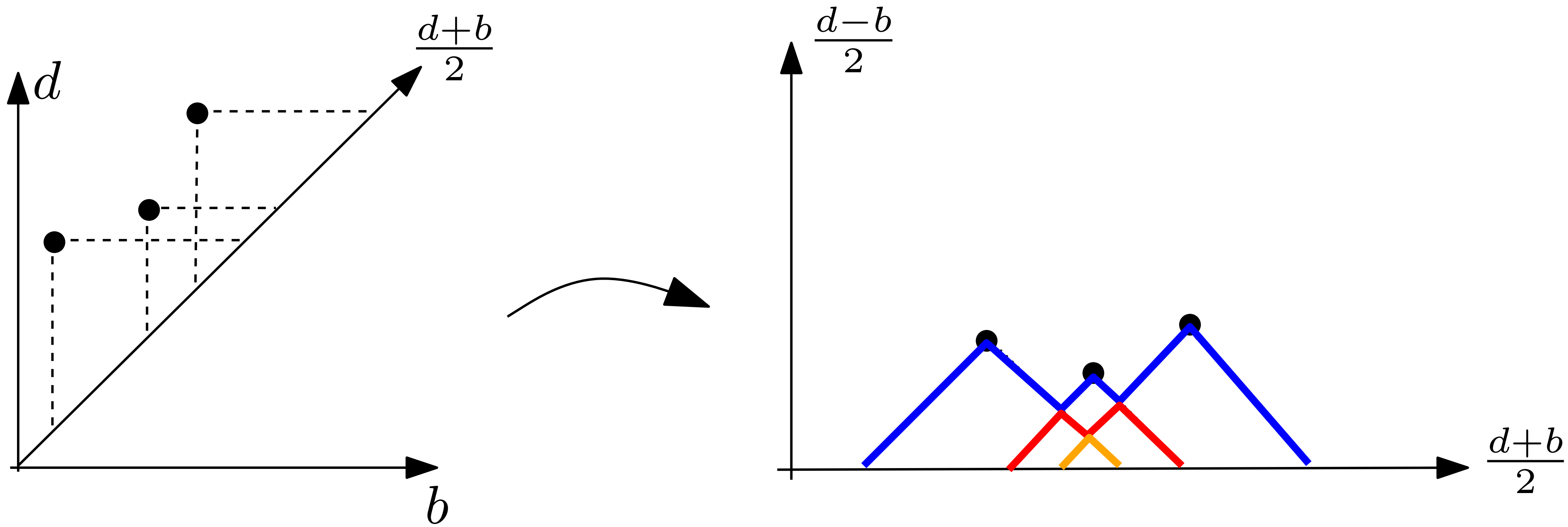
$$\Lambda_p(t) \;=\; \begin{cases} t-b & t \in [b, \frac{b+d}{2}] \\ d-t & t \in (\frac{b+d}{2}, d] \\ 0 & \text{otherwise.} \end{cases}$$

Persistence landscape [Bubenik 2012]:

$$\lambda_D(k,t) = \underset{p \in \mathsf{dgm}}{\mathsf{kmax}}\, \Lambda_p(t), \quad t \in \mathbb{R}, k \in \mathbb{N},$$

where kmax is the kth largest value in the set.

Persistence landscapes



Persistence landscape [Bubenik 2012]:

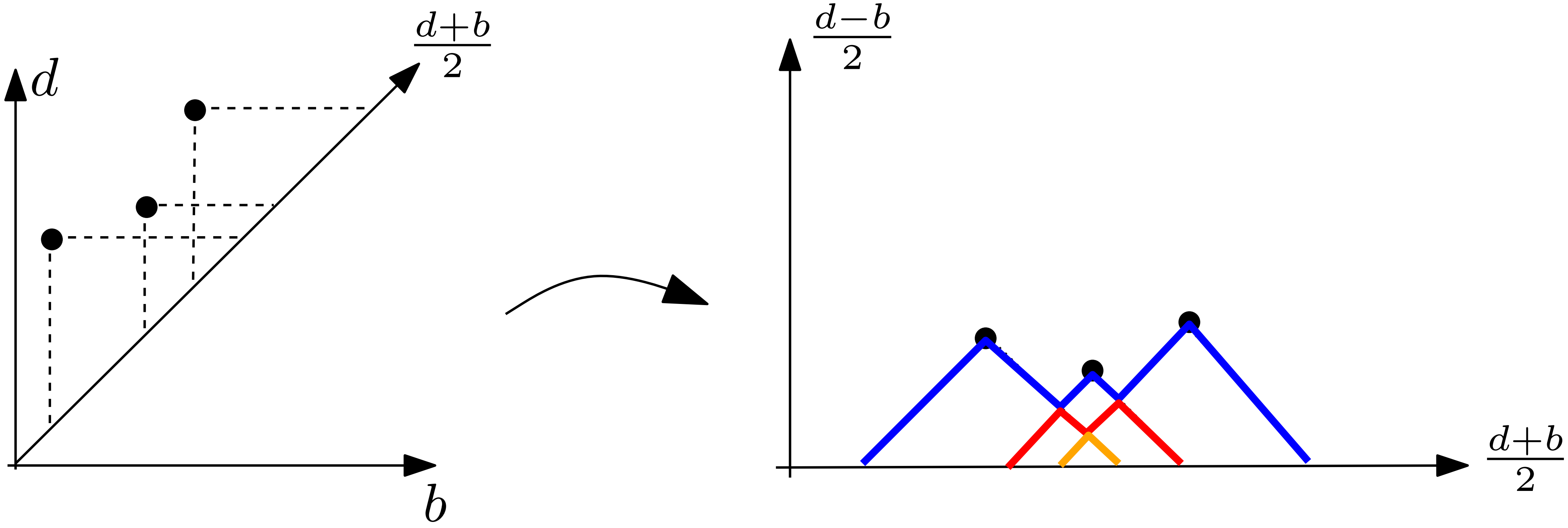
$$\lambda_D(k,t) = \underset{p \in \mathrm{dgm}}{\mathrm{kmax}}\ \Lambda_p(t), \quad t \in \mathbb{R}, k \in \mathbb{N},$$

Properties

- For any $t \in \mathbb{R}$ and any $k \in \mathbb{N}$, $0 \leq \lambda_D(k,t) \leq \lambda_D(k+1,t)$.
- For any $t \in \mathbb{R}$ and any $k \in \mathbb{N}$, $|\lambda_D(k,t) - \lambda_{D'}(k,t)| \leq d_B(D,D')$ where $d_B(D,D')$ denotes the bottleneck distance between D and D'.

stability properties of persistence landscapes

Persistence landscapes



- Persistence encoded as an element of a functional space (vector space!).
- Expectation of distribution of landscapes is well-defined and can be approximated from average of sampled landscapes.
- process point of view: convergence results and convergence rates $\rightarrow$ confidence intervals can be computed using bootstrap.

[C., Fasy, Lecci, Rinaldo, Wasserman SoCG 2014]

Weak convergence of landscapes

Let $\mathcal{L}_T$ be the space of landscapes with support contained in $[0, T]$.

Let P be a probability distribution on $\mathcal{L}_T$, and let $\lambda_1, \ldots, \lambda_n \sim P$. Let μ be the mean landscape:

$$\mu(t) = \mathbb{E}[\lambda_i(t)], \quad t \in [0, T].$$

We estimate μ with the sample average

$$\overline{\lambda}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i(t), \quad t \in [0, T].$$

Since $\mathbb{E}(\overline{\lambda}_n(t)) = \mu(t)$, $\overline{\lambda}_n$ is a point-wise unbiased estimator of μ.

For fixed t: pointwise convergence of $\lambda_n(t)$ to $\mu(t)$ + CLT

Here, convergence of the process

$$\left\{ \sqrt{n} \left(\overline{\lambda}_n(t) - \mu(t) \right) \right\}_{t \in [0,T]}$$

Weak convergence of landscapes

Let

$$\mathcal{F} = \{f_t\}_{0 \leq t \leq T}$$

where $f_t : \mathcal{L}_T \to \mathbb{R}$ is defined by $f_t(\lambda) = \lambda(t)$.

Empirical process indexed by $f_t \in \mathcal{F}$:

$$\mathbb{G}_n(t) = \mathbb{G}_n(f_t) := \sqrt{n}\left(\overline{\lambda}_n(t) - \mu(t)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f_t(\lambda_i) - \mu(t)) = \sqrt{n}(P_n - P)(f_t)$$

Theorem [Weak convergence of landscapes]. Let $\mathbb{G}$ be a Brownian bridge with covariance function $\kappa(t,s) = \int f_t(\lambda) f_s(\lambda) dP(\lambda) - \int f_t(\lambda) dP(\lambda) \int f_s(\lambda) dP(\lambda)$, for $t, s \in [0, T]$. Then $\mathbb{G}_n \rightsquigarrow \mathbb{G}$.

Weak convergence of landscapes

Let

$$\mathcal{F} = \{f_t\}_{0 \le t \le T}$$

where $f_t : \mathcal{L}_T \to \mathbb{R}$ is defined by $f_t(\lambda) = \lambda(t)$.

Empirical process indexed by $f_t \in \mathcal{F}$:

$$\mathbb{G}_n(t) = \mathbb{G}_n(f_t) := \sqrt{n}\left(\overline{\lambda}_n(t) - \mu(t)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f_t(\lambda_i) - \mu(t)) = \sqrt{n}(P_n - P)(f_t)$$

For $t \in [0, T]$, let $\sigma(t)$ be the standard deviation of $\sqrt{n}\,\overline{\lambda}_n(t)$, i.e. $\sigma(t) = \sqrt{n\mathsf{Var}(\overline{\lambda}_n(t))} = \sqrt{\mathsf{Var}(f_t(\lambda_1))}$.

Theorem [Uniform CLT]. Suppose that $\sigma(t) > c > 0$ in an interval $[t_*\,, t^*] \subset [0, T]$, for some constant c. Then there exists a random variable $W \stackrel{d}{=} \sup_{t \in [t_*\,, t^*]} |\mathbb{G}(f_t)|$ such that

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}\left(\sup_{t \in [t_*\,, t^*]} |\mathbb{G}_n(t)| \le z \right) - \mathbb{P}\left(W \le z\right) \right| = O\left(\frac{(\log n)^{7/8}}{n^{1/8}} \right).$$

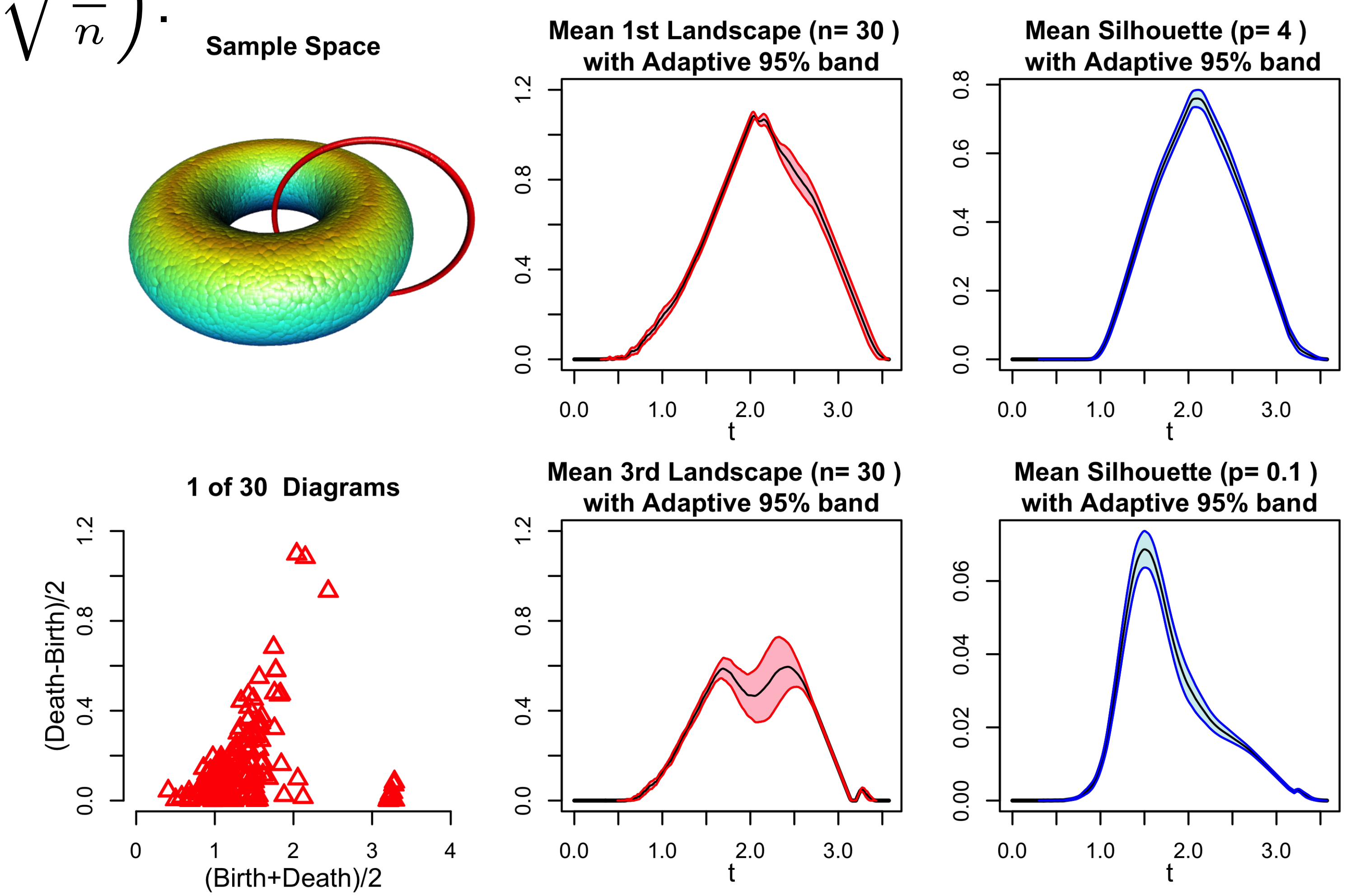
Some consequences

Bootstrap for landscapes $\rightarrow$ confidence bands for landscapes.

Theorem. Suppose that $\sigma(t) > c > 0$ in an interval $[t_*, t^*] \subset [0, T]$, for some constant c. Then, given a confidence level $1 - \alpha$, one can construct confidence functions $\ell_n(t)$ and $u_n(t)$ such that

$$\mathbb{P}\Big(\ell_n(t) \leq \mu(t) \leq u_n(t) \text{ for all } t \in [t_*, t^*]\Big) \geq 1 - \alpha - O\left(\frac{(\log n)^{7/8}}{n^{1/8}}\right).$$

Also, $\sup_t (u_n(t) - \ell_n(t)) = O_P\left(\sqrt{\frac{1}{n}}\right)$.



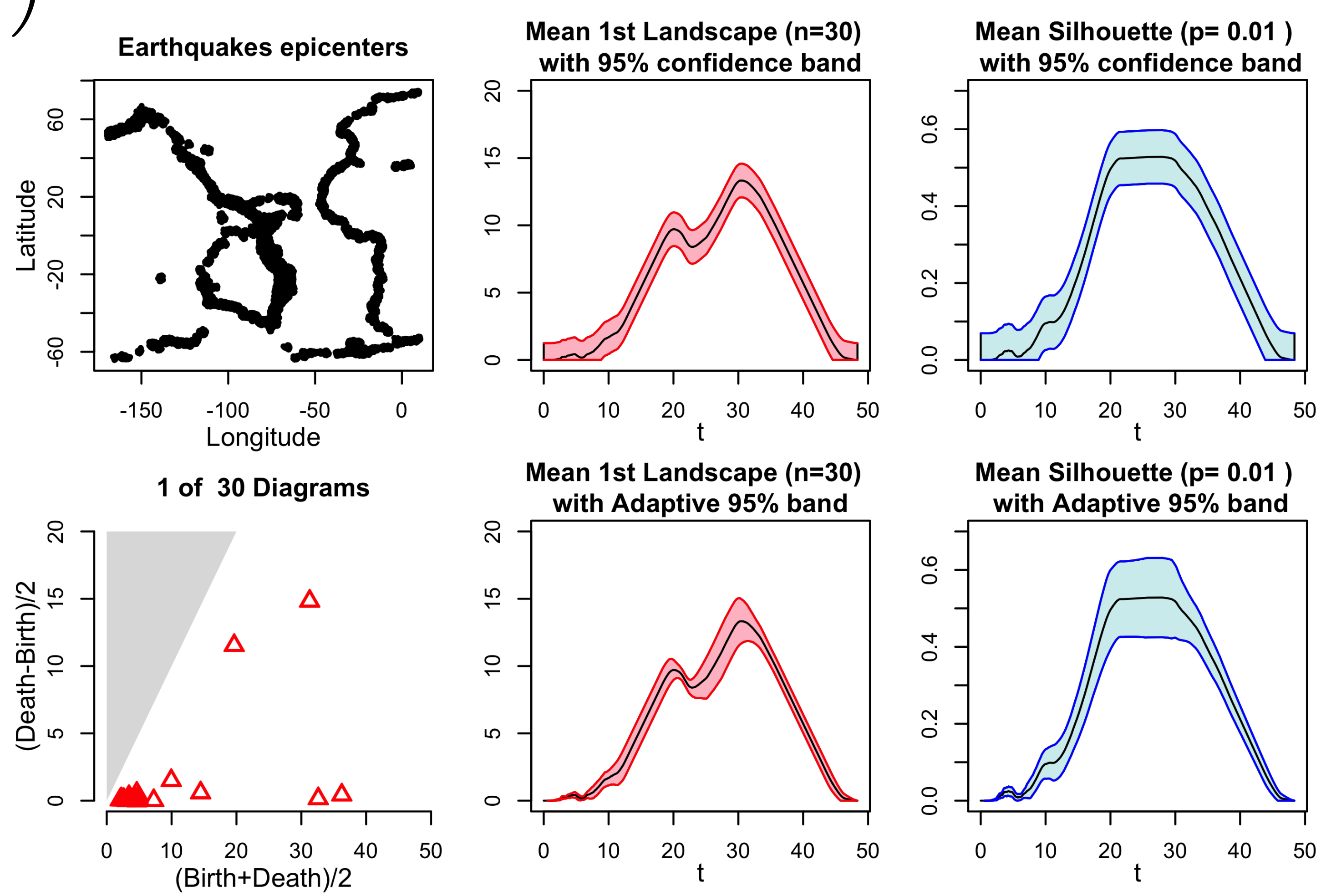
Some consequences

Bootstrap for landscapes $\to$ confidence bands for landscapes.

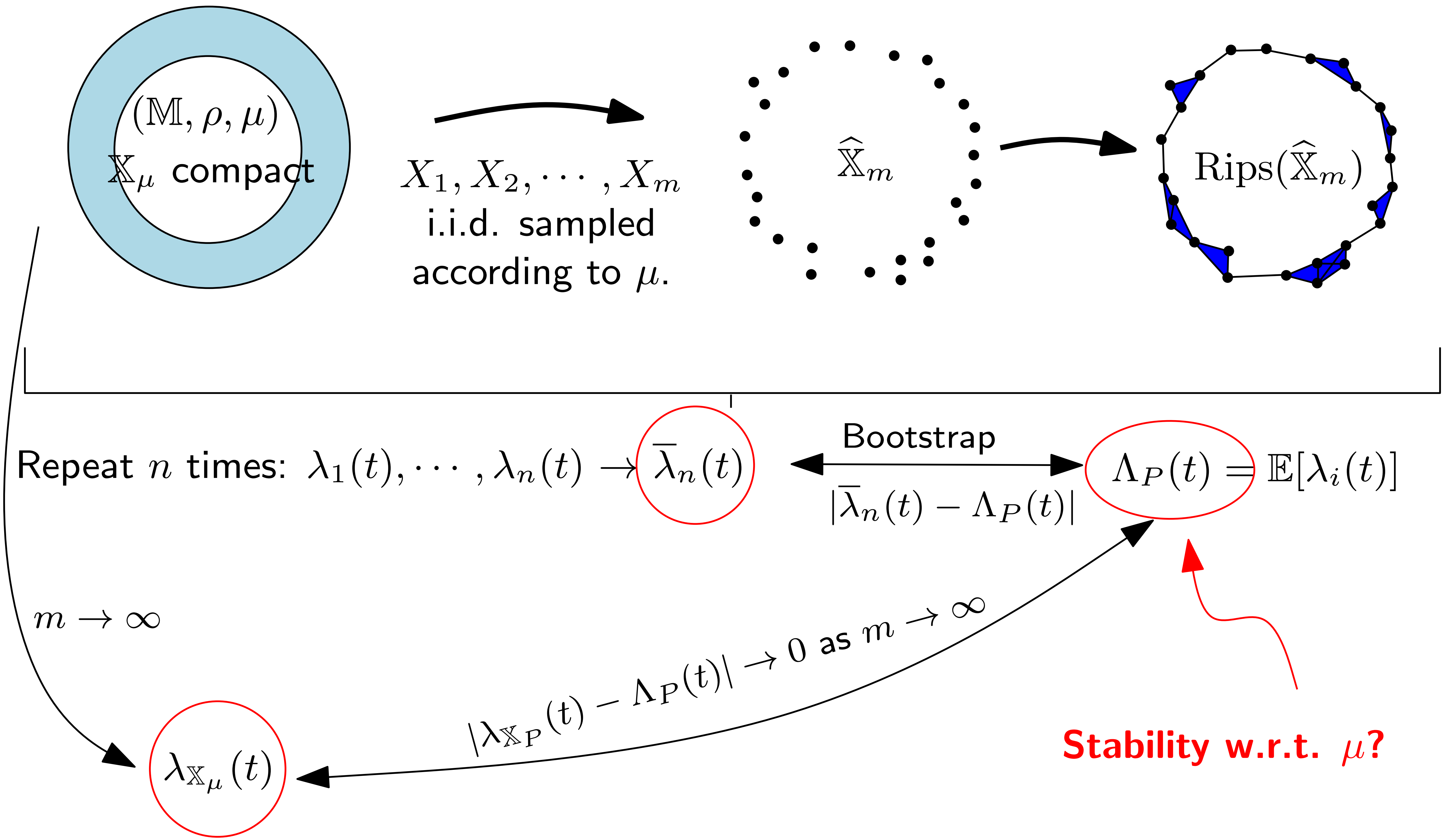
Theorem. Suppose that $\sigma(t) > c > 0$ in an interval $[t_*, t^*] \subset [0, T]$, for some constant c. Then, given a confidence level $1-\alpha$, one can construct confidence functions $\ell_n(t)$ and $u_n(t)$ such that

$$\mathbb{P}\Big(\ell_n(t) \le \mu(t) \le u_n(t) \text{ for all } t \in [t_*, t^*]\Big) \ge 1 - \alpha - O\left(\frac{(\log n)^{7/8}}{n^{1/8}}\right).$$

Also, $\sup_t (u_n(t) - \ell_n(t)) = O_P\left(\sqrt{\frac{1}{n}}\right)$.



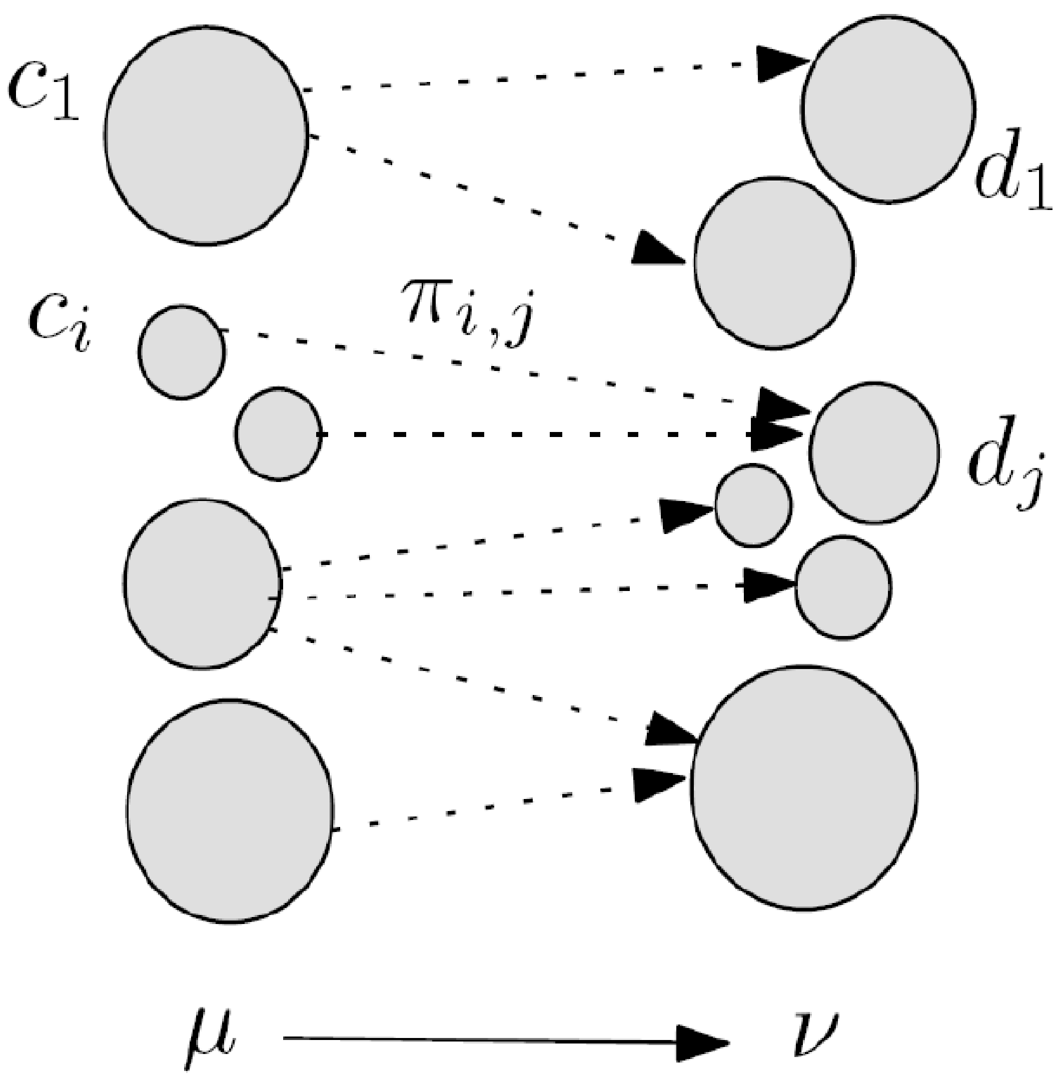
To summarize



Wasserstein distance

Let $(\mathbb{M}, \rho)$ be a metric space and let μ, ν be probability measures on $\mathbb{M}$ with finite p-moments ($p \geq 1$).

"The" Wasserstein distance $W_p(\mu, \nu)$ quantifies the optimal cost of pushing μ onto ν, the cost of moving a small mass dx from x to y being $\rho(x, y)^p dx$.

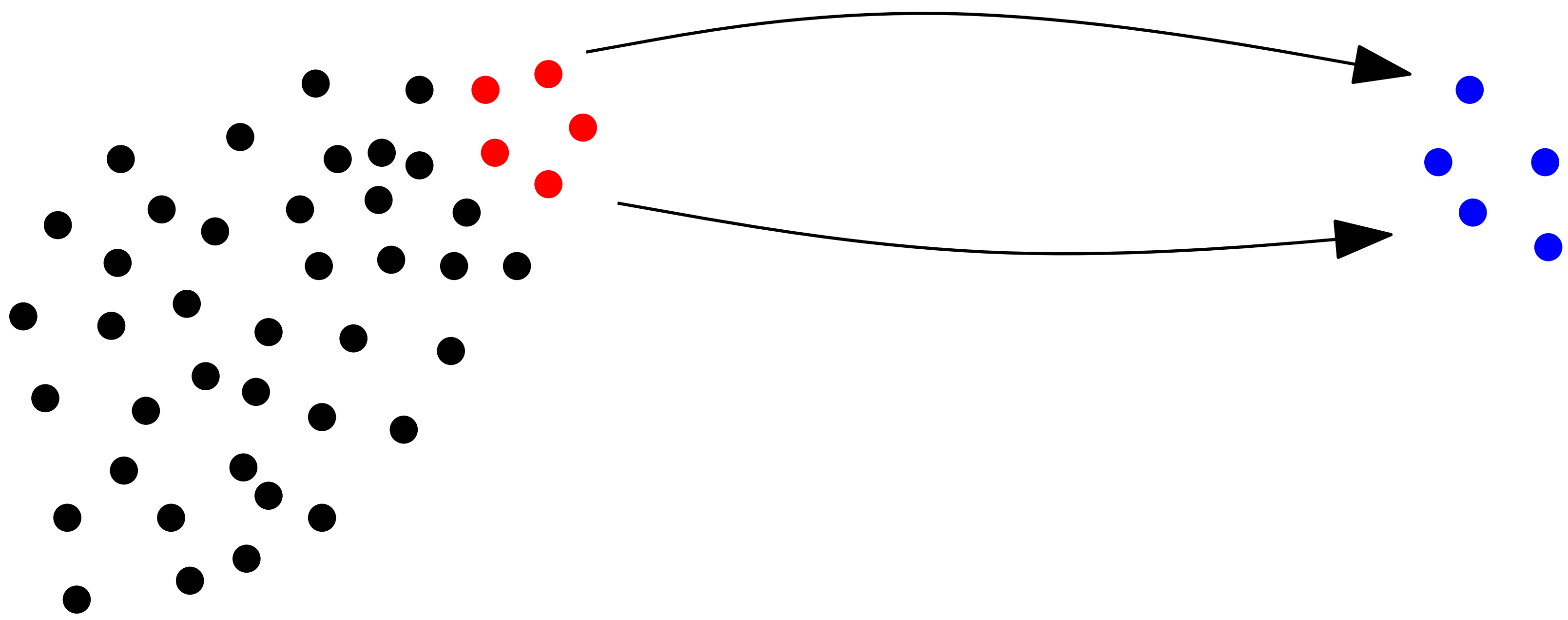


- Transport plan: Π a proba measure on $M \times M$ such that $\Pi(A \times \mathbb{R}^d) = \mu(A)$ and $\Pi(\mathbb{R}^d \times B) = \nu(B)$ for any borelian sets $A, B \subset M$.
- Cost of a transport plan:

$$C(\Pi) = \left(\int_{M \times M} \rho(x, y)^p d\Pi(x, y) \right)^{\frac{1}{p}}$$

- $W_p(\mu, \nu) = \inf_\Pi C(\Pi)$

Wasserstein distance



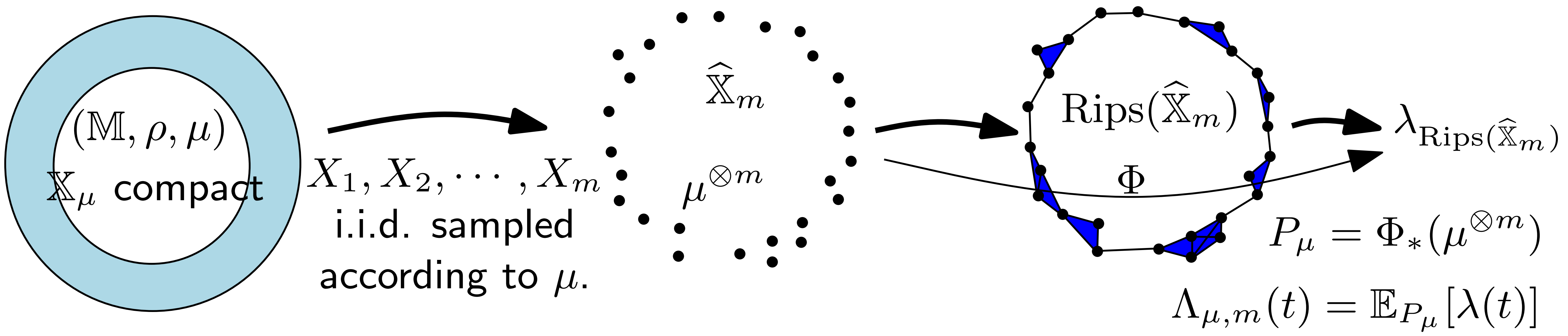
Example:

- If $P = \{p_1, \ldots, p_n\}$ is a point cloud, and $P' = \{p_1, \ldots, p_{n-k-1}, o_1, \ldots, o_k\}$ with $d(o_i, P) = R$, then

$$d_H(C, C') \geq R \qquad \text{but} \qquad W_2(\mu_C, \mu_{C'}) \leq \sqrt{\frac{k}{n}}(R + \mathrm{diam}(C))$$

(Sub)sampling and stability of expected landscapes

[C., Fasy, Lecci, Michel, Rinaldo, Wasserman ICML 2015]



Theorem: Let $(\mathbb{M}, \rho)$ be a metric space and let μ, ν be proba measures on $\mathbb{M}$ with compact supports. We have

$$\boxed{\|\Lambda_{\mu,m} - \Lambda_{\nu,m}\|_\infty \leq m^{\frac{1}{p}} W_p(\mu, \nu)}$$

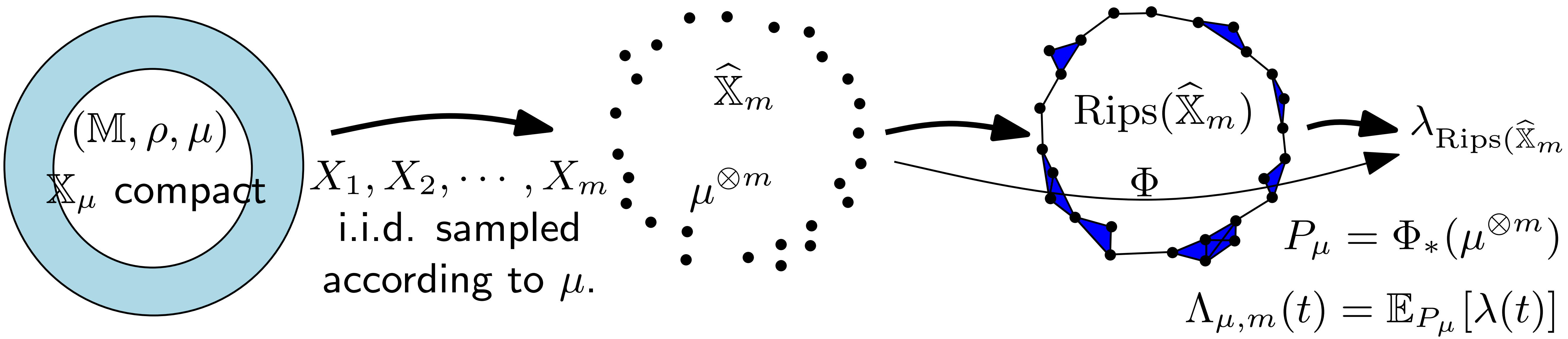
where W_p denotes the Wasserstein distance with cost function $\rho(x,y)^p$.

Remarks:

- similar results by Blumberg et al (2014) in the (Gromov-)Prokhorov metric (for distributions, not for expectations) ;
- also work with “Gromov-Wasserstein” metric;
- $m^{\frac{1}{p}}$ cannot be replaced by a constant.

(Sub)sampling and stability of expected landscapes

[C., Fasy, Lecci, Michel, Rinaldo, Wasserman ICML 2015]



Theorem: Let $(\mathbb{M}, \rho)$ be a metric space and let μ, ν be proba measures on $\mathbb{M}$ with compact supports. We have

$$\boxed{\|\Lambda_{\mu,m} - \Lambda_{\nu,m}\|_\infty \leq m^{\frac{1}{p}} W_p(\mu, \nu)}$$

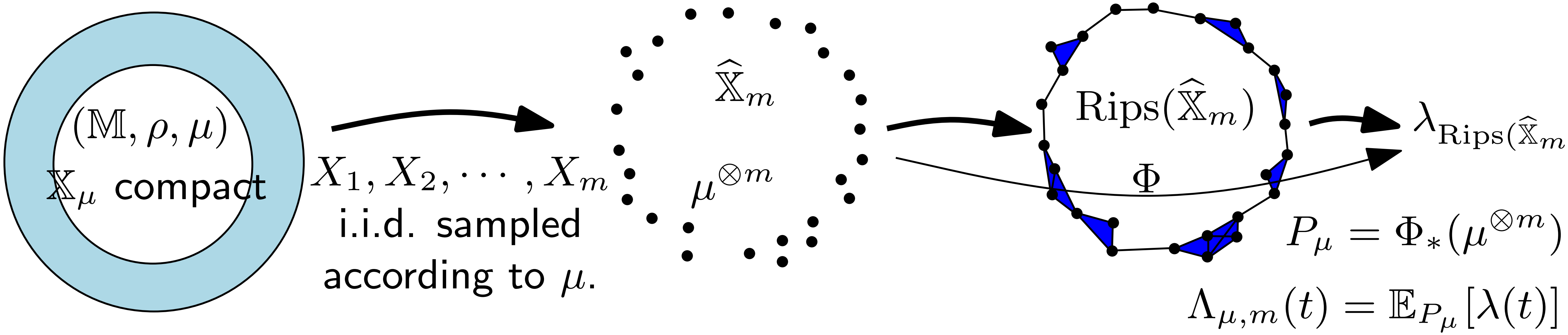
where W_p denotes the Wasserstein distance with cost function $\rho(x, y)^p$.

Consequences:

- Subsampling: efficient and easy to parallelize algorithm to infer topol. information from huge data sets.
- Robustness to outliers.
- R package TDA +Gudhi library: https://project.inria.fr/gudhi/software/

(Sub)sampling and stability of expected landscapes

[C., Fasy, Lecci, Michel, Rinaldo, Wasserman ICML 2015]



Theorem: Let $(\mathbb{M}, \rho)$ be a metric space and let μ, ν be proba measures on $\mathbb{M}$ with compact supports. We have

$$\boxed{\|\Lambda_{\mu,m} - \Lambda_{\nu,m}\|_\infty \leq m^{\frac{1}{p}} W_p(\mu, \nu)}$$

where W_p denotes the Wasserstein distance with cost function $\rho(x,y)^p$.

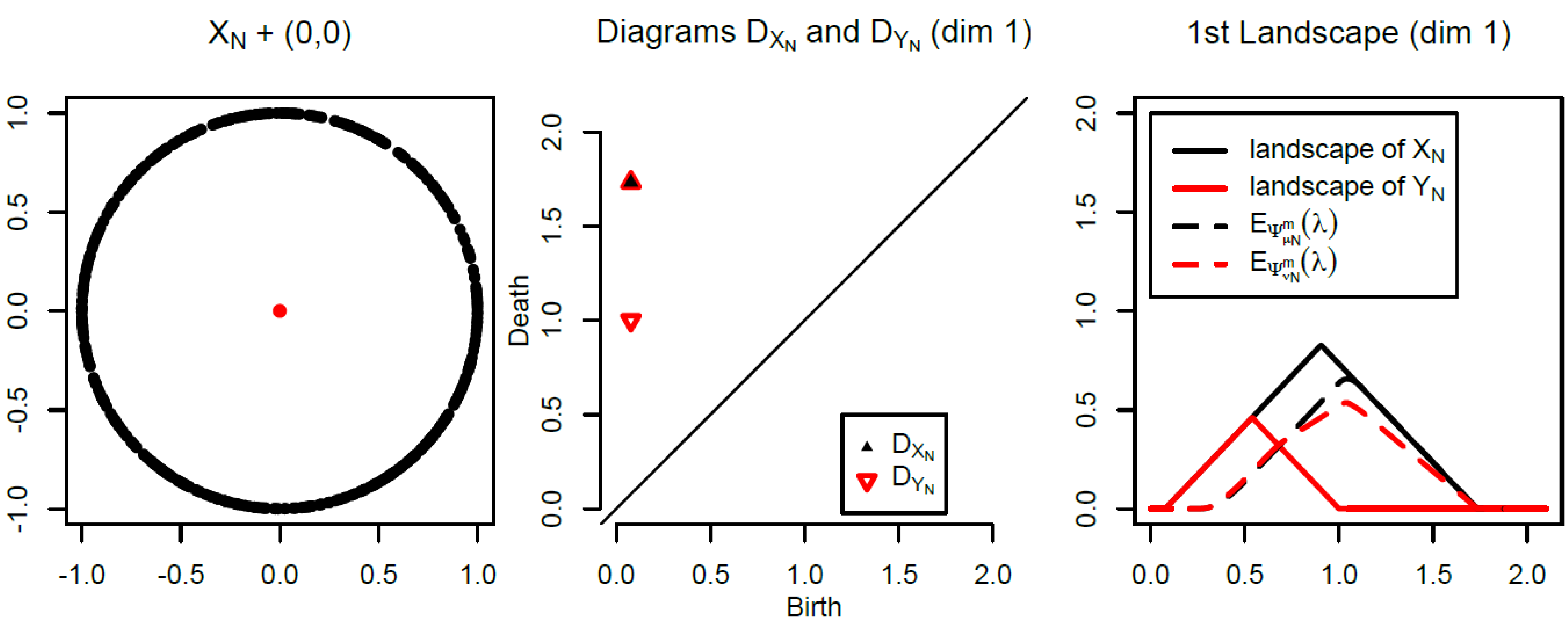
Proof:

1. $W_p(\mu^{\otimes m}, \nu^{\otimes m}) \leq m^{\frac{1}{p}} W_p(\mu, \nu)$
2. $W_p(P_\mu, P_\nu) \leq W_p(\mu^{\otimes m}, \nu^{\otimes m})$ (stability of persistence!)
3. $\|\Lambda_{\mu,m} - \Lambda_{\nu,m}\|_\infty \leq W_p(P_\mu, P_\nu)$ (Jensen's inequality)

(Sub)sampling and stability of expected landscapes

[C., Fasy, Lecci, Michel, Rinaldo, Wasserman ICML 2015]

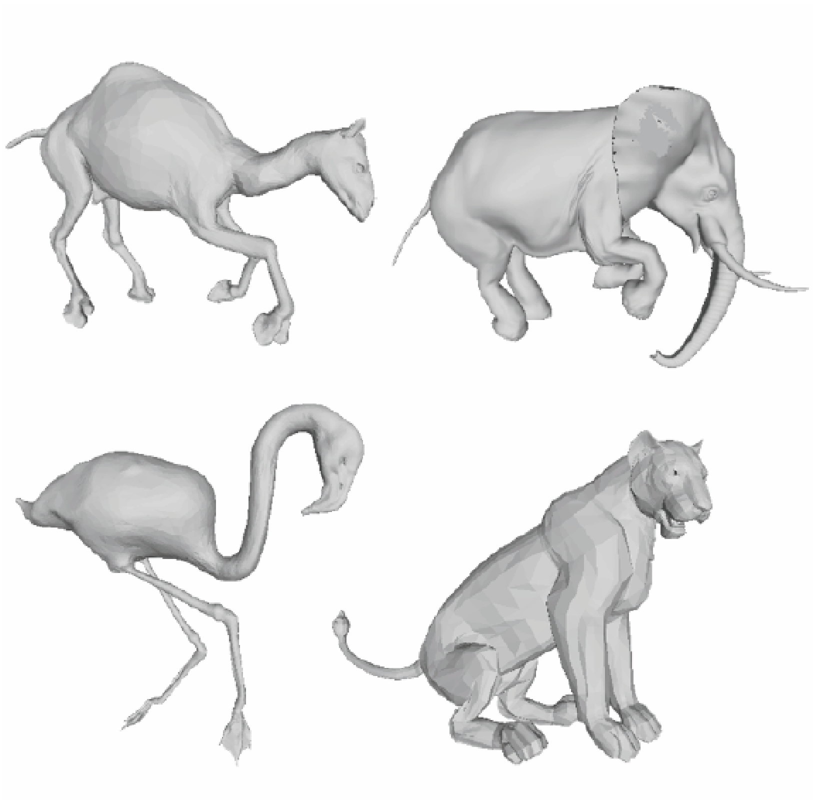
Example: Circle with one outlier.

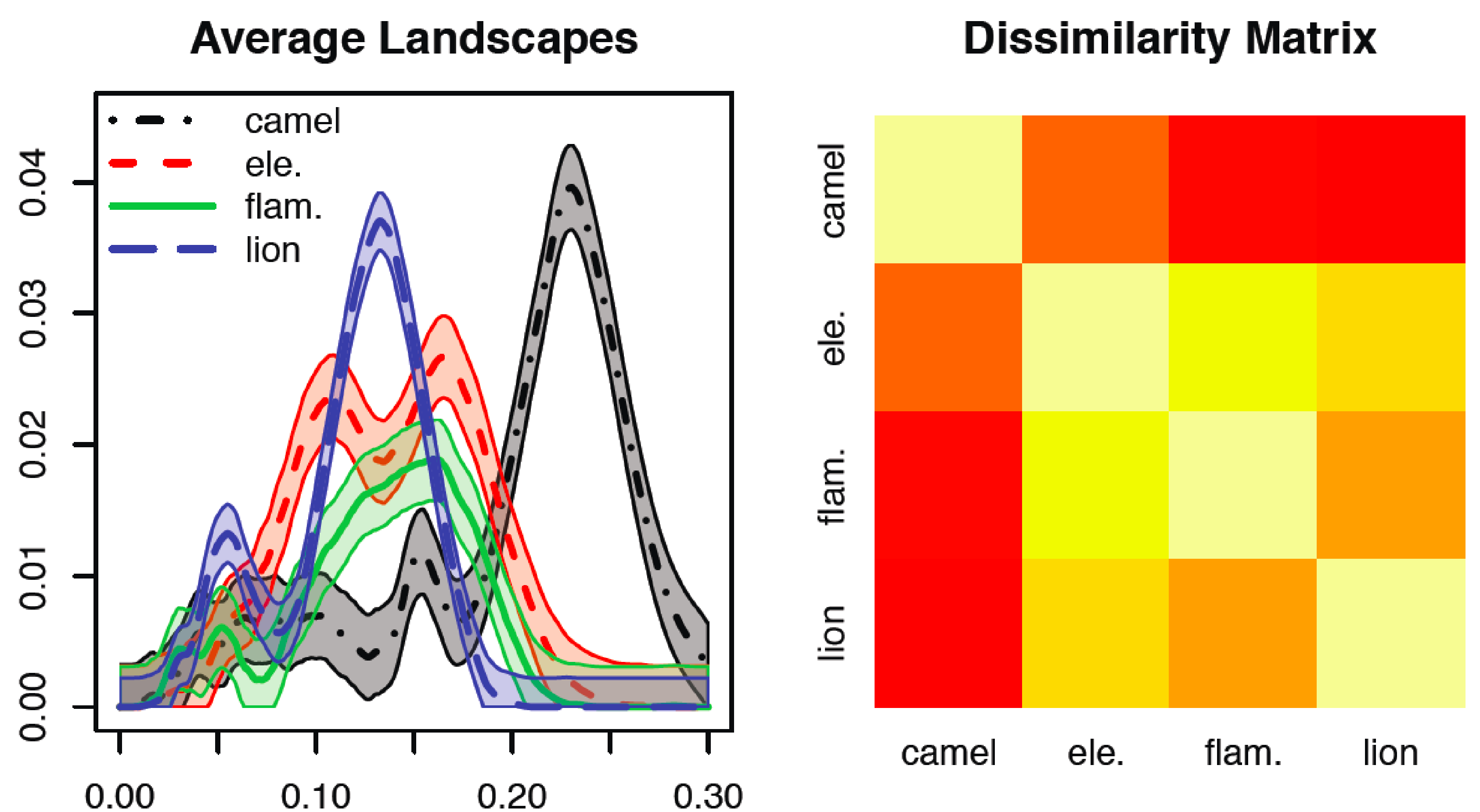


(Sub)sampling and stability of expected landscapes

[C., Fasy, Lecci, Michel, Rinaldo, Wasserman ICML 2015]

Example: 3D shapes



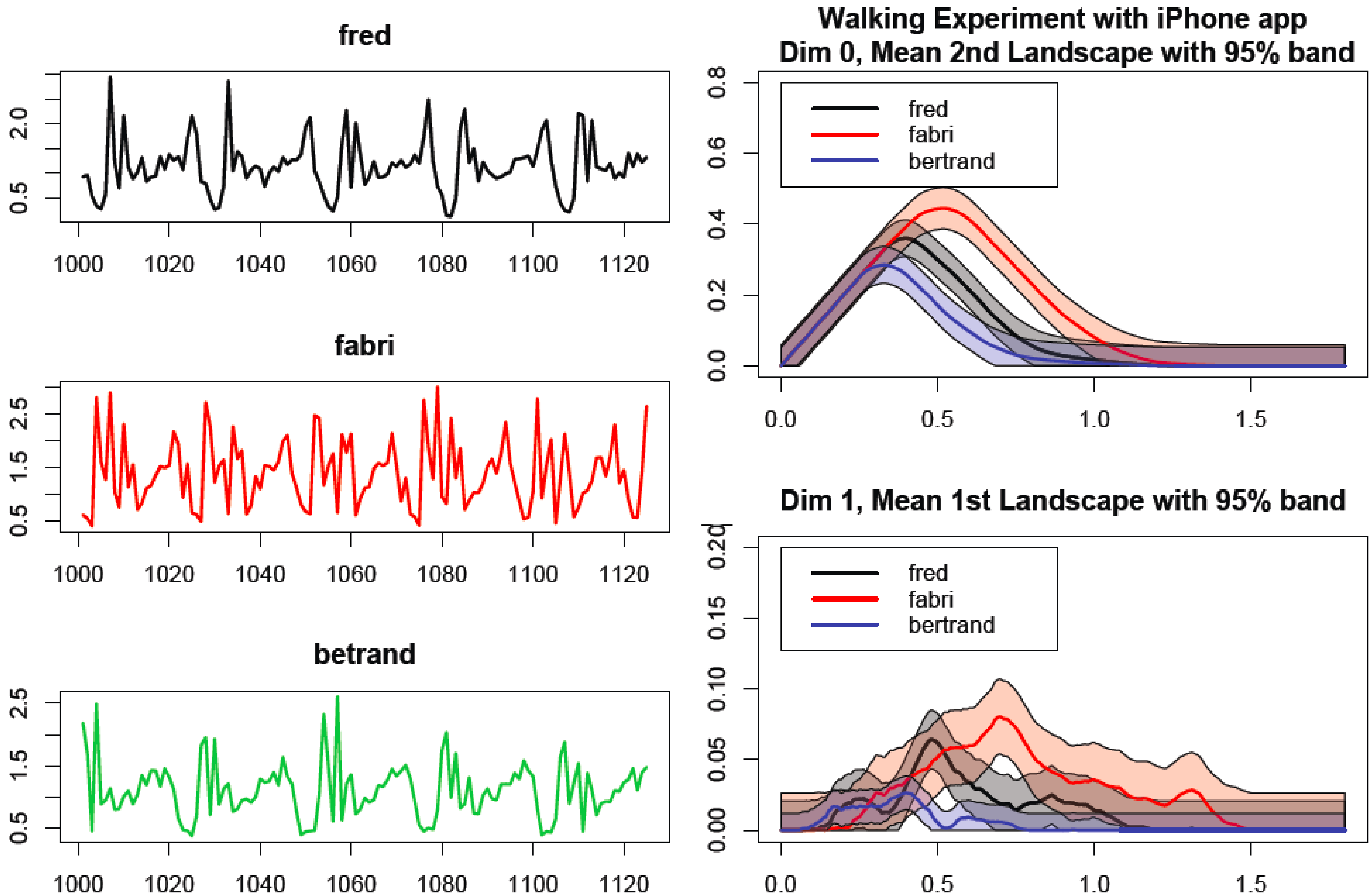


From $n = 100$ subsamples of size $m = 300$

(Sub)sampling and stability of expected landscapes

[C., Fasy, Lecci, Michel, Rinaldo, Wasserman ICML 2015]

(Toy) Example: Accelerometer data from smartphone.



- spatial time series (accelerometer data from the smarphone of users).
- no registration/calibration preprocessing step needed to compare!

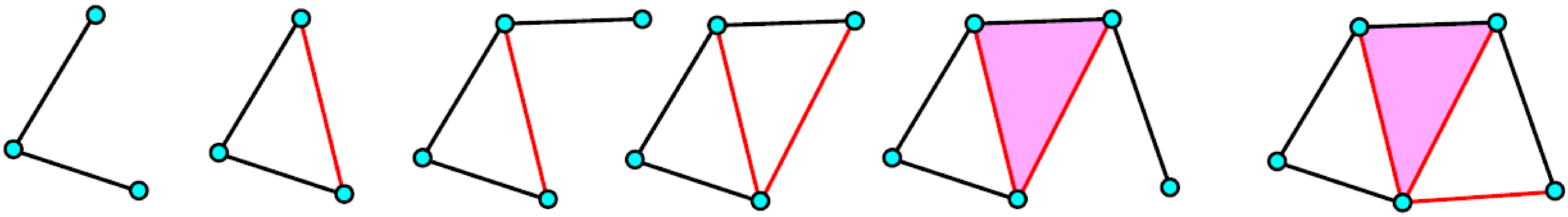
References

- F. Chazal, M. Glisse, C. Labruère, B. Michel, Convergence rates for persistence diagram estimation in TDA, in Journal of Machine Learn. Research 2015.
- C. Li, M. Ovsjanikov, F. Chazal, Persistence-based Structural Recognition, in proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2014).
- F. Chazal, B. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, Stochastic Convergence of Persistence Landscapes and Silhouettes, in ACM Symposium on Computational Geometry 2014 and Journal of Comp. Geom. 2015.
- F. Chazal, B. Fasy, F. Lecci, B. Michel, A. Rinaldo, L. Wasserman, Subsampling Methods for Persistent Homology, in Int. Conf. on Mach. Learning 2015 (ICML 2015).
- F. Chazal, V. de Silva, S. Oudot, Persistence Stability for Geometric complexes, Geometria Dedicata 2014.
- F. Chazal, V. de Silva, M. Glisse, S. Oudot, The Structure and Stability of Persistence Modules, arXiv:1207.3674, July 2012.
- P. Skraba, M. Ovsjanikov, F. Chazal, L. Guibas, Persistence-Based Segmentation of Deformable Shapes, Proc. Workshop on Nonrigid Shape Analysis and Deformable Image Alignment (NORDIA), Proc. CVPR 2010.

Software:

- The Gudhi library (C++): https://project.inria.fr/gudhi/software/
- R package TDA

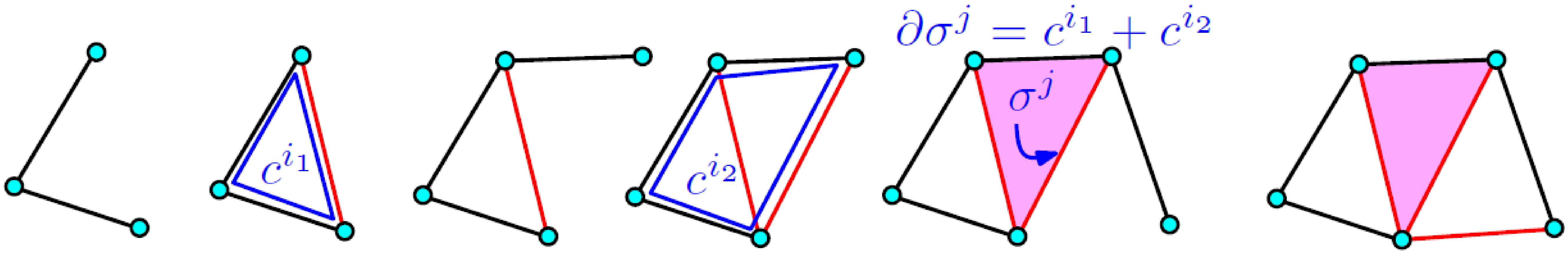
Cycle associated to a positive simplex



Lemma: If σ^i is a positive k-cycle, then there exists a k-cycle c_σ s.t.:
- c_σ is not a boundary in K^i,
- c_σ contains σ^i but no other positive k-simplex.
The cycle c^σ is unique.

Proof:
By induction on the order of appearence of the simplices in the filtration.

Homology basis



- At the beginning: the basis of H_k^0 is empty.
- If a basis of H_k^{i-1} has been built and σ^i is a positive k-simplex then one adds the homology class of the cycle c^i associated to σ^i to the basis of $H_k^{i-1} \Rightarrow$ basis of H_k^i.
- If a basis of H_k^{j-1} has been built and σ^j is a negative $(k+1)$-simplex:
  - let $c^{i_1}, \cdots, c^{i_p}$ be the cycles associated to the positive simplices $\sigma^{i_1}, \cdots, \sigma^{i_p}$ that form a basis of H_k^{j-1}
  - $d = \partial\sigma^j = \sum_{k=1}^{p} \varepsilon_k c^{i_k} + b$
  - $l(j) = \max\{i_k : \varepsilon_k = 1\}$
  - Remove the homology class of $c^{l(j)}$ from the basis of $H_k^{j-1} \Rightarrow$ basis of H_k^j.

Pairing simplices

If a basis of H_k^{j-1} has been built and σ^j is a negative $(k+1)$-simplex:

- let $c^{i_1}, \cdots, c^{i_p}$ be the cycles associated to the positive simplices $\sigma^{i_1}, \cdots, \sigma^{i_p}$ that form a basis of H_k^{j-1}
- $d = \partial \sigma^j = \sum_{k=1}^{p} \varepsilon_k c^{i_k} + b$
- $l(j) = \max\{i_k : \varepsilon_k = 1\}$
- Remove the homology class of $c^{l(j)}$ from the basis of $H_k^{j-1} \Rightarrow$ basis of H_k^j.

The simplices $\sigma^{l(j)}$ and σ^j are paired to form a persistent pair $(\sigma^{l(j)}, \sigma^j)$.
$\rightarrow$ The homology class created by $\sigma^{l(j)}$ in $K^{l(j)}$ is killed by σ^j in K^j. The persistence (or life-time) of this cycle is : $j - l(j) - 1$.

Remark: filtrations of K can be indexed by increasing sequences α_i of real numbers (useful when working with a function defined on the vertices of a simplicial complex).

The persistence algorithm: first version

Input: $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ a d-dimensional filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

$L_0 = L_1 = \cdots = L_{d-1} = \emptyset$
For $j = 0$ to m
  $k = \dim \sigma^j - 1$;
  if σ^j is a negative simplex
    $l(j) =$ highest index of the positive simplices associated to $\partial\sigma^j$;
    $L_k = L_k \cup \{(\sigma^{l(j)}, \sigma^j)\}$;
  end if
end for

Output: $L_0, L_1, \cdots, L_{d-1}$;

The persistence algorithm: first version

Input: $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ a d-dimensional filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

$L_0 = L_1 = \cdots = L_{d-1} = \emptyset$
For $j = 0$ to m
    $k = \dim \sigma^j - 1$;
    if σ^j is a negative simplex
        $l(j) =$ highest index of the positive simplices associated to $\partial\sigma^j$;
        $L_k = L_k \cup \{(\sigma^{l(j)}, \sigma^j)\}$;
    end if
end for

Output: $L_0, L_1, \cdots, L_{d-1}$;

How to test this condition?