MVA, Fall 2024

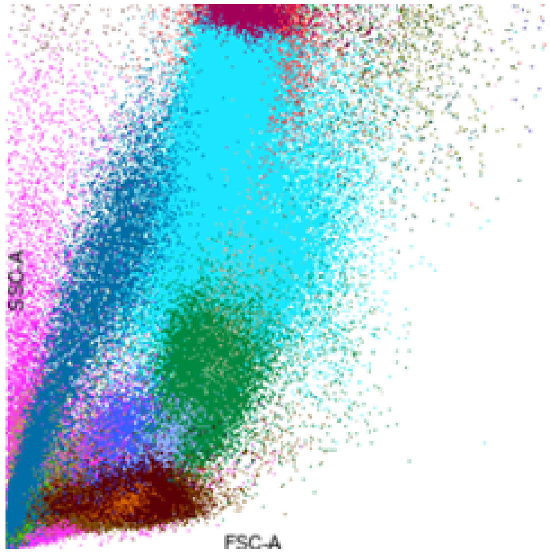
Persistent homology for TDA

Frédéric Chazal
DataShape team
Inria & Laboratoire de Mathématiques d'Orsay Institut
DATAIA Université Paris-Saclay

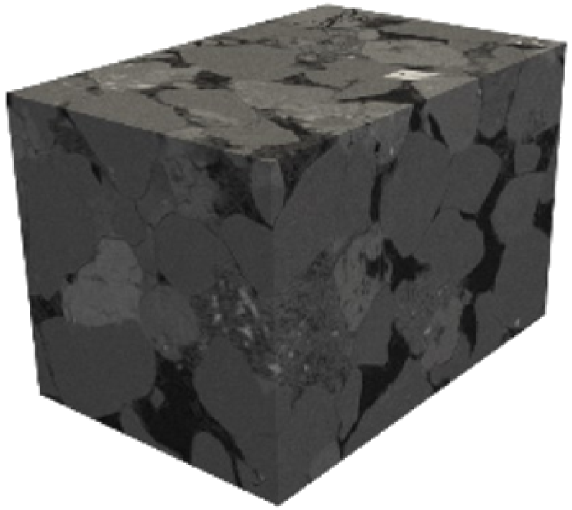
For slides and practical classes:
https://geometrica.saclay.inria.fr/team/Fred.Chazal/MVA2024.htm

Ref (for lectures 1 and 2): J.-D. Boissonnat, F. Chazal, M. Yvinec, Geometric and Topological Inference, Cambridge University Press, 2018.

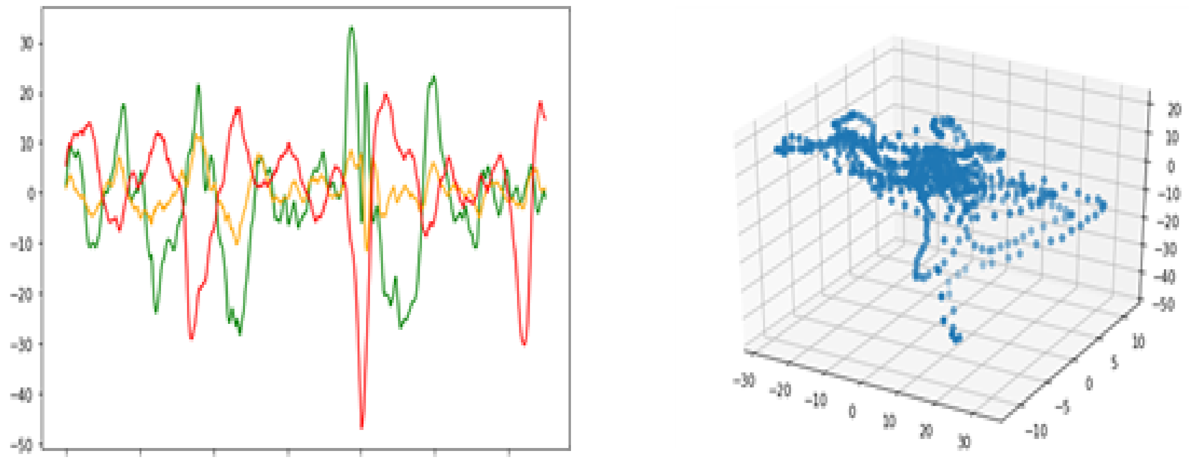
What is Topological Data Analysis (TDA)?



[Cell population - cytometry - MetaFora courtesy]



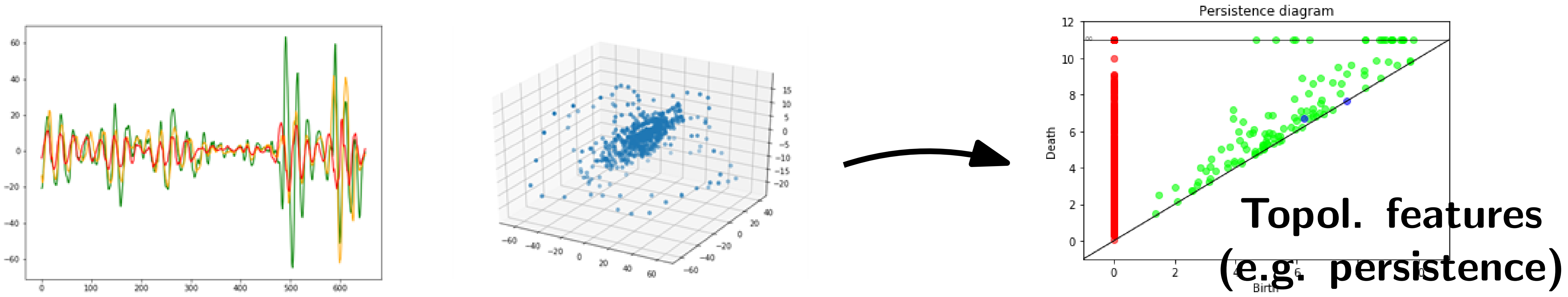
[Porous material (IFPEN courtesy)]



[Sensors (Sysnav courtesy)]

Modern data carry complex, but important, geometric/topological structure!

What is Topological Data Analysis (TDA)?



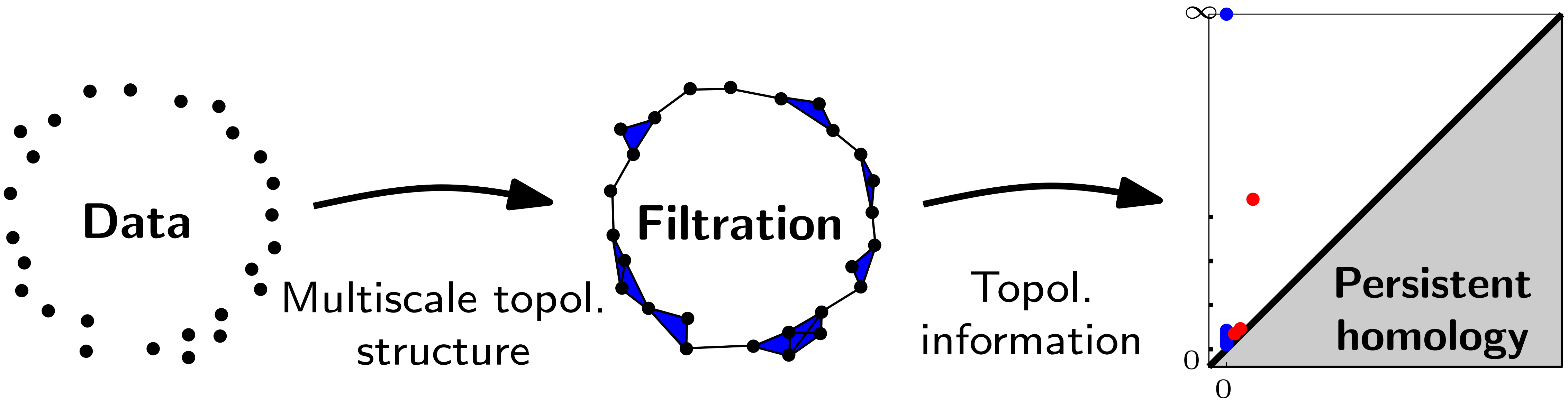
Data

Topol. features (e.g. persistence)

Topological Data Analysis (TDA) is a recent field whose aim is to:

- infer relevant topological and geometric features from complex data,
- take advantage of topological/geometric information for further Data Analysis, Machine Learning and AI tasks:
  - using topological features in ML pipelines,
  - taking advantage of topological information to improve ML pipelines.

A classical TDA pipeline



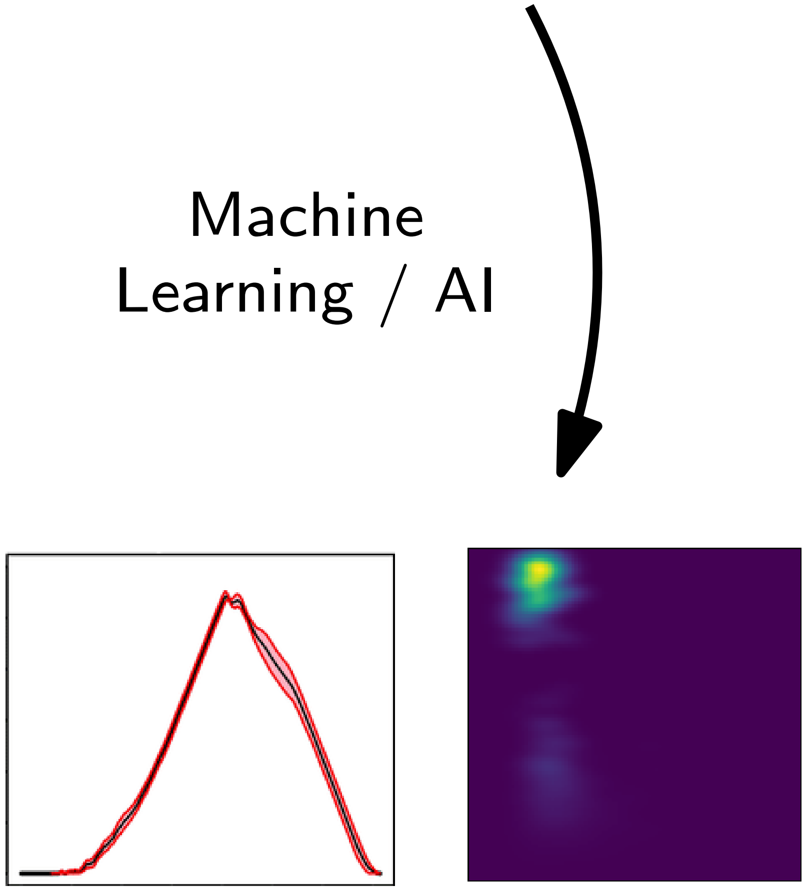
Data

Multiscale topol. structure

Filtration

Topol. information

Persistent homology

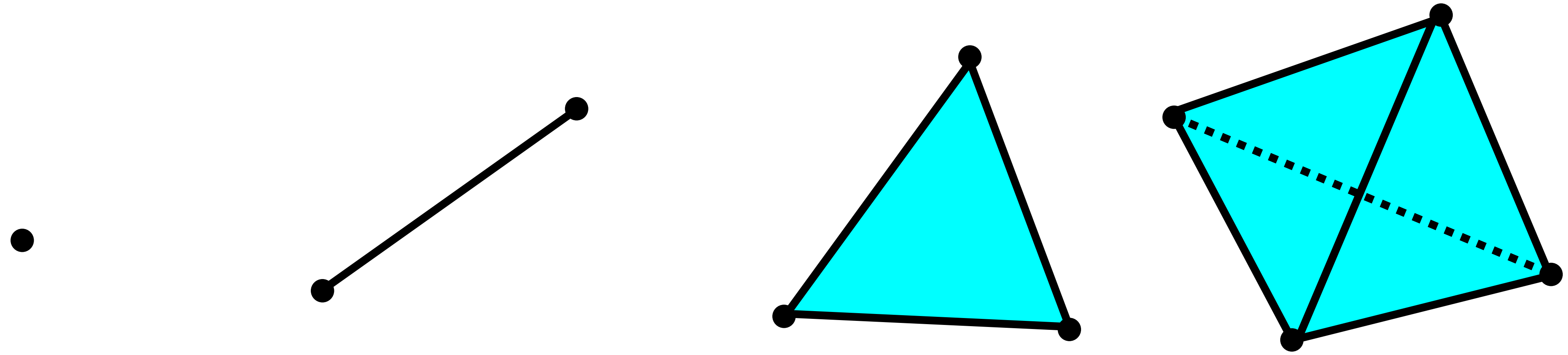


Machine Learning / AI

Representations of persistence

1. Build a multiscale topol. structure on top of data: filtrations.
2. Compute multiscale topol. signatures: persistent homology
3. Take advantage of the signature for further Machine Learning and AI tasks: Statistical aspects and representations of persistence

Simplicial complexes



0-simplex: vertex

1-simplex: edge

2-simplex: triangle

3-simplex: tetrahedron

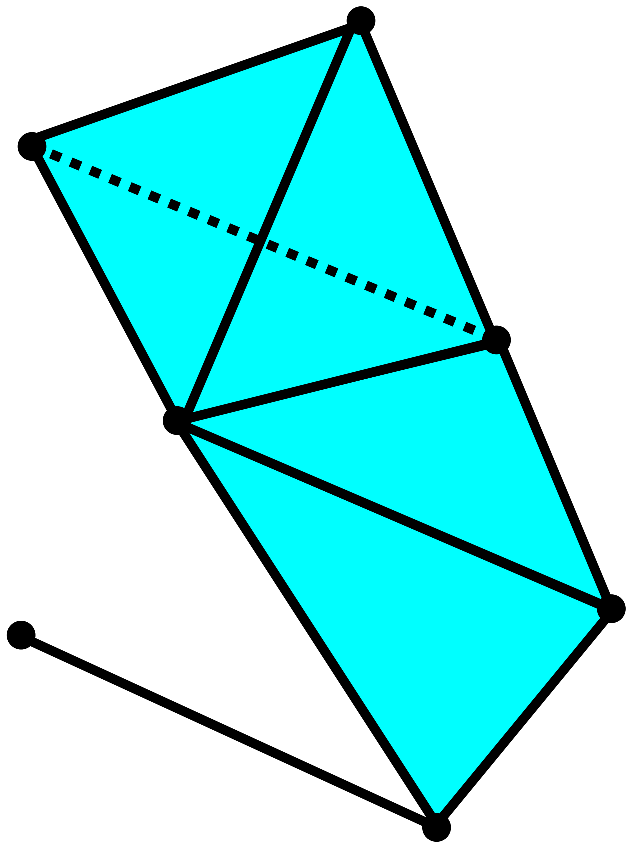
etc...

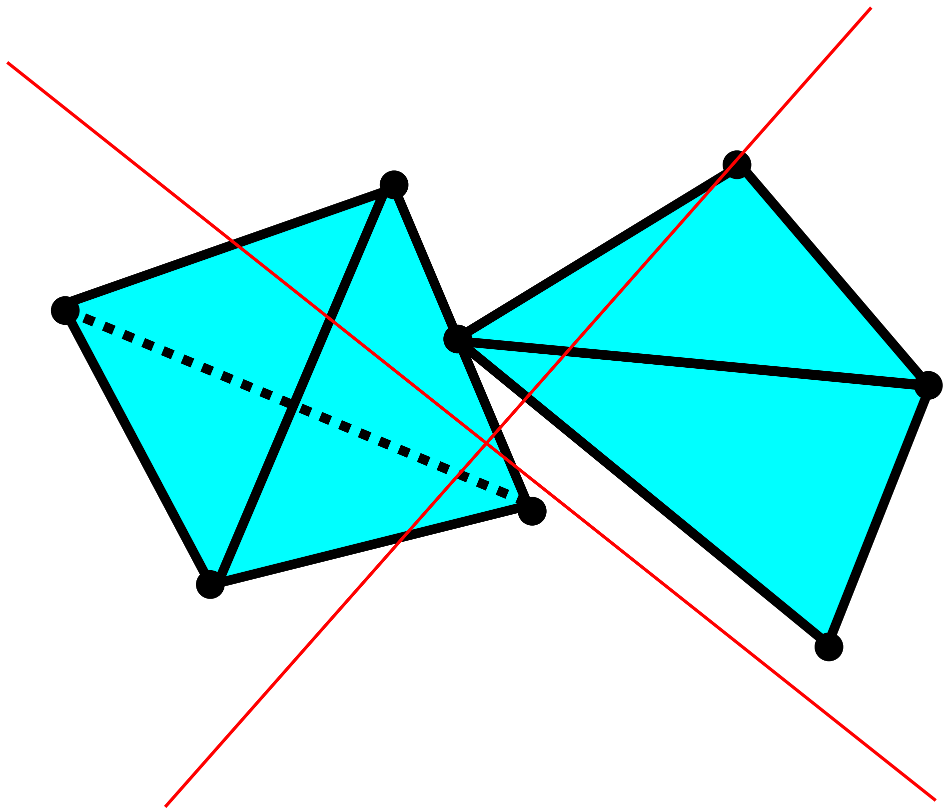
Given a set $P = \{p_0, \ldots, p_k\} \subset \mathbb{R}^d$ of $k+1$ affinely independent points, the k-dimensional simplex σ, or k-simplex for short, spanned by P is the set of convex combinations

$$\sum_{i=0}^{k} \lambda_i \, p_i, \quad \text{with} \quad \sum_{i=0}^{k} \lambda_i = 1 \quad \text{and} \quad \lambda_i \geq 0.$$

The points $p_0, \ldots, p_k$ are called the vertices of σ.

Simplicial complexes



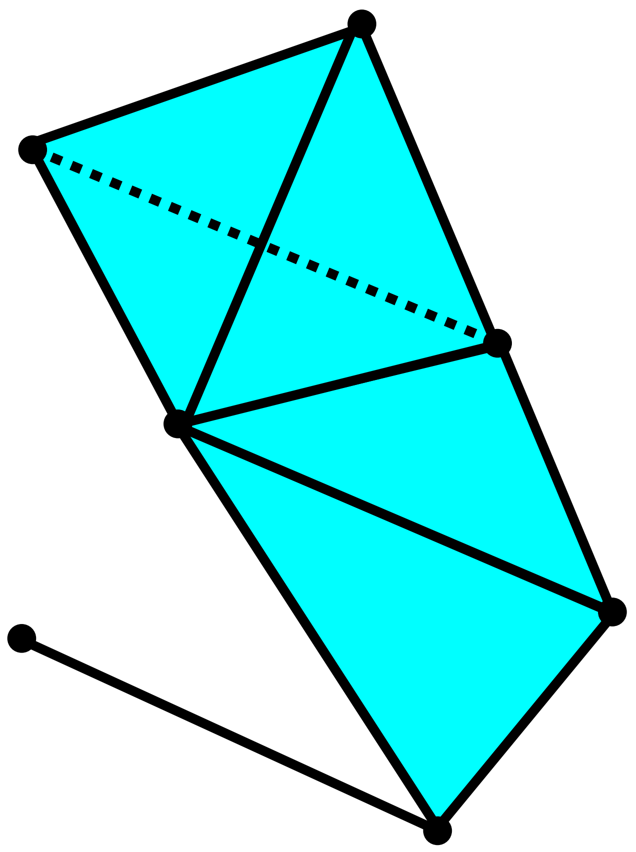


A (finite) simplicial complex K in $\mathbb{R}^d$ is a (finite) collection of simplices such that:

1. any face of a simplex of K is a simplex of K,
2. the intersection of any two simplices of K is either empty or a common face of both.

The underlying space of K, denoted by $|K| \subset \mathbb{R}^d$ is the union of the simplices of K.

Abstract simplicial complexes



Let P be a set. An abstract simplicial complex K with vertex set P is a set of finite subsets of P satisfying the two conditions :

1. The elements of P belong to K.
2. If $\tau \in K$ and $\sigma \subseteq \tau$, then $\sigma \in K$.

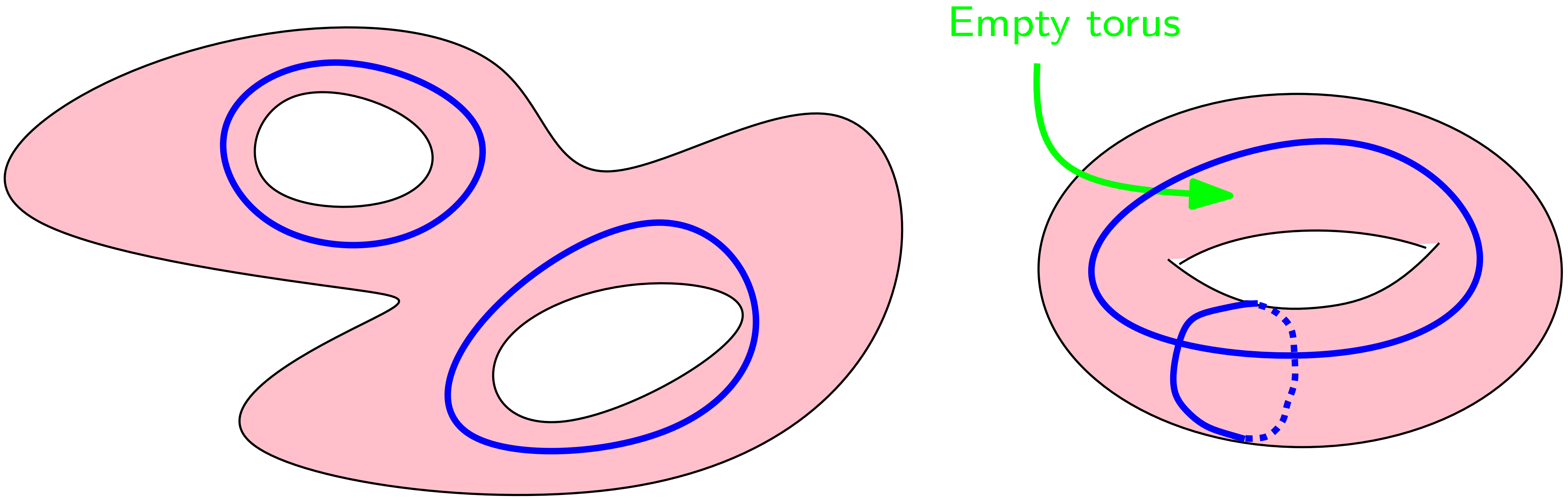
The elements of K are the simplices.

IMPORTANT

Simplicial complexes can be seen at the same time as geometric/topological spaces (good for top./geom. inference) and as combinatorial objects (abstract simplicial complexes, good for computations).

Homology in a nutshell (with coeff. in $\mathbb{Z}/2\mathbb{Z}$)

Formalize the notion of connected components, cycles/holes, voids... in a topological space (here we will restrict to simplicial complexes).



- 2 connected components (0-dim homology)
- 4 cycles (1-dim homology)
- 1 void (2-dim homology)

Homology in a nutshell (with coeff. in $\mathbb{Z}/2\mathbb{Z}$)

The space of k-chains:

Let K be a d-dimensional simplicial complex. Let $k \in \{0, 1, \cdots, d\}$ and $\{\sigma_1, \cdots, \sigma_p\}$ be the set of k-simplices of K.

k-chain:

$$c = \sum_{i=1}^{p} \varepsilon_i \sigma_i \quad \text{with} \quad \varepsilon_i \in \mathbb{Z}/2\mathbb{Z} = \{0, 1\}$$

Sum of k-chains:

$$c + c' = \sum_{i=1}^{p} (\varepsilon_i + \varepsilon_i')\sigma_i \quad \text{and} \quad \lambda.c = \sum_{i=1}^{p} (\lambda \varepsilon_i')\sigma_i$$

where the sums $\varepsilon_i + \varepsilon_i'$ and the products $\lambda \varepsilon_i$ are modulo 2.

Homology in a nutshell (with coeff. in $\mathbb{Z}/2\mathbb{Z}$)

The boundary operator:

The boundary $\partial\sigma$ of a k-simplex σ is the sum of its $(k-1)$-faces. This is a $(k-1)$-chain.

$$If \sigma = [v_0, \cdots, v_k] \ \text{ then } \ \partial_k \sigma = \sum_{i=0}^{k} (-1)^i [v_0 \cdots \hat{v}_i \cdots v_k]$$

The boundary operator is the linear map defined by

$$\begin{array}{rlcl} \partial_k : & \mathcal{C}_k(K) & \rightarrow & \mathcal{C}_{k-1}(K) \\ & c & \rightarrow & \partial_k c = \sum_{\sigma \in c} \partial_k \sigma \end{array}$$

$$\partial_k \partial_{k+1} := \partial_k \circ \partial_{k+1} = 0$$

Homology in a nutshell (with coeff. in $\mathbb{Z}/2\mathbb{Z}$)

Cycles and boundaries:

The chain complex associated to a complex K of dimension d

$$\emptyset \to \mathcal{C}_d(K) \overset{\partial}{\to} \mathcal{C}_{d-1}(K) \overset{\partial}{\to} \cdots \mathcal{C}_{k+1}(K) \overset{\partial}{\to} \mathcal{C}_k(K) \overset{\partial}{\to} \cdots \mathcal{C}_1(K) \overset{\partial}{\to} \mathcal{C}_0(K) \overset{\partial}{\to}$$

k-cycles:

$$Z_k(K) := \ker(\partial : \mathcal{C}_k \to \mathcal{C}_{k-1}) = \{c \in \mathcal{C}_k : \partial c = \emptyset\}$$

k-boundaries:

$$B_k(K) := im(\partial : \mathcal{C}_{k+1} \to \mathcal{C}_k) = \{c \in \mathcal{C}_k : \exists c' \in \mathcal{C}_{k+1}, c = \partial c'\}$$

$$B_k(K) \subset Z_k(K) \subset \mathcal{C}_k(K)$$

Homology in a nutshell (with coeff. in $\mathbb{Z}/2\mathbb{Z}$)

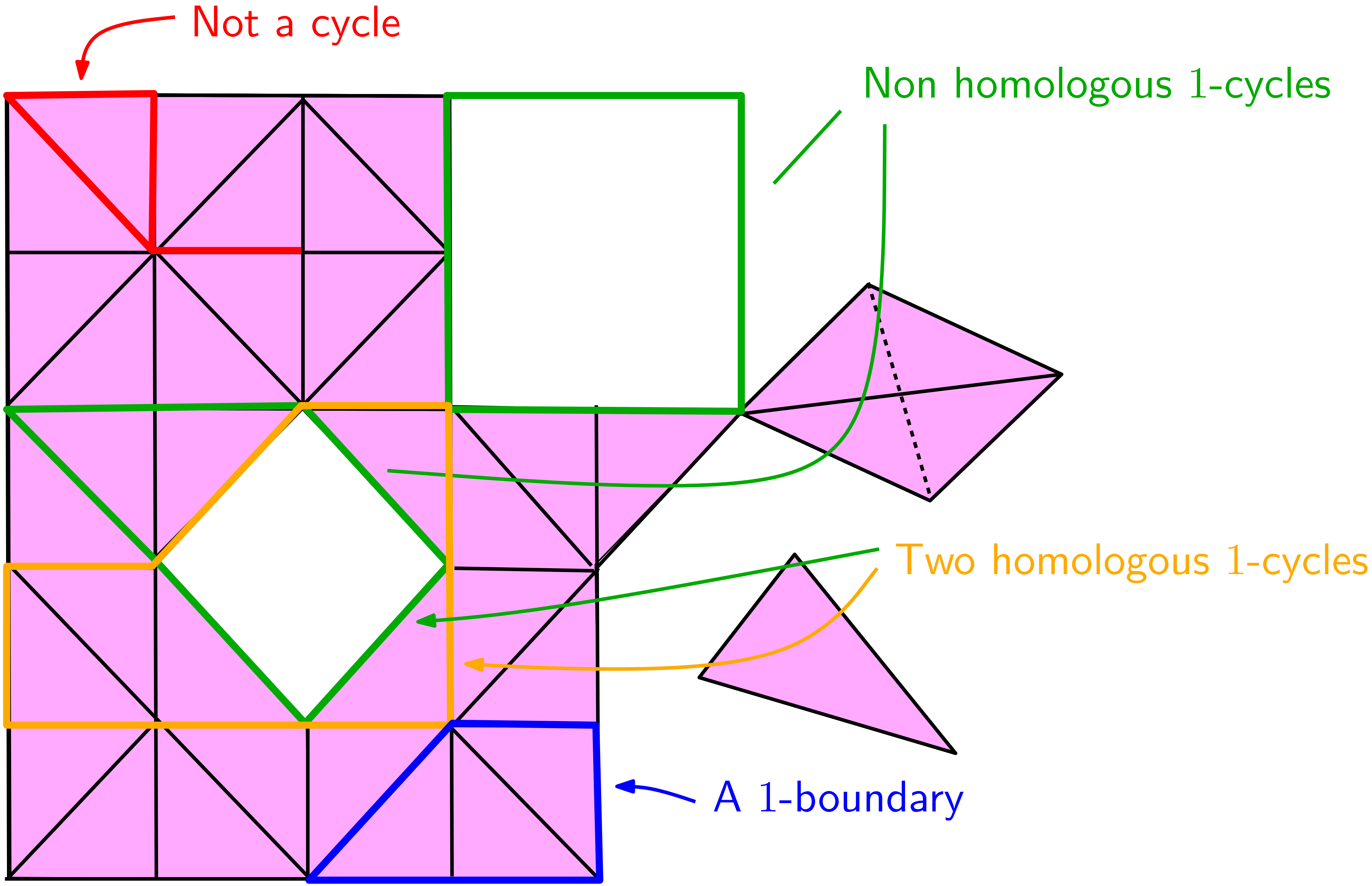
Homology groups and Betti numbers:

$$B_k(K) \subset Z_k(K) \subset \mathcal{C}_k(K)$$

- The k^{th} homology group of K: $H_k(K) = Z_k/B_k$
- Tout each cycle $c \in Z_k(K)$ corresponds its homology class $c + B_k(K) = \{c + b : b \in B_k(K)\}$.
- Two cycles c, c' are homologous if they are in the same homology class: $\exists b \in B_k(K)$ s. t. $b = c' - c (= c' + c)$.
- The k^{th} Betti number of K: $\beta_k(K) = \dim(H_k(K))$.

  Remark: $\beta_0(K) =$ number of connected components of K.

Cycles and boundaries



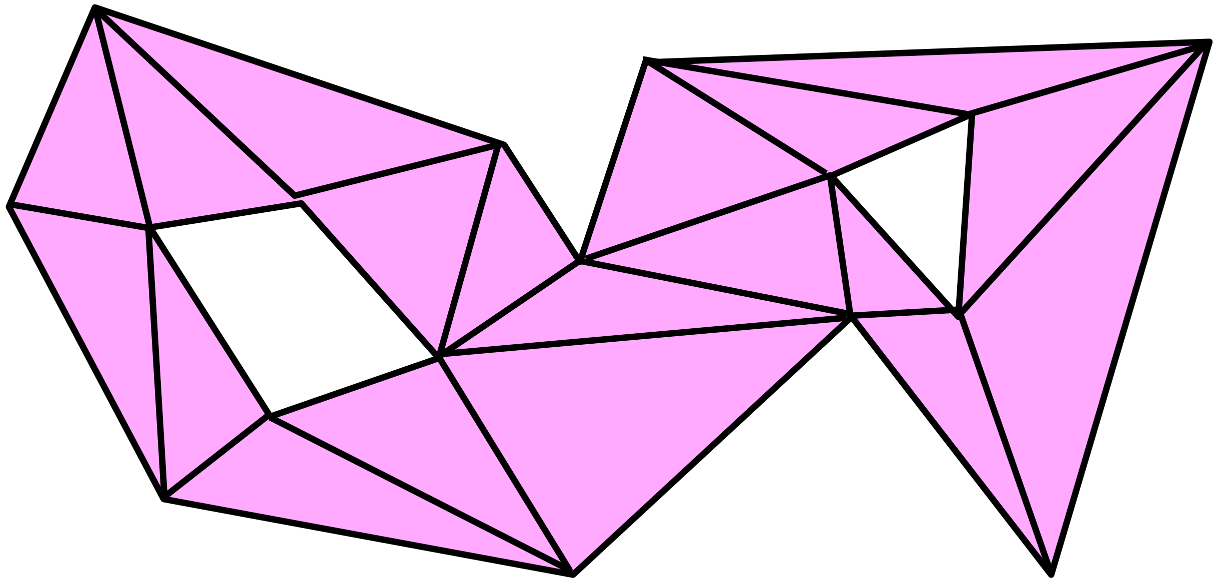
Not a cycle

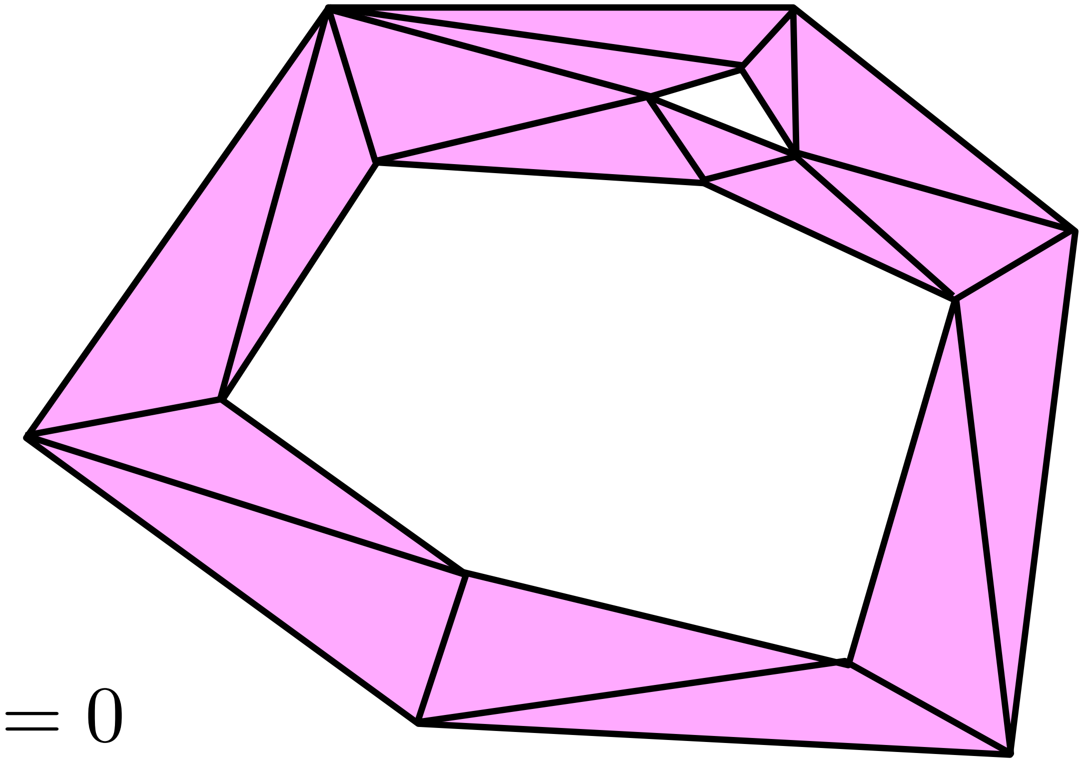
Non homologous 1-cycles

Two homologous 1-cycles

A 1-boundary

Topological invariance and singular homology



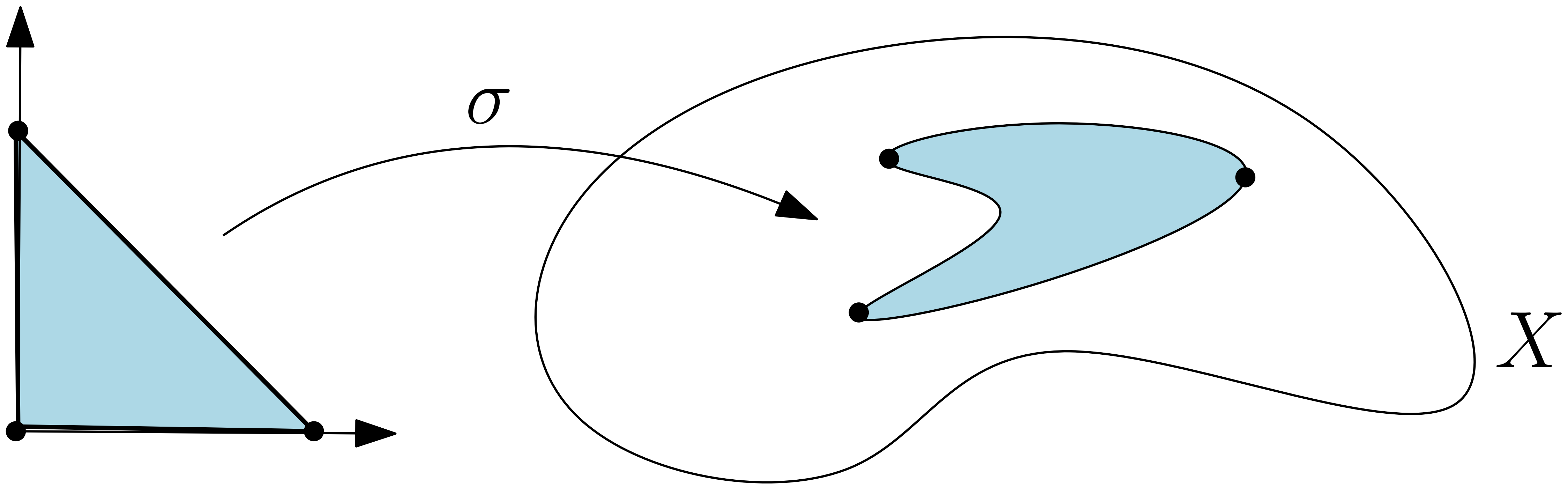


$\beta_0 = 1,\ \beta_1 = 2, \beta_2 = 0$

Theorem: If K and K' are two simplicial complexes with homeomorphic supports then their homology groups are isomorphic and their Betti numbers are equal.

- This is a classical result in algebraic topology but the proof is not obvious.
- Rely on the notion of singular homology → defined for any topological space.

Topological invariance and singular homology



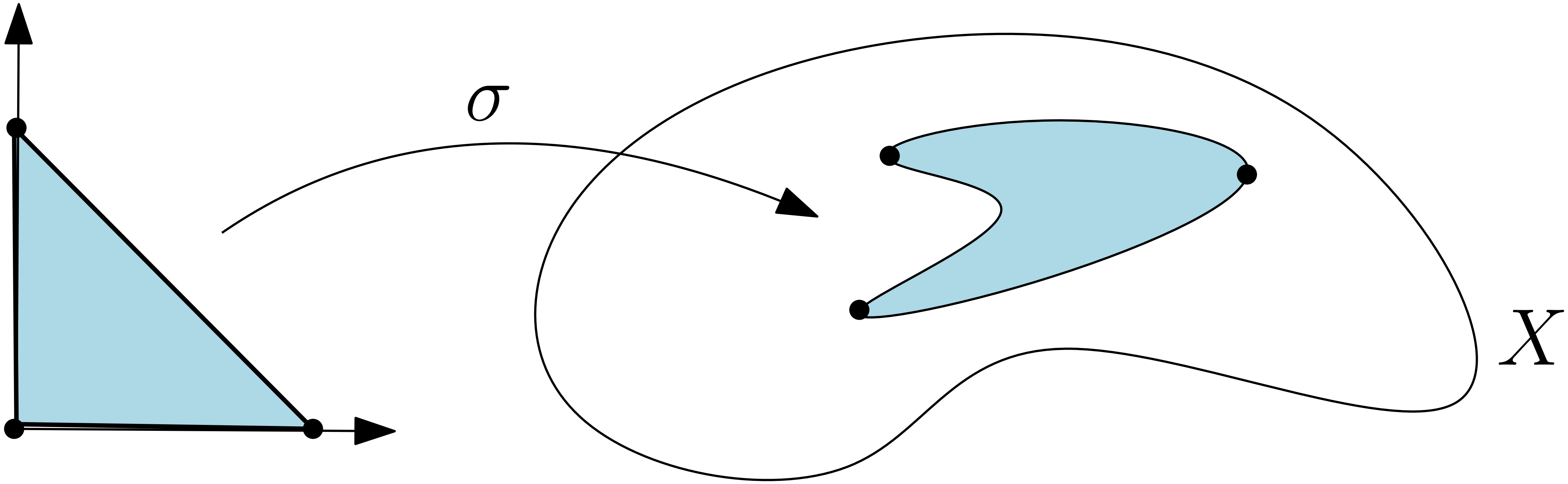
Let Δ_k be the standard simplex in $\mathbb{R}^k$. A singular k-simplex in a topological space X is a continuous map $\sigma : \Delta_k \rightarrow X$.

The same construction as for simplicial homology can be done with singular complexes → Singular homology

Important properties:

- Singular homology is defined for any topological space X.
- If X is homotopy equivalent to the support of a simplicial complex, then the singular and simplicial homology coincide!

Topological invariance and singular homology



Homology and continuous maps:

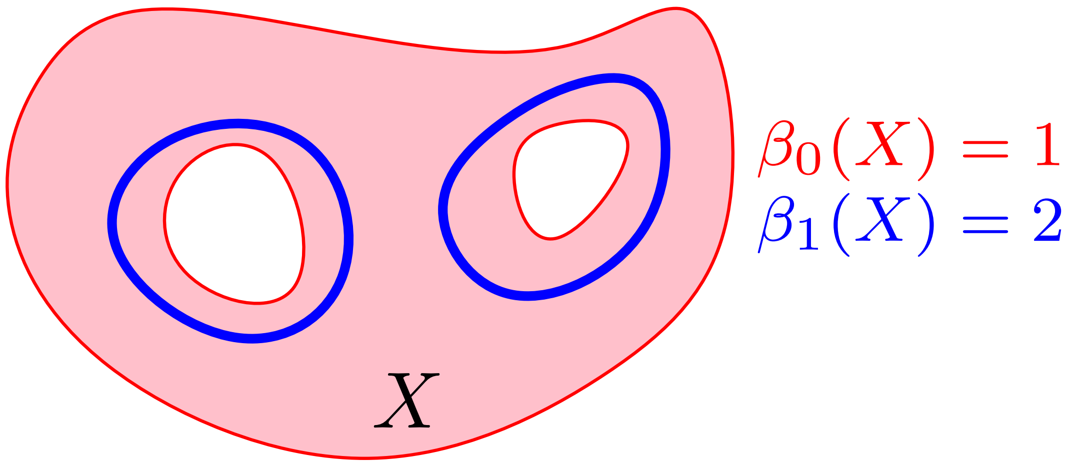
- if $f : X \to Y$ is a continuous map and $\sigma : \Delta_k \to X$ a simplex in X, then $f \circ \sigma : \Delta_k \to Y$ is a simplex in $Y \Rightarrow f$ induces a linear maps between homology groups:

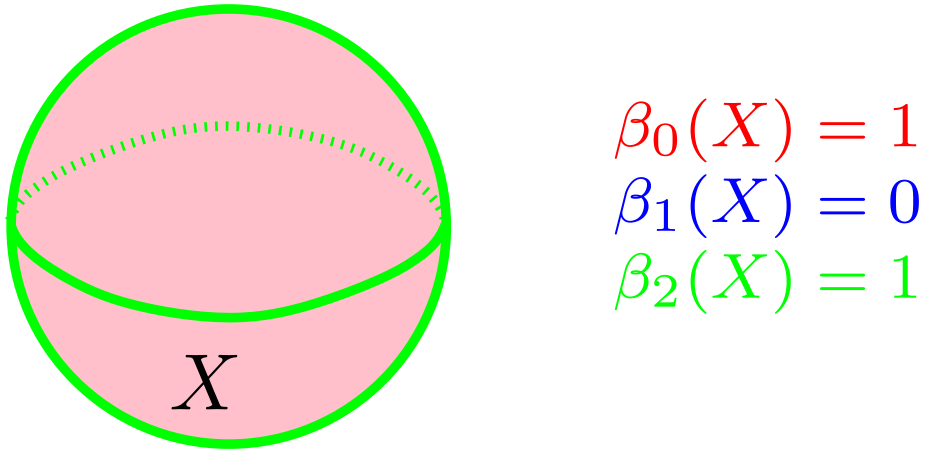
$$f_\sharp : H_k(X) \to H_k(Y)$$

- if $f : X \to Y$ is an homeomorphism or an homotopy equivalence then $f_\sharp$ is an isomorphism.

Homology (to summarize)

- X a topological space + $\mathbb{K}$ (e.g. $\mathbb{K} = \mathbb{Z}/2, \mathbb{Z}/p, \mathbb{R} \cdots$) a field + k a non-negative integer.
- The k-th homology group $H_k(X, \mathbb{K})$: a vector space with coefficients in $\mathbb{K}$.
- Elements of $H_k(X, \mathbb{K})$: represent the k-dimensional cycles in X.
- Betti numbers: $\beta_k(X) = \dim(H_k(X, \mathbb{K}))$.

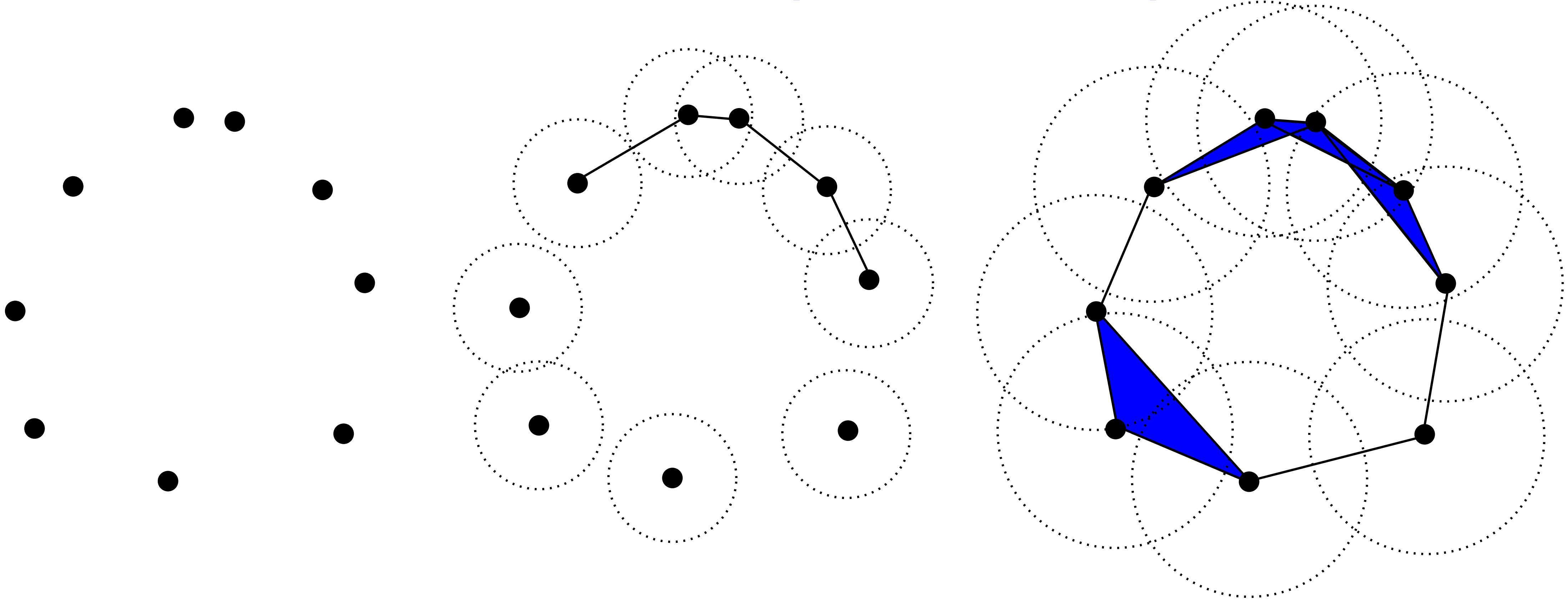




(Some) properties:

- $\beta_0(X) =$ number of connected components of X.
- If $f : X \to Y$ is continuous, then f induces a linear map $f_\sharp : H_k(X) \to H_k(Y)$.
- In particular, if $X \subset Y$, then the inclusion map $i : X \to Y$ induces a linear map $H_k(X) \to H_k(Y)$.

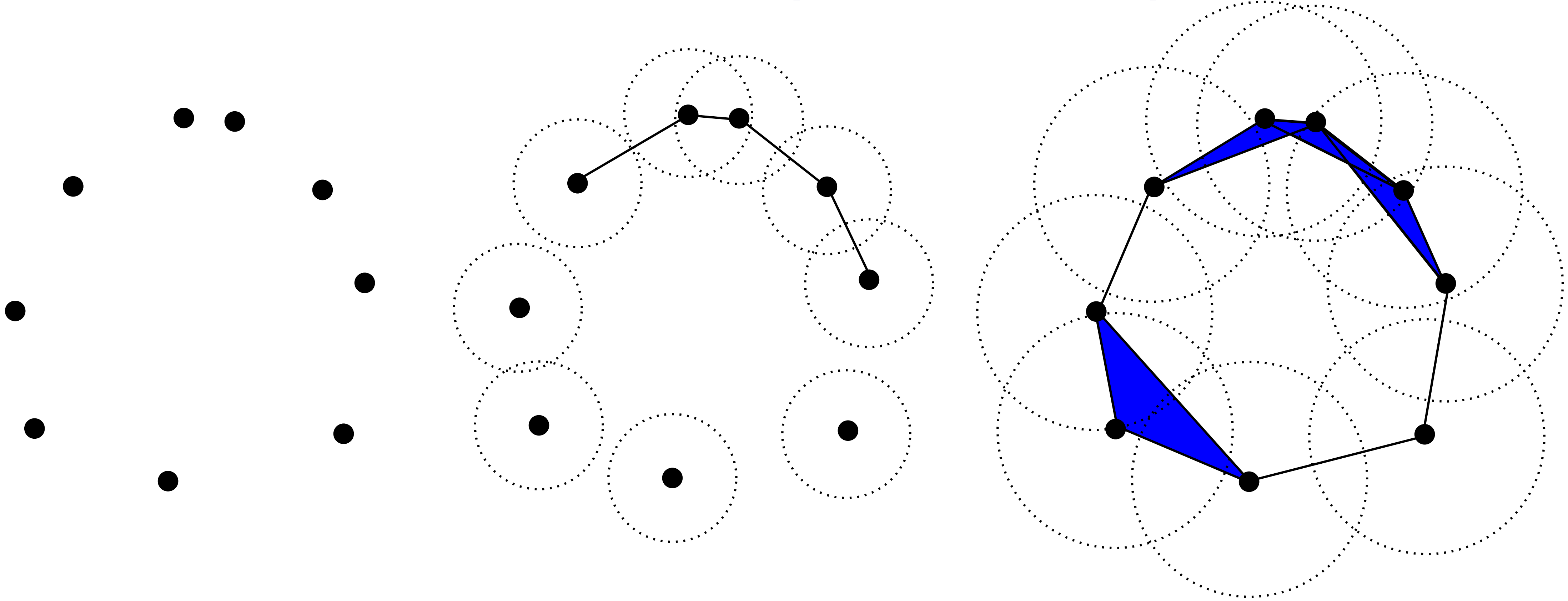
Filtrations of simplicial complexes



- A filtered simplicial complex (or a filtration) $\mathbb{K}$ built on top of a set X is a family $(K_a \mid a \in \mathbf{T})$, $\mathbf{T} \subseteq \mathbb{R}$, of subcomplexes of some fixed simplicial complex K with vertex set X s. t. $K_a \subseteq K_b$ for any $a \leq b$.
- More generaly, filtration $=$ nested family of topological spaces indexed by $\mathbf{T}$.

Persistent homology of a filtered simplicial complexe encodes the evolution of the homology of the subcomplexes.

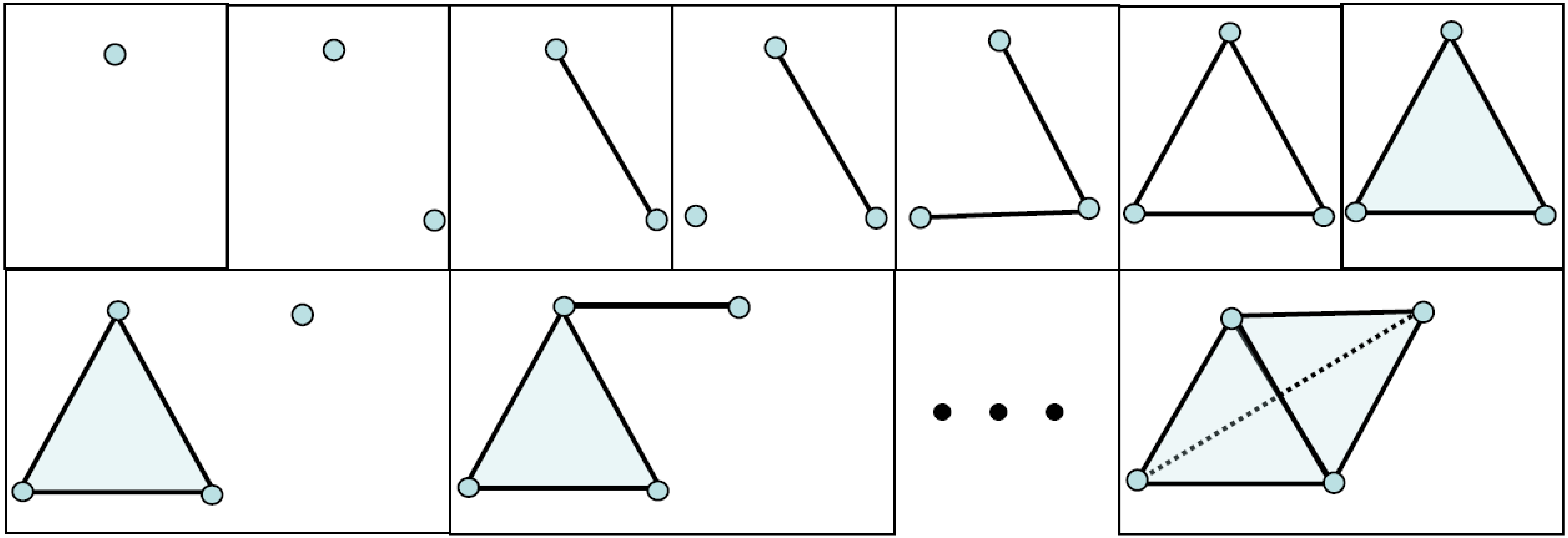
Filtrations of simplicial complexes



- A filtered simplicial complex (or a filtration) $\mathbb{K}$ built on top of a set X is a family $(K_a \mid a \in \mathbf{T})$, $\mathbf{T} \subseteq \mathbb{R}$, of subcomplexes of some fixed simplicial complex K with vertex set X s. t. $K_a \subseteq K_b$ for any $a \leq b$.
- More generaly, filtration $=$ nested family of topological spaces indexed by $\mathbf{T}$.

Many examples and ways to design filtrations depending on the application and targeted objectives : sublevel and upperlevel sets, Čech complex,...

Filtrations of simplicial complexes



A filtration of a (finite) simplicial complex K is a sequence of subcomplexes such that

i) $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$,

ii) $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

There are many ways to build filtrations - see next lesson.

An algorithm to compute (simplicial) homology

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

Output: 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 $\sigma^i$ 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 $(\beta_0, \beta_1, \cdots, \beta_d)$;
```

An algorithm to compute (simplicial) homology

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

Output: 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 $\sigma^i$ 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 $(\beta_0, \beta_1, \cdots, \beta_d)$;
```

Remark: At the i^{th} step of the algorithm, the vector $(\beta_0, \cdots, \beta_d)$ stores the Betti numbers of K^i.

Proof

- If σ^i is contained in a $(k+1)$-cycle in K^i, this cycle is not a boundary in K^i.

- If σ^i is contained in a $(k+1)$-cycle c in K^i, then c cannot be homologous to a cycle in K^{i-1}

$$\Rightarrow \beta_{k+1}(K^i) \geq \beta_{k+1}(K^{i-1}) + 1$$

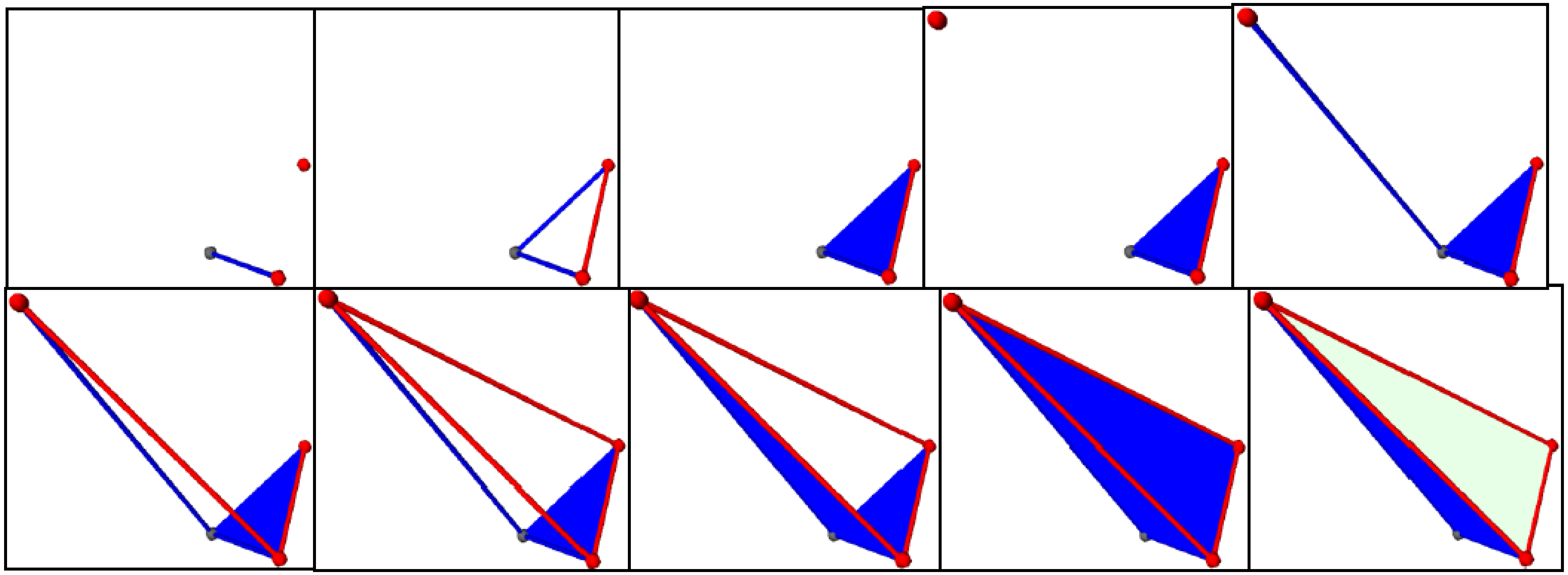
- If σ^i is not contained in a $(k+1)$-cycle c in K^i, then $\partial\sigma^i$ is not a boundary in K^{i-1}

$$\Rightarrow \beta_k(K^i) \leq \beta_k(K^{i-1}) - 1$$

- the previous inequalities are equalities.

Positive and negative simplicies

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.



Definition: A $(k+1)$-simplex σ^i is positive if it is contained in a $(k+1)$-cycle in K^i. It is negative otherwise.

Create a new $(k+1)$-cycle in K^i

Destroy a k-cycle in K^i

$$\beta_k(K) = \sharp(\text{positive simplices}) \;-\; \sharp(\text{negative simplices})$$

Getting more information

Definition: A $(k+1)$-simplex σ^i is positive if it is contained in a $(k+1)$-cycle in K^i. It is negative otherwise.

Create a new $(k+1)$-cycle in K^i

Destroy a k-cycle in K^i

$$\beta_k(K) = \sharp(\text{positive simplices}) \; - \; \sharp(\text{negative simplices})$$

- How to keep track of the evolution of the topology all along the filtration?
- What are the created/destroyed cycles?
- What is the lifetime of a cycle?
- How to compute rank$(H_k(K^i) \to H_k(K^j))$?

Getting more information

Definition: A $(k+1)$-simplex σ^i is positive if it is contained in a $(k+1)$-cycle in K^i. It is negative otherwise.

Create a new $(k+1)$-cycle in K^i

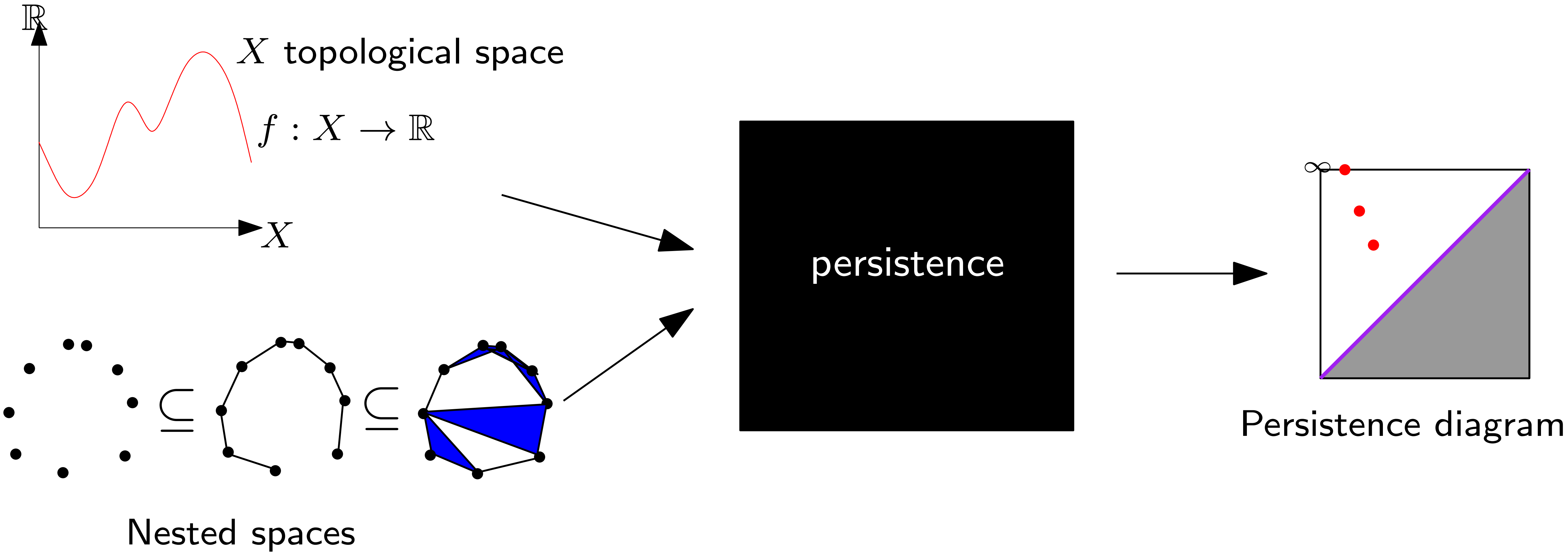
Destroy a k-cycle in K^i

$$\beta_k(K) = \sharp(\text{positive simplices}) \; - \; \sharp(\text{negative simplices})$$

- How to keep track of the evolution of the topology all along the filtration?
- What are the created/destroyed cycles?
- What is the lifetime of a cycle?
- How to compute rank$(H_k(K^i) \to H_k(K^j))$?

This is where persistent homology comes into play!

Persistent homology



X topological space

$f : X \to \mathbb{R}$

persistence

Persistence diagram

Nested spaces

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

The theory of persistence

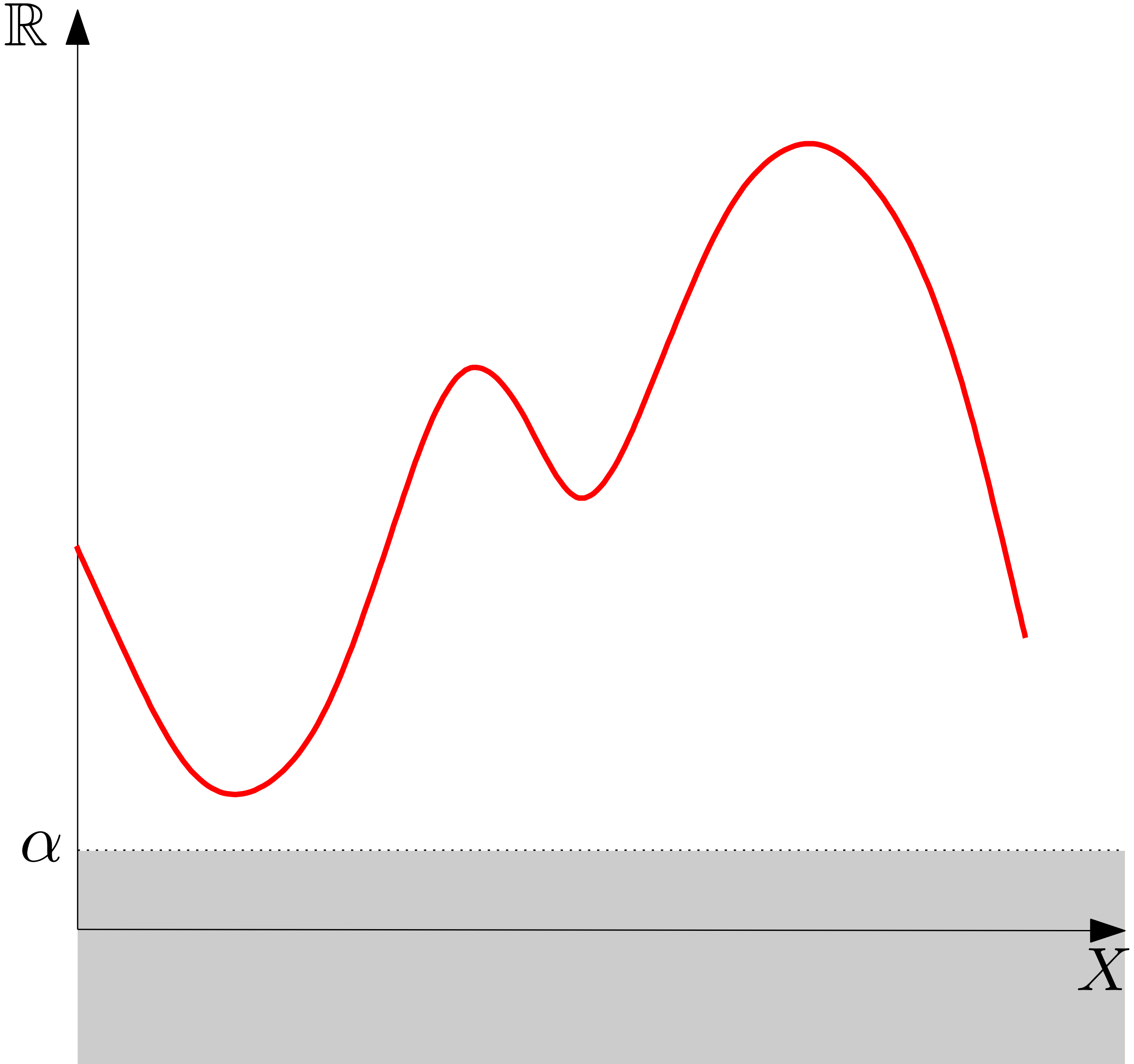
A general mathematical framework to encode the evolution of the topology (homology) of families of nested spaces (filtrations).

Historical landmarks:

- $90's$: size theory (P. Frosini et al), framed Morse complex and stability (S.A. Barannikov).
- $2002-2005$: persistent homology (H. Edelsbrunner et al, Carlsson et al).
- 2005: stability of persistence for continuous functions (D. Cohen-Steiner et al).
- $2009-2012$: algebraic stability of persistence modules (F.C. et al).
- 2014: the GUDHI software plateform (J.-D. Boissonnat et al). Also several other softs since 2005: Dionysus, (J)Plex, PHAT,...
- Last 5 years: statistical aspects of persistence and machine learning.

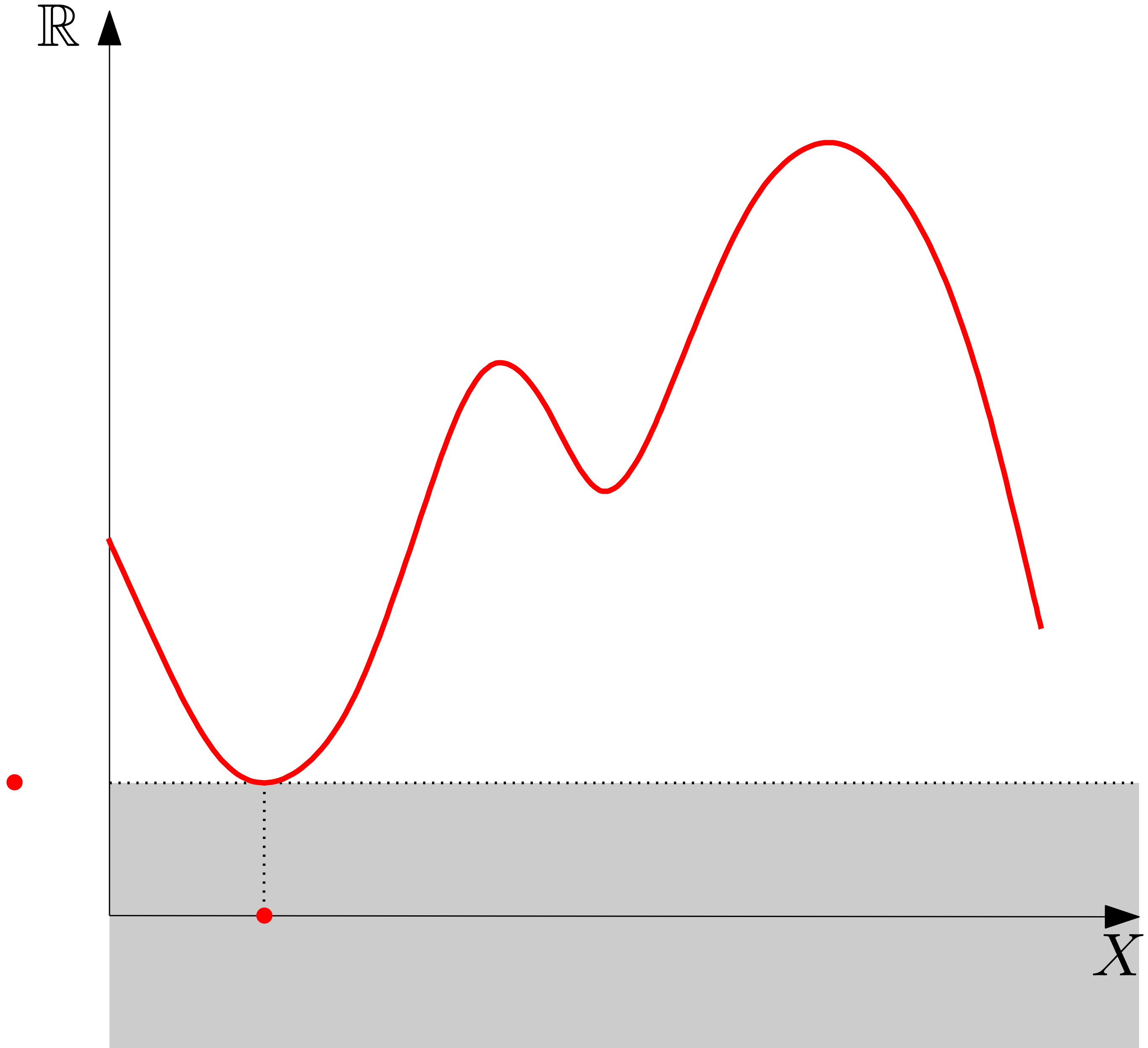
Persistent homology for functions

- Tracking and encoding the evolution of the connected components (0-dimensional homology) of the sublevel sets of a function
- The family of sublevel sets of a function is an example of filtration.



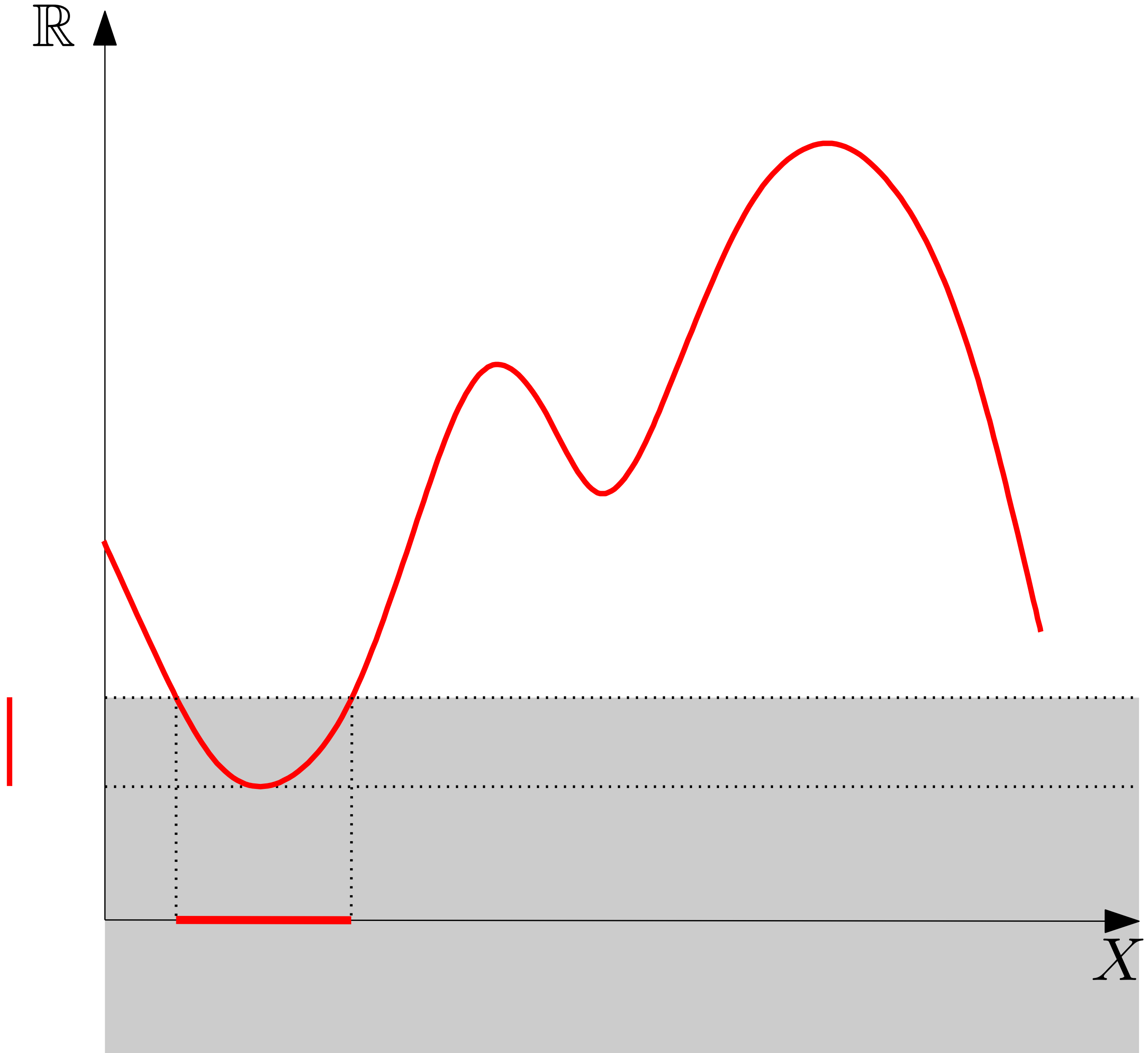
Persistent homology for functions

- Tracking and encoding the evolution of the connected components (0-dimensional homology) of the sublevel sets of a function
- The family of sublevel sets of a function is an example of filtration.



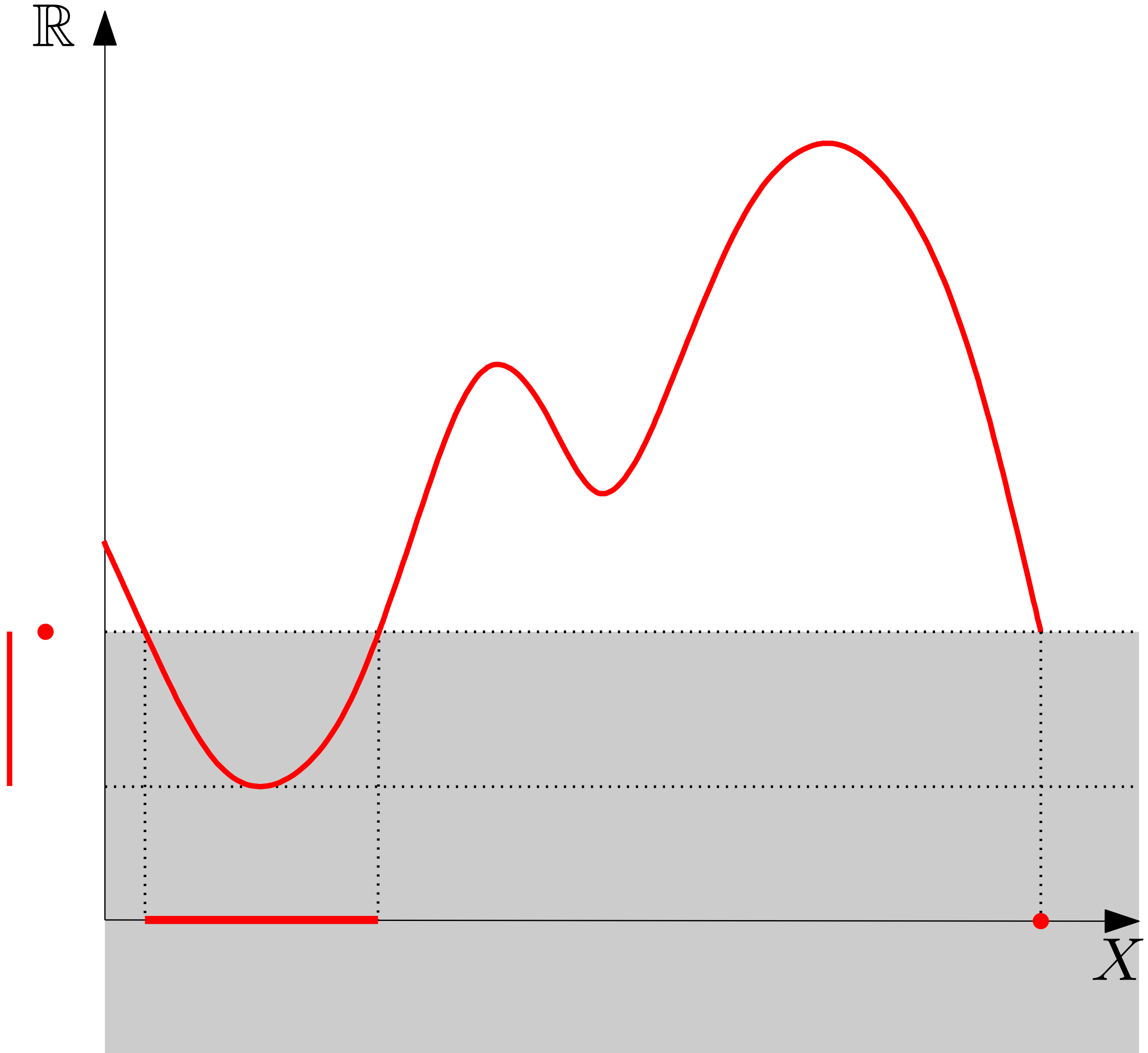
Persistent homology for functions

- Tracking and encoding the evolution of the connected components (0-dimensional homology) of the sublevel sets of a function
- The family of sublevel sets of a function is an example of filtration.



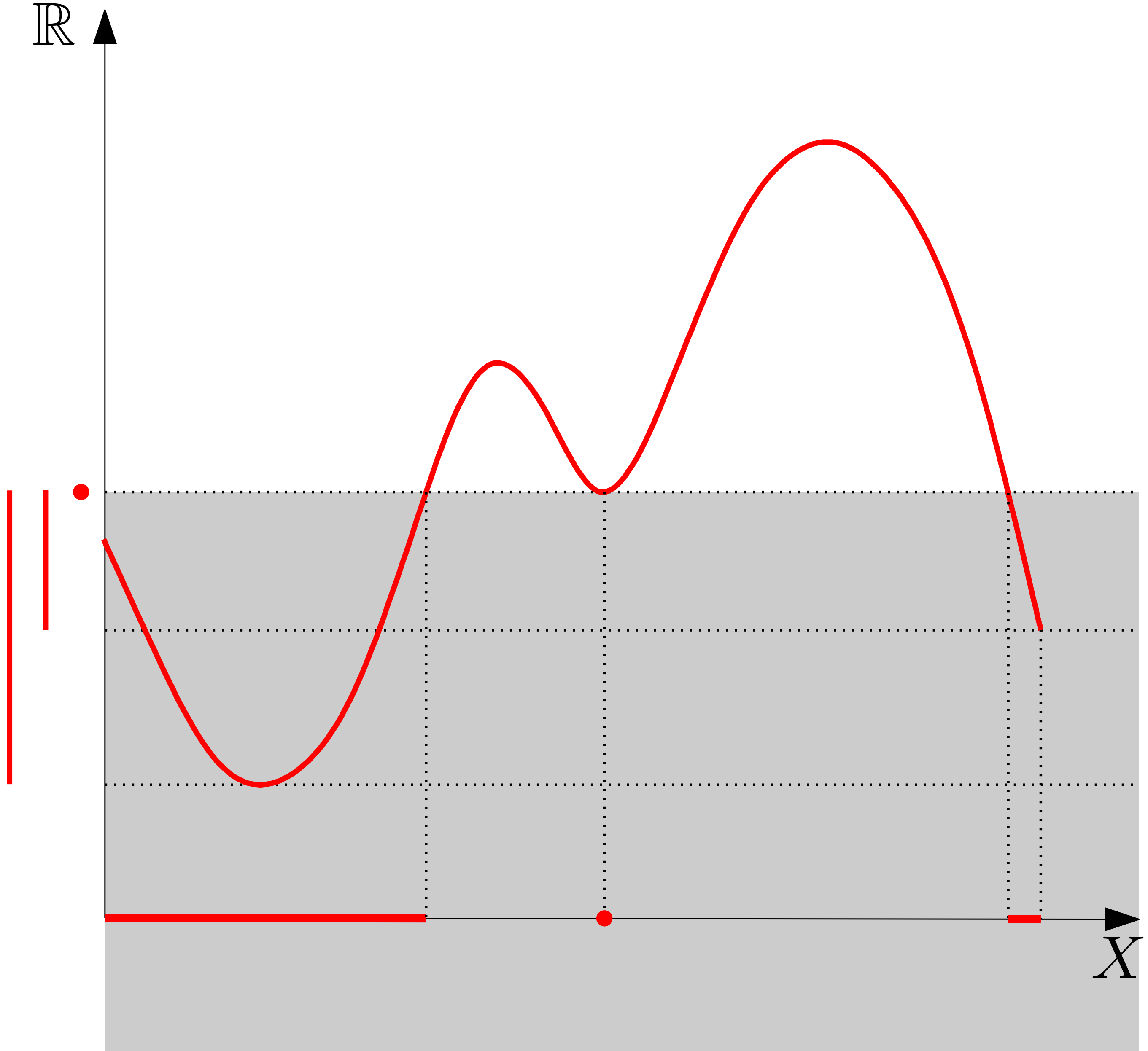
Persistent homology for functions

- Tracking and encoding the evolution of the connected components (0-dimensional homology) of the sublevel sets of a function
- The family of sublevel sets of a function is an example of filtration.



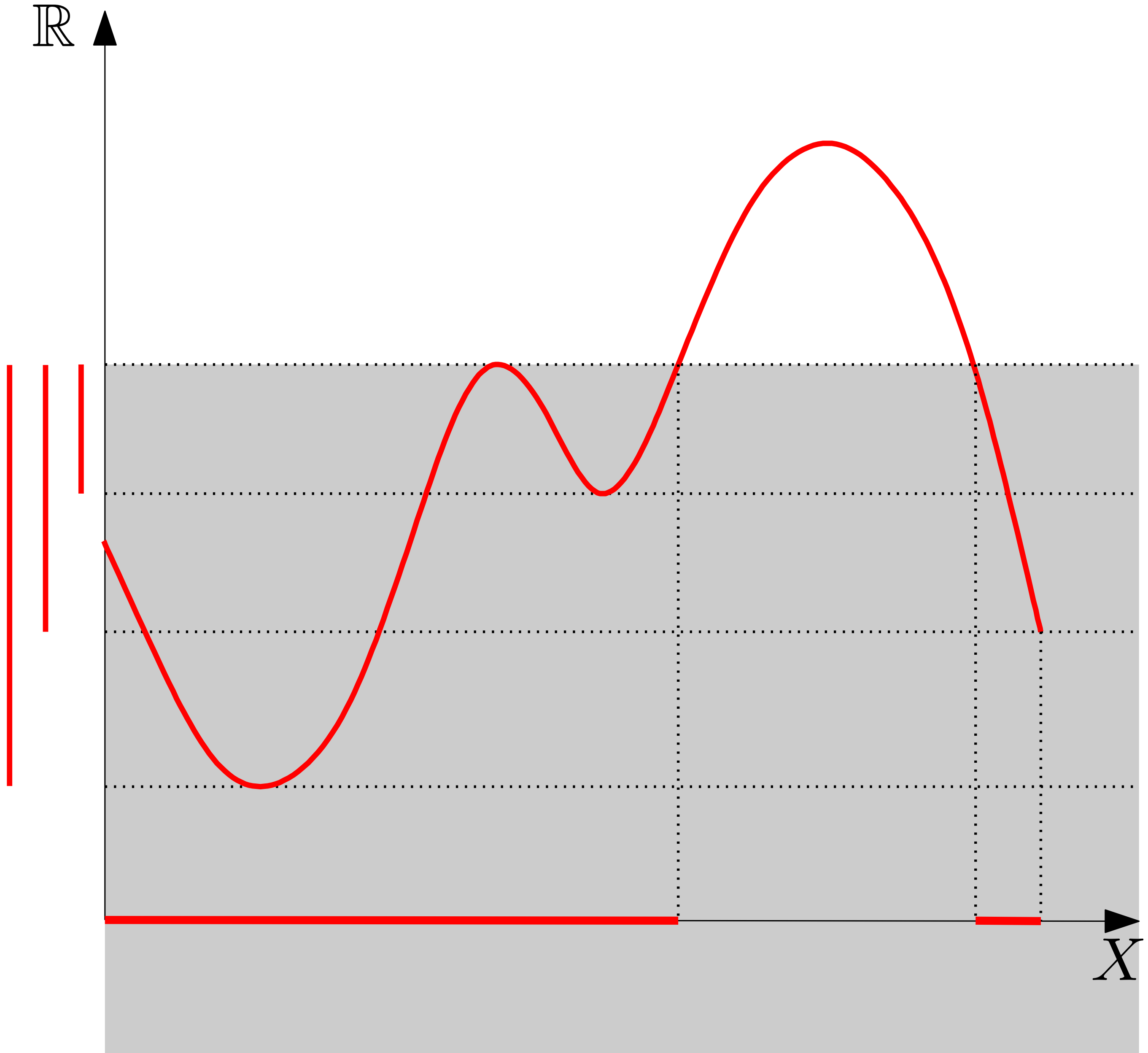
Persistent homology for functions

- Tracking and encoding the evolution of the connected components (0-dimensional homology) of the sublevel sets of a function
- The family of sublevel sets of a function is an example of filtration.



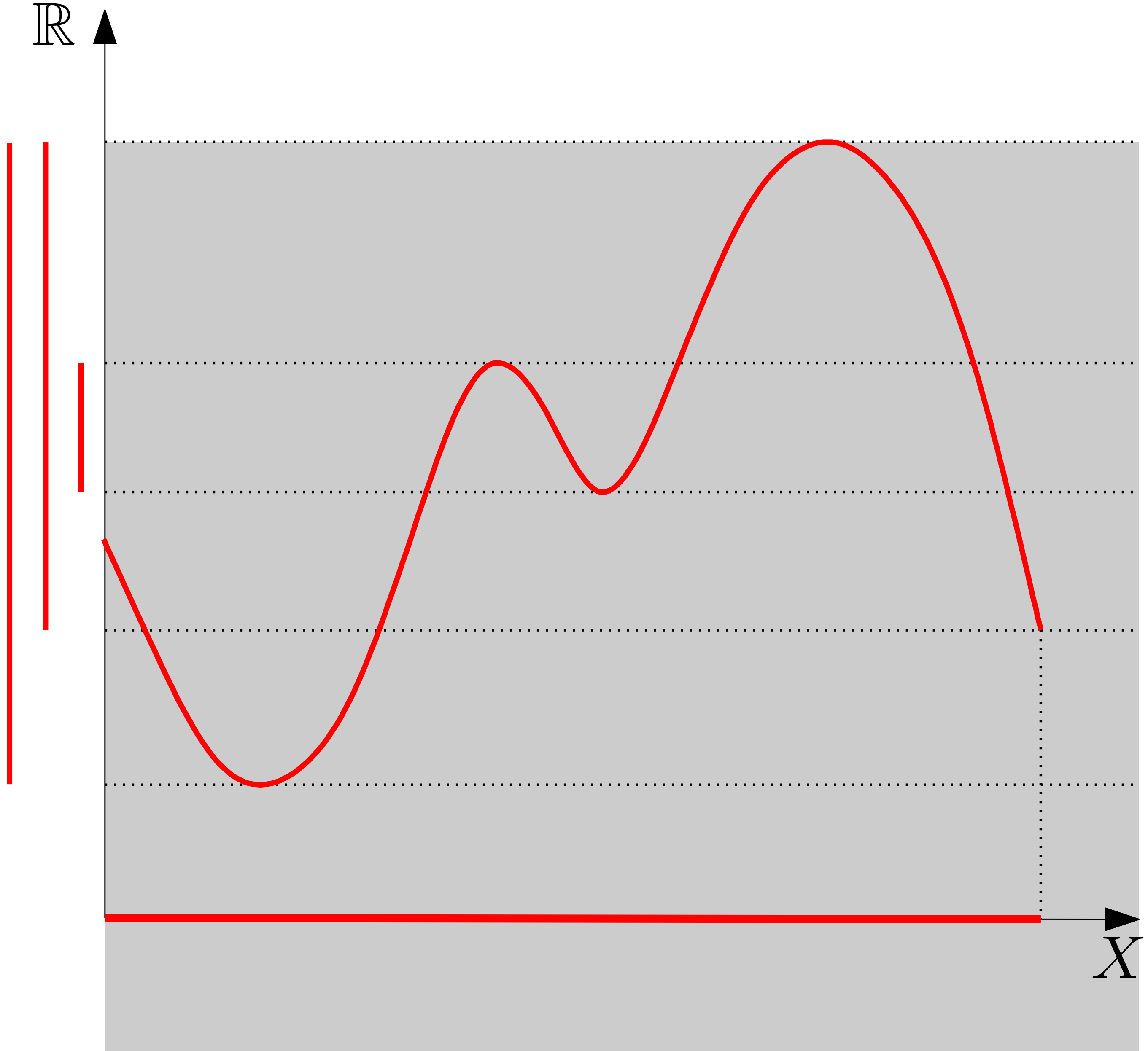
Persistent homology for functions

- Tracking and encoding the evolution of the connected components (0-dimensional homology) of the sublevel sets of a function
- The family of sublevel sets of a function is an example of filtration.



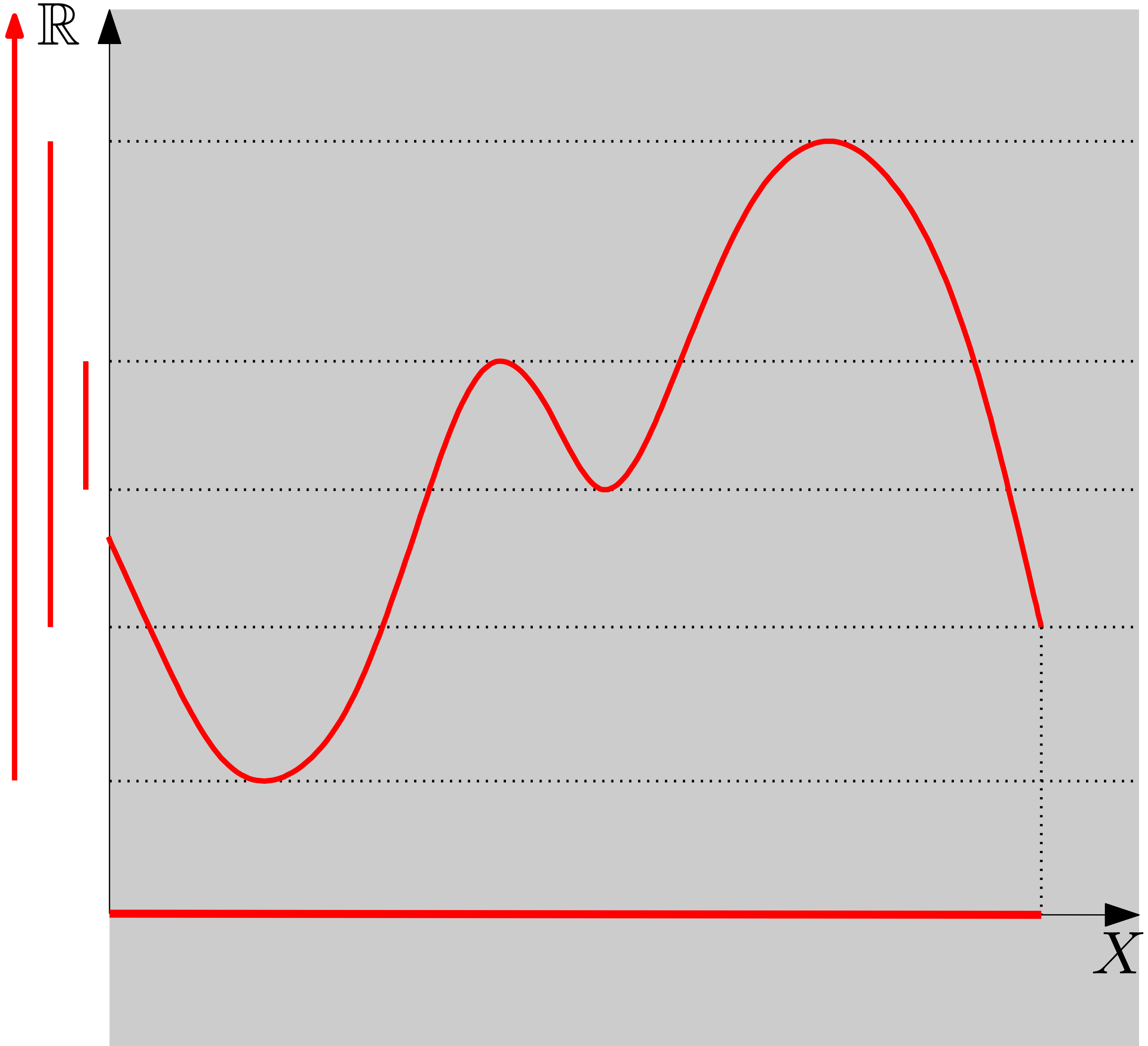
Persistent homology for functions

- Tracking and encoding the evolution of the connected components (0-dimensional homology) of the sublevel sets of a function
- The family of sublevel sets of a function is an example of filtration.



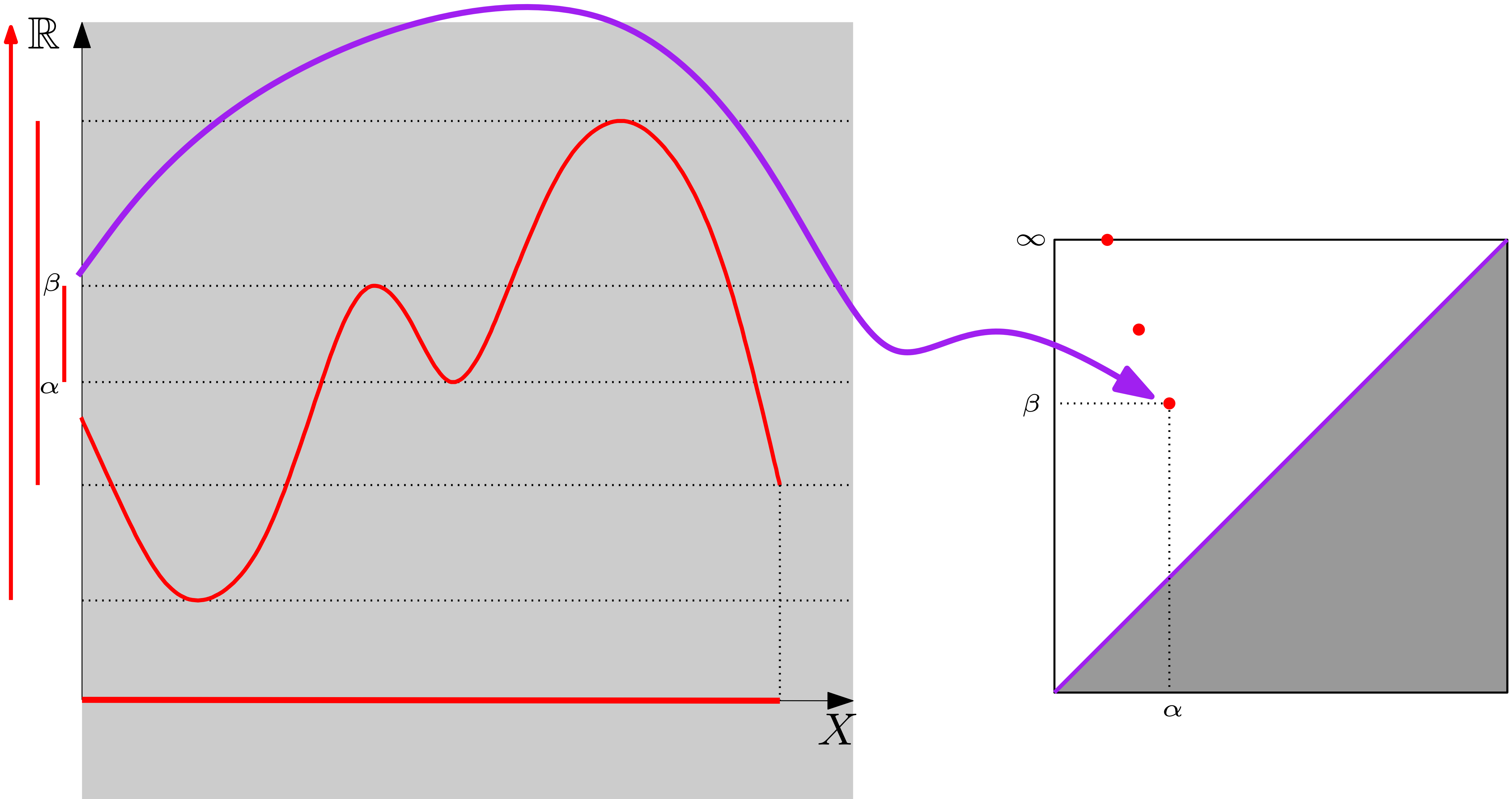
Persistent homology for functions

- The family of sublevel sets of a function is an example of filtration.



Persistent homology for functions

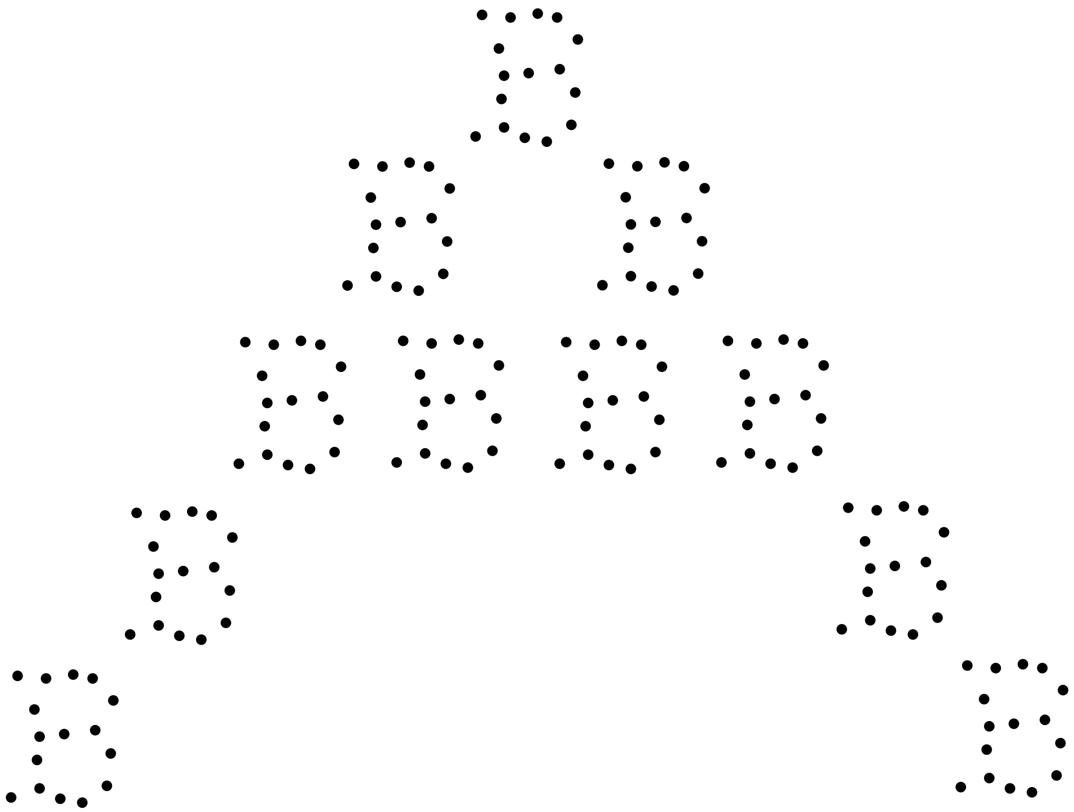
- Tracking and encoding the evolution of the connected components (0-dimensional homology) of the sublevel sets of a function
- The family of sublevel sets of a function is an example of filtration.

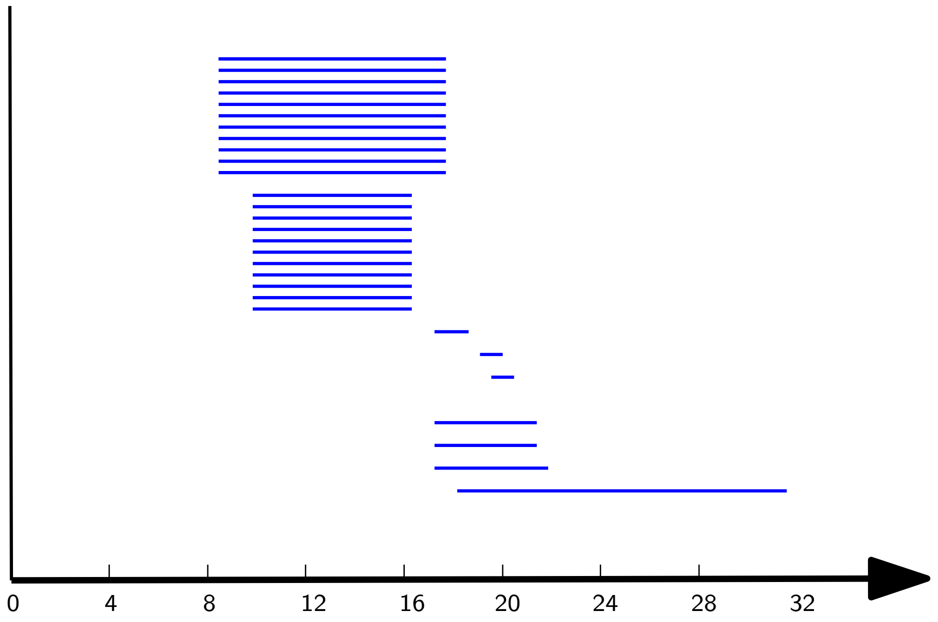


Persistent homology for functions

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

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



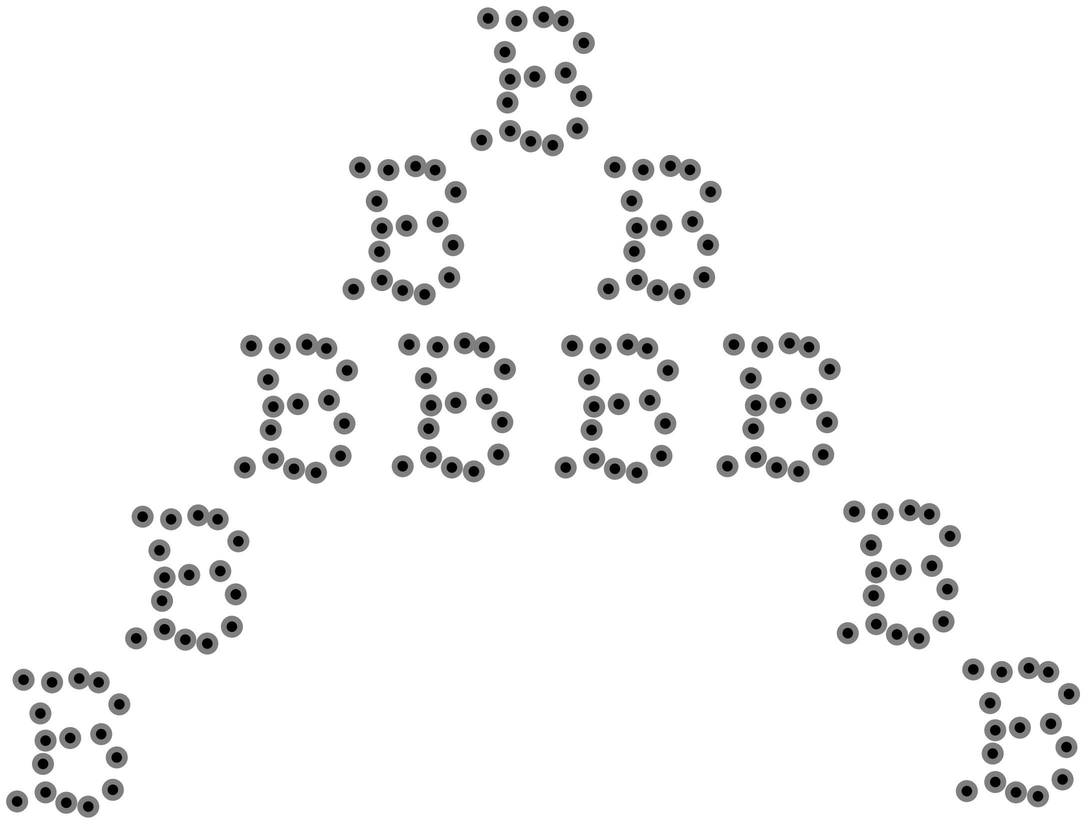


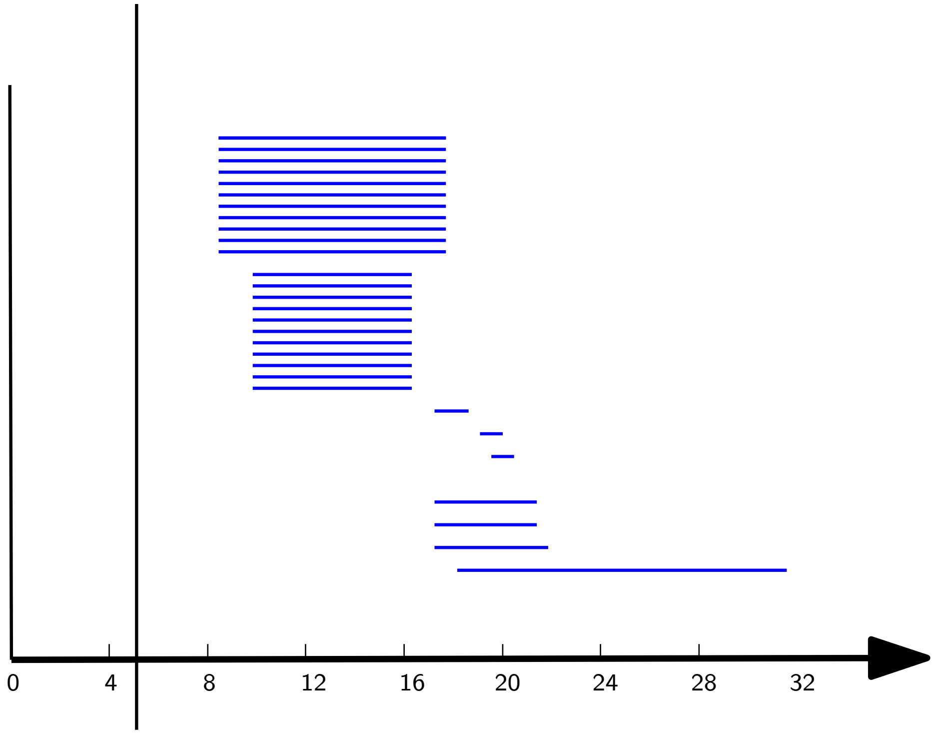
barcode for holes (1-d homology)

Persistent homology for functions

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

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



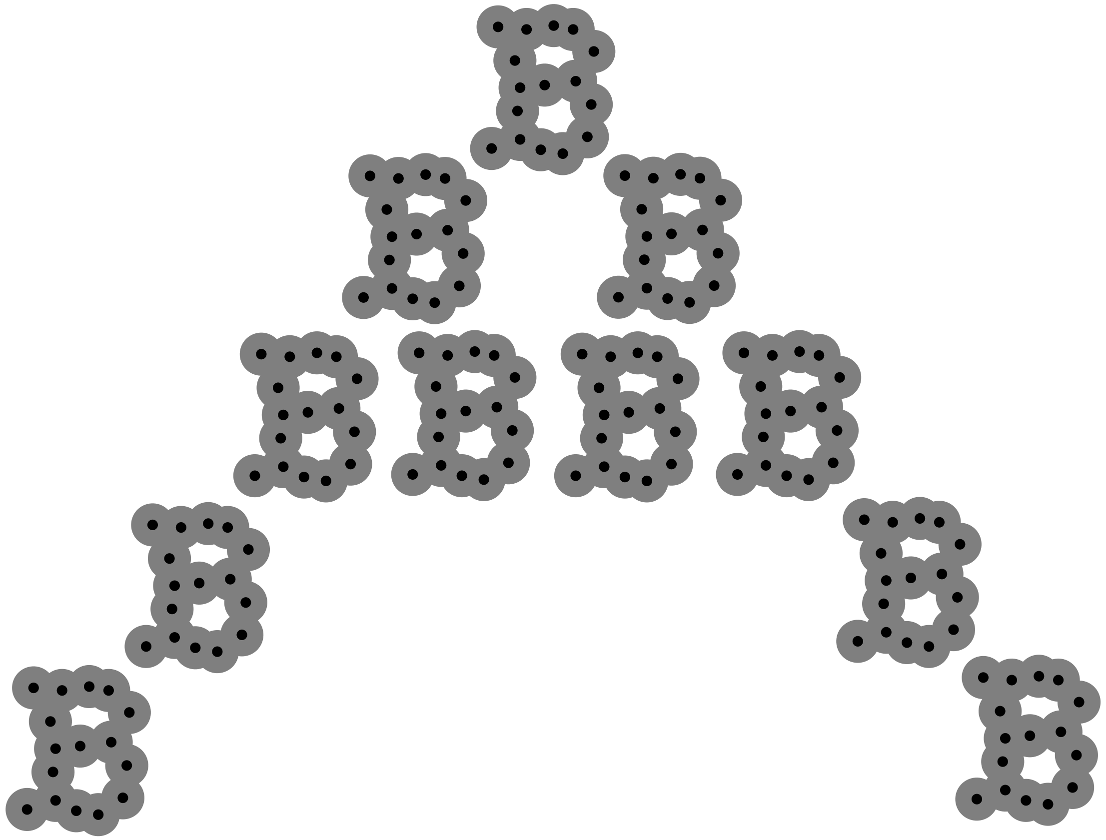


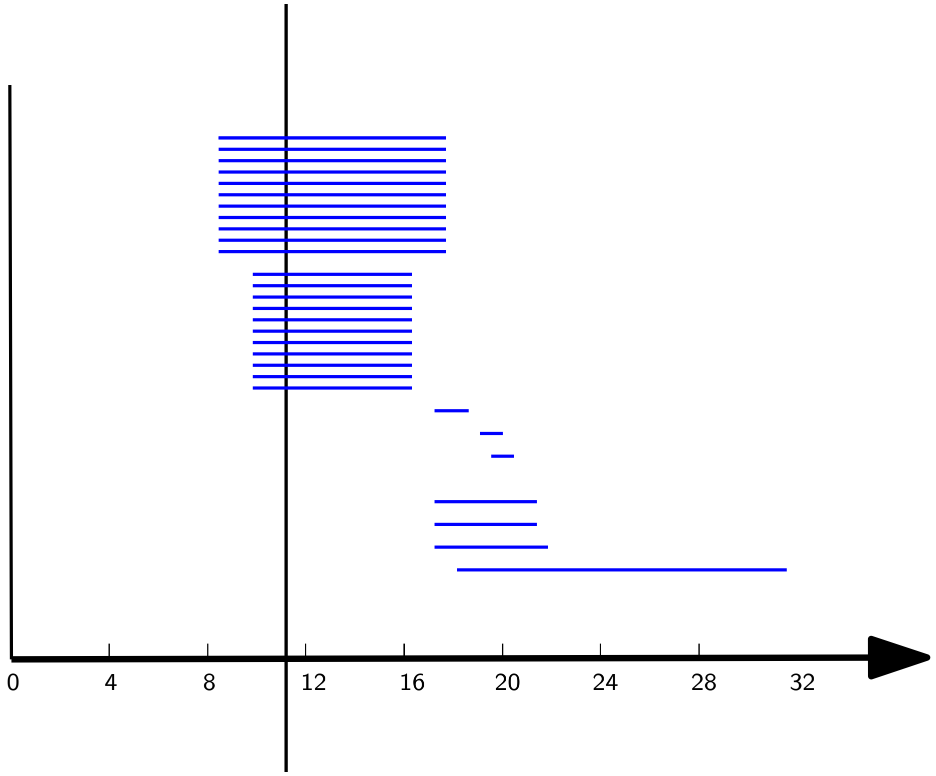
barcode for holes (1-d homology)

Persistent homology for functions

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

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



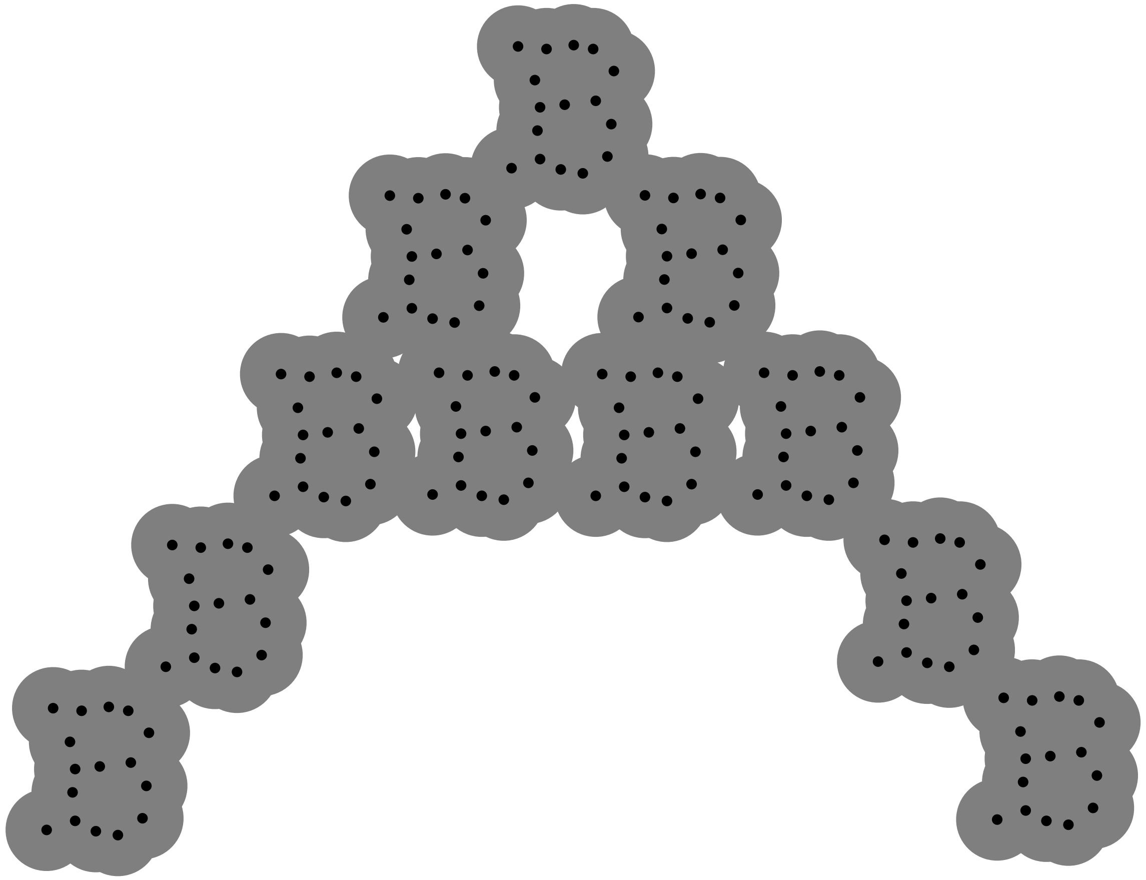


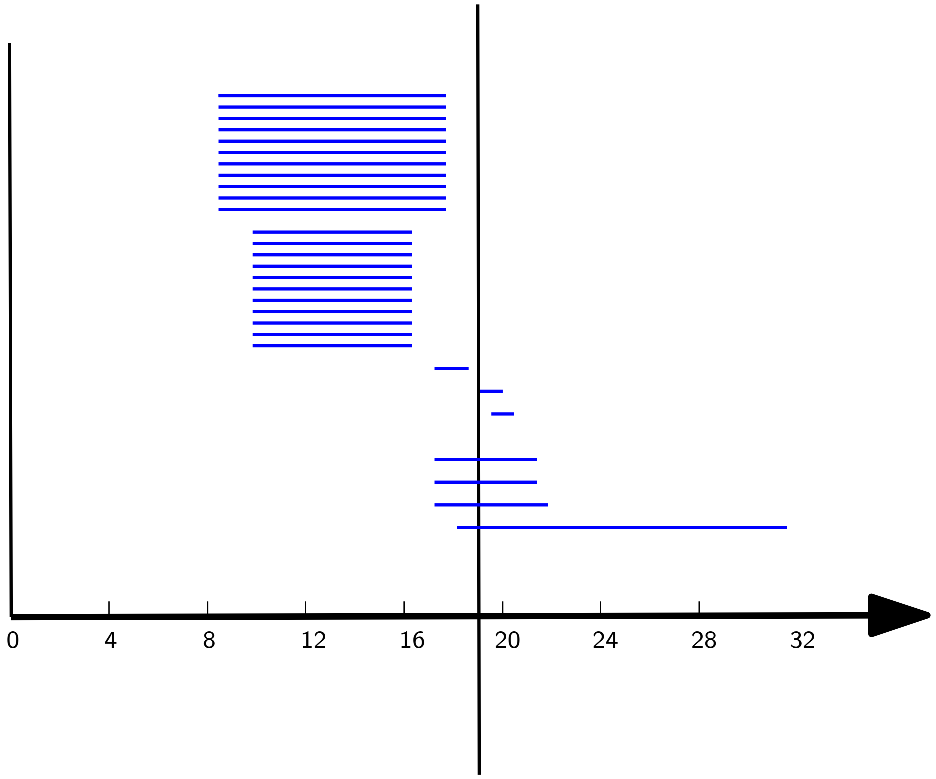
barcode for holes (1-d homology)

Persistent homology for functions

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

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



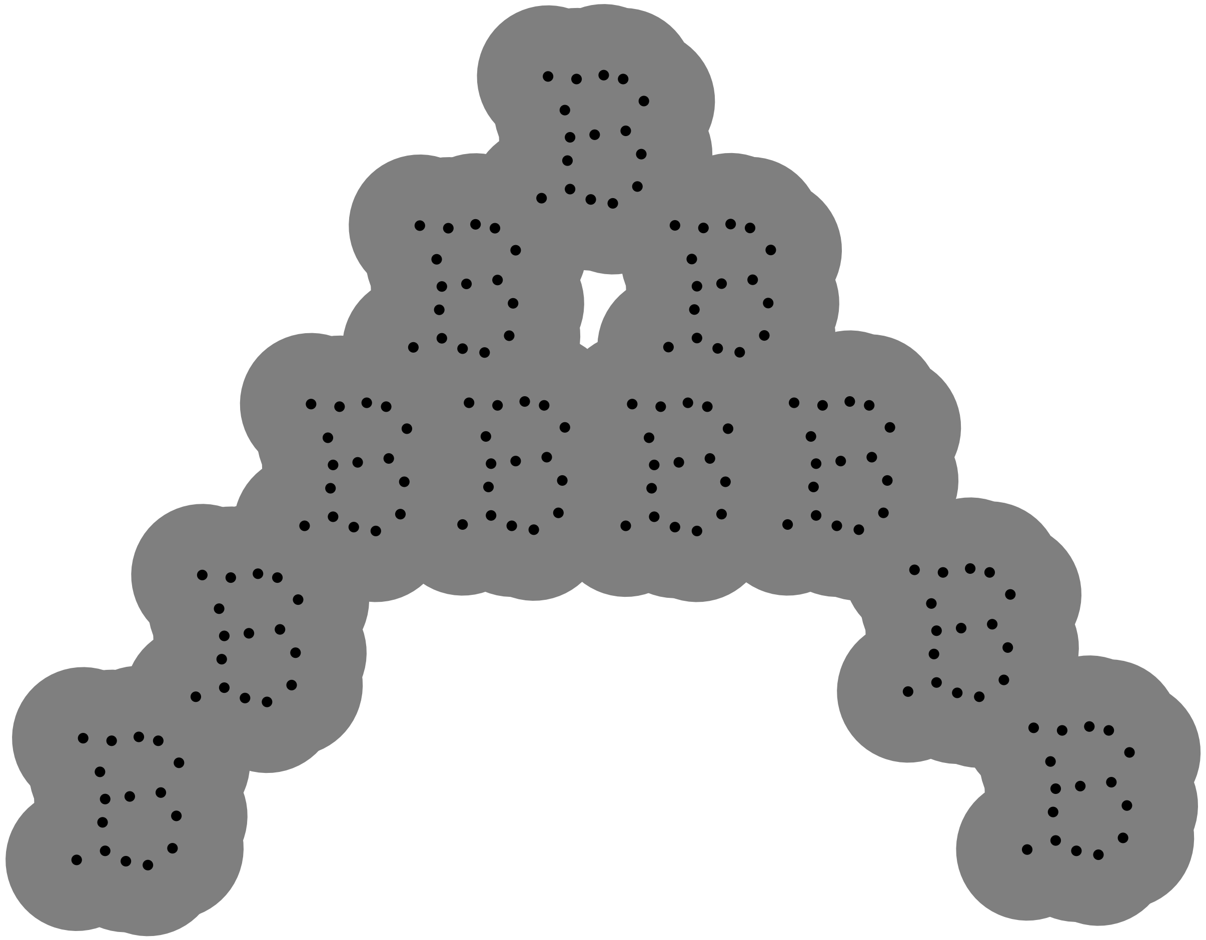


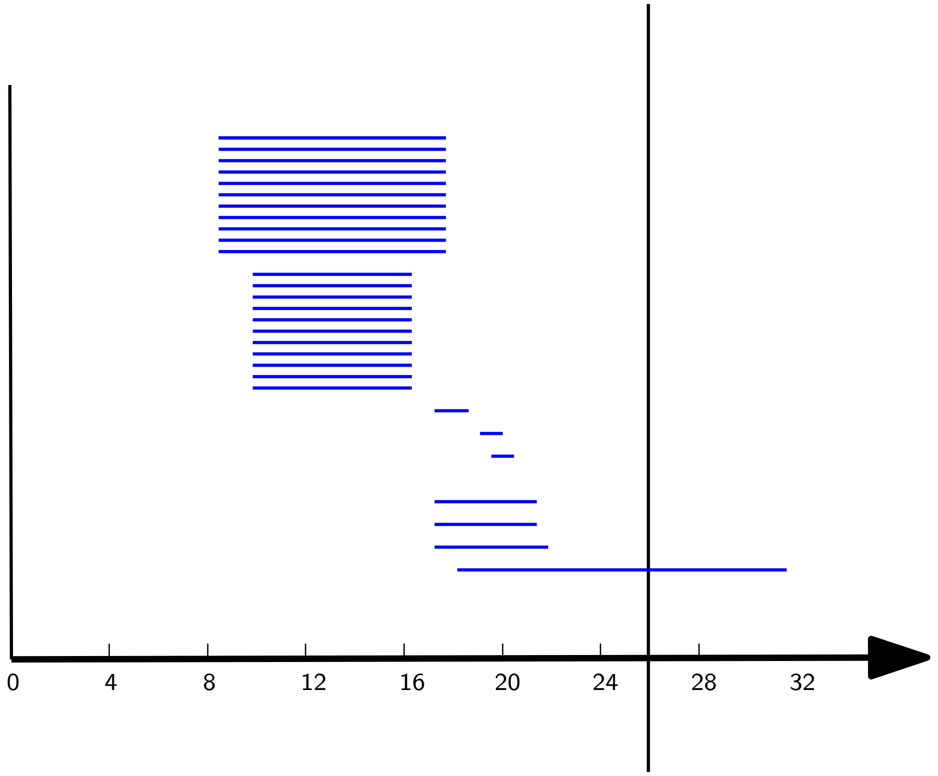
barcode for holes (1-d homology)

Persistent homology for functions

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

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



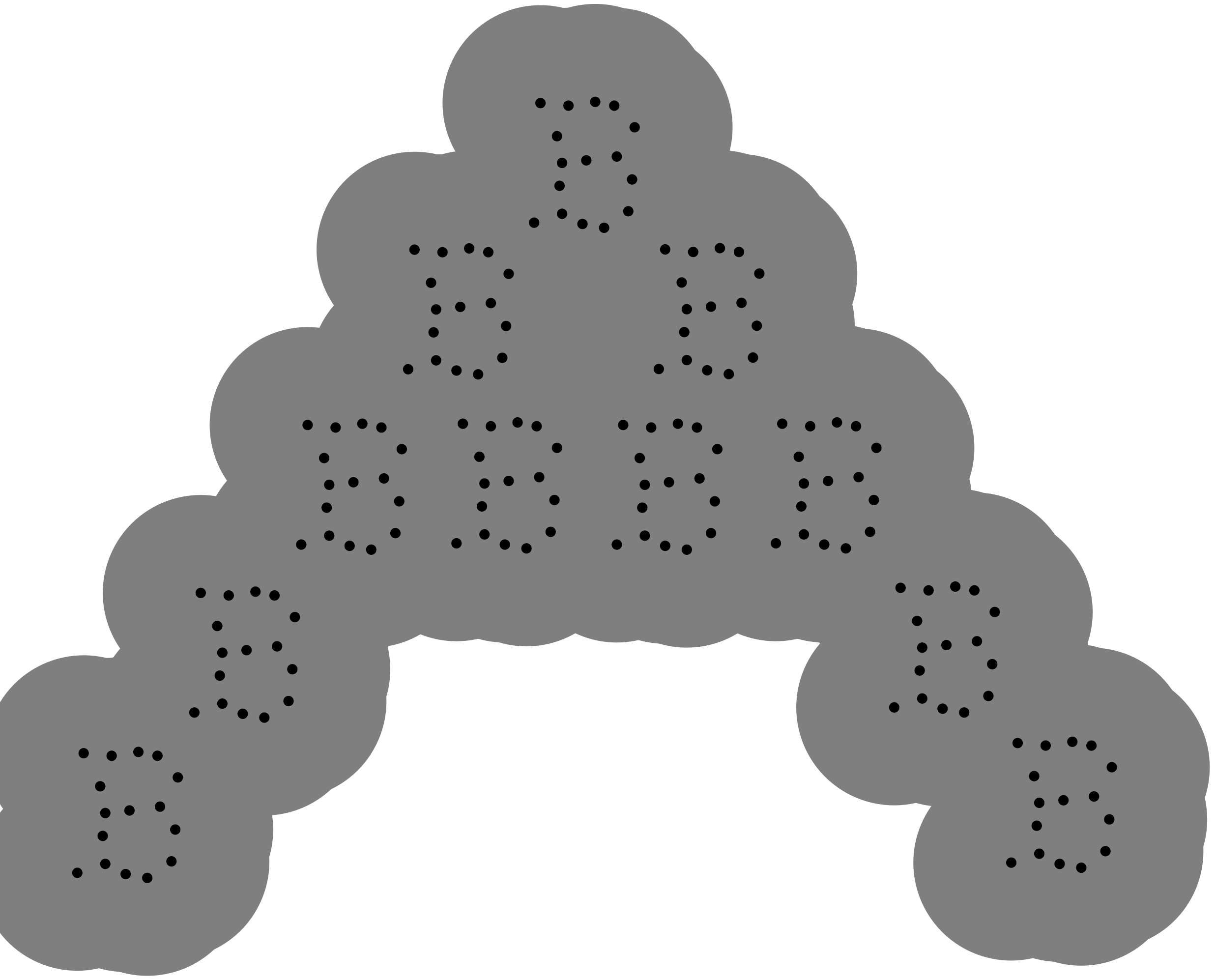


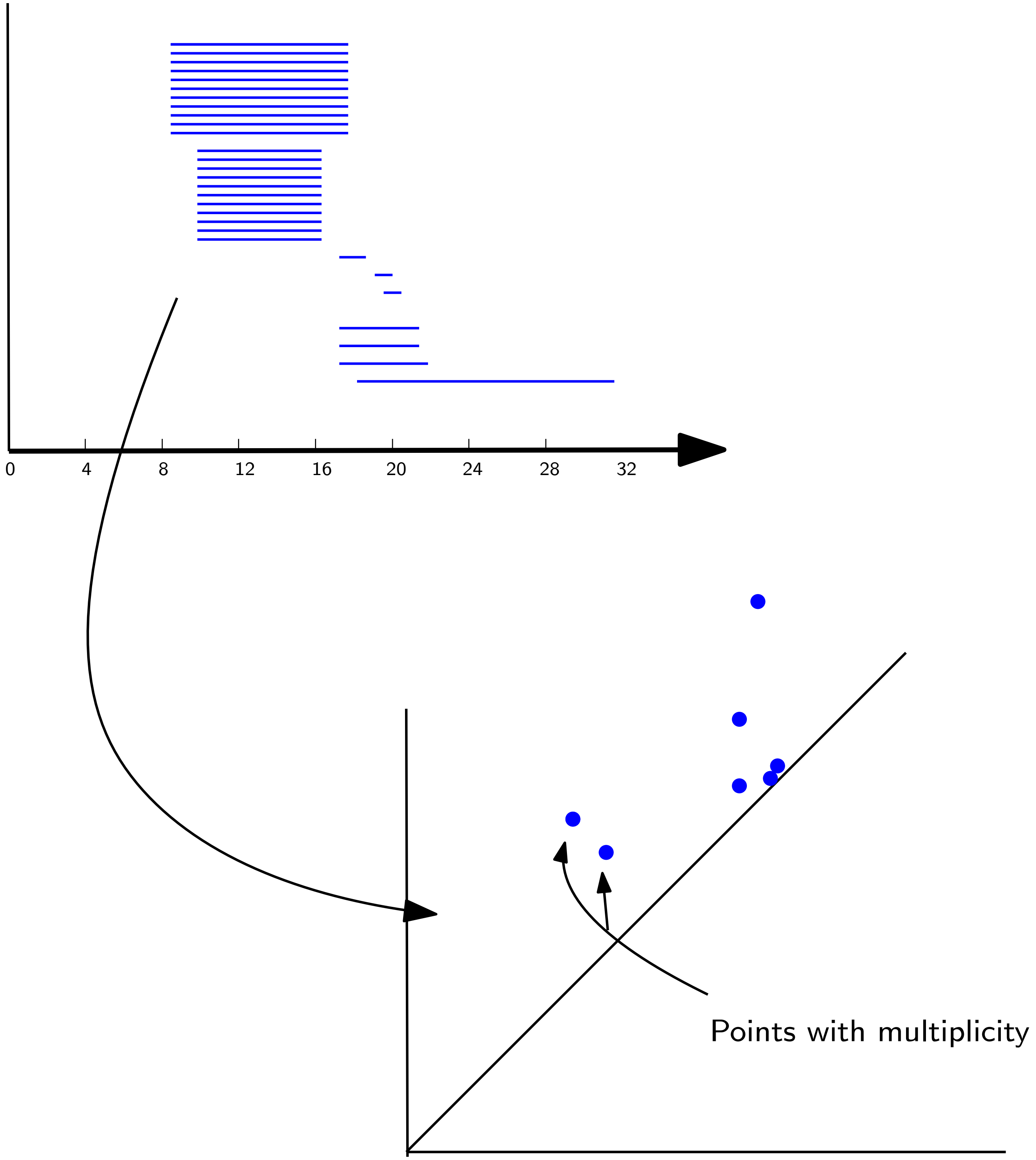
barcode for holes (1-d homology)

Persistent homology for functions

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

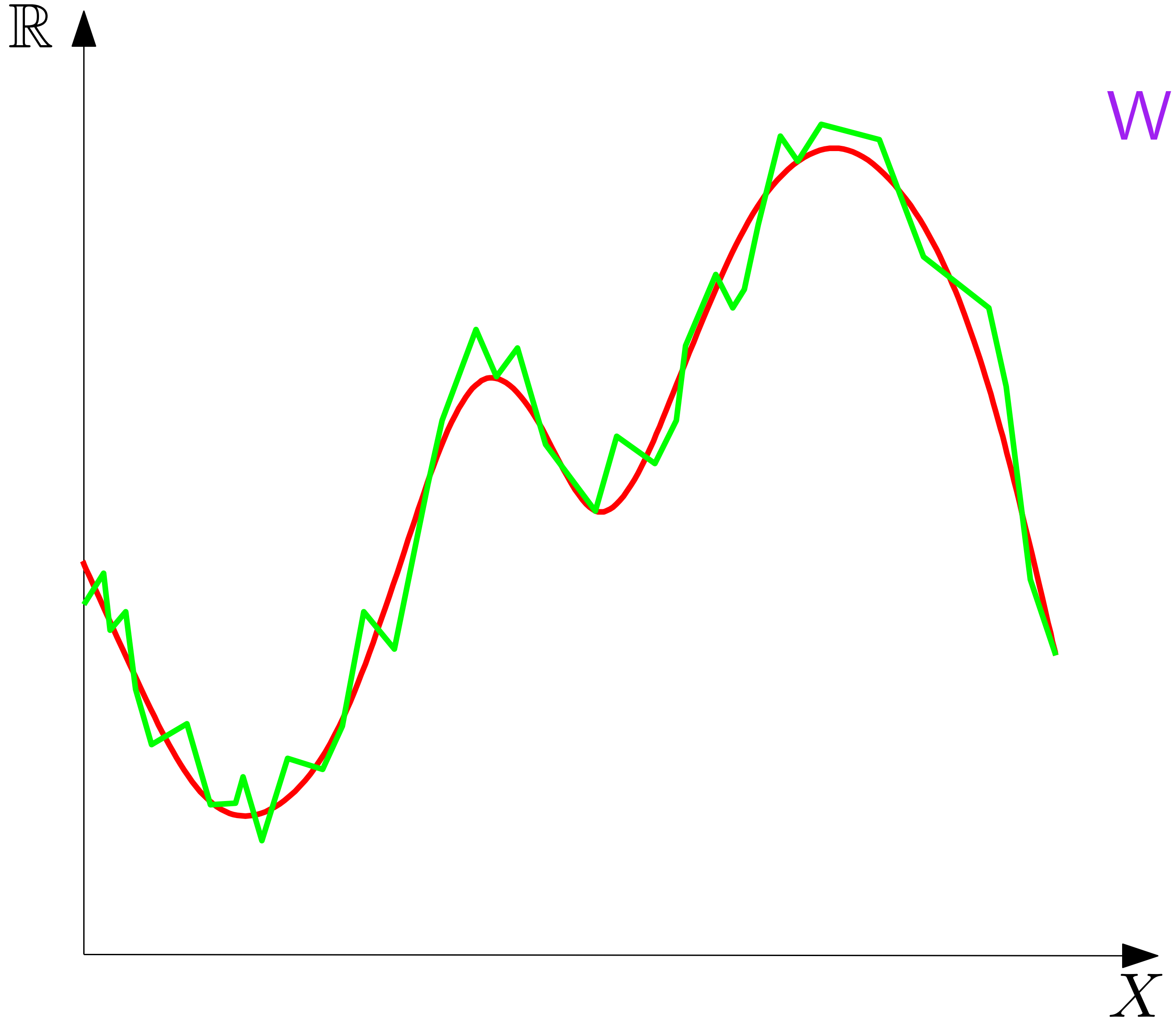
$$x \to \min_{p \in P} \|x - p\|_2$$



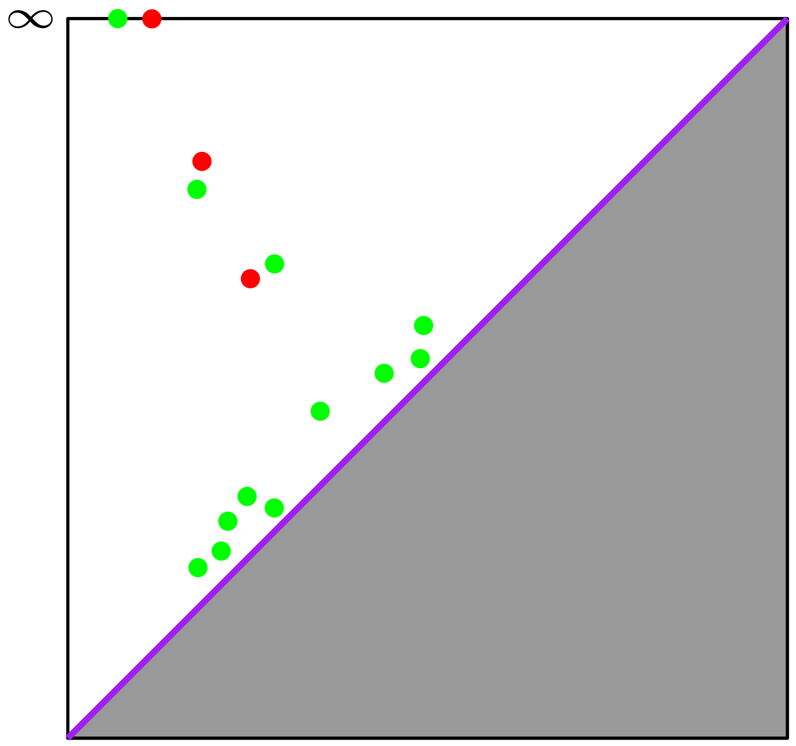


barcode for holes (1-d homology)

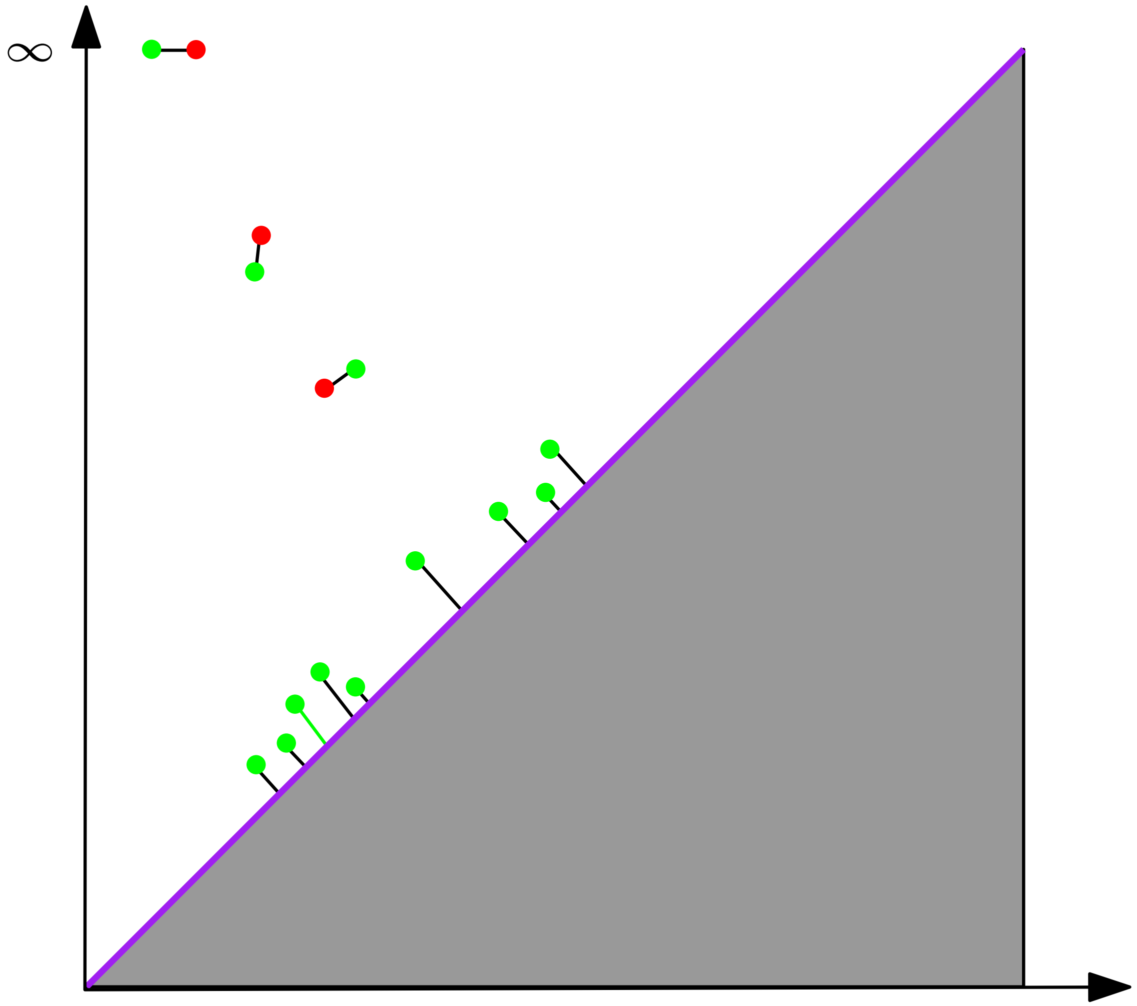
Stability properties



What if f is slightly perturbed?



Distance between persistence diagrams



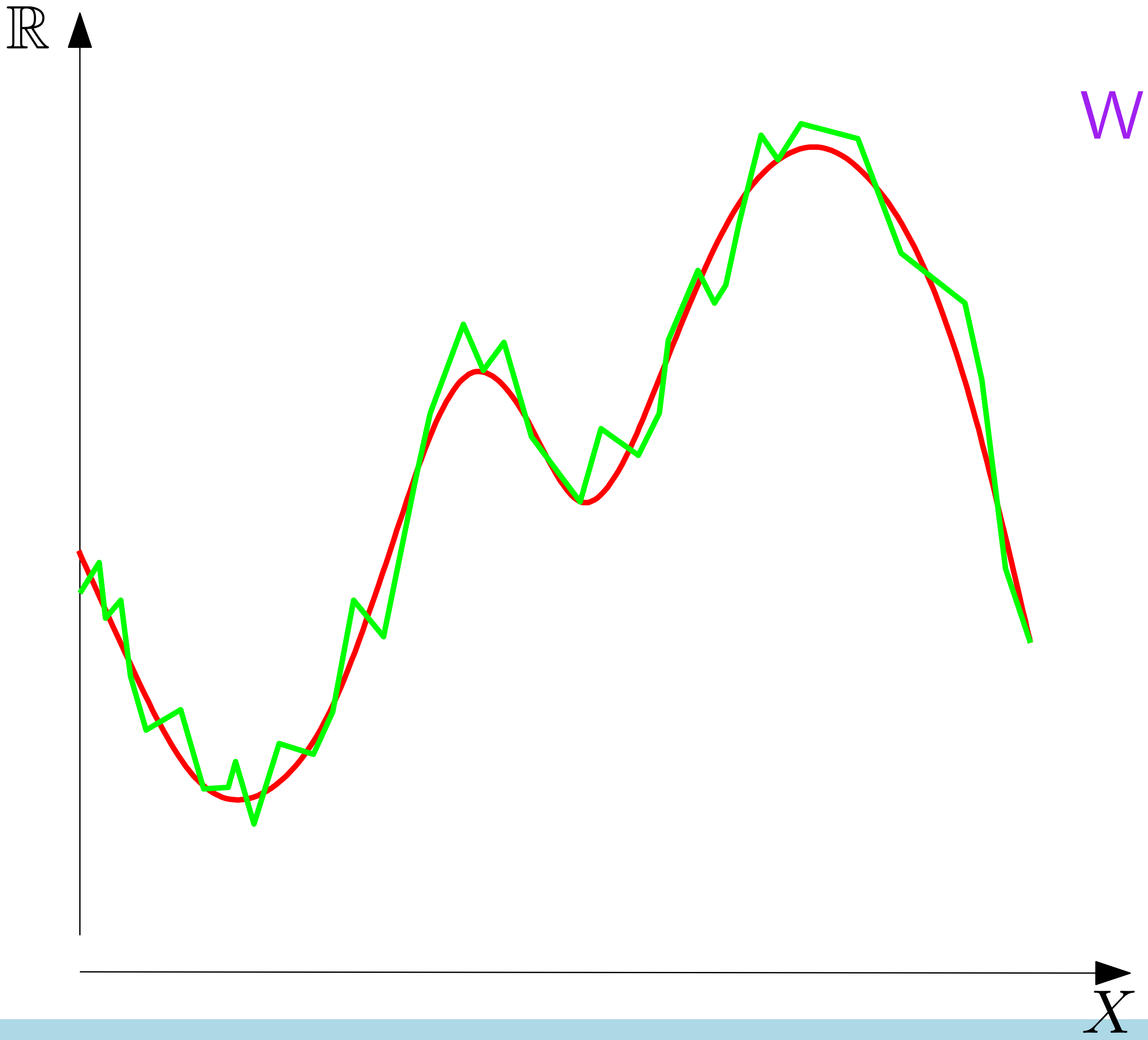
The bottleneck distance between two diagrams D_1 and D_2 is

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

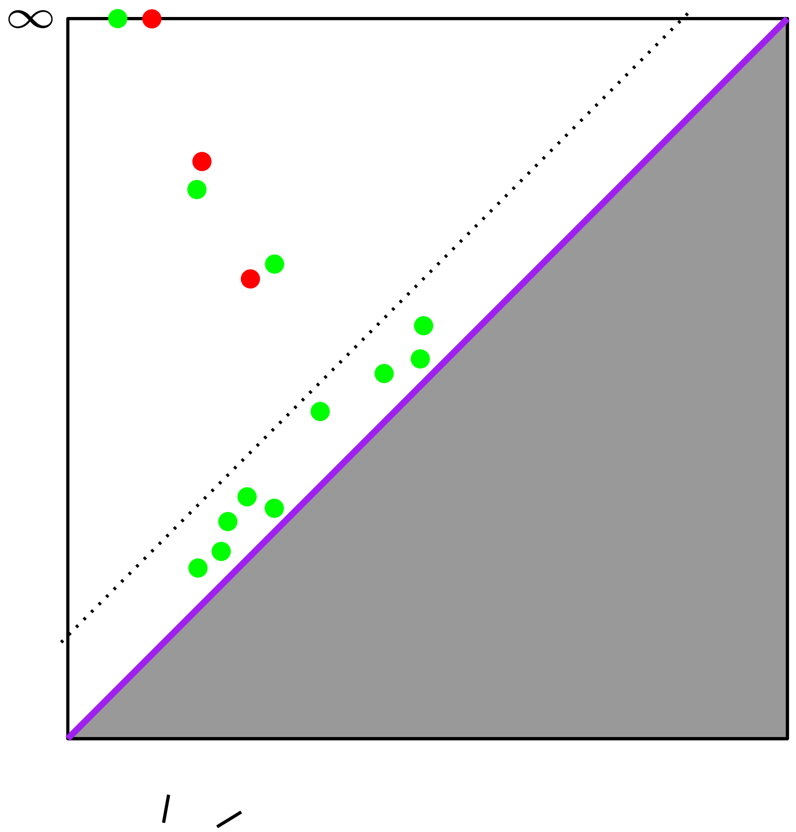
where Γ is the set of all the bijections between D_1 and D_2 and $\|p - q\|_\infty = \max(|x_p - x_q|, |y_p - y_q|)$.

Important Remark: There is one persistence diagram per homology dimension. In general, are compared diagrams corresponding to same homology dim.

Stability properties



What if f is slightly perturbed?



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

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

Important Remark: if $\phi : \mathbb{X} \to \mathbb{X}$ is an homeomorphism, then $D_{f \circ \phi} = D_f$.

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:

- $\forall t \leq t' \in \mathbb{R}$, $f^{-1}((-\infty, t]) \subseteq f^{-1}((-\infty, t'])$ → 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.

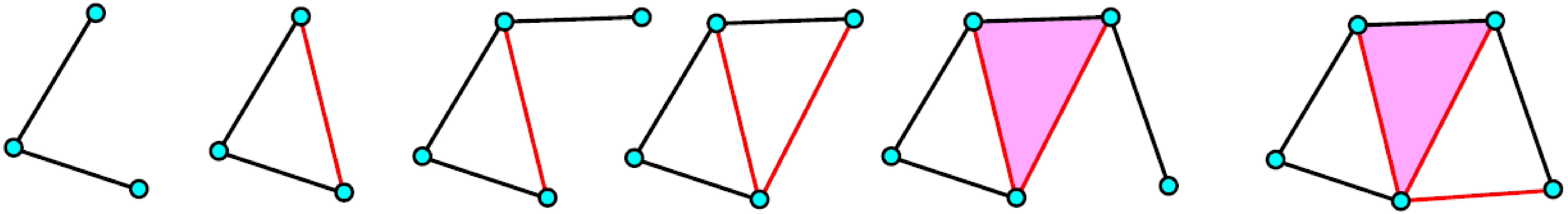
Algorithm to compute the Betti numbers $\beta_0, \beta_1, \cdots, \beta_d$ of K:

```
$\beta_0 = \beta_1 = \cdots = \beta_d = 0$;
for $i = 1$ to $m$
  $k = \dim \sigma^i - 1$;
  if $\sigma^i$ 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

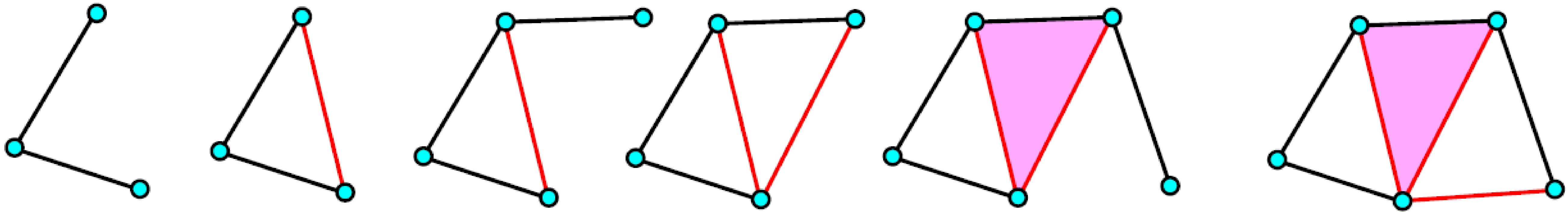


Lemma: If σ^i is a positive k-simplex, then there exists a k-cycle c_σ s.t.:
- c_σ is not a boundary in K^i,
- c_σ contains σ^i but no other positive k-simplex.

The cycle c^σ is unique.

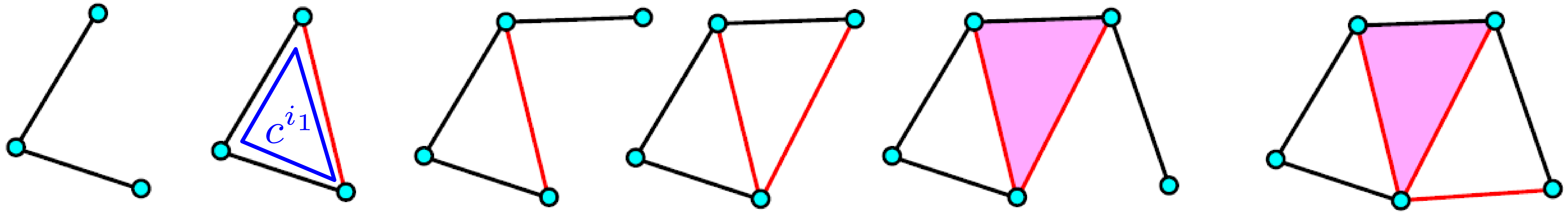
Proof:
By induction on the order of appearence of the simplices in the filtration.

Homology basis



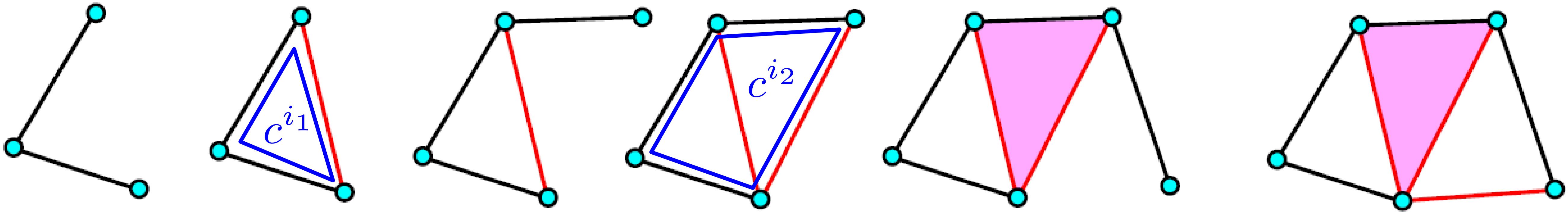
- At the beginning: the basis of H_k^0 is empty.
- If a basis of H_k^{i-1} has been built and σ^i is a positive k-simplex then one adds the homology class of the cycle c^i associated to σ^i to the basis of H_k^{i-1} $\Rightarrow$ basis of H_k^i.
- If a basis of H_k^{j-1} has been built and σ^j is a negative $(k+1)$-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



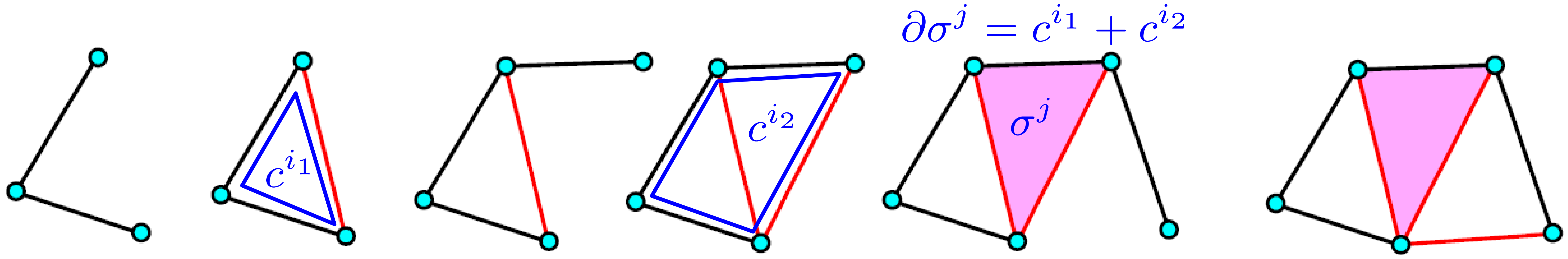
- At the beginning: the basis of H_k^0 is empty.
- If a basis of H_k^{i-1} has been built and σ^i is a positive k-simplex then one adds the homology class of the cycle c^i associated to σ^i to the basis of H_k^{i-1} $\Rightarrow$ basis of H_k^i.
- If a basis of H_k^{j-1} has been built and σ^j is a negative $(k+1)$-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 $(\sigma^{l(j)}, \sigma^j)$.
→ 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 $\sigma^j$ 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.

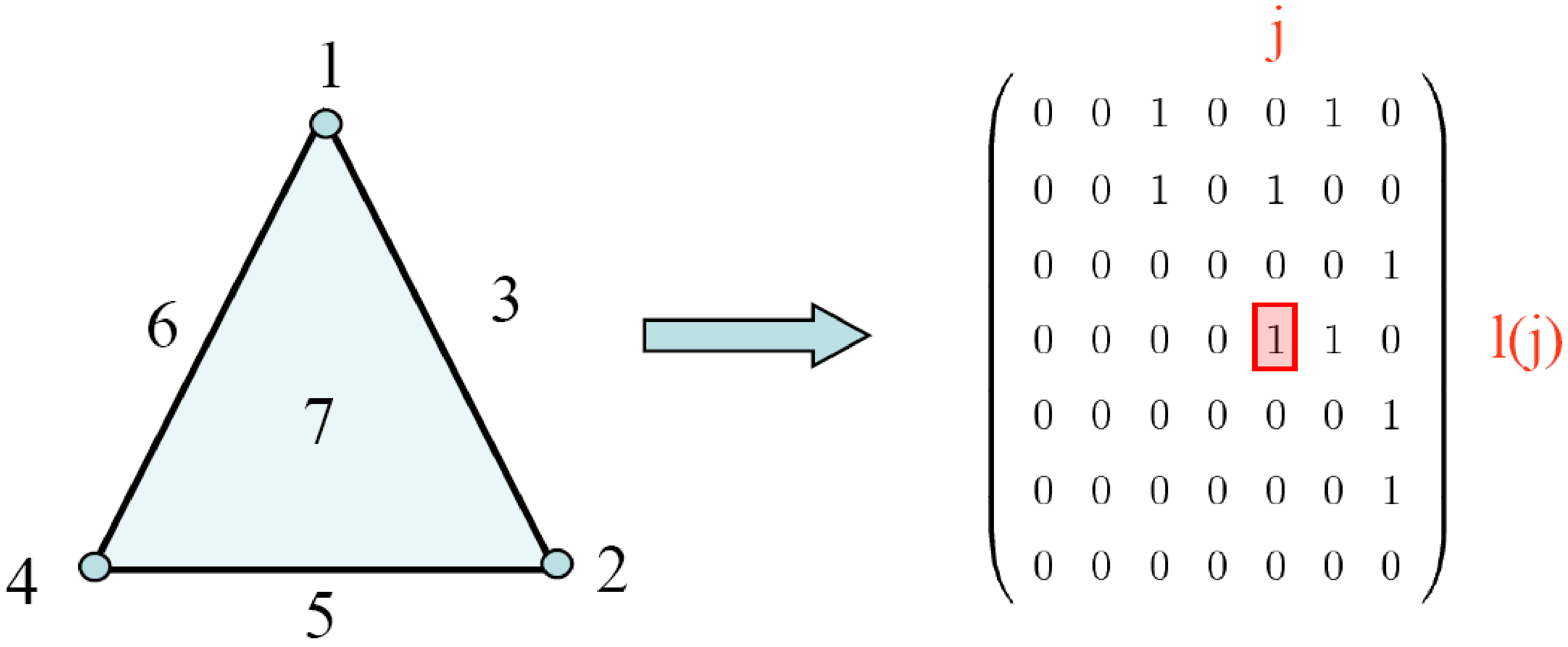
```
$L_0 = L_1 = \cdots = L_{d-1} = \emptyset$
For $j = 0$ to $m$
    $k = \dim \sigma^j - 1$;
    if $\sigma^j$ is a negative simplex
        $l(j)$ = highest index of the positive simplices associated to $\partial\sigma^j$;
        $L_k = L_k \cup \{(\sigma^{l(j)}, \sigma^j)\}$;
    end if
end for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



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

- $M = (m_{ij})_{i,j=1,\cdots,m}$ 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

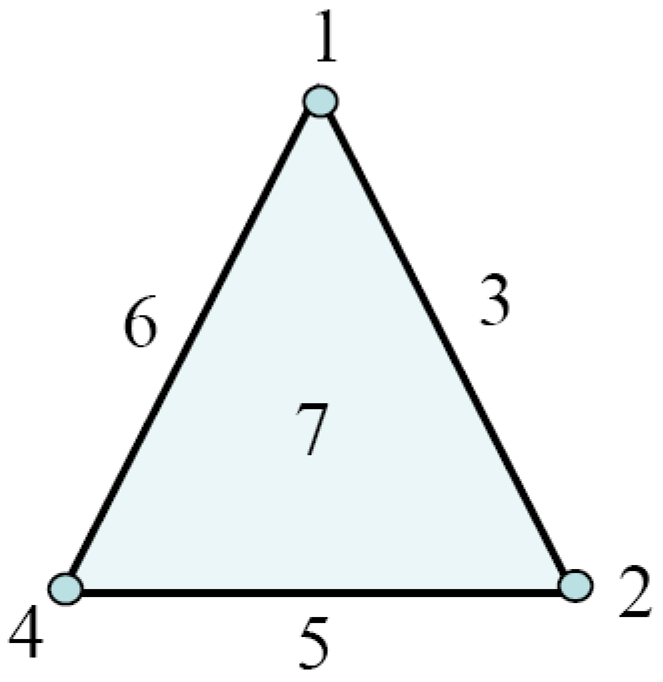
Input: $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ a d-dimensional filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.

```
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:



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

$$\Longrightarrow \begin{array}{c} \begin{array}{ccccccc} & & & & & \overset{C_5 + C_6}{\downarrow} & \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & \boxed{1} & 0 & 1 & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

$$\Longrightarrow \begin{array}{c} \begin{array}{ccccccc} & & & & & \overset{C_3 + C_6}{\downarrow} & \end{array} \\ \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{1} & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

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

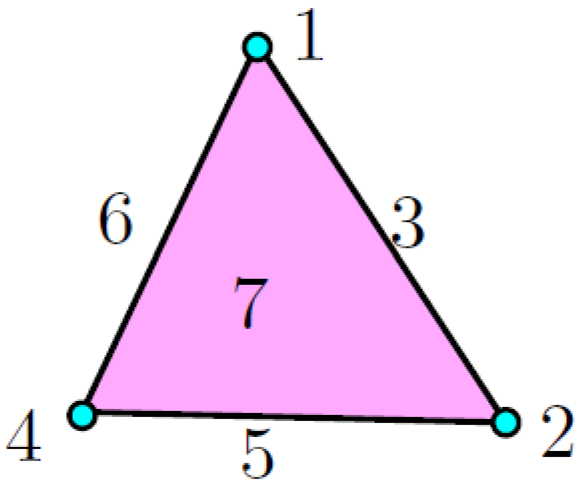
Correctness of the algorithm

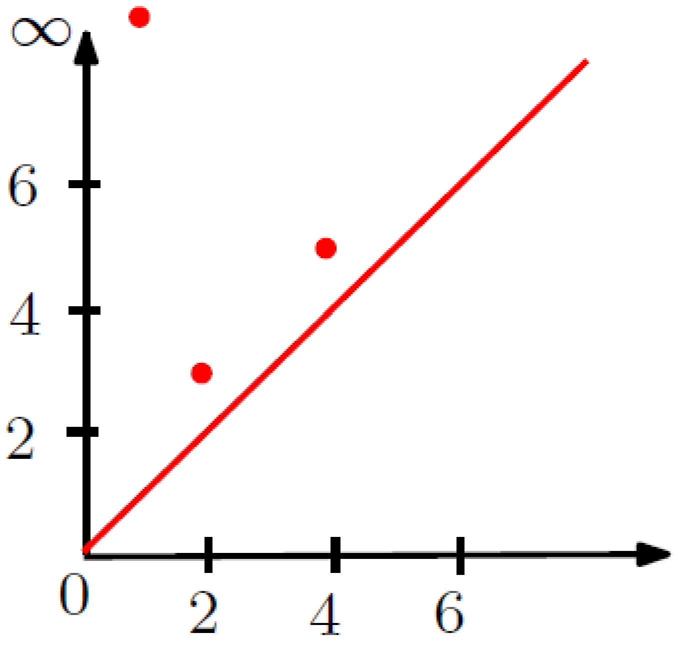
Proposition: the second algorithm (matrix version) outputs the persistence pairs.

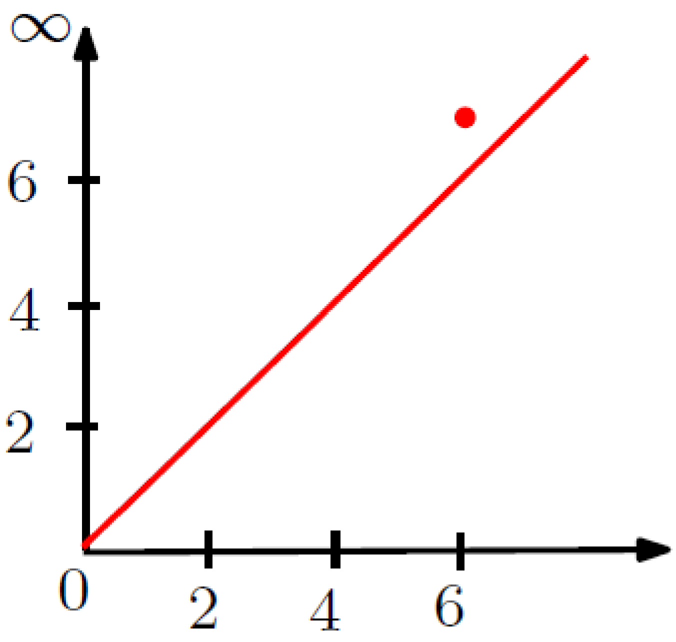
Proof: follows from the four remarks below.

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 the 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 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

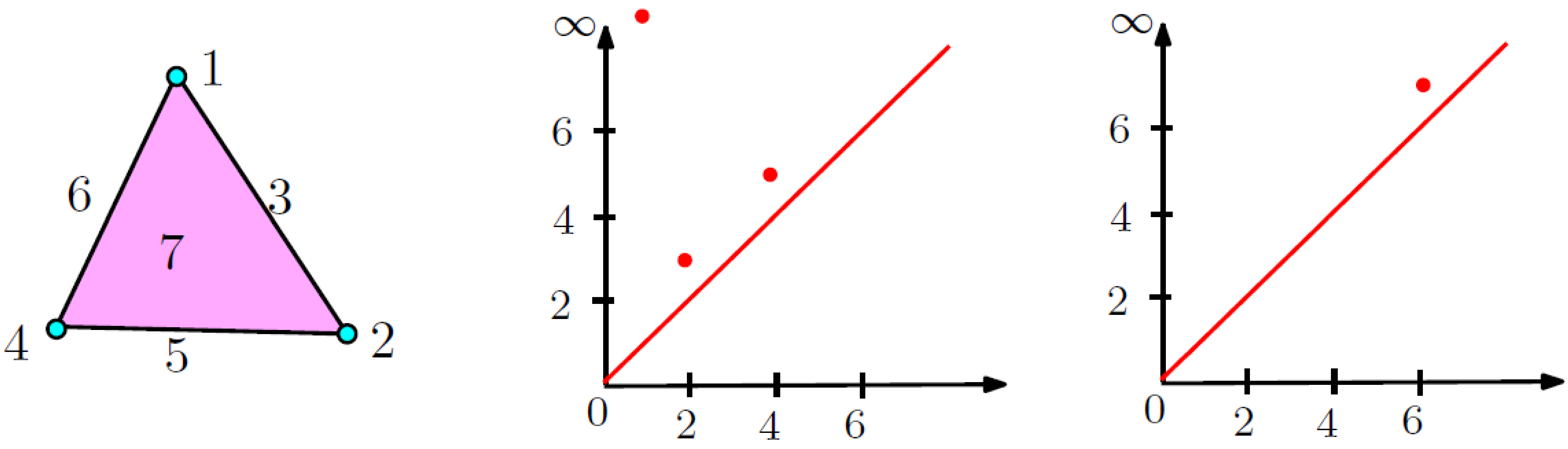






- each pair $(\sigma^{l(j)}, \sigma^j)$ is represented by $(l(j), j)$ or $(f(\sigma^{l(j)}), f(\sigma^j)) \in \mathbb{R}^2$ when considering filtrations induced by functions, or $(\alpha_{l(j)}, \alpha_j)$ if the filtration is indexed by a real valued sequence $(\alpha_i)_{i \in I}$.
- The diagonal $\{y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

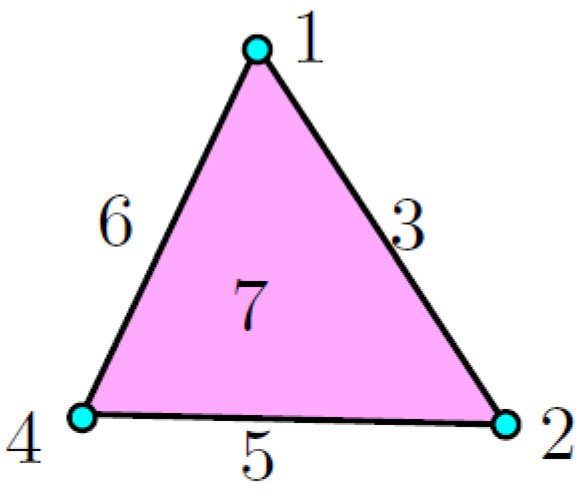
Persistence diagram

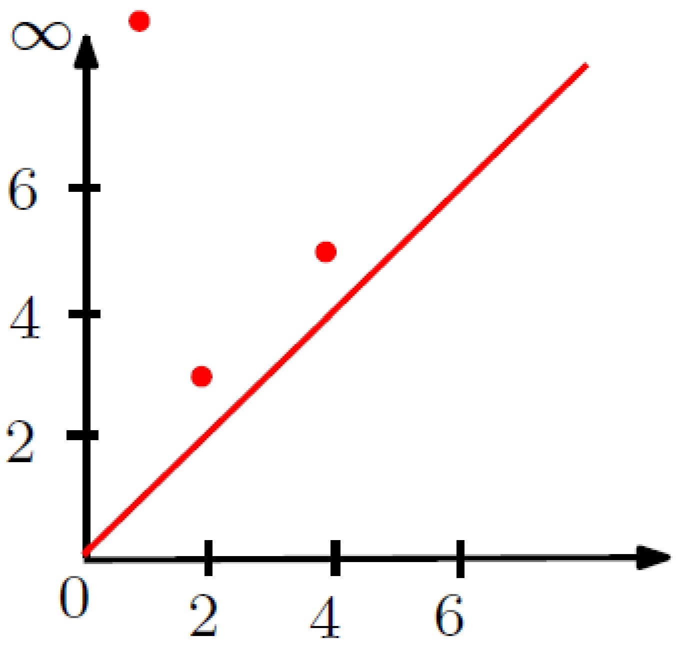


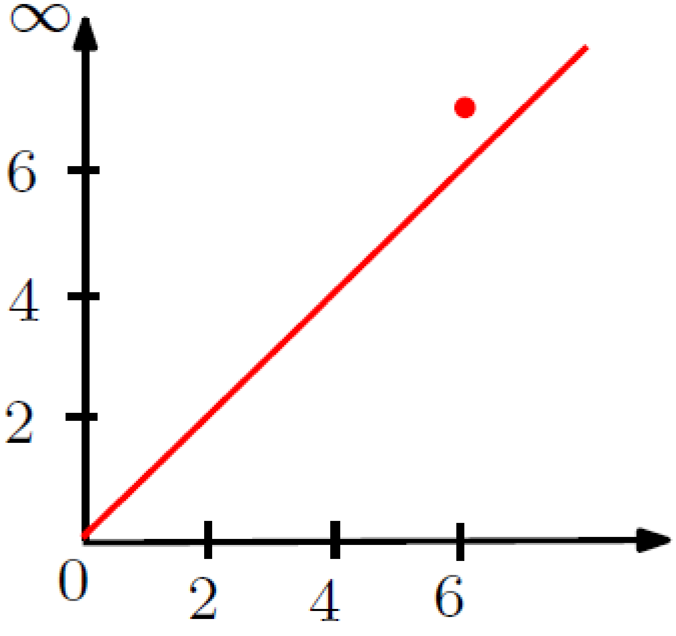
- each pair $(\sigma^{l(j)}, \sigma^j)$ is represented by $(l(j), j)$ or $(f(\sigma^{l(j)}), f(\sigma^j)) \in \mathbb{R}^2$ when considering filtrations induced by functions, or $(\alpha_{l(j)}, \alpha_j)$ if the filtration is indexed by a real valued sequence $(\alpha_i)_{i \in I}$.
- The diagonal $\{y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

Points may have multiplicity

Persistence diagram

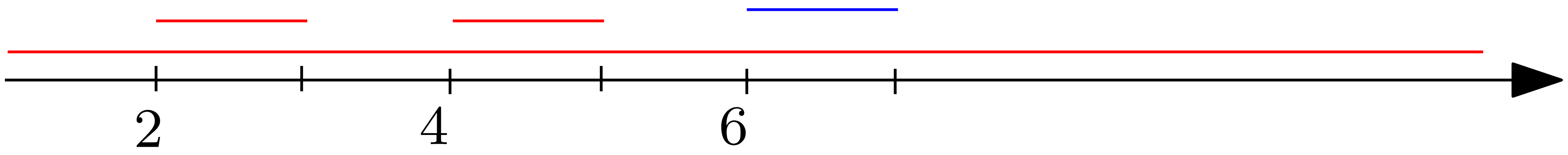




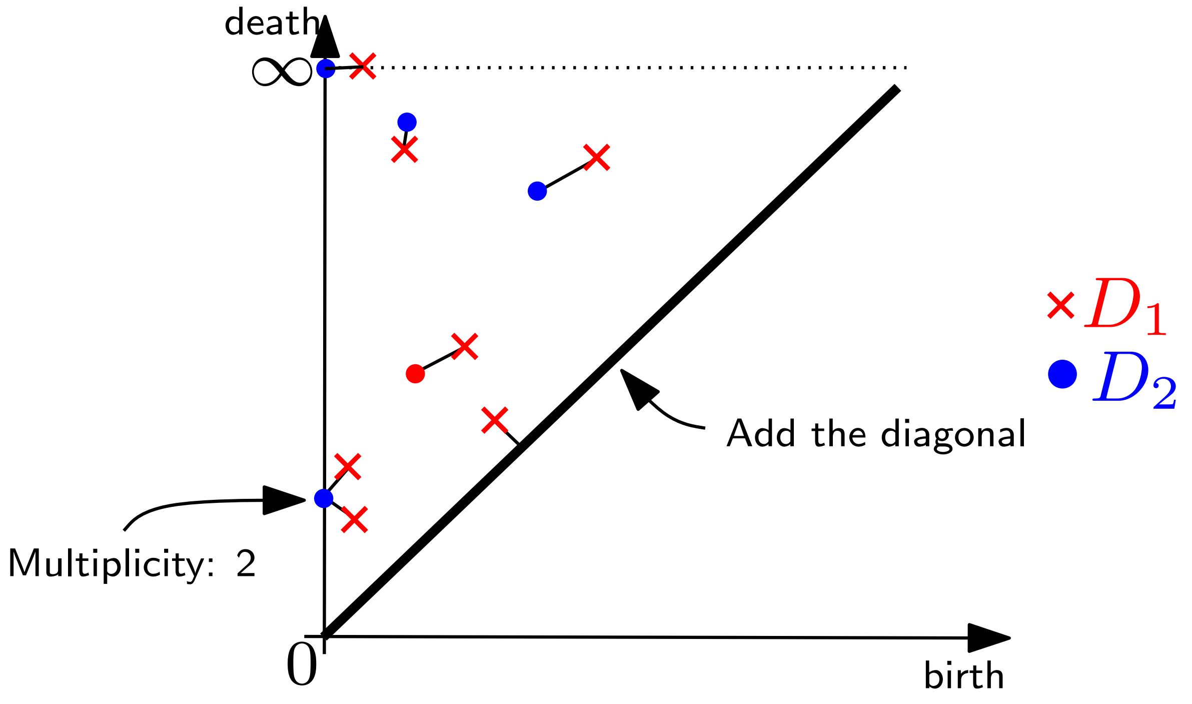


- each pair $(\sigma^{l(j)}, \sigma^j)$ is represented by $(l(j), j)$ or $(f(\sigma^{l(j)}), f(\sigma^j)) \in \mathbb{R}^2$ when considering filtrations induced by functions, or $(\alpha_{l(j)}, \alpha_j)$ if the filtration is indexed by a real valued sequence $(\alpha_i)_{i \in I}$.
- The diagonal $\{y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



Distances between persistence diagrams



The bottleneck distance between two diagrams D_1 and D_2 is

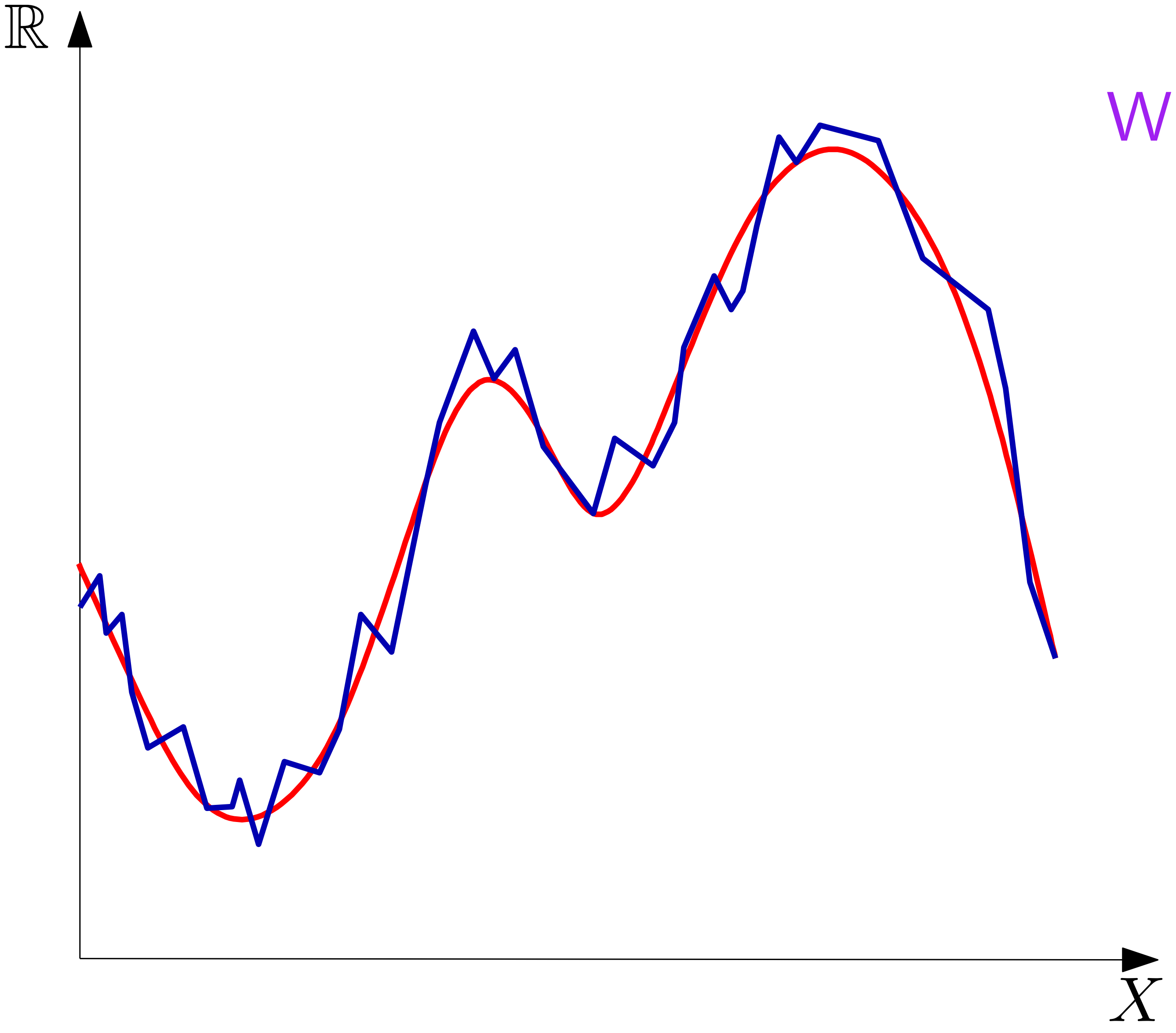
$$d_B(D_1, D_2) = \inf_{\gamma \in \Gamma} \sup_{p \in D_1} \|p - \gamma(p)\|_\infty$$

and the “p-Wasserstein” distance ($p \geq 1$) is

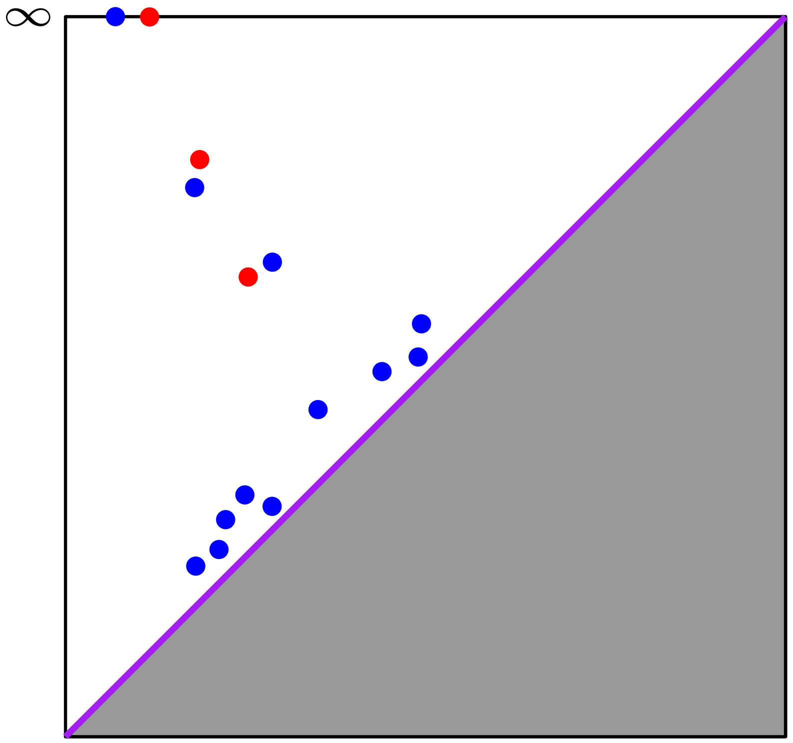
$$W_p(D_1, D_2) = \inf_{\gamma \in \Gamma} \left(\sum_{p \in D_1} \|p - \gamma(p)\|_p^p \right)^{\frac{1}{p}}$$

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



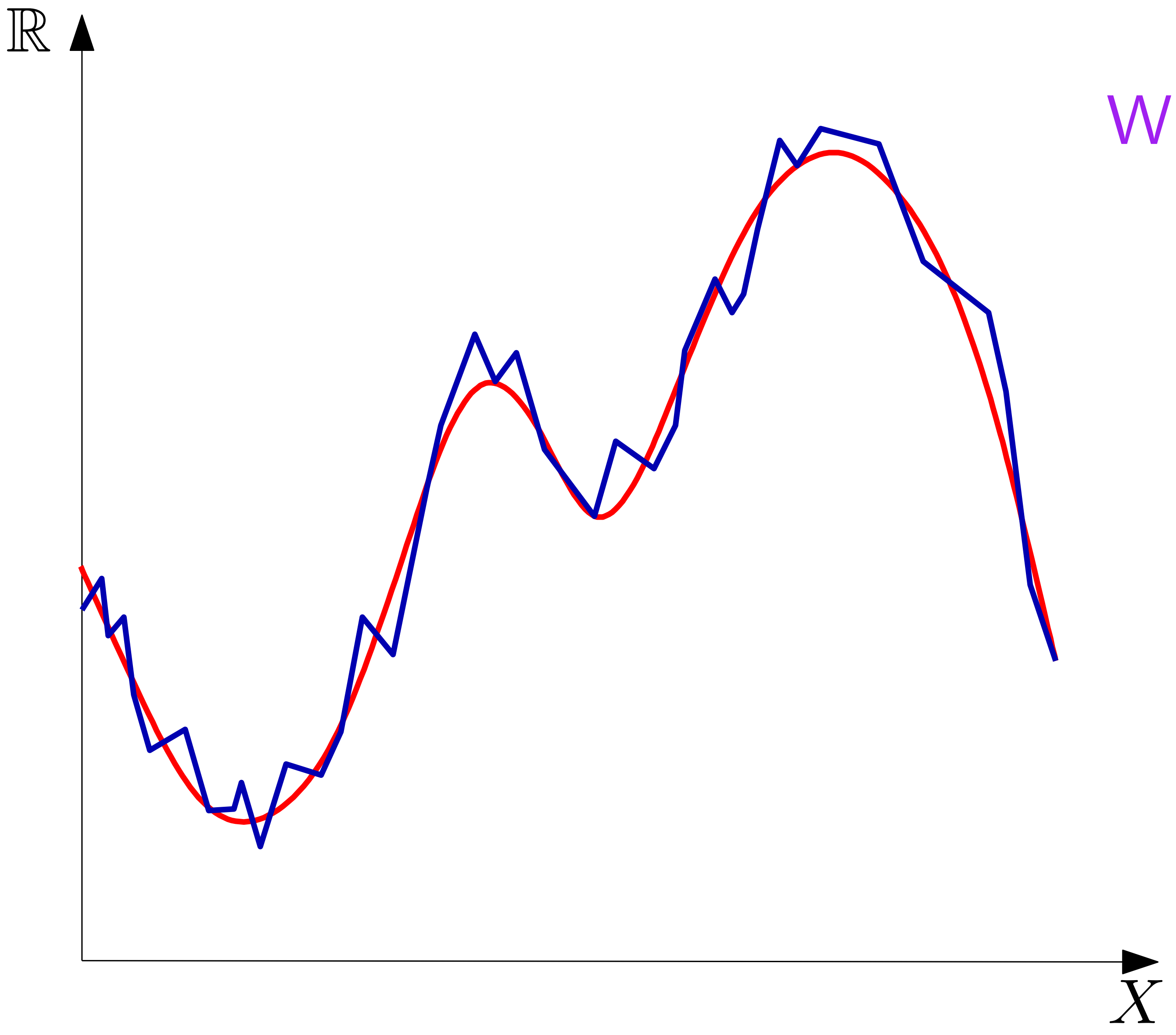
What if f is slightly perturbed?



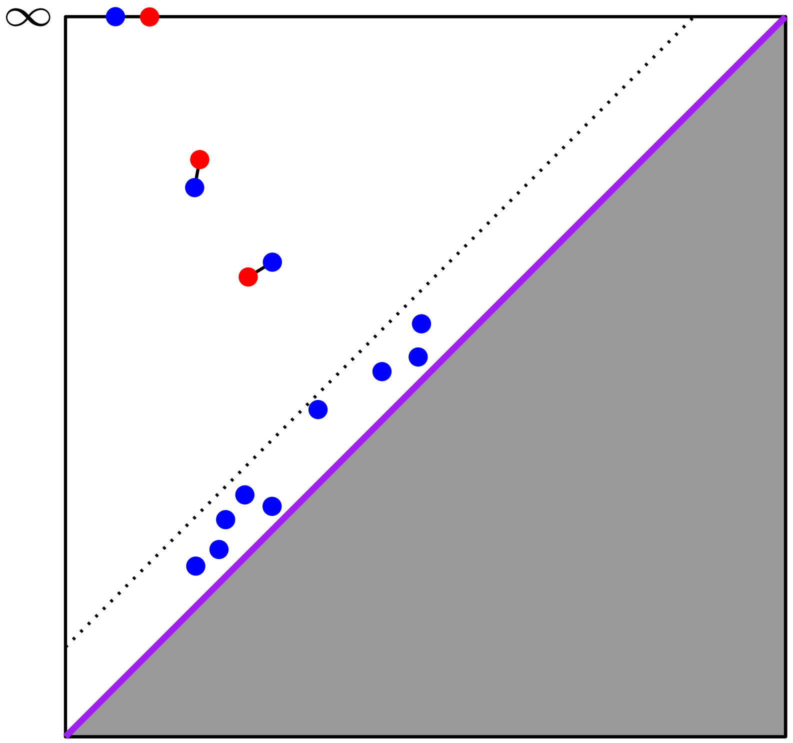
Stability properties

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]



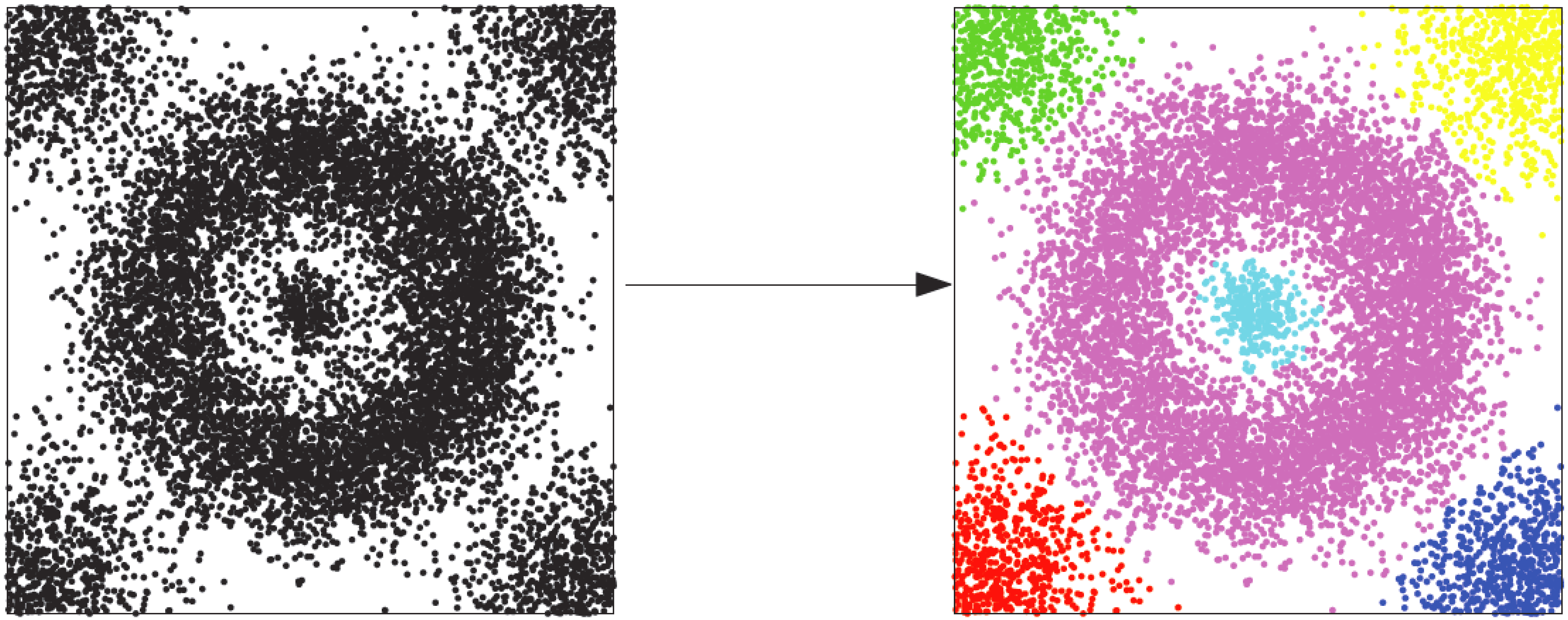
What if f is slightly perturbed?



Persistence-based clustering

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

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



Input:

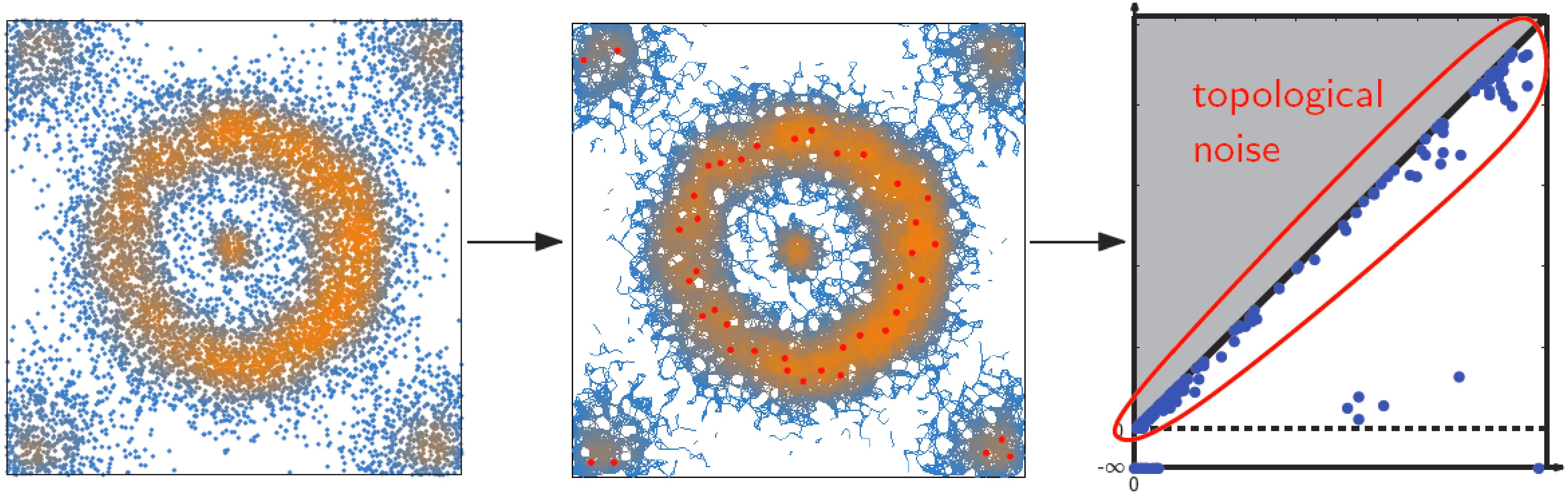
1. A finite set X of observations (point cloud with coordinates or pairwise distance matrix),
2. A real valued function f defined on the observations (e.g. density estimate).

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]



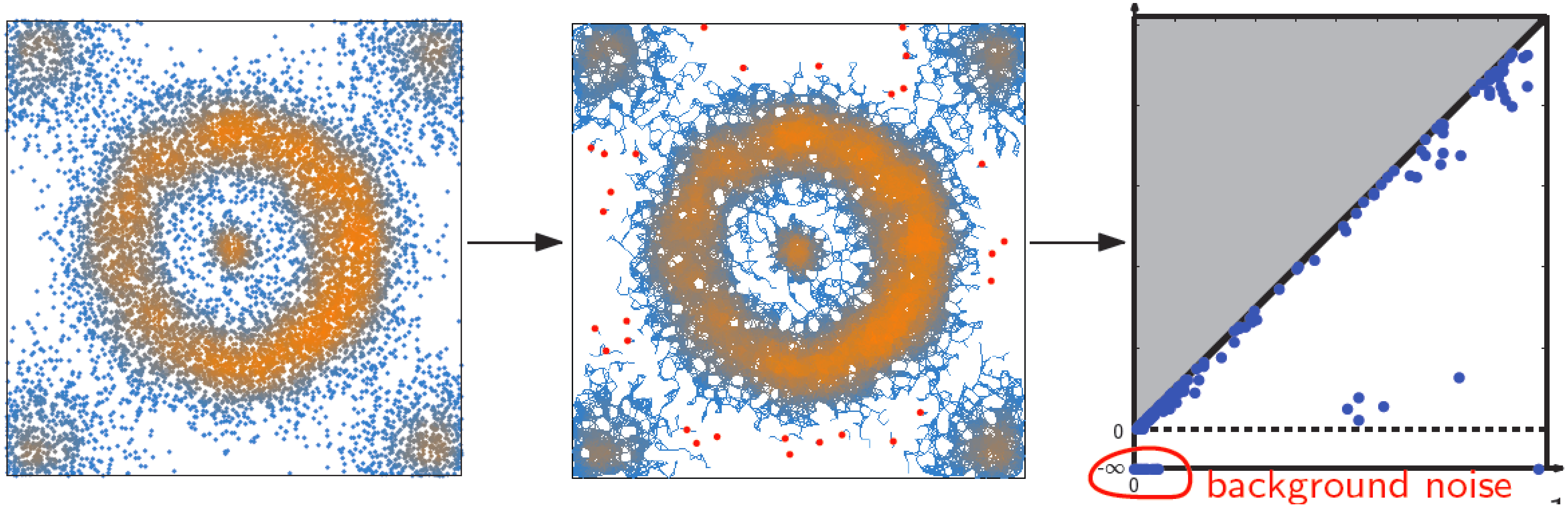
1. Build a neighborhing graph G on top of X.

2. Compute the (0-dim) persistence of f to identify prominent peaks → 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]



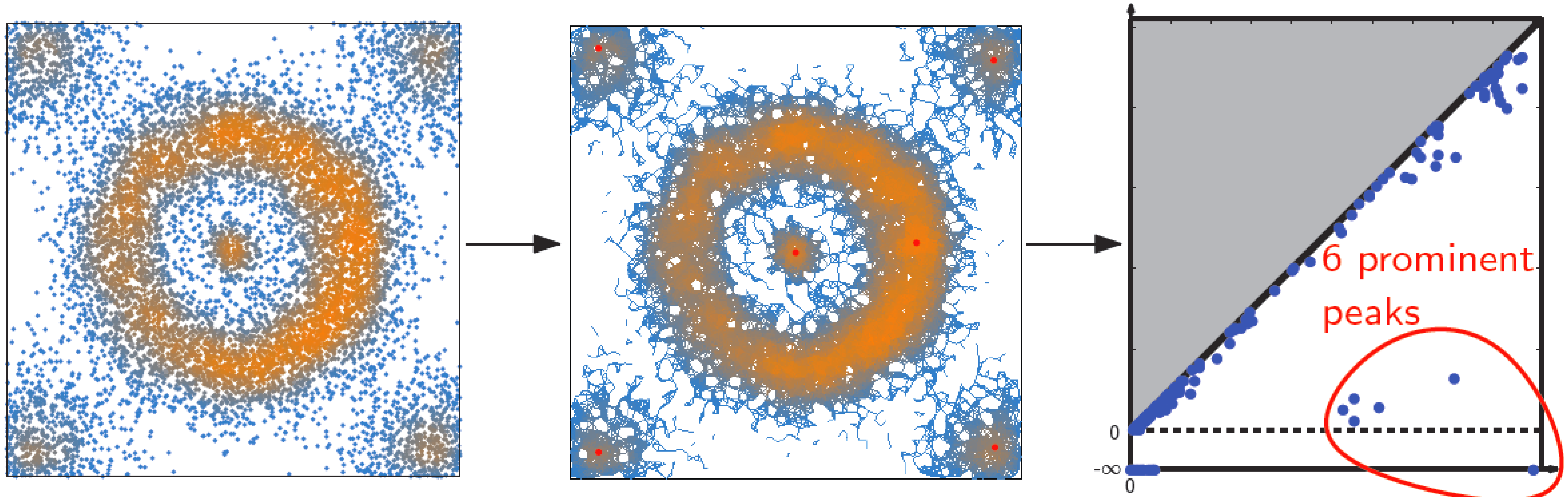
1. Build a neighborhing graph G on top of X.

2. Compute the (0-dim) persistence of f to identify prominent peaks → 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]



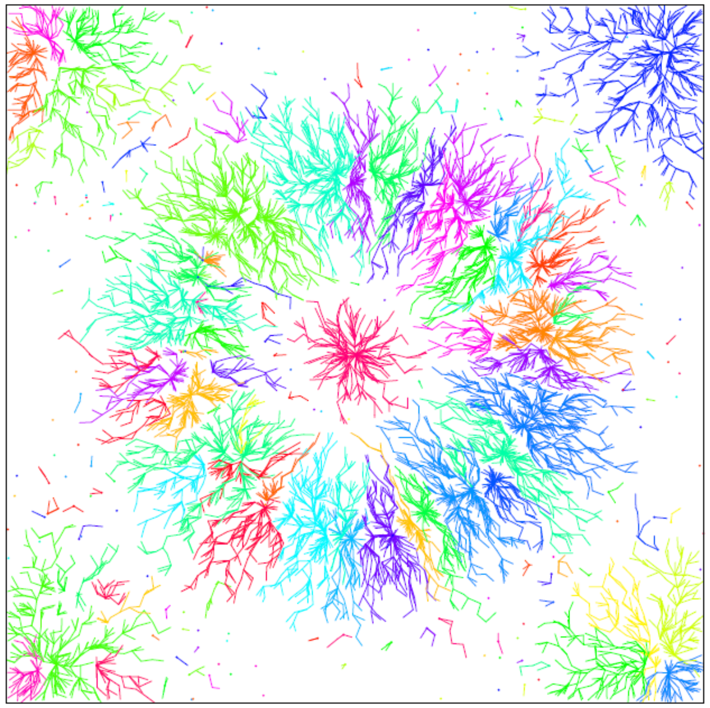
1. Build a neighborhing graph G on top of X.

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

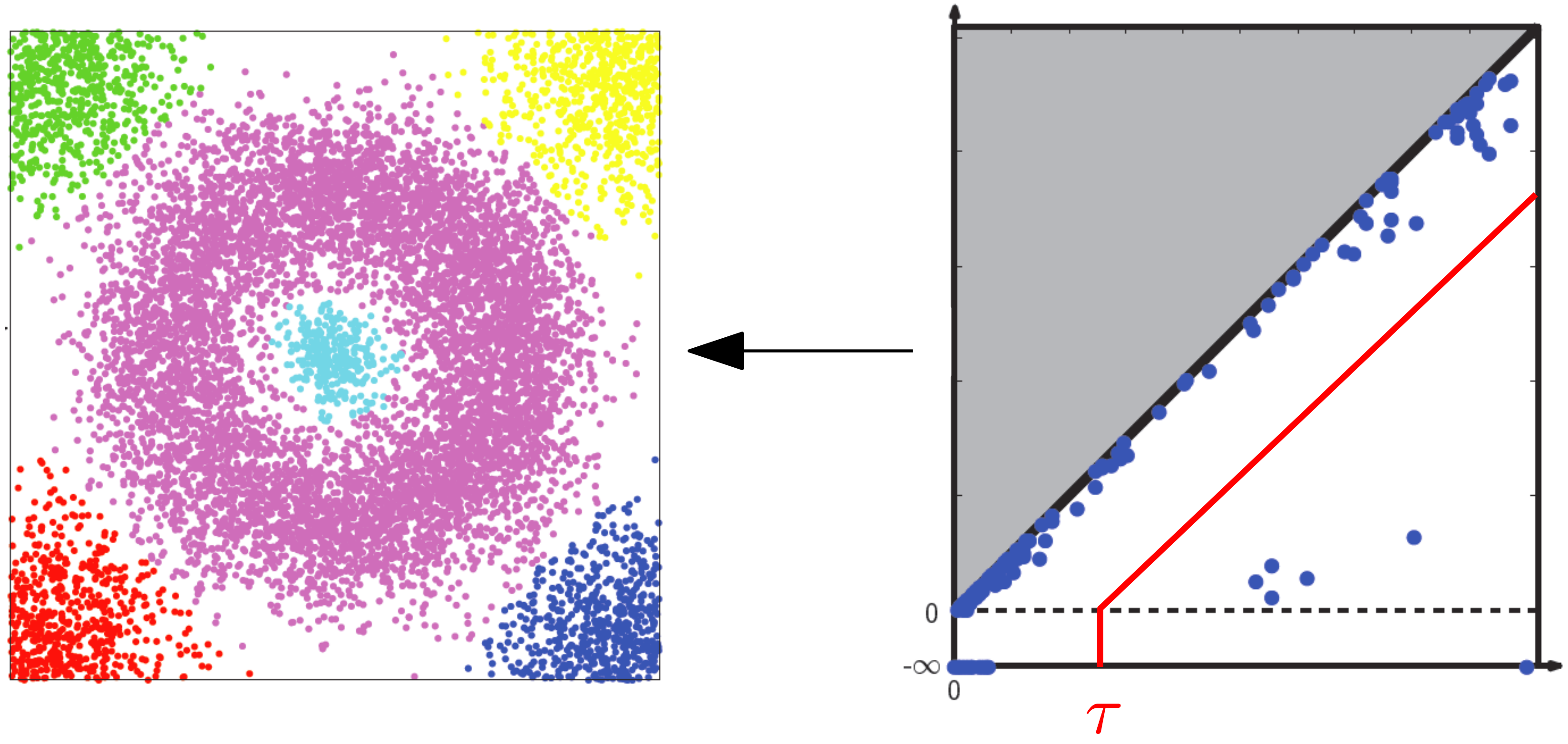
Persistence-based clustering

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.

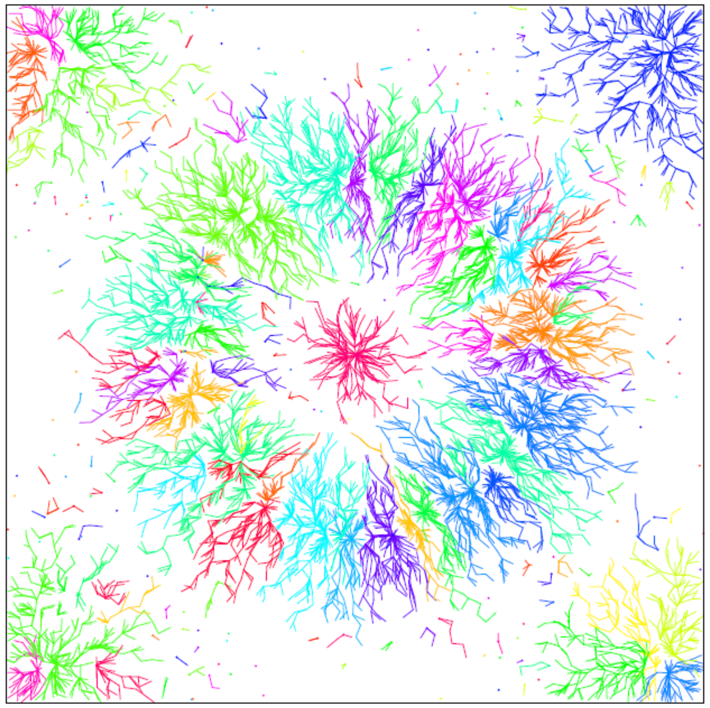
2. Compute the (0-dim) persistence of f to identify prominent peaks → number of clusters (union-find algorithm).

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

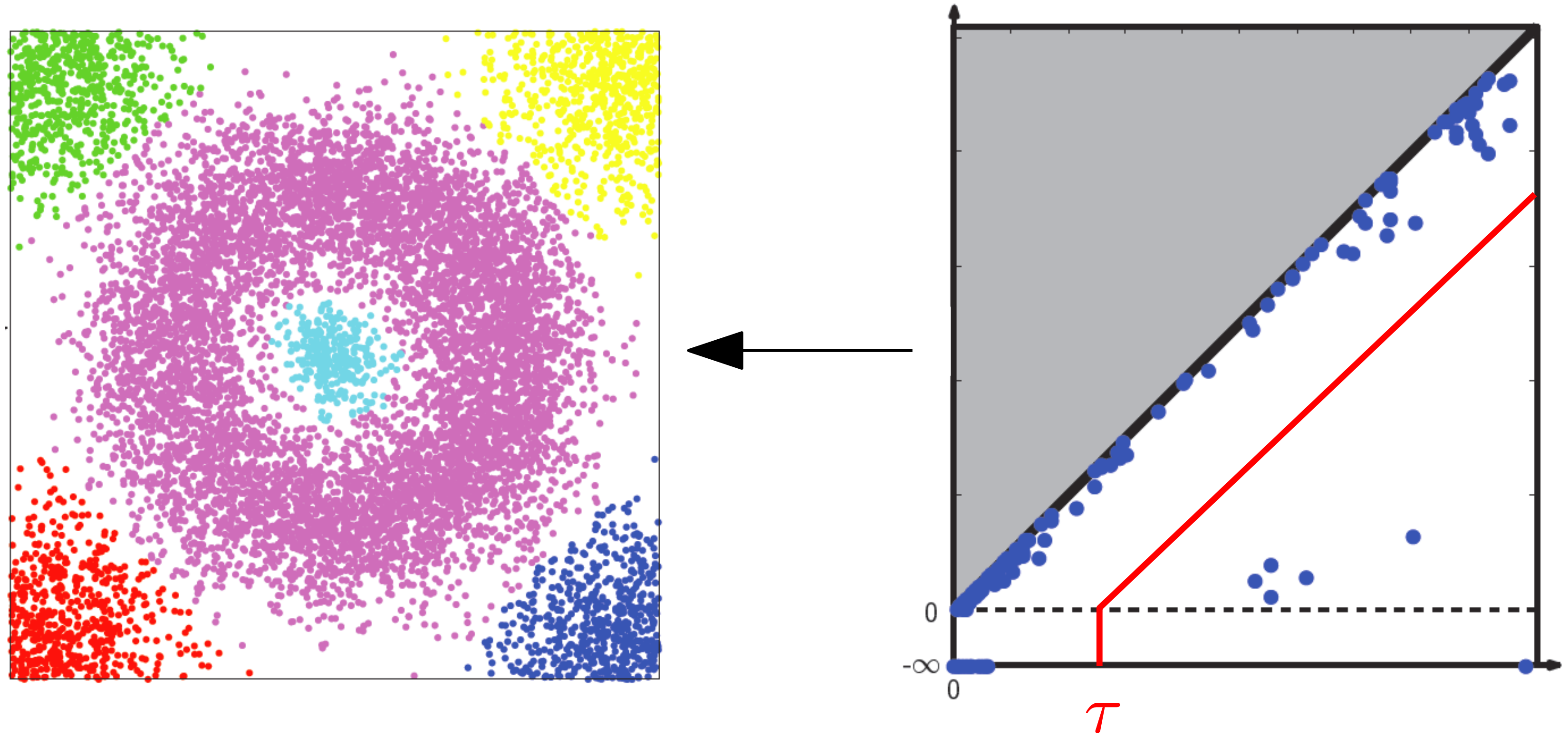
Persistence-based clustering

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

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



$\tau = 0$



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

Theoretical guarantees:

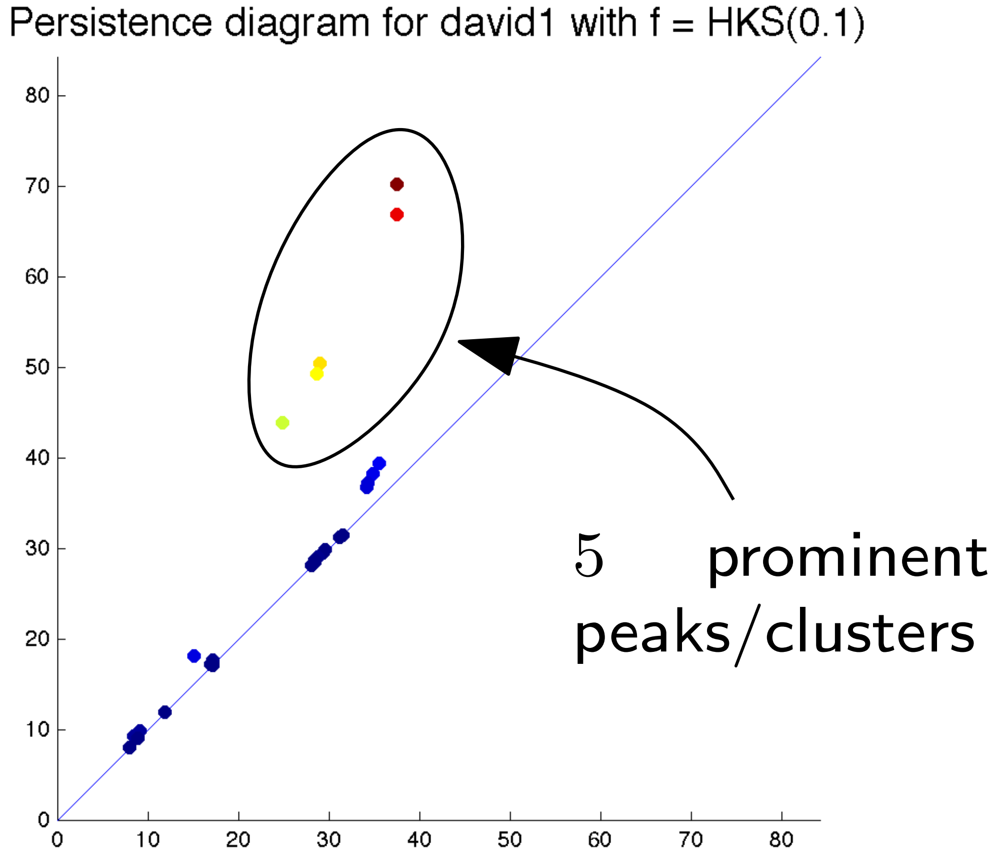
- Stability of the number of clusters (w.r.t. perturbations of X and f).
- Partial stability of clusters: well identified stable parts in each cluster.

→ "soft " clustering

Application to non-rigid shape segmentation

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

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



Persistence diagram for david1 with f = HKS(0.1)

5 prominent peaks/clusters

Problem: some part of clusters are unstable → dirty segments

Application to non-rigid shape segmentation

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

Problem: some part of clusters are unstable → dirty segments

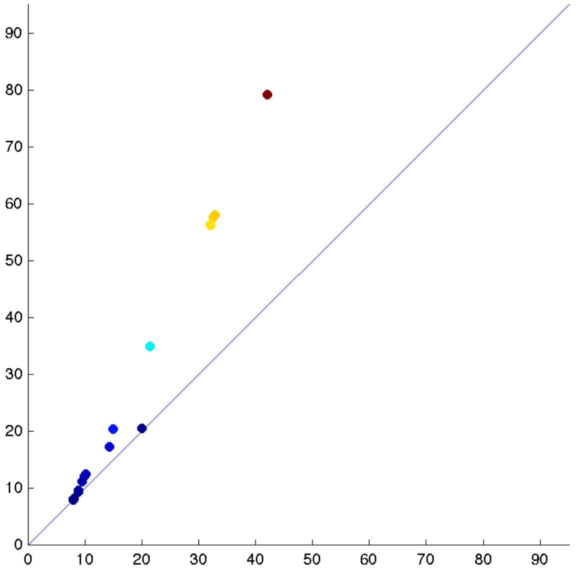
Idea:

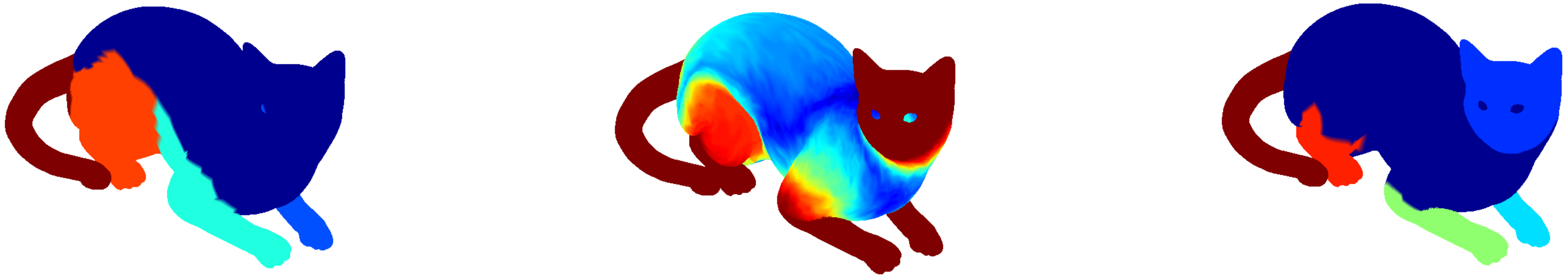
- Run the persistence based algorithm several times on random perturbations of f (size bounded by the "persistence" gap).

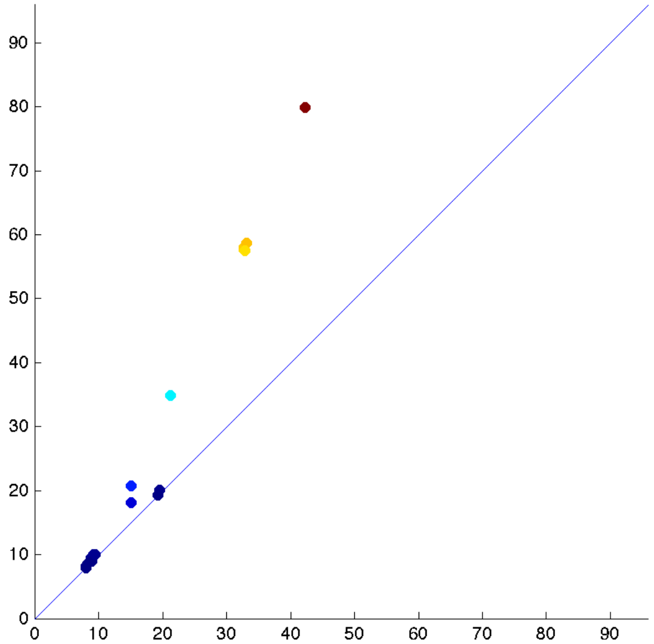
- Partial stability of clusters allows to establish correspondences between clusters across the different runs → for any $x \in X$, a vector giving the probability for x to belong to each cluster.

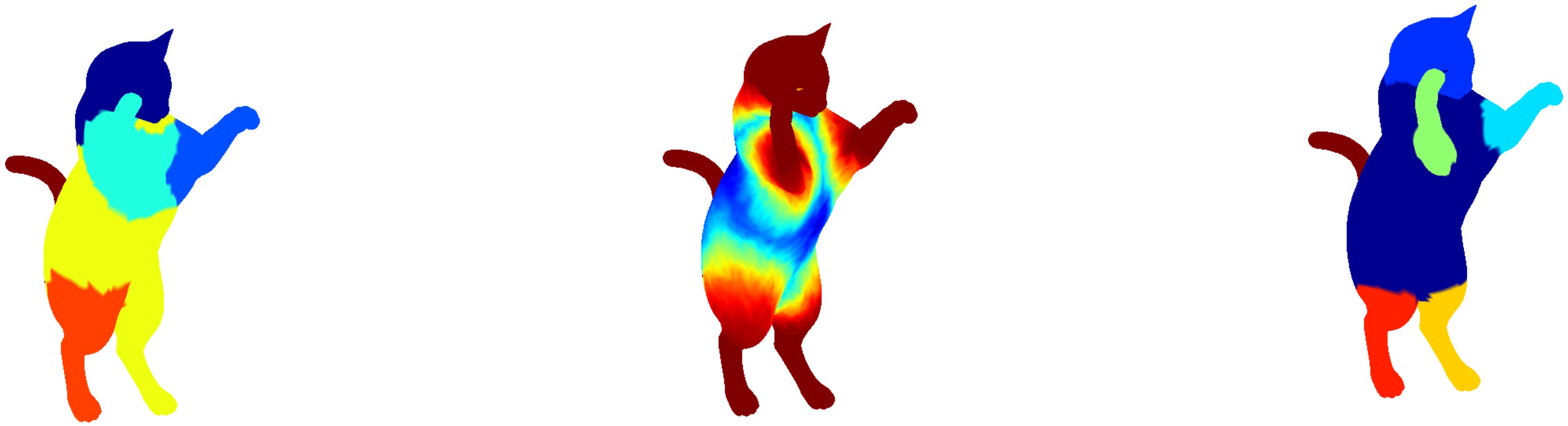
Application to non-rigid shape segmentation

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


Persistence diagram for cat7 with f = HKS(0.1)




Persistence diagram for cat1 with f = HKS(0.1)



An example of application

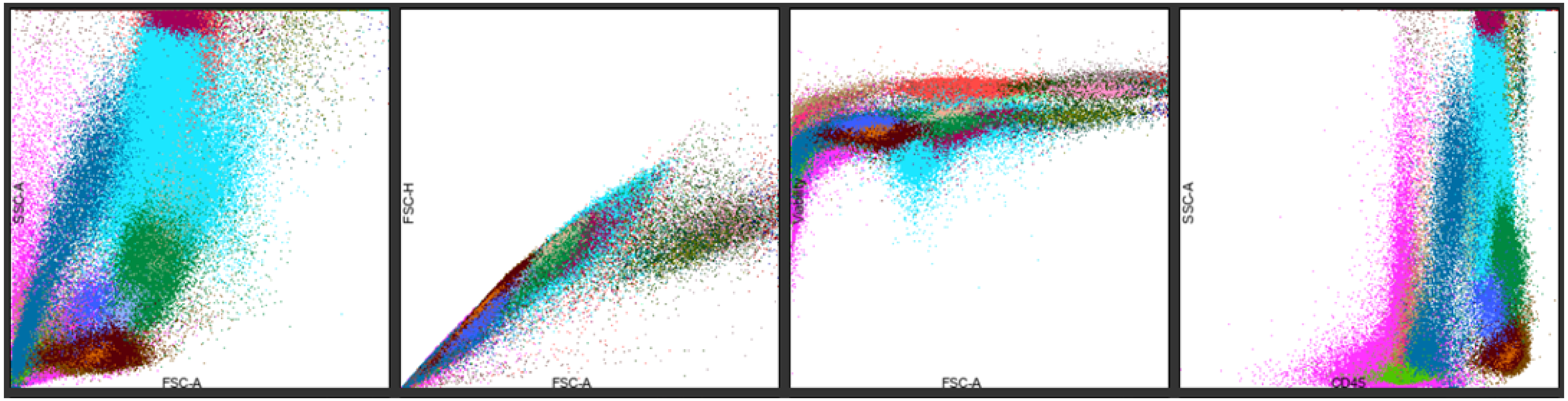
Topology-based unsupervised classification [C, Guibas, Oudot, Skraba 2013]

Segmentation of cytometry data for medical diagnosis

[M. Glisse, L. Pujol et al 2020]

METAFORA biosystems

An innovative start-up specialized in biological diagnosis from cytometry data.

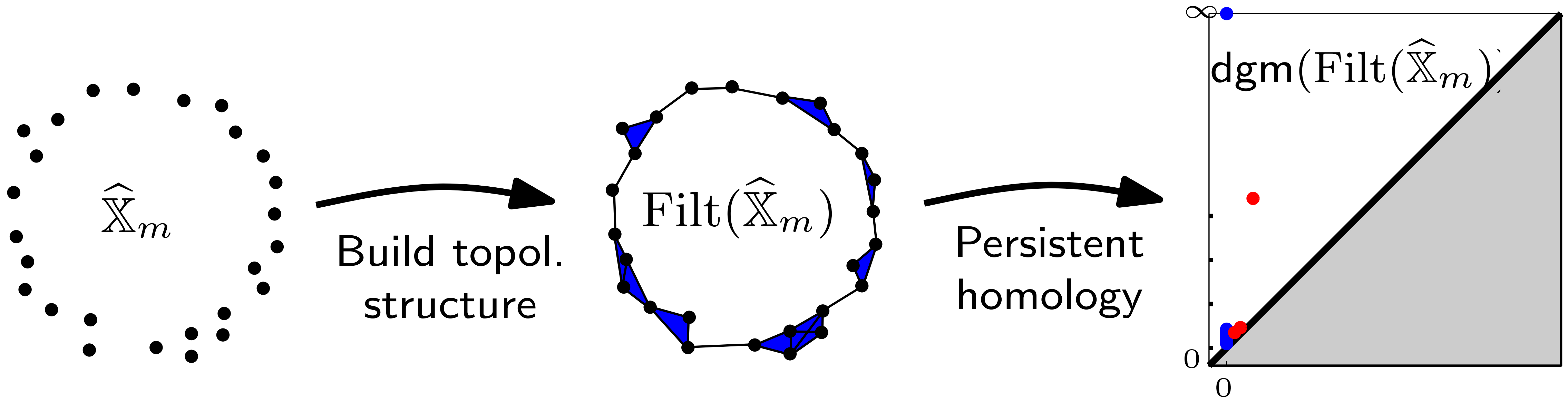


Objective: unsupervised learning in large point clouds (several millions) in medium/high dimensions ($\approx 4 \to 80$)

Applications: medical diagnosis from blood samples (1 point $=$ 1 blood cell)

Methodology: Persistence-based clustering to robustly identify relevant clusters.

Persistent homology for (point cloud) data

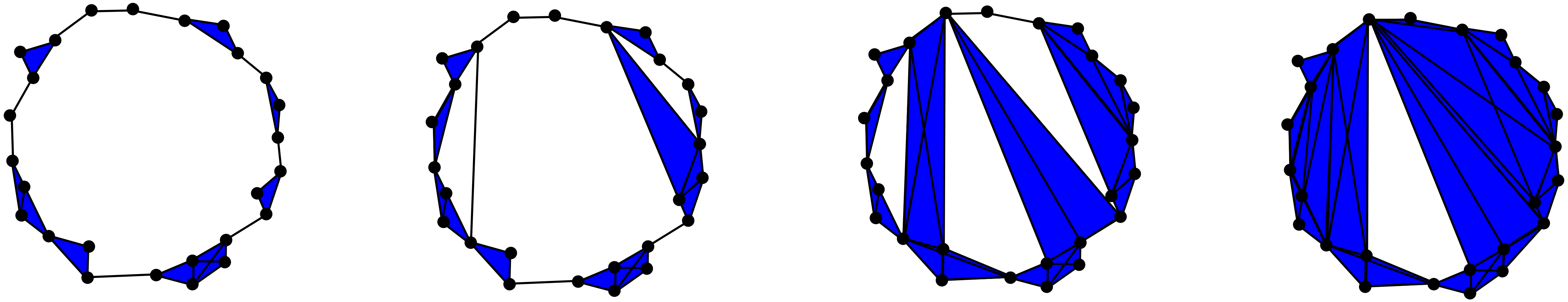


Build topol. structure

Persistent homology

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

Filtered complexes and filtrations



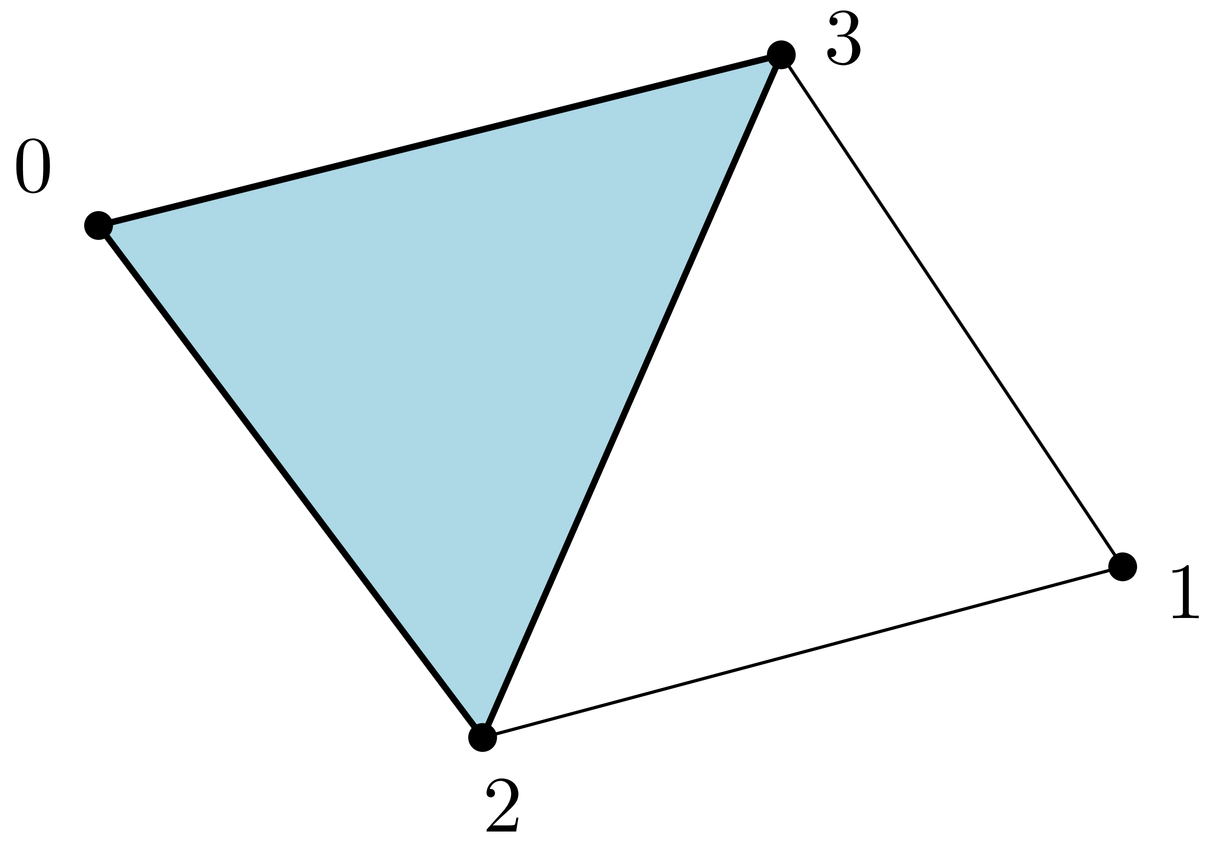
A filtered simplicial complex $\mathbb{S}$ built on top of a set X is a family $(\mathbb{S}_a \mid a \in \mathbf{R})$ of subcomplexes of some fixed simplicial complex $\overline{\mathbb{S}}$ with vertex set X s. t. $\mathbb{S}_a \subseteq \mathbb{S}_b$ for any $a \leq b$.

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

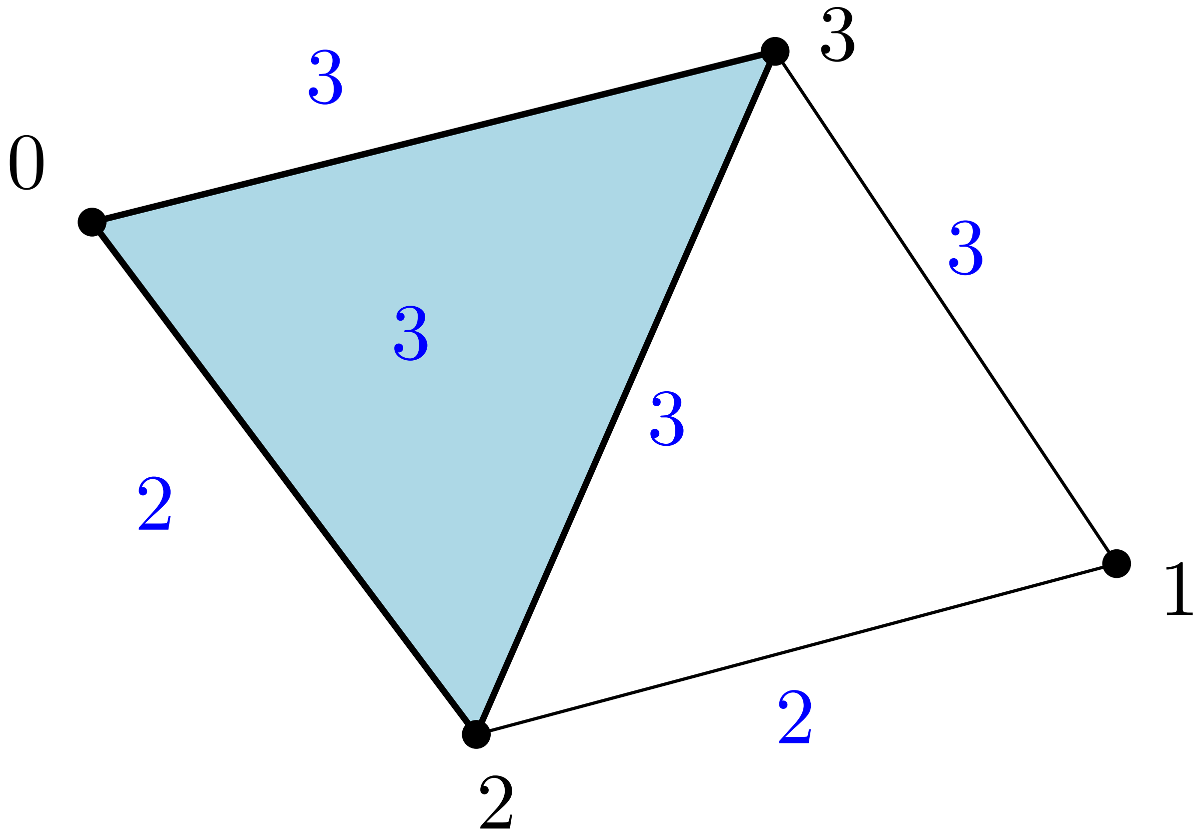
Sublevel set filtration associated to a function



- f a real valued function defined on the vertices of 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).

Exercise: show that this is a filtration

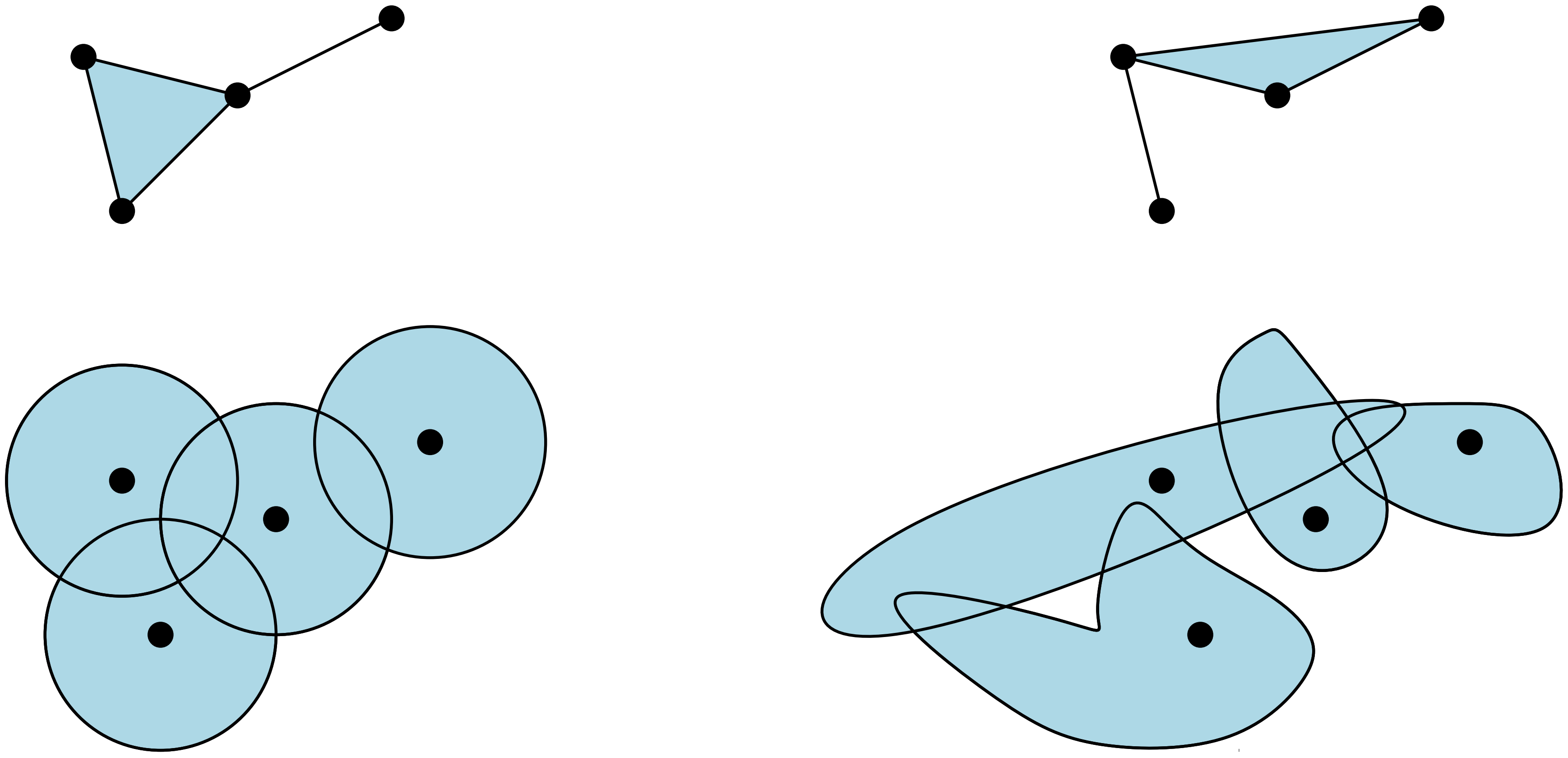
Sublevel set filtration associated to a function



- f a real valued function defined on the vertices of 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).

Exercise: show that this is a filtration

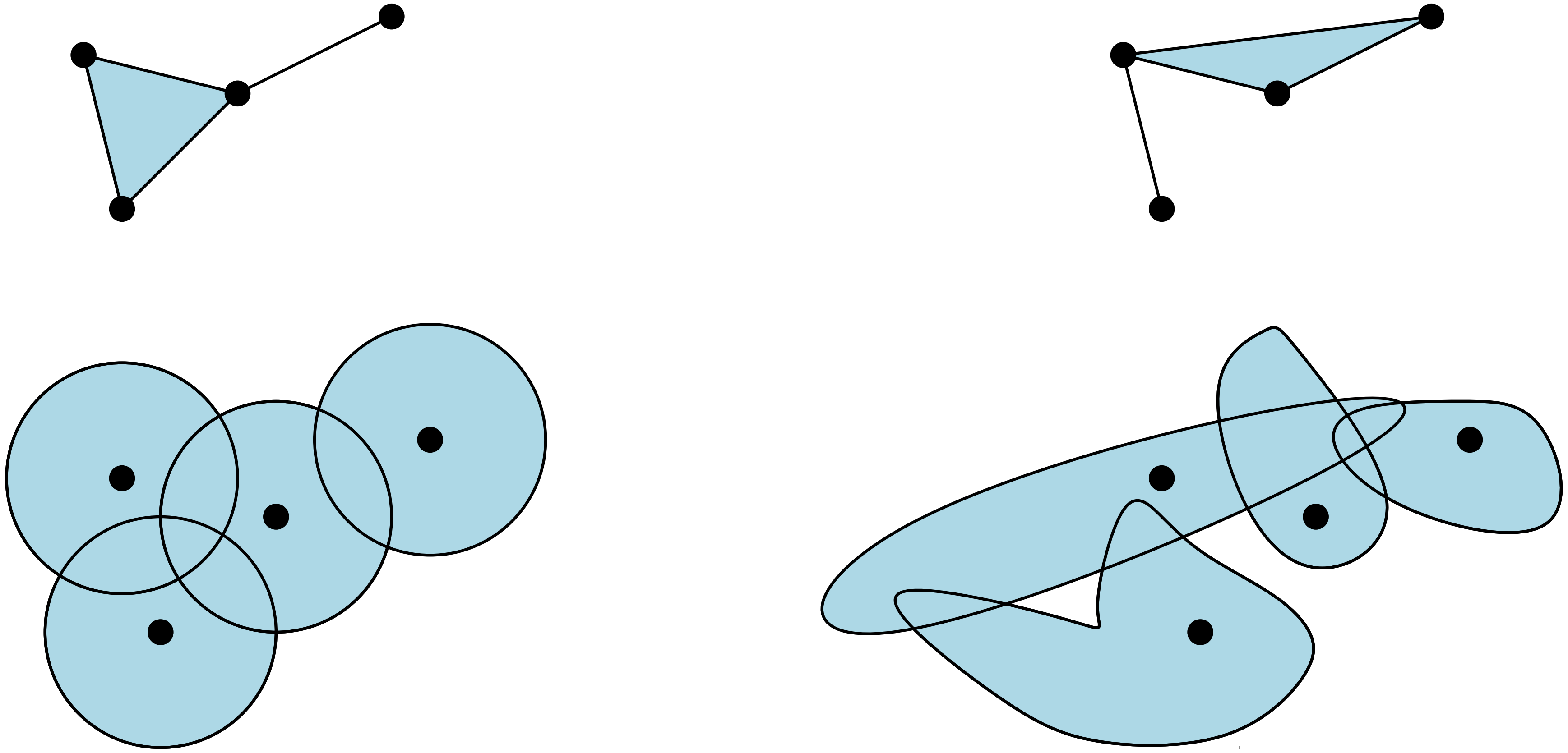
The Čech complex and filtration



- Let $\mathcal{U} = (U_i)_{i \in I}$ be a cover of a topological space X by open sets: $X = \cup_{i \in I} U_i$.

- The Cěch complex $C(\mathcal{U})$ associated to the covering $\mathcal{U}$ is the simplicial complex defined by:
  - the vertex set of $C(\mathcal{U})$ is the set of the open sets U_i
  - $[U_{i_0}, \cdots, U_{i_k}]$ is a k-simplex in $C(\mathcal{U})$ iff $\cap_{j=0}^{k} U_{i_j} \neq \emptyset$.

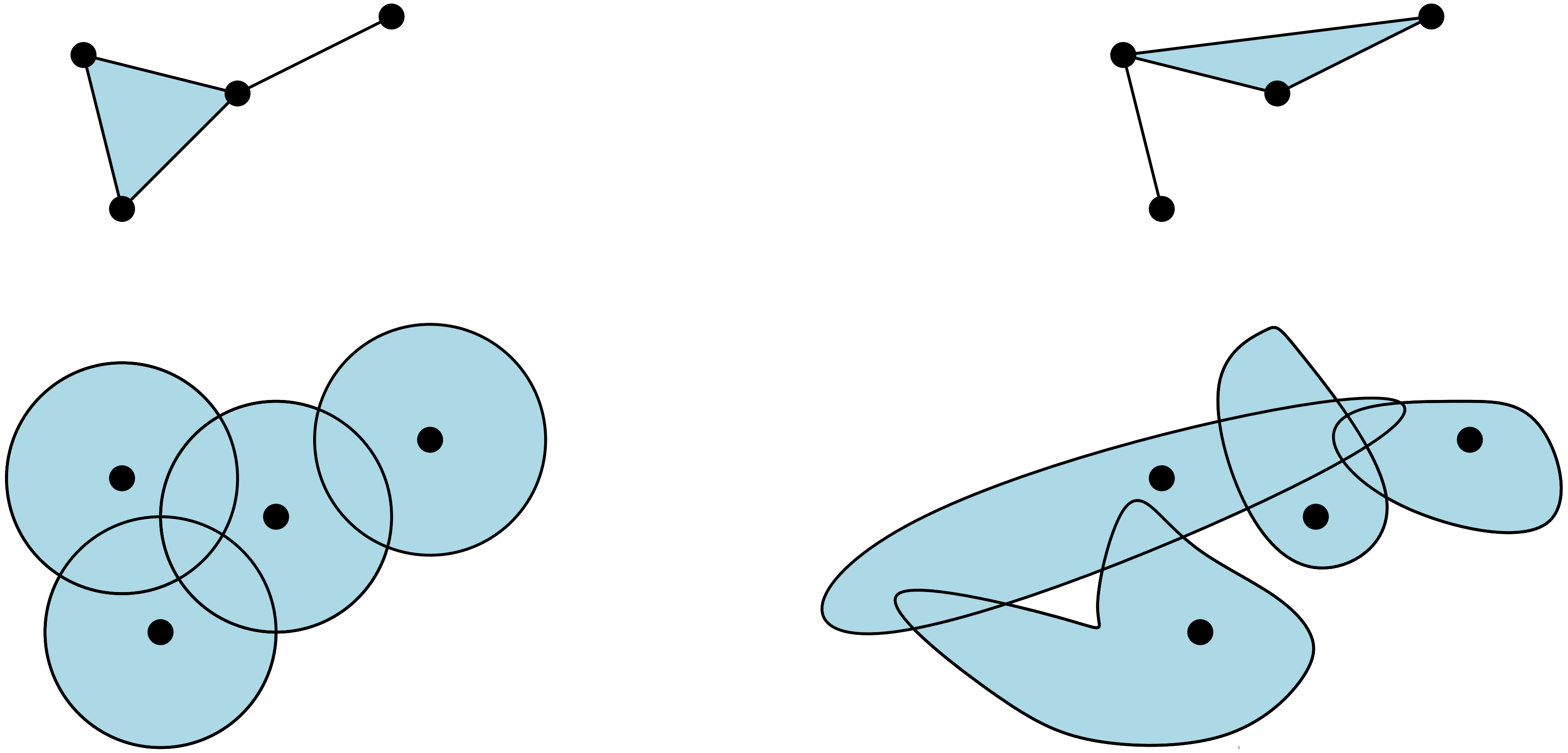
The Čech complex and filtration



Nerve theorem (Leray): If all the intersections between opens in $\mathcal{U}$ are either empty or contractible then $C(\mathcal{U})$ and $X = \cup_{i \in I} U_i$ are homotopy equivalent.

$\Rightarrow$ The combinatorics of the covering (a simplicial complex) carries the topology of the space.

The Čech complex and filtration

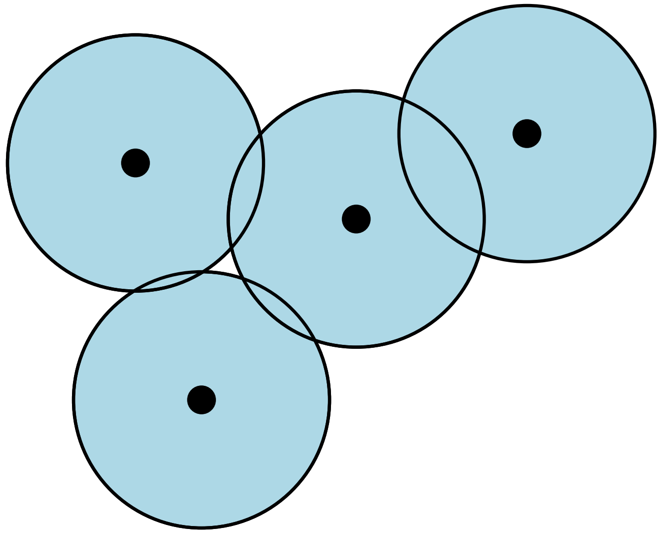


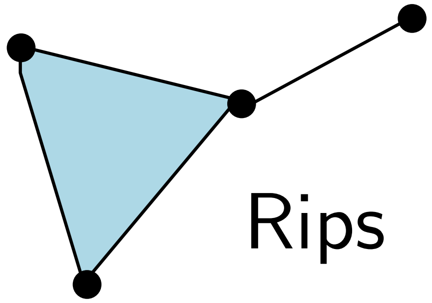
Let V be a point cloud (in a metric space).

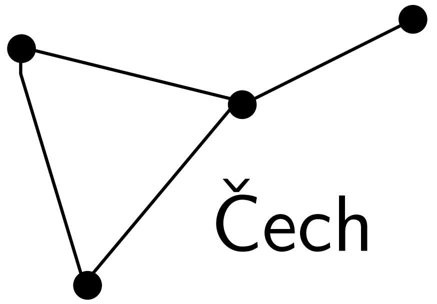
The **Čech complex** $\check{\mathrm{C}}\mathrm{ech}(V)$ is the filtered simplicial complex indexed by $\mathbb{R}$ whose vertex set is V and defined by:

$$\sigma = [p_0 p_1 \cdots p_k] \in \check{\mathrm{C}}\mathrm{ech}(V, \alpha) \ \text{ iff } \ \cap_{i=0}^{k} B(p_i, \alpha) \neq \emptyset$$

The Vietoris-Rips filtration







Let V be a point cloud (in a metric space (X, d)).

The Vietoris-Rips complex $\mathrm{Rips}(V)$ is the filtered simplicial complex indexed by $\mathbb{R}$ whose vertex set is V and defined by:

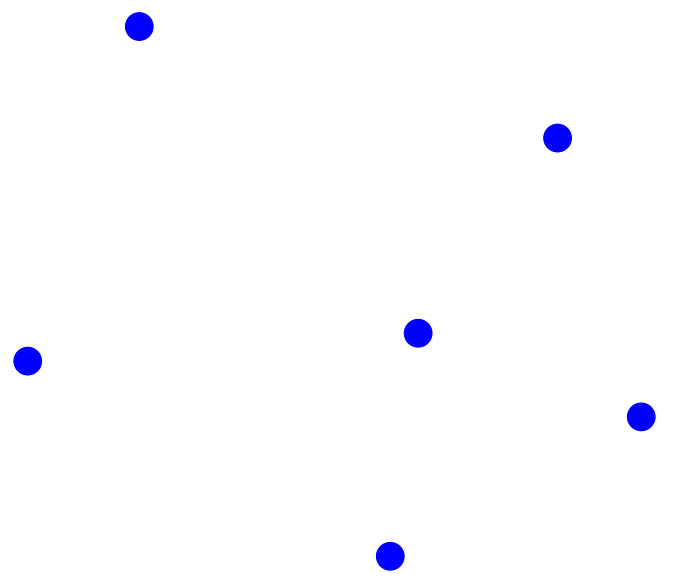
$$\sigma = [p_0 p_1 \cdots p_k] \in \mathrm{Rips}(V, \alpha) \text{ iff } \forall i, j \in \{0, \cdots, k\},\ d(p_i, p_j) \leq \alpha$$

- Easy to compute and fully determined by its 1-skeleton
- Rips-Čech interleaving: for any $\alpha > 0$,

$$\check{\mathrm{C}}\mathrm{ech}(L, \frac{\alpha}{2}) \subseteq \mathrm{Rips}(L, \alpha) \subseteq \check{\mathrm{C}}\mathrm{ech}(L, \alpha)$$

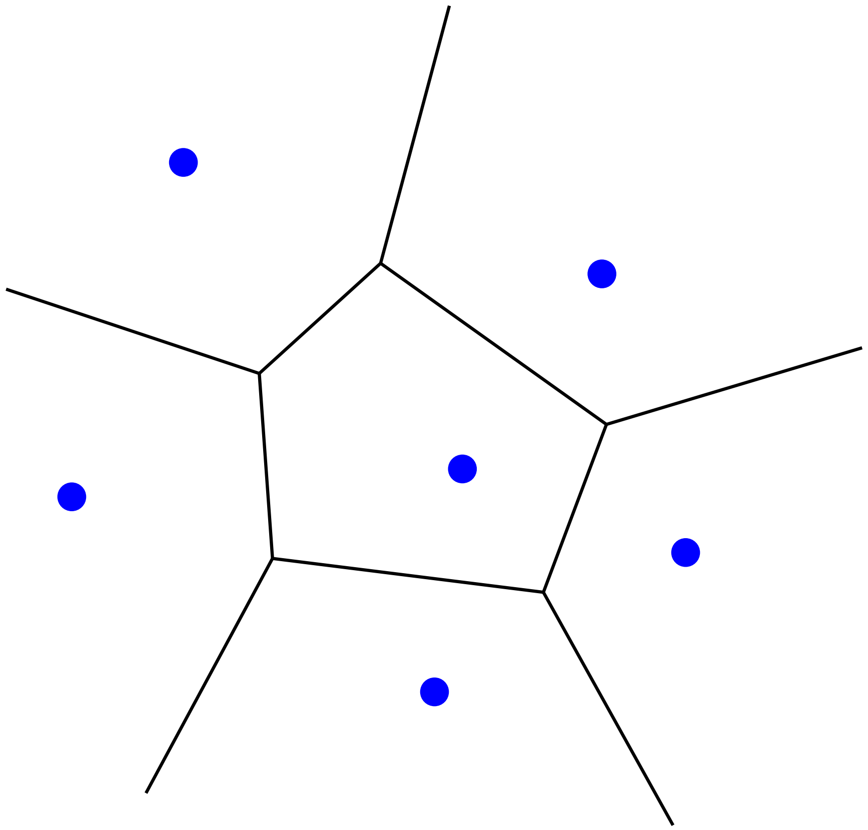
Voronoï diagrams, Delaunay triangulations and α-complexes

$P = \{p_1, \cdots, p_n\} \in \mathbb{R}^d$



Voronoï diagrams, Delaunay triangulations and α-complexes

$P = \{p_1, \cdots, p_n\} \in \mathbb{R}^d$

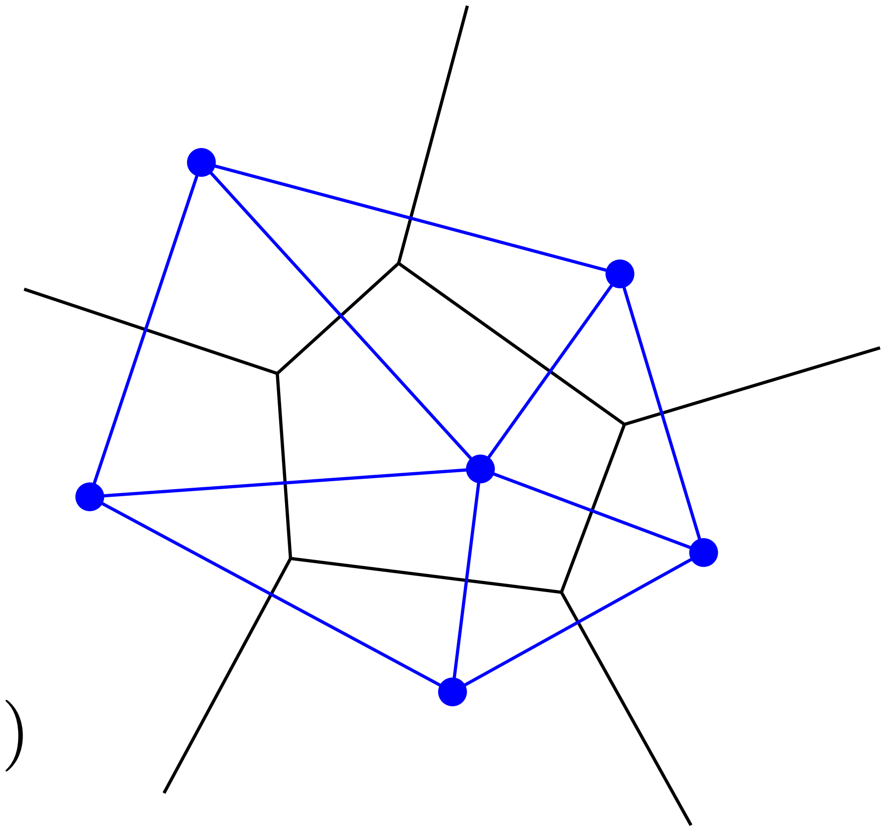


Voronoï cells:

$$Vor(p_i) = \{x \in \mathbb{R}^d : \forall j, \|x - p_i\| \leq \|x - p_j\|\}$$

Voronoï diagrams, Delaunay triangulations and α-complexes

$P = \{p_1, \cdots, p_n\} \in \mathbb{R}^d$



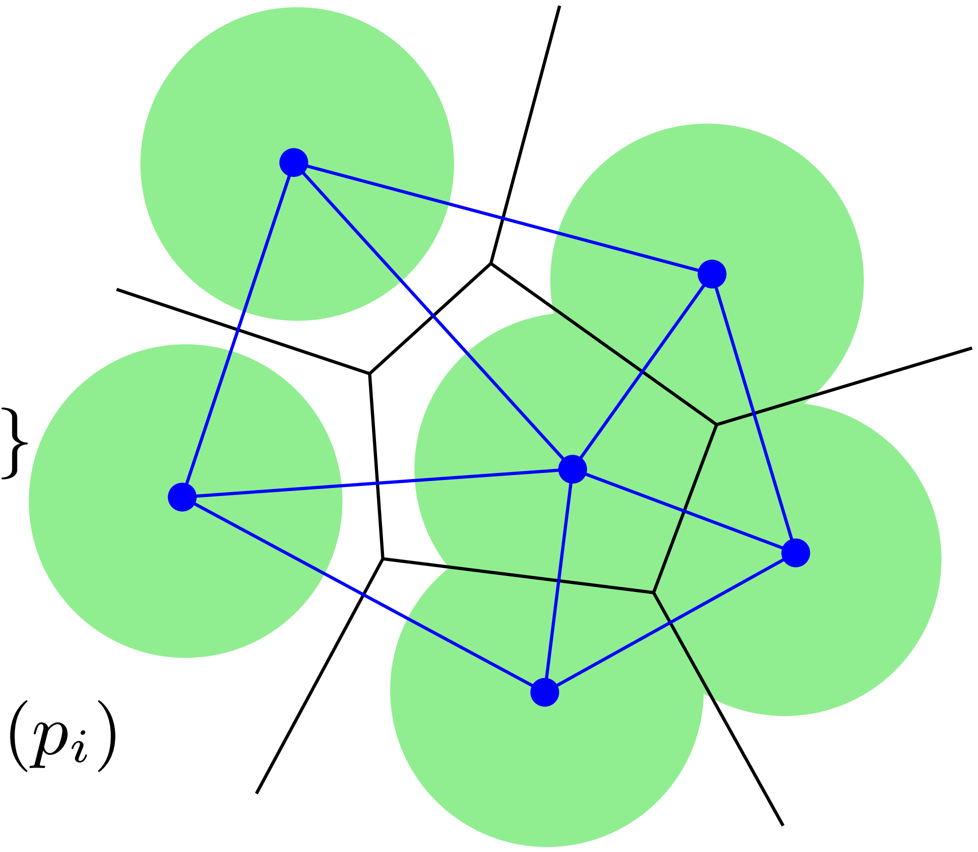
Voronoï cells:

$$Vor(p_i) = \{x \in \mathbb{R}^d : \forall j, \|x - p_i\| \leq \|x - p_j\|\}$$

Delaunay complex $\mathcal{D}(P)$:
nerve of the cover made by the Voronoï cells $Vor(p_i)$

Voronoï diagrams, Delaunay triangulations and α-complexes

$P = \{p_1, \cdots, p_n\} \in \mathbb{R}^d$



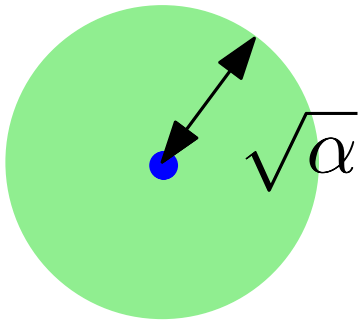
Voronoï cells:

$$Vor(p_i) = \{x \in \mathbb{R}^d : \forall j, \|x - p_i\| \leq \|x - p_j\|\}$$

Delaunay complex $\mathcal{D}(P)$:
nerve of the cover made by the Voronoï cells $Vor(p_i)$

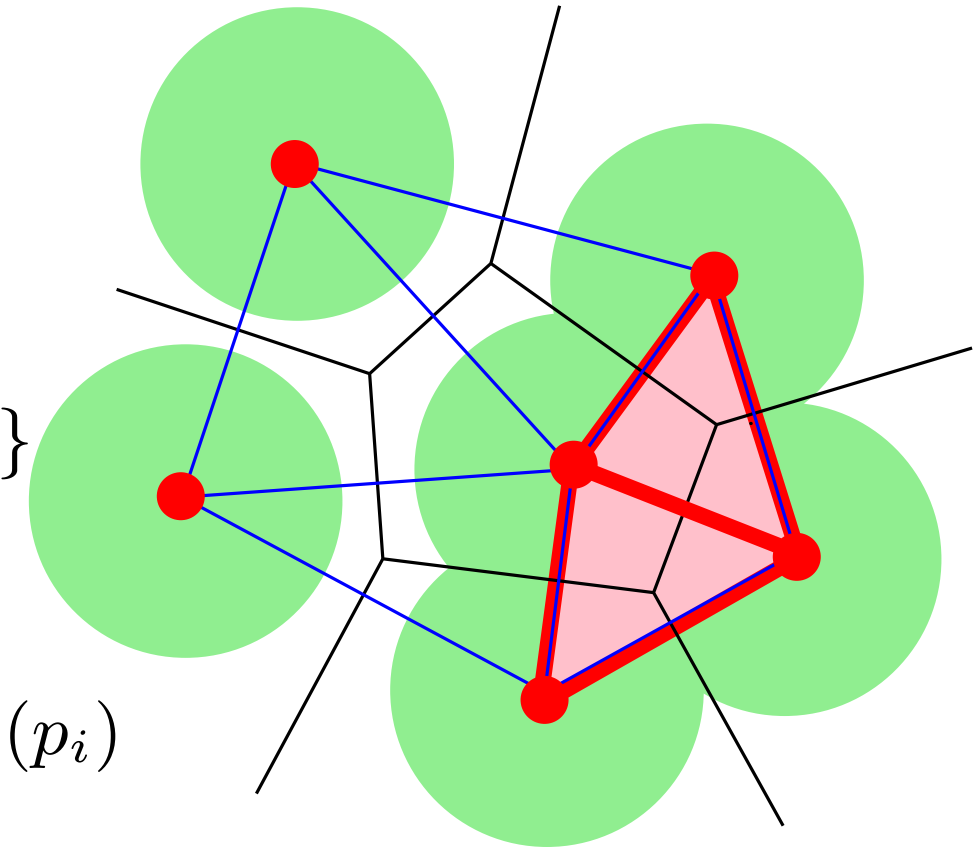
Alpha complex $\mathcal{A}(P, \alpha)$:
For $\alpha \geq 0$, nerve of the family

$$\left(Vor(p_i) \cap B(p_i, \sqrt{\alpha})\right)_{i=1,\cdots,n}$$



Voronoï diagrams, Delaunay triangulations and α-complexes

$P = \{p_1, \cdots, p_n\} \in \mathbb{R}^d$

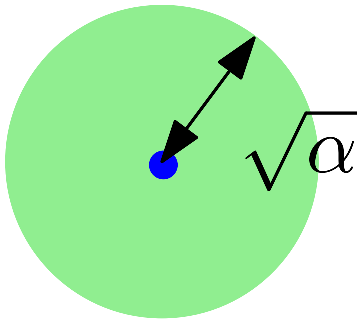


Voronoï cells:

$$Vor(p_i) = \{x \in \mathbb{R}^d : \forall j, \|x - p_i\| \leq \|x - p_j\|\}$$

Delaunay complex $\mathcal{D}(P)$:
nerve of the cover made by the Voronoï cells $Vor(p_i)$

Alpha complex $\mathcal{A}(P, \alpha)$:
For $\alpha \geq 0$, nerve of the family



$$\left(Vor(p_i) \cap B(p_i, \sqrt{\alpha})\right)_{i=1,\cdots,n}$$

Theorem:
$\mathcal{A}(P, \alpha)$ is homotopy equivalent to $\cup_{i=1}^{n} B(p_i, \sqrt{\alpha})$.

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

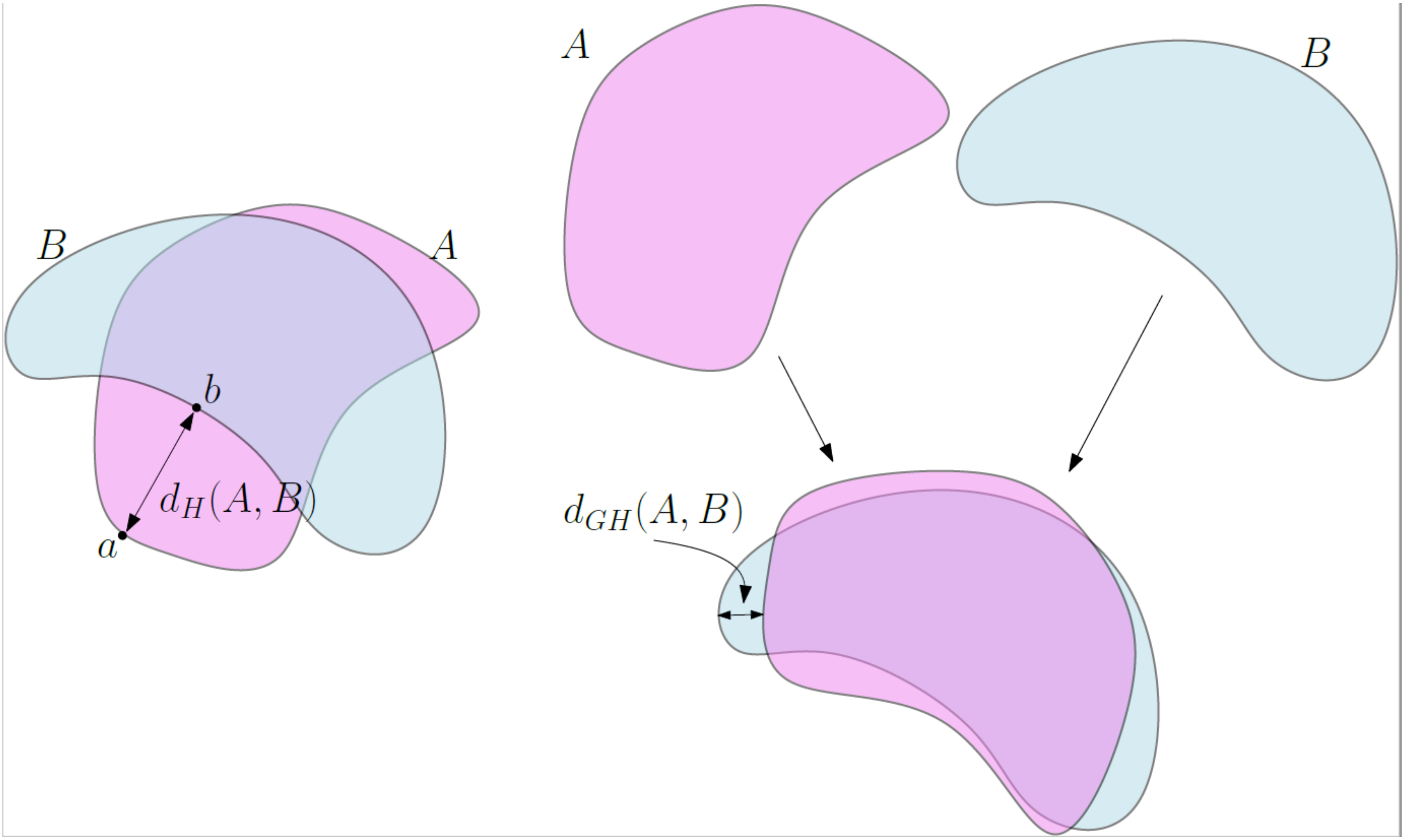
Gromov-Hausdorff distance

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

$\mathbb{Z}$ 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).

Hausdorff distance



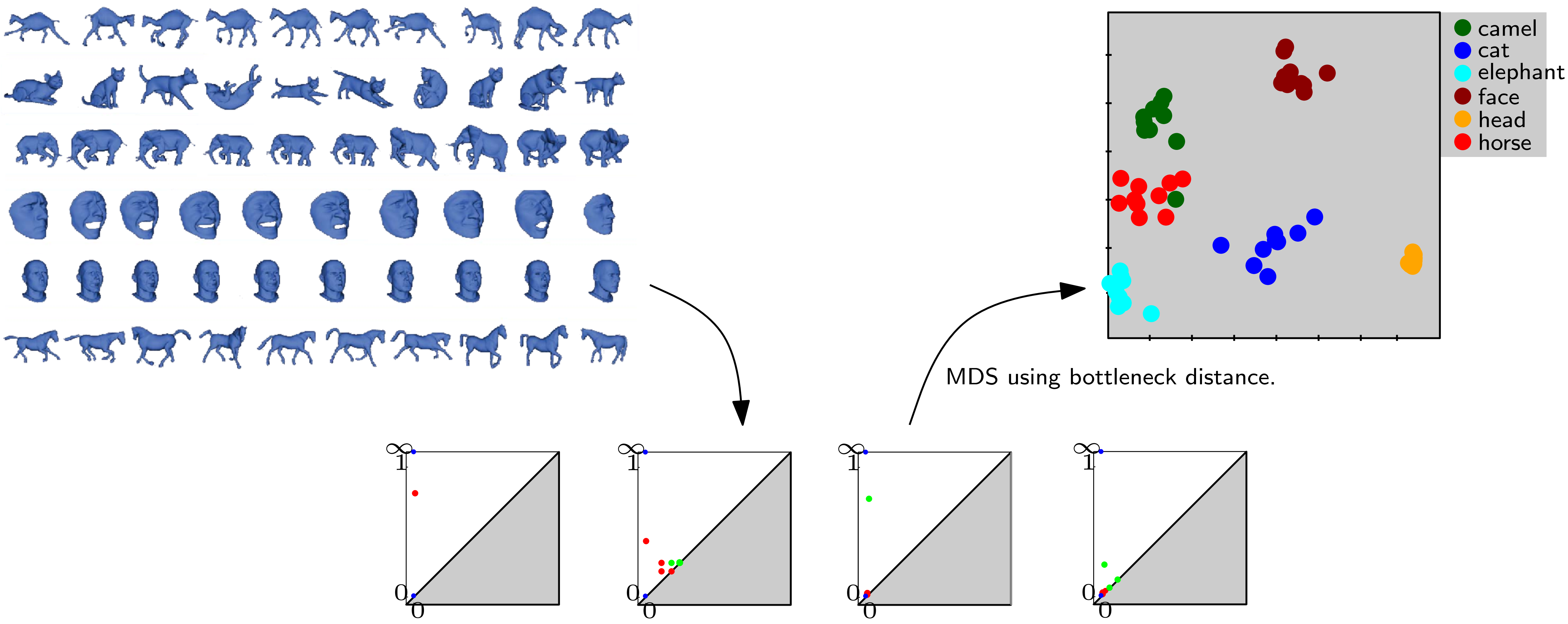
Let $A, B \subset M$ be two compact subsets of a metric space (M, d)

$$d_H(A, B) \quad = \quad \max\{\sup_{b \in B} d(b, A), \sup_{a \in A} d(a, B)\}$$

where $d(b, A) = \sup_{a \in A} d(b, a)$.

Application: non rigid shape classification

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



MDS using bottleneck distance.

- 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.

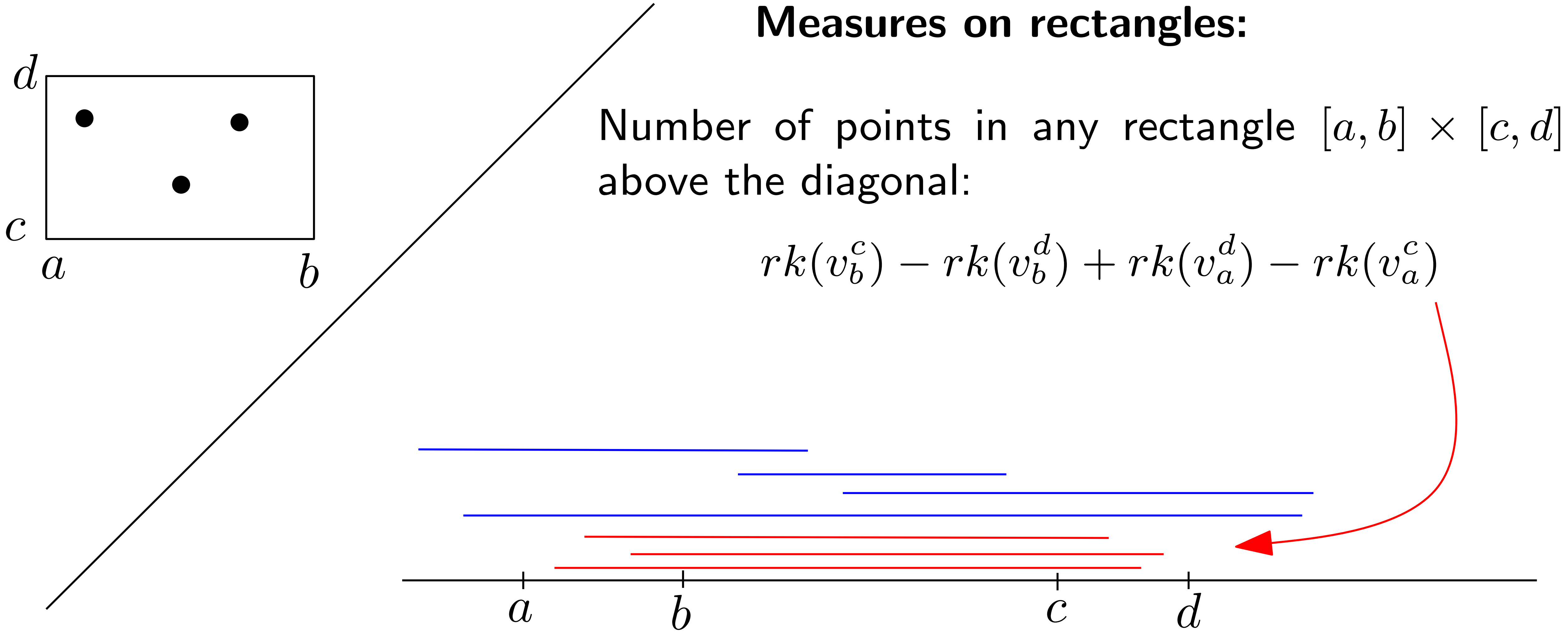
Theorem: [C., Cohen-Steiner, Glisse, Guibas, Oudot - SoCG'09], [C., de Silva, Glisse, Oudot 12]

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 $a \leq b$.

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

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 $\mathbb{V}$.

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

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

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

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

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

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*.

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]

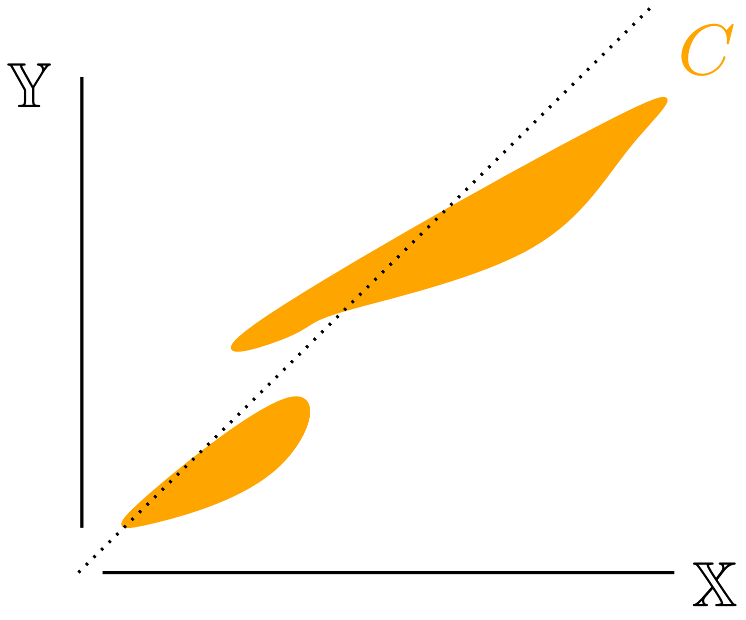
If $\mathbb{U}$ and $\mathbb{V}$ are q-tame and ϵ-interleaved for some $\epsilon \geq 0$ then

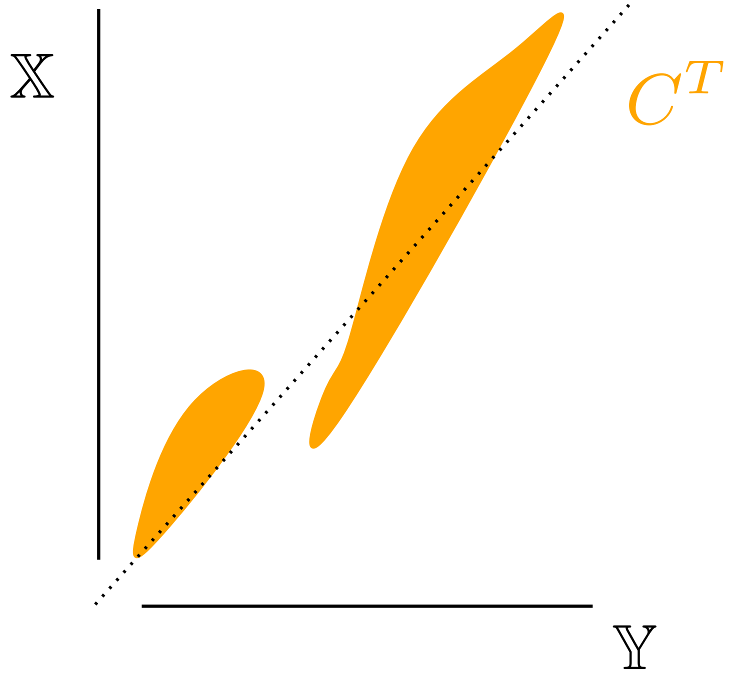
$$d_B(\mathrm{dgm}(\mathbb{U}), \mathrm{dgm}(\mathbb{V})) \leq \epsilon$$

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*.

Exercise: Prove the stability theorem for functions.

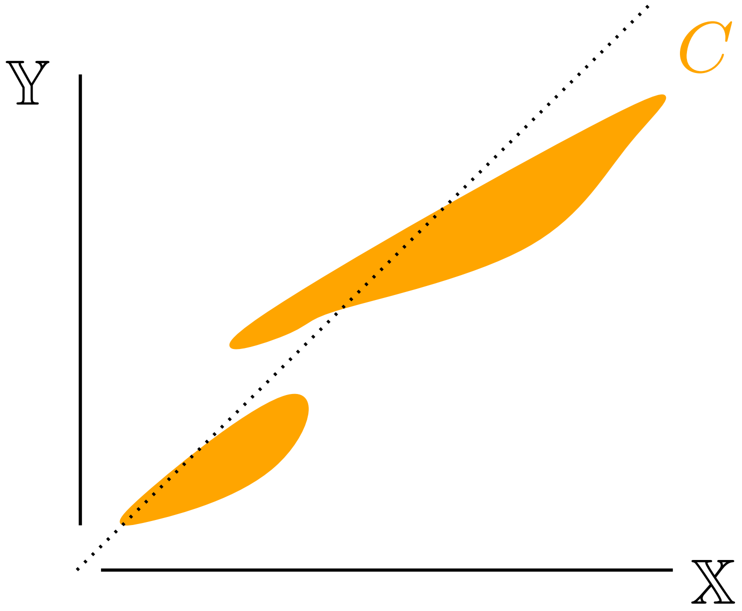
Multivalued maps and correspondences

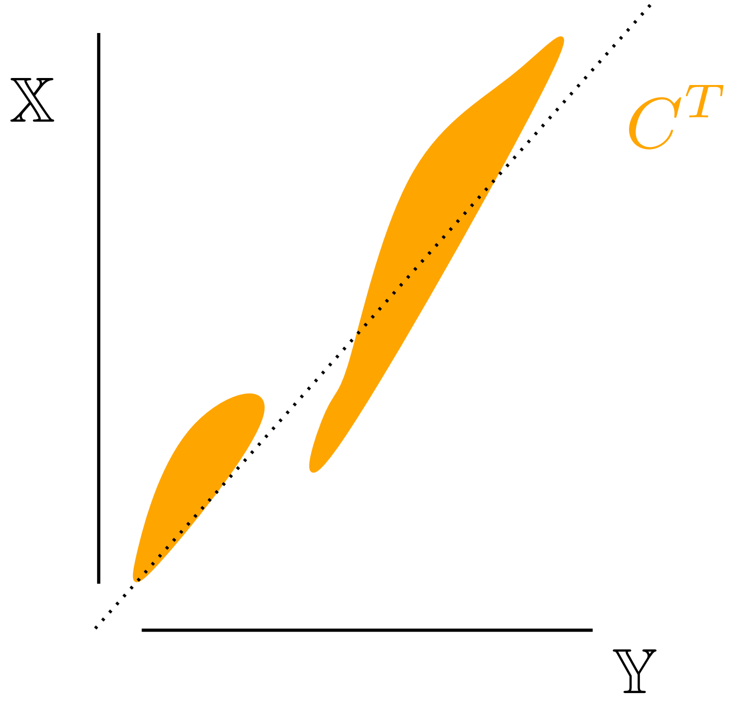




A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ from a set $\mathbb{X}$ to a set $\mathbb{Y}$ is a subset of $\mathbb{X} \times \mathbb{Y}$, also denoted C, that projects surjectively onto $\mathbb{X}$ through the canonical projection $\pi_{\mathbb{X}} : \mathbb{X} \times \mathbb{Y} \to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



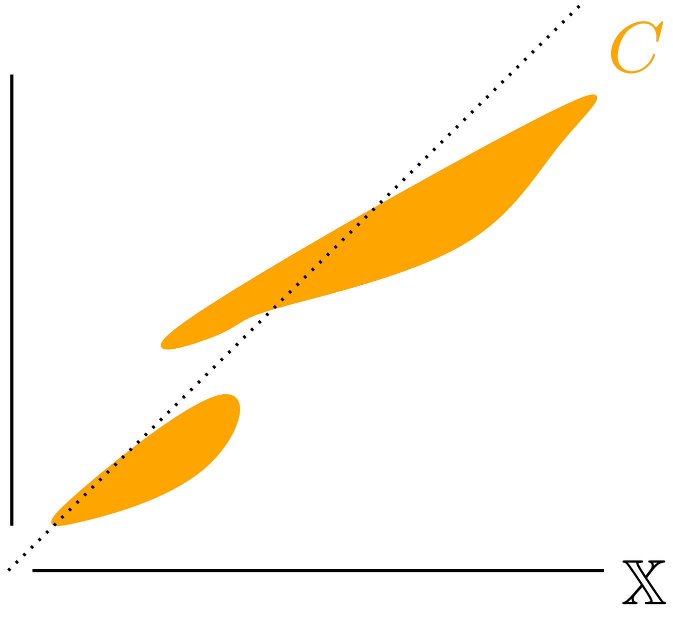


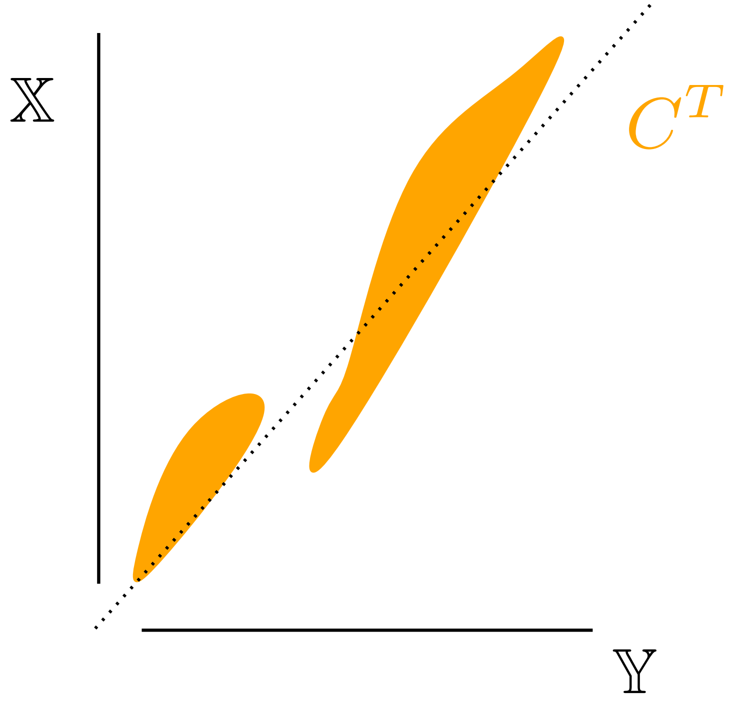
A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ from a set $\mathbb{X}$ to a set $\mathbb{Y}$ is a subset of $\mathbb{X} \times \mathbb{Y}$, also denoted C, that projects surjectively onto $\mathbb{X}$ through the canonical projection $\pi_{\mathbb{X}} : \mathbb{X} \times \mathbb{Y} \to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

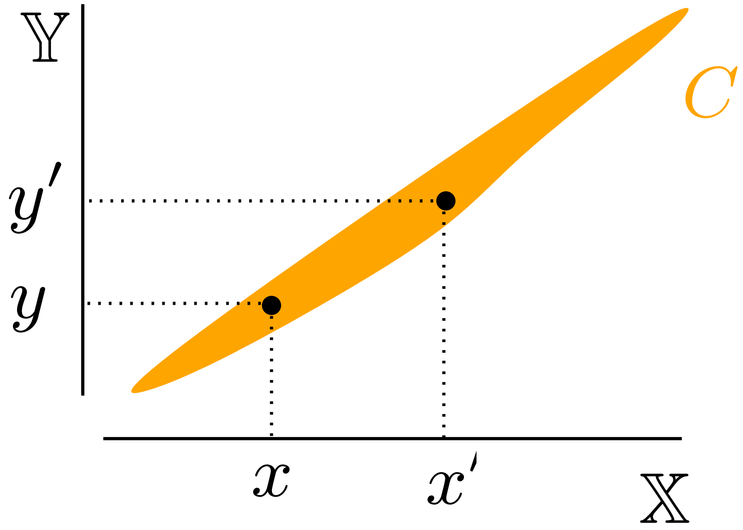




A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ from a set $\mathbb{X}$ to a set $\mathbb{Y}$ is a subset of $\mathbb{X} \times \mathbb{Y}$, also denoted C, that projects surjectively onto $\mathbb{X}$ through the canonical projection $\pi_{\mathbb{X}} : \mathbb{X} \times \mathbb{Y} \to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

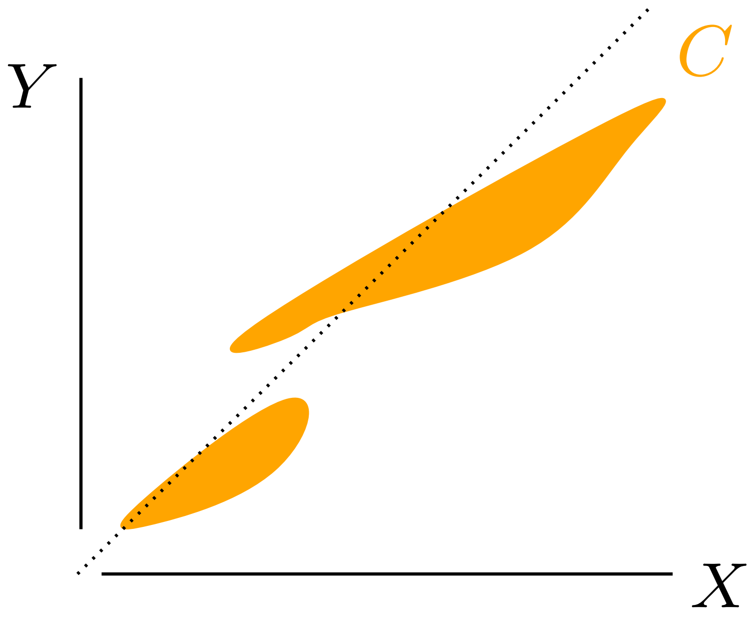
Example: ϵ-correspondence and Gromov-Hausdorff distance.

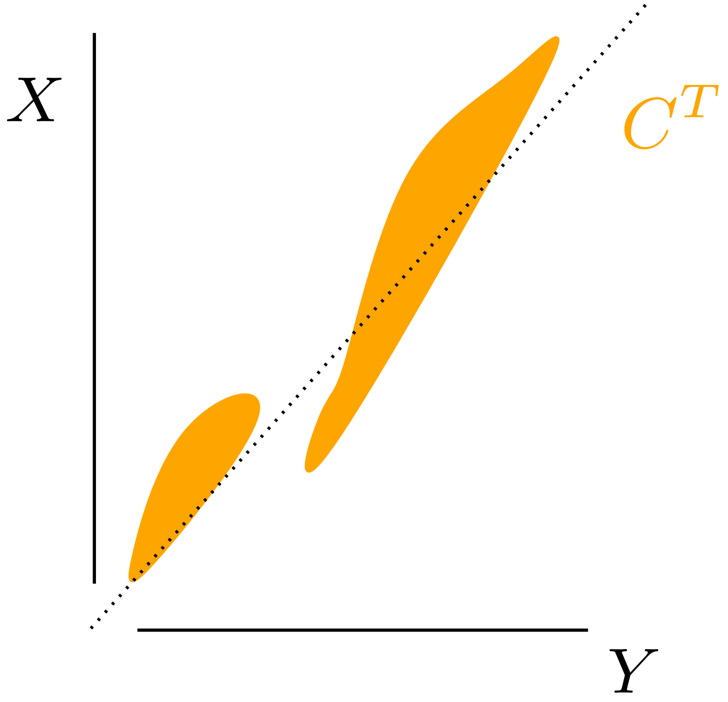
Let $(\mathbb{X}, \rho_{\mathbb{X}})$ and $(\mathbb{Y}, \rho_{\mathbb{Y}})$ be compact metric spaces.
A correspondence $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is an ϵ-correspondence if $\forall (x, y), (x', y') \in C$, $|\rho_{\mathbb{X}}(x, x') - \rho_{\mathbb{Y}}(y, y')| \leq \varepsilon$.



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

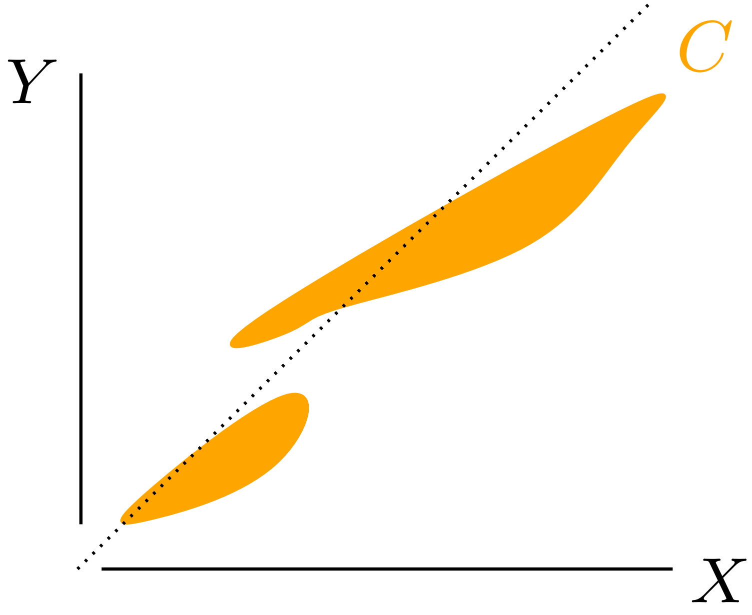
Multivalued simplicial maps

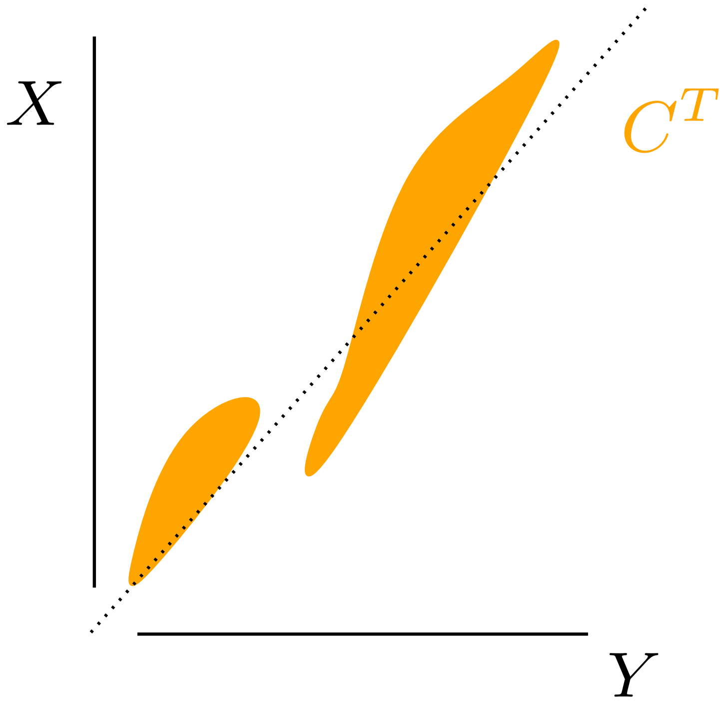




Let $\mathbb{S}$ and $\mathbb{T}$ be two filtered simplicial complexes with vertex sets $\mathbb{X}$ and $\mathbb{Y}$ respectively. A multivalued map $C : \mathbb{X} \rightrightarrows \mathbb{Y}$ is ε-simplicial from $\mathbb{S}$ to $\mathbb{T}$ if for any $a \in \mathbf{R}$ and any simplex $\sigma \in \mathbb{S}_a$, 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 $\epsilon > 2\mathrm{d}_{\mathrm{GH}}(\mathbb{X}, \mathbb{Y})$ 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.

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!

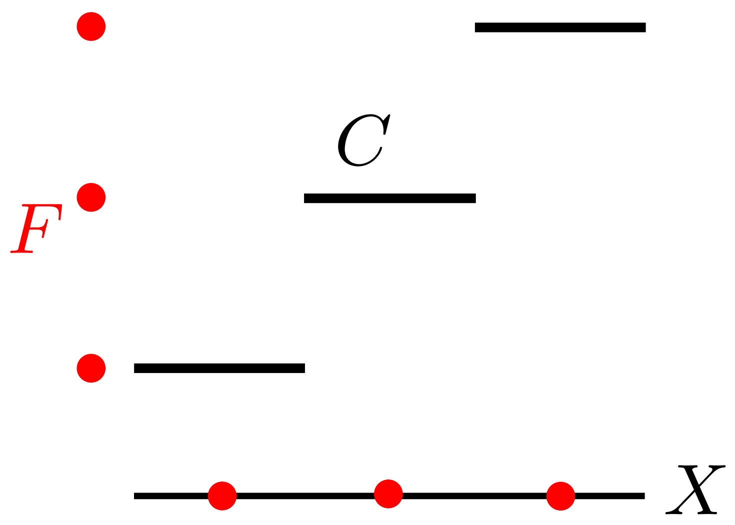
Tameness of the Rips and Čech filtrations

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

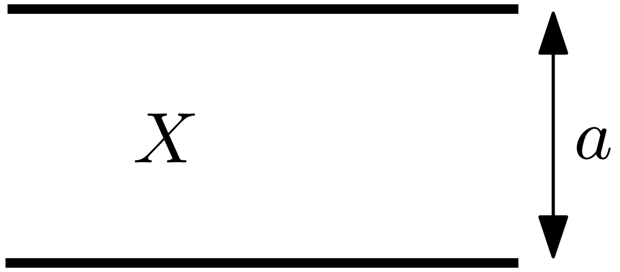
$$\mathrm{d_b}(\mathsf{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{X}))), \mathsf{dgm}(\mathrm{H}(\check{\mathrm{C}}\mathrm{ech}(\mathbb{Y})))) \leq 2\mathrm{d_{GH}}(\mathbb{X}, \mathbb{Y}),$$

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

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:

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

- 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]).

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]).

- 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 $O(\delta)$.

Computational issues and robustness to noise
A statistical perspective

Some weaknesses

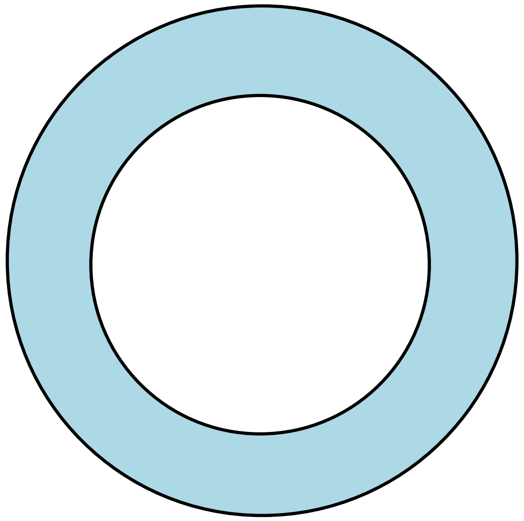
If $\mathbb{X}$ and $\mathbb{Y}$ are pre-compact metric spaces, then

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

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

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

Statistical setting

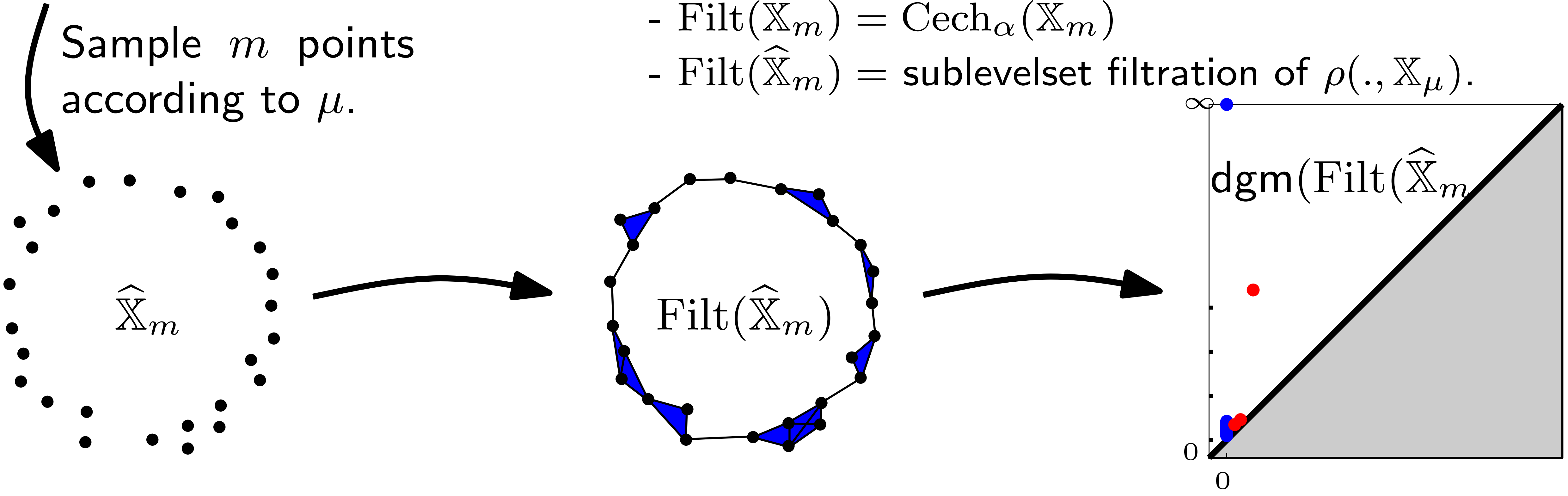


$(\mathbb{M}, \rho)$ metric space

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 $\rho(., \mathbb{X}_\mu)$.

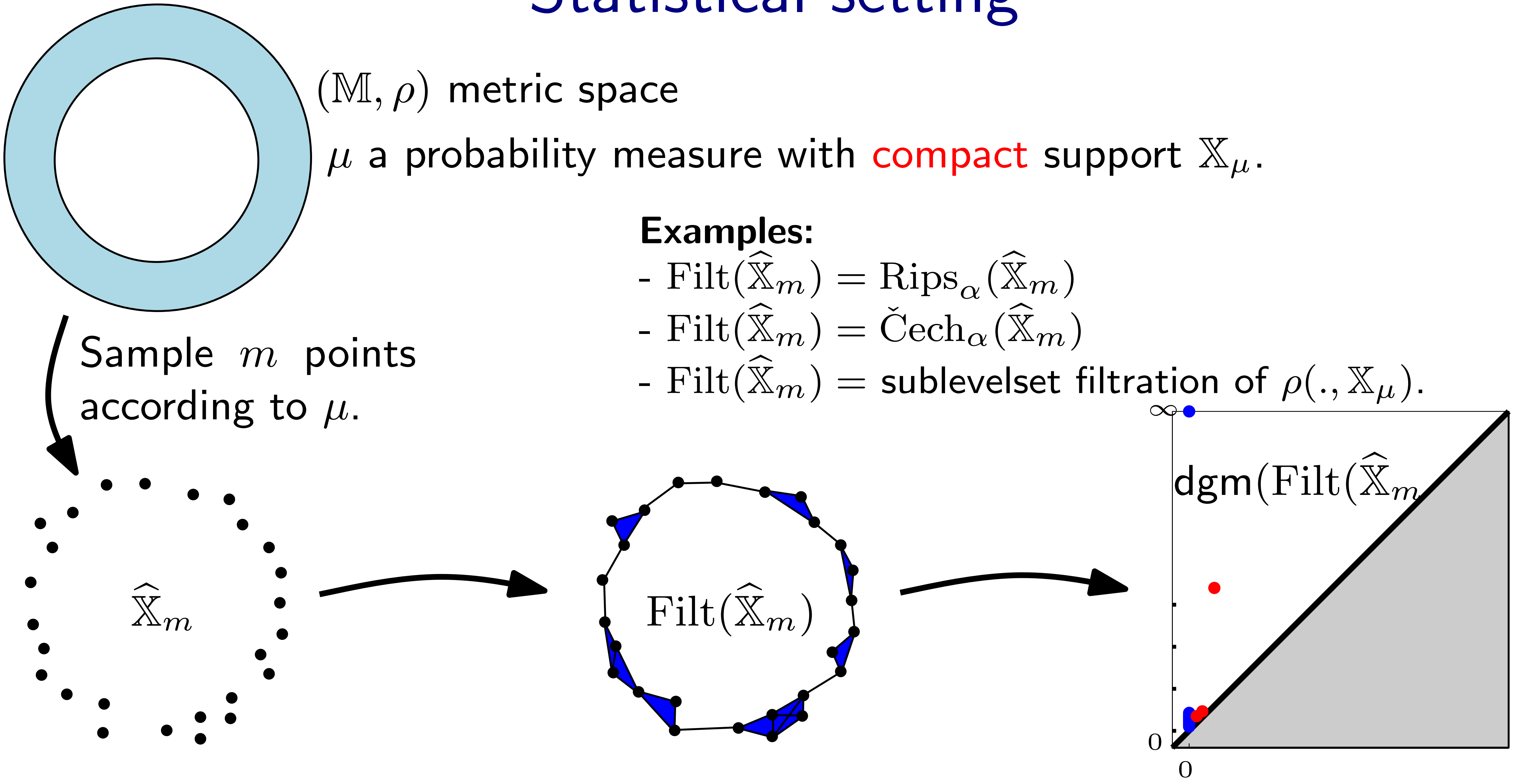


Sample m points according to μ.

Questions:

- Statistical properties of dgm$(\mathrm{Filt}(\widehat{\mathbb{X}}_m))$? 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 $\mathbb{X}_\mu$.

Sample m points according to μ.

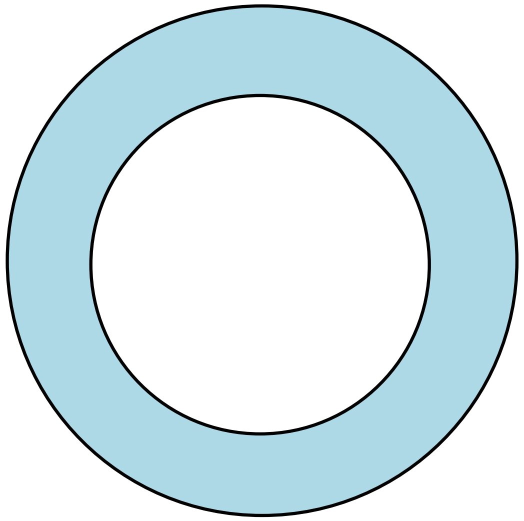
Examples:

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

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?

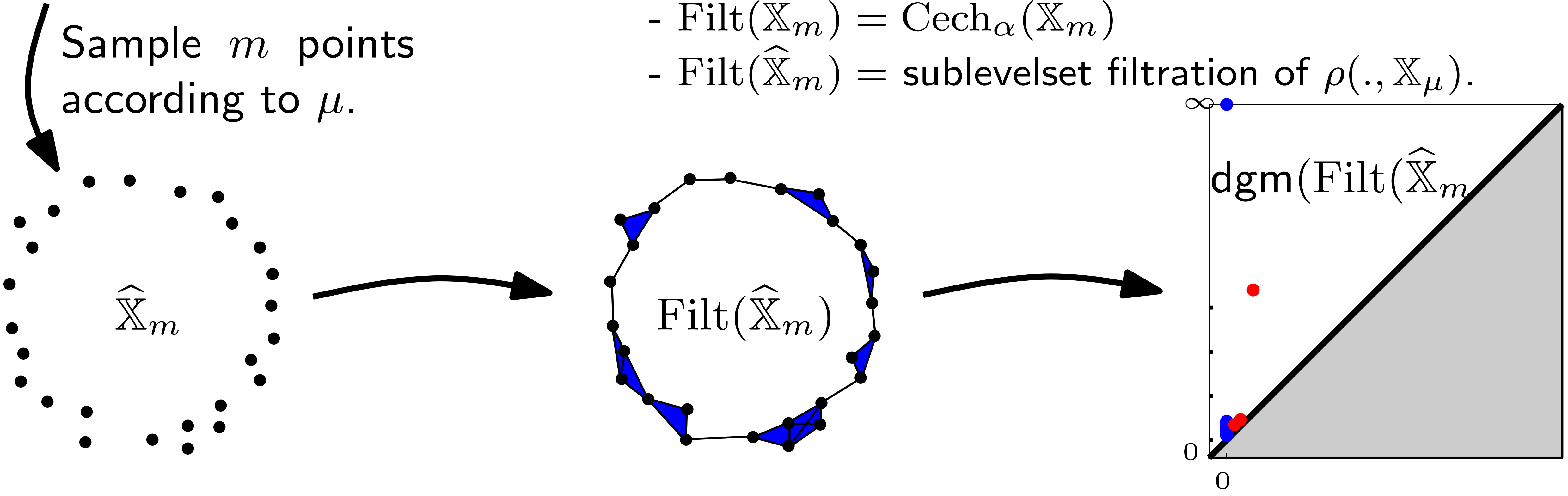
Statistical setting



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

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

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



Sample m points according to μ.

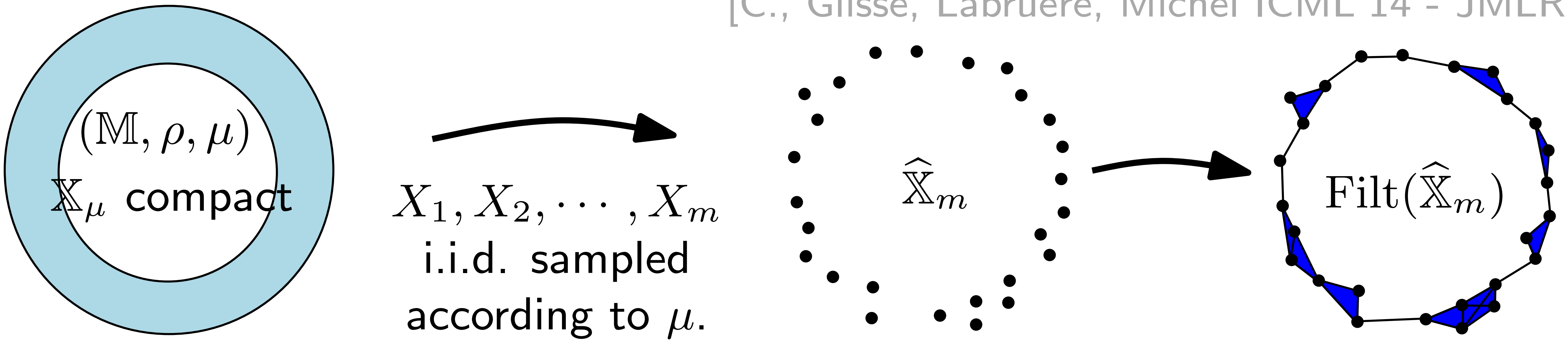
Stability thm: $\mathrm{d_b}(\text{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \text{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_m))) \leq 2d_{GH}(\mathbb{X}_\mu, \widehat{\mathbb{X}}_m)$

So, for any $\varepsilon > 0$,

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

Deviation inequality

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



$(\mathbb{M}, \rho, \mu)$

$\mathbb{X}_\mu$ compact

$X_1, X_2, \cdots, X_m$
i.i.d. sampled
according to μ.

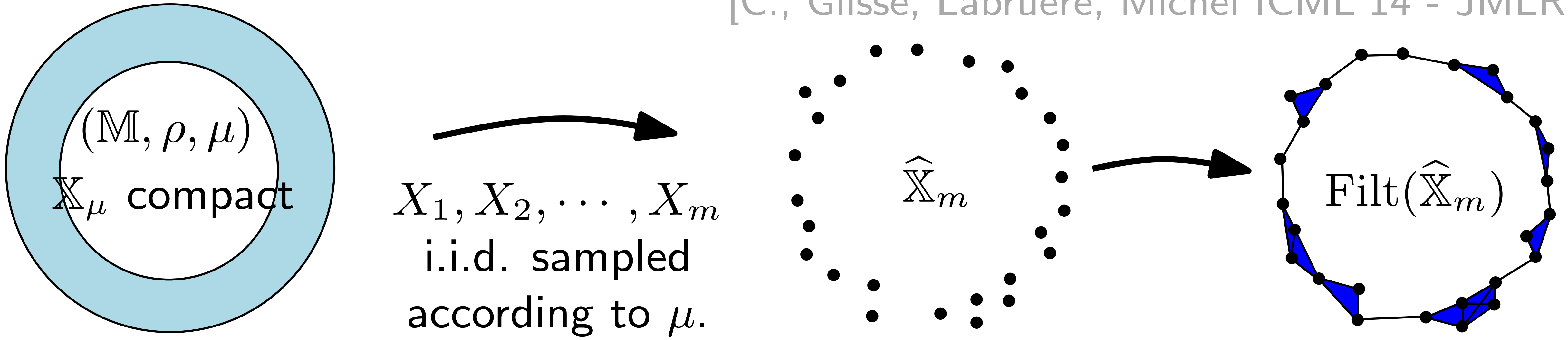
$\widehat{\mathbb{X}}_m$

$\mathrm{Filt}(\widehat{\mathbb{X}}_m)$

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]



$(\mathbb{M}, \rho, \mu)$

$\mathbb{X}_\mu$ compact

$X_1, X_2, \cdots, X_m$
i.i.d. sampled
according to μ.

$\widehat{\mathbb{X}}_m$

$\mathrm{Filt}(\widehat{\mathbb{X}}_m)$

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 $\varepsilon > 0$:

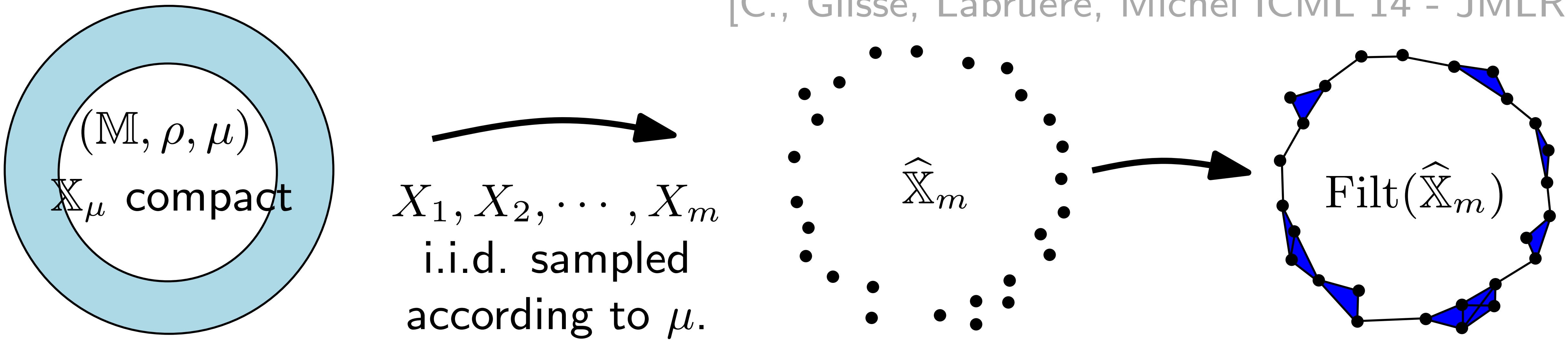
$$\mathbb{P}\left(\mathrm{d_b}\left(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathsf{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}}_m))\right) > \varepsilon\right) \leq \min(\frac{8^b}{a\varepsilon^b}\exp(-ma\varepsilon^b), 1).$$

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

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

Deviation inequality

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



$(\mathbb{M}, \rho, \mu)$

$\mathbb{X}_\mu$ compact

$X_1, X_2, \cdots, X_m$
i.i.d. sampled
according to μ.

$\widehat{\mathbb{X}}_m$

$\mathrm{Filt}(\widehat{\mathbb{X}}_m)$

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:

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

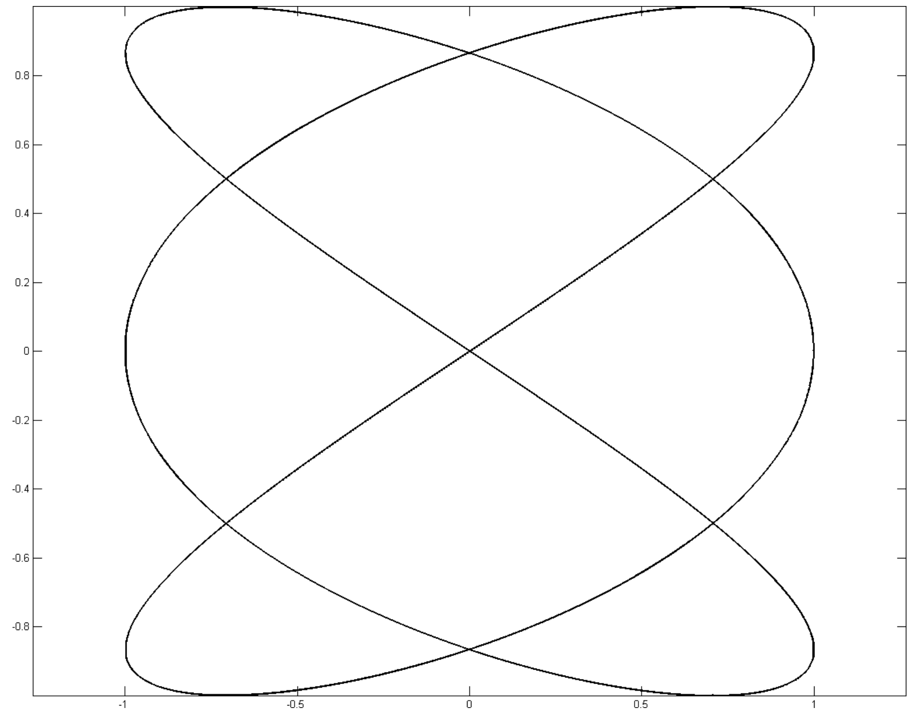
where the constant C only depends on a and b (not on $\mathbb{M}$!). Assume moreover that there exists a non isolated point x in $\mathbb{M}$ and let x_m be a sequence in $\mathbb{M} \setminus \{x\}$ such that $\rho(x, x_m) \leq (am)^{-1/b}$. Then for any estimator $\widehat{\mathsf{dgm}}_m$ of $\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu))$:

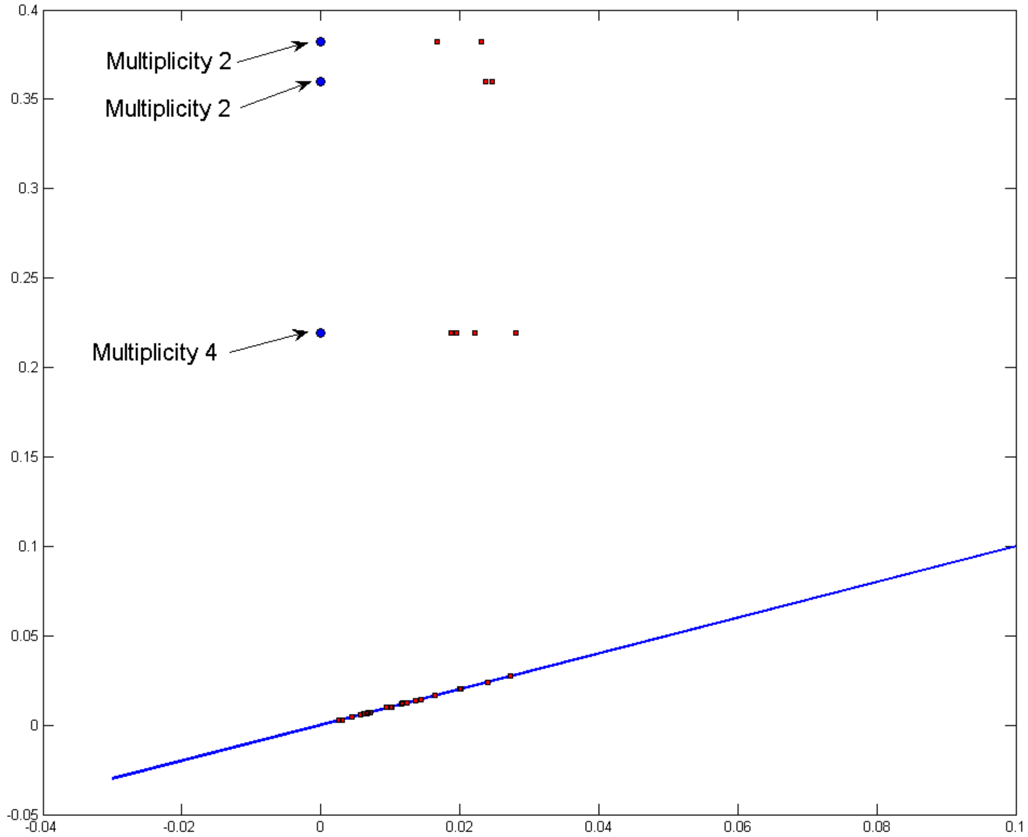
$$\liminf_{m \to \infty} \rho(x, x_m)^{-1} \sup_{\mu \in \mathcal{P}(a,b,\mathbb{M})} \mathbb{E}\left[\mathrm{d_b}(\mathsf{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \widehat{\mathsf{dgm}}_m)\right] \geq C'$$

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

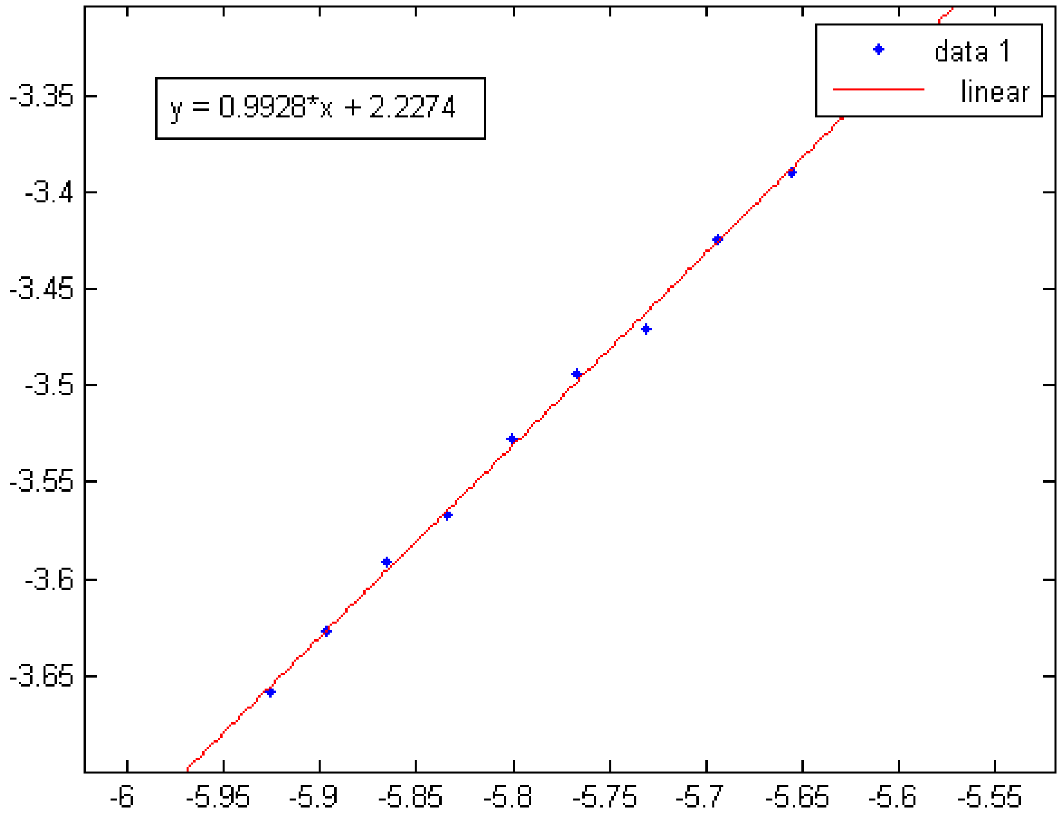




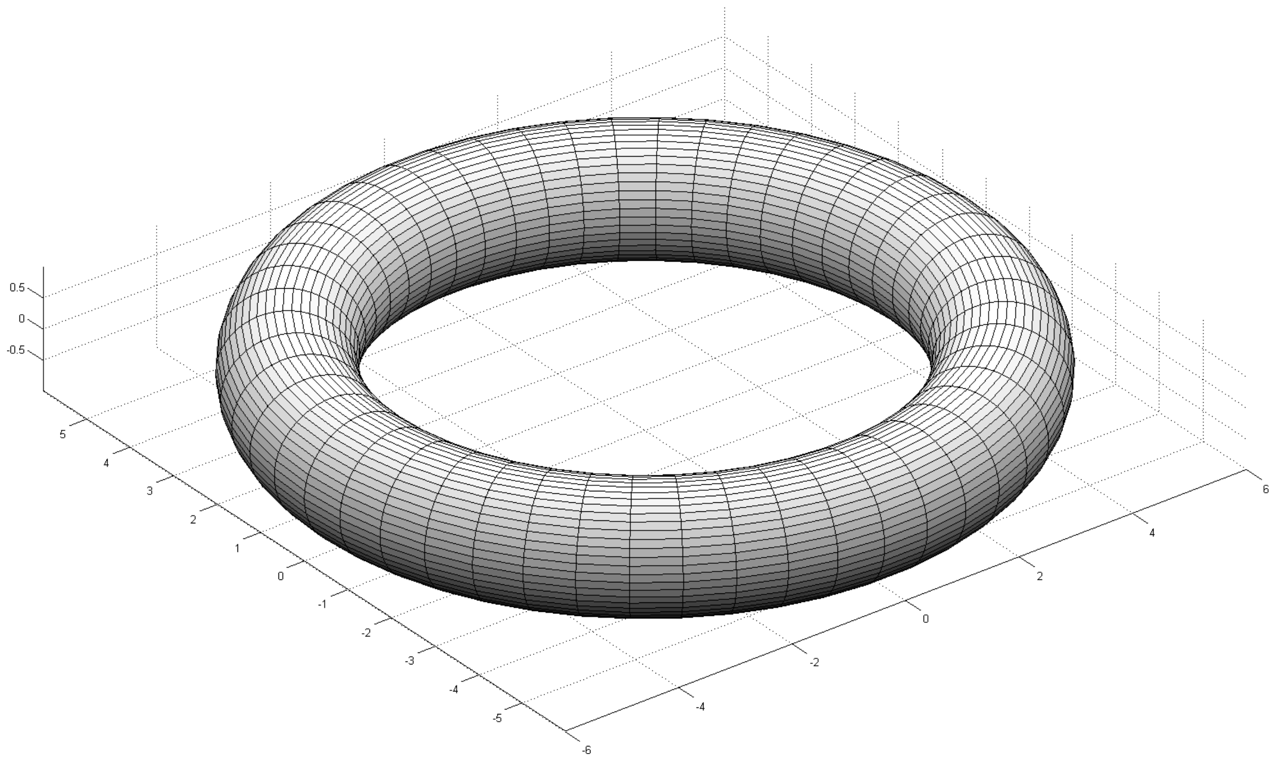
- μ: unif. measure on Lissajous curve $\mathbb{X}_\mu$.
- Filt: distance to $\mathbb{X}_\mu$ in $\mathbb{R}^2$.
- sample $k = 300$ sets of m points for $m = [2100 : 100 : 3000]$.
- compute

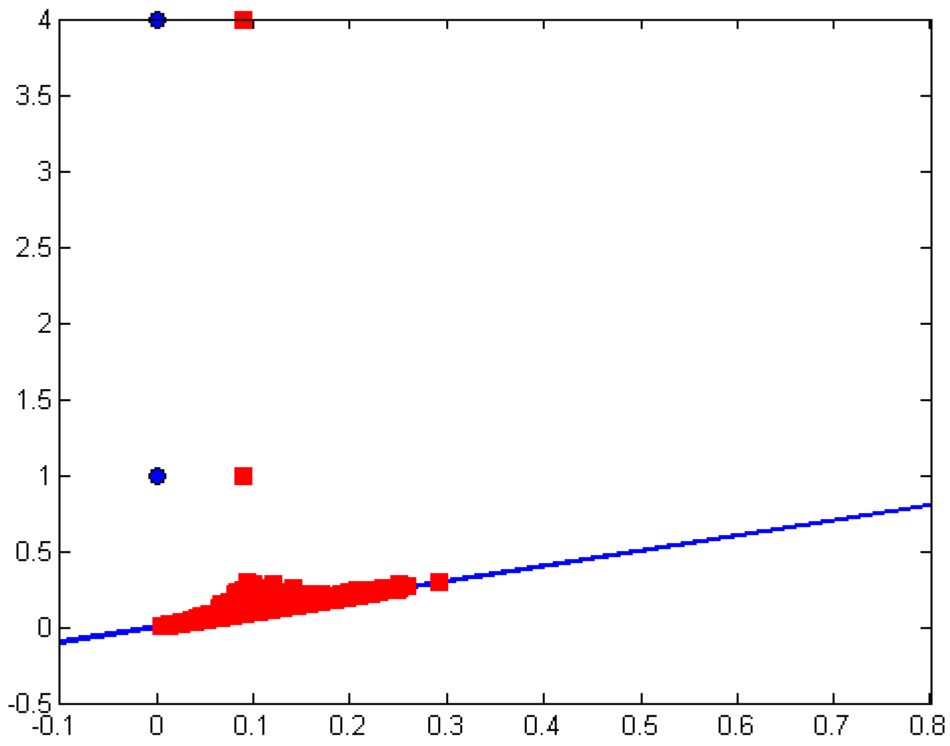
$$\widehat{\mathbb{E}}_m = \widehat{\mathbb{E}}[d_B(\mathrm{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathrm{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}_n})))]$$

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



Numerical illustrations

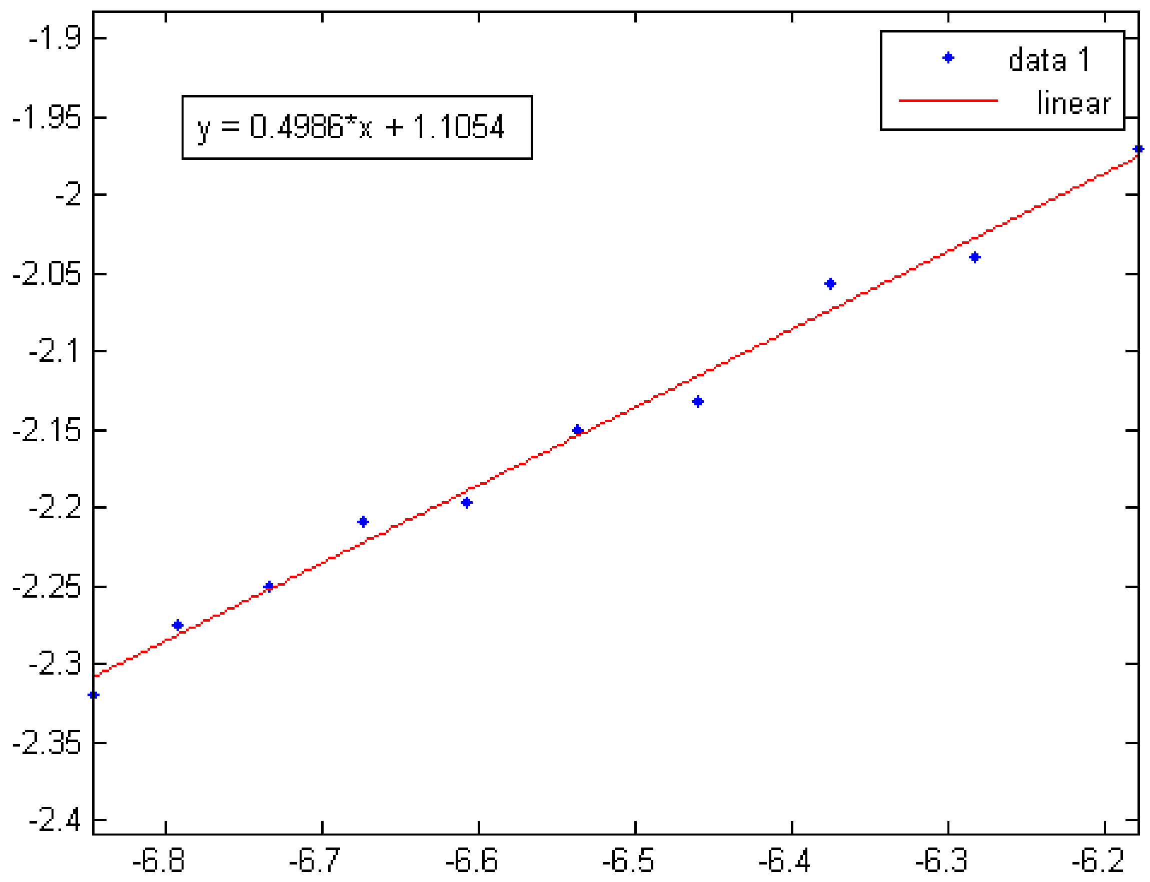




- μ: unif. measure on a torus $\mathbb{X}_\mu$.
- Filt: distance to $\mathbb{X}_\mu$ in $\mathbb{R}^3$.
- sample $k = 300$ sets of n points for $m = [12000 : 1000 : 21000]$.
- compute

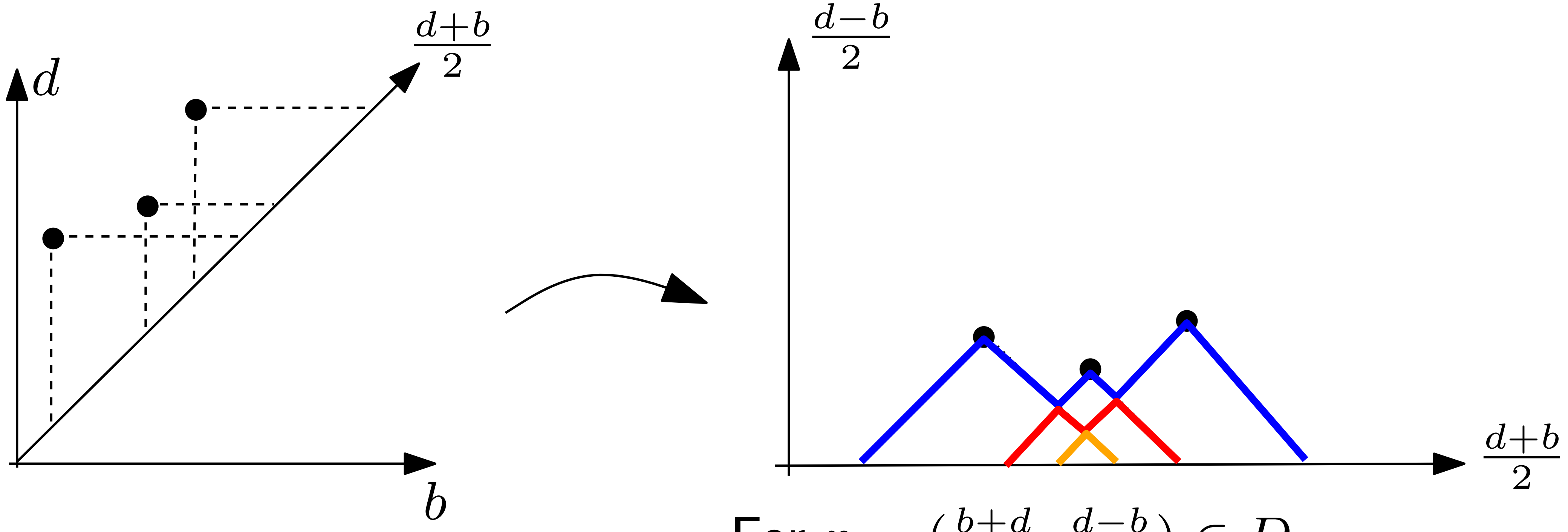
$$\widehat{\mathbb{E}}_m = \widehat{\mathbb{E}}[d_B(\mathrm{dgm}(\mathrm{Filt}(\mathbb{X}_\mu)), \mathrm{dgm}(\mathrm{Filt}(\widehat{\mathbb{X}_m})))$$

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



Persistence landscapes

[Bubenik 2012]



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

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

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

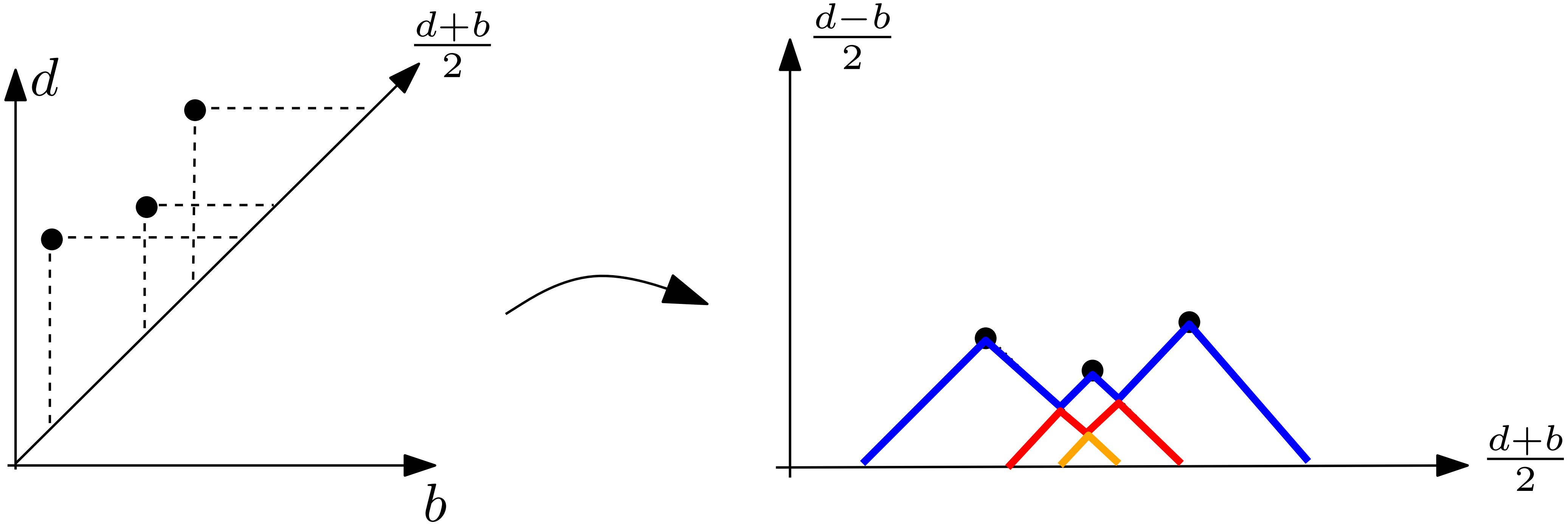
Persistence landscape [Bubenik 2012]:

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

where kmax is the kth largest value in the set.

Persistence landscapes

[Bubenik 2012]



Persistence landscape [Bubenik 2012]:

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

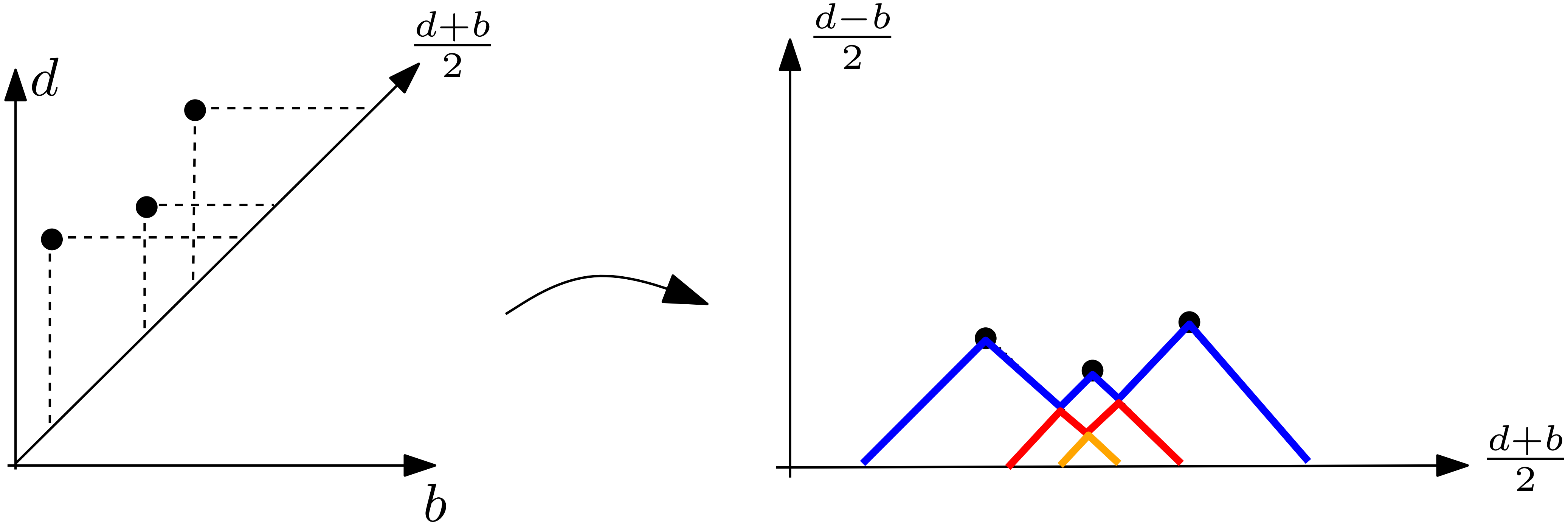
Properties

- 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

[Bubenik 2012]



- 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 → confidence intervals can be computed using bootstrap.

  [C., Fasy, Lecci, Rinaldo, Wasserman SoCG 2014]
- Provide a convenient way to process persistence information in deep neural networks.

  [Kim, Kim, Zaheer, Kim, C., Wasserman NeurIPS 2020, Carrière, C., Ike, Lacombe, Royer, Umeda AISTAT 2020]

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. $\sigma(t) = \sqrt{n\text{Var}(\overline{\lambda}_n(t))} = \sqrt{\text{Var}(f_t(\lambda_1))}$.

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 $W \overset{d}{=} \sup_{t \in [t_*, t^*]} |\mathbb{G}(f_t)|$ such that

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

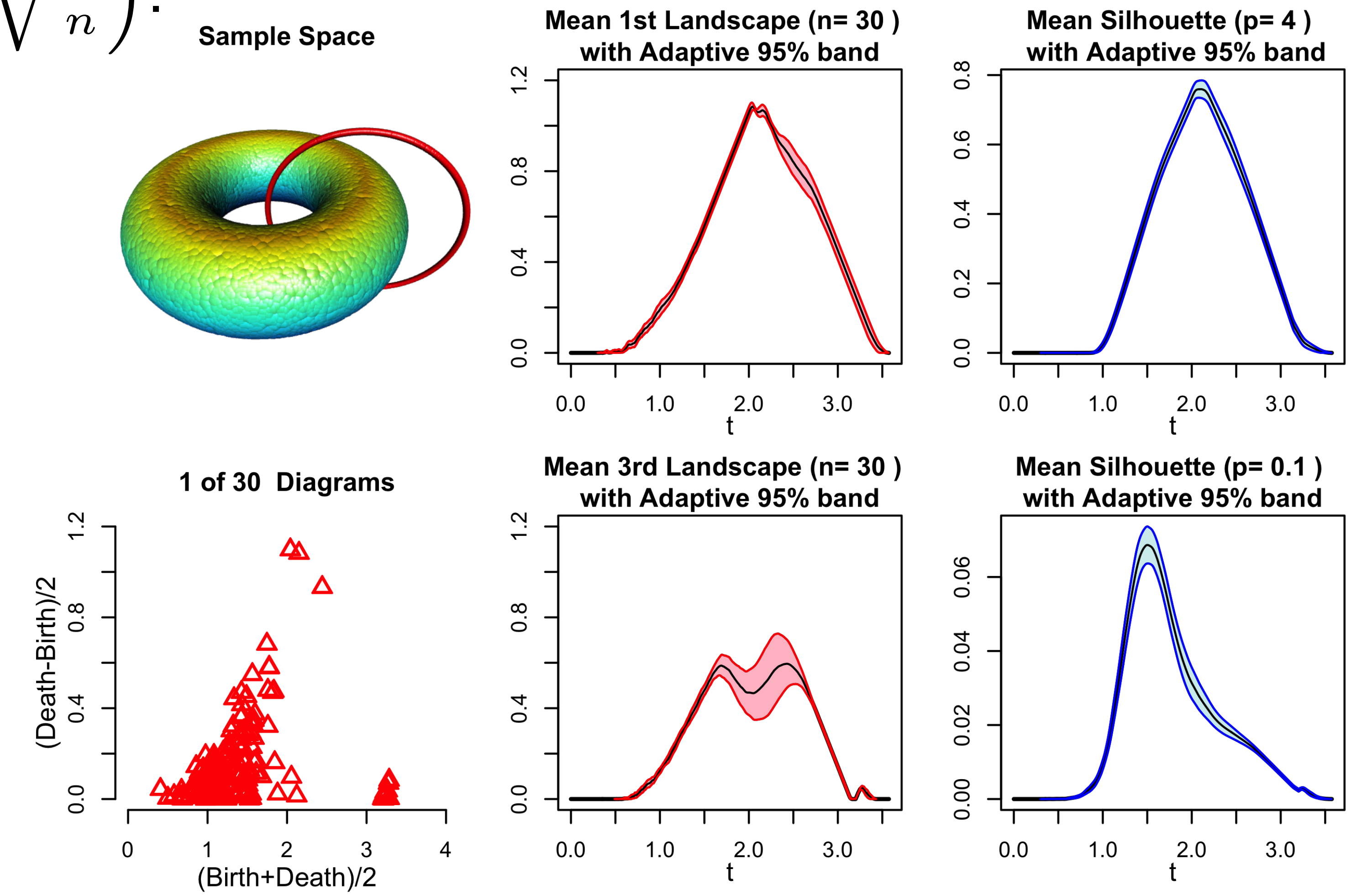
Some consequences

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



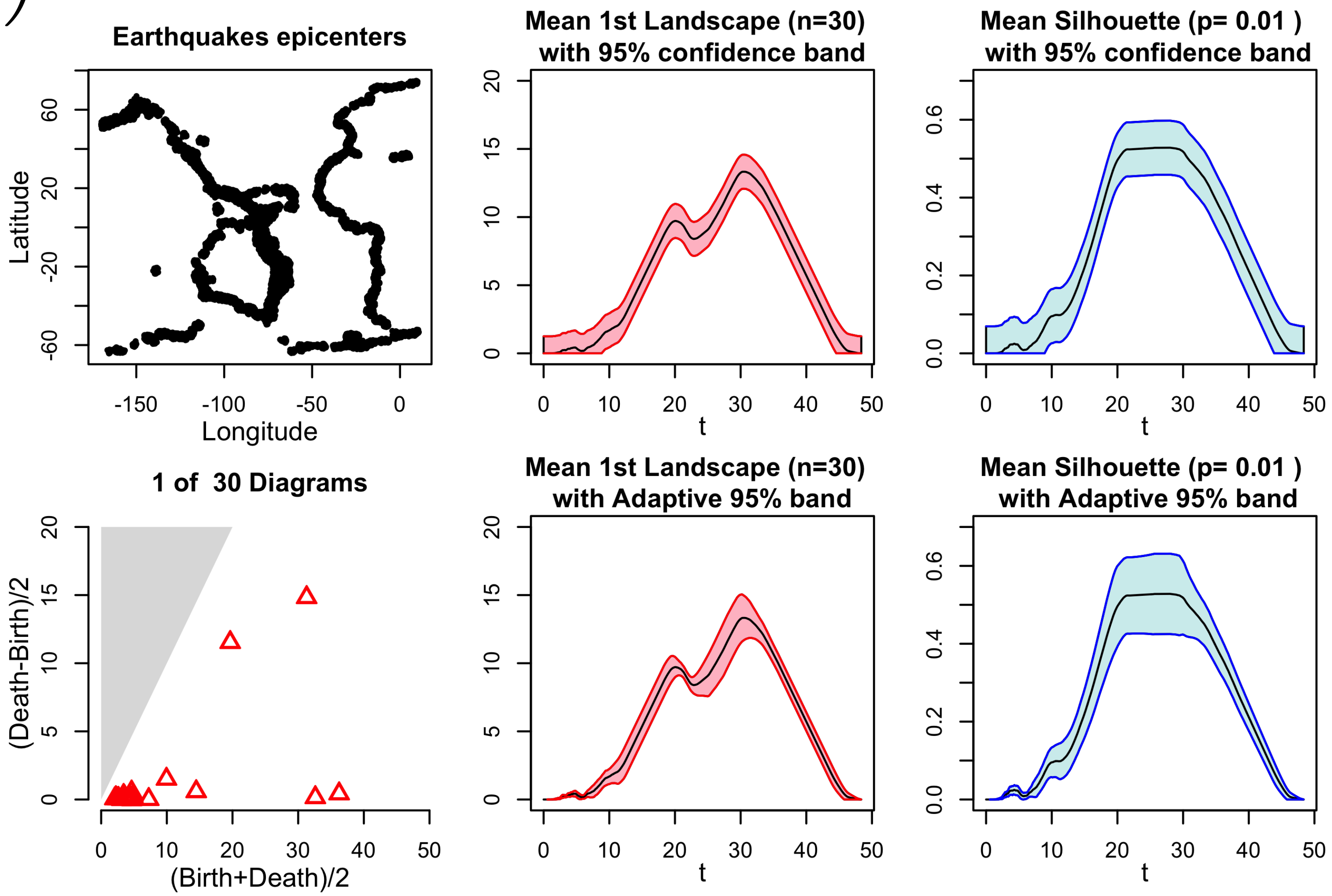
Some consequences

Bootstrap for landscapes → confidence bands for landscapes.

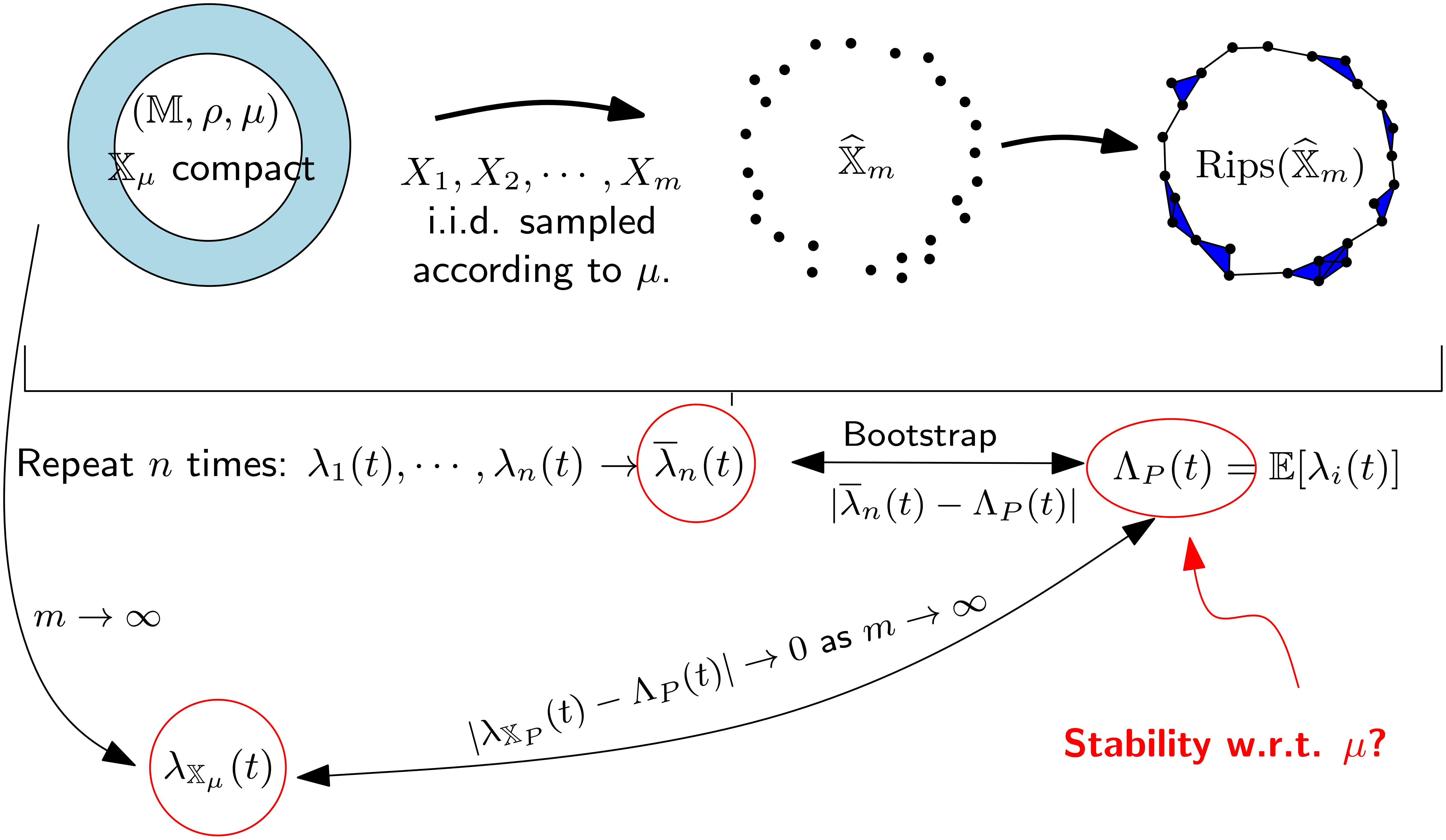
Theorem. Suppose that $\sigma(t) > c > 0$ in an interval $[t_*, t^*] \subset [0, T]$, for some constant c. Then, given a confidence level $1-\alpha$, one can construct confidence functions $\ell_n(t)$ and $u_n(t)$ such that

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

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



To summarize



$(\mathbb{M}, \rho, \mu)$

$\mathbb{X}_\mu$ compact

$X_1, X_2, \cdots, X_m$
i.i.d. sampled
according to μ.

$\widehat{\mathbb{X}}_m$

$\text{Rips}(\widehat{\mathbb{X}}_m)$

Repeat n times: $\lambda_1(t), \cdots, \lambda_n(t) \to \overline{\lambda}_n(t)$

Bootstrap

$|\overline{\lambda}_n(t) - \Lambda_P(t)|$

$\Lambda_P(t) = \mathbb{E}[\lambda_i(t)]$

$m \to \infty$

$|\lambda_{\mathbb{X}_P}(t) - \Lambda_P(t)| \to 0$ as $m \to \infty$

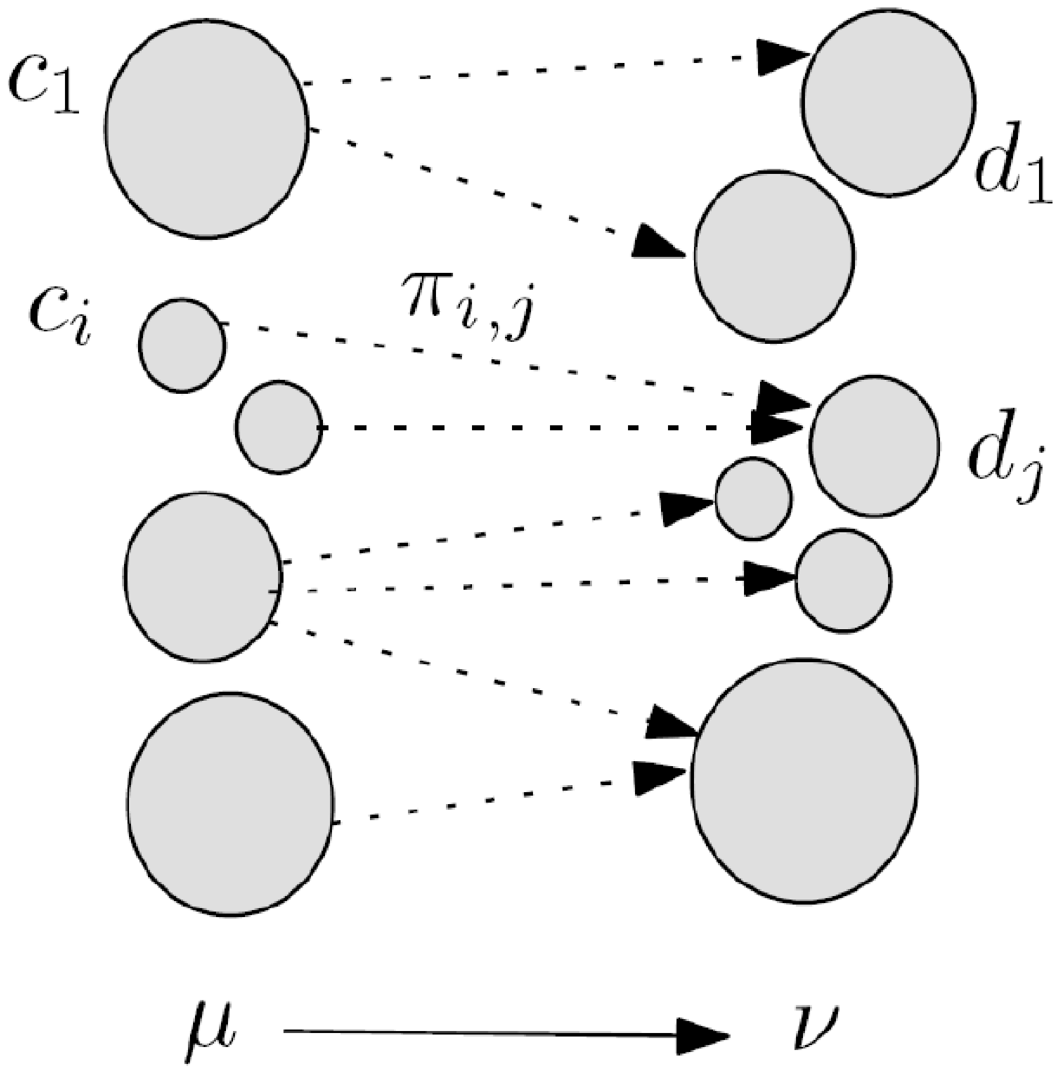
$\lambda_{\mathbb{X}_\mu}(t)$

Stability w.r.t. μ?

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 $\rho(x,y)^p dx$.

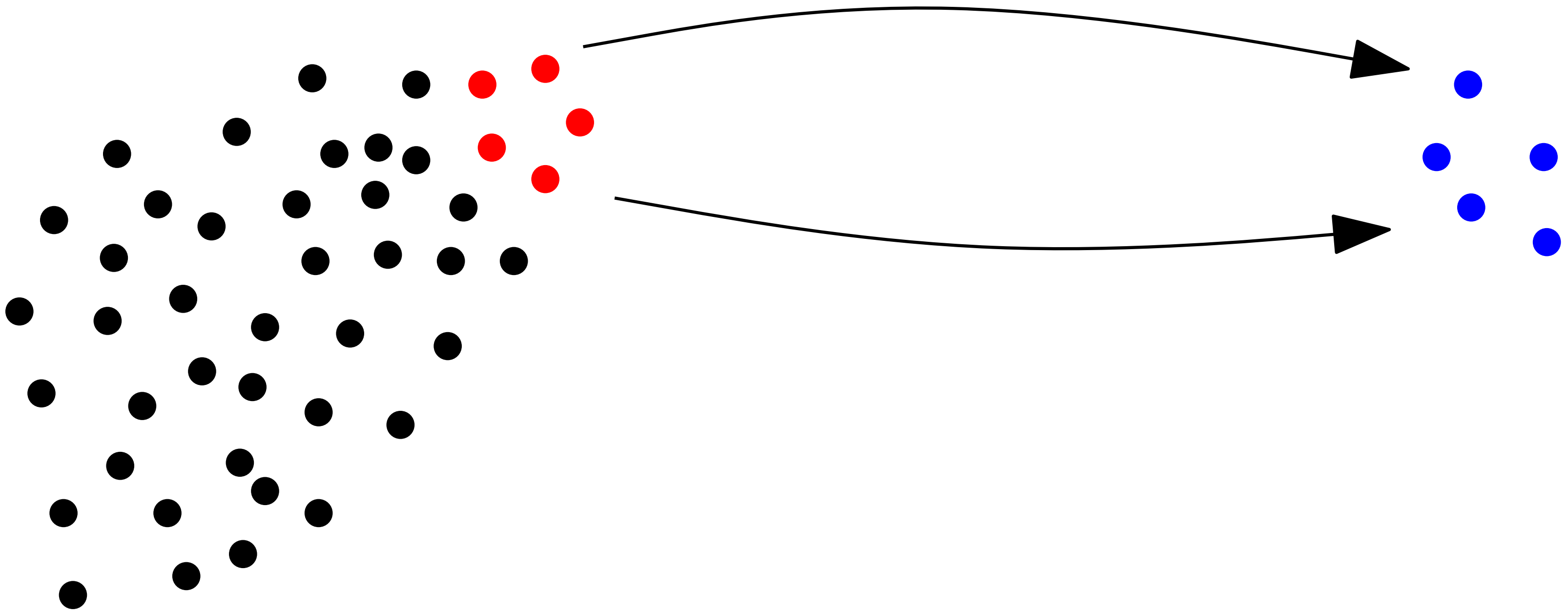


- Transport plan: Π a proba measure on $M \times M$ such that $\Pi(A \times \mathbb{R}^d) = \mu(A)$ and $\Pi(\mathbb{R}^d \times B) = \nu(B)$ for any borelian sets $A, B \subset M$.
- Cost of a transport plan:

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

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

Wasserstein distance



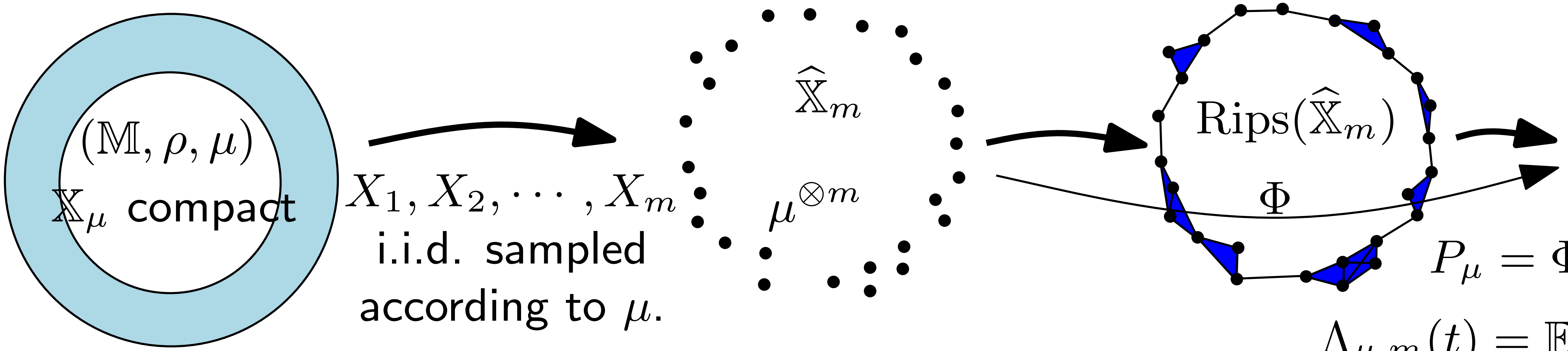
Example:

- 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]



$(\mathbb{M}, \rho, \mu)$

$\mathbb{X}_\mu$ compact

$X_1, X_2, \cdots, X_m$ i.i.d. sampled according to μ.

$\widehat{\mathbb{X}}_m$

$\mu^{\otimes m}$

$\mathrm{Rips}(\widehat{\mathbb{X}}_m)$

Φ

$\lambda_{\mathrm{Rips}(\widehat{\mathbb{X}}_r}$

$P_\mu = \Phi_*(\mu^{\otimes m})$

$\Lambda_{\mu,m}(t) = \mathbb{E}_{P_\mu}[\lambda(t)]$

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

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

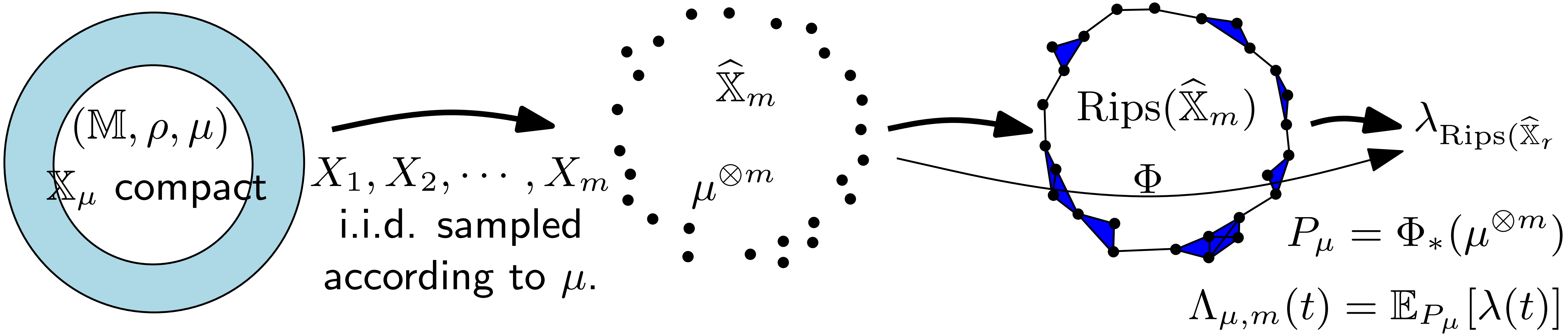
where W_p denotes the Wasserstein distance with cost function $\rho(x, y)^p$.

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]



$(\mathbb{M}, \rho, \mu)$

$\mathbb{X}_\mu$ compact

$X_1, X_2, \cdots, X_m$ i.i.d. sampled according to μ.

$\widehat{\mathbb{X}}_m$

$\mu^{\otimes m}$

$\mathrm{Rips}(\widehat{\mathbb{X}}_m)$

$\lambda_{\mathrm{Rips}(\widehat{\mathbb{X}}_r}$

Φ

$P_\mu = \Phi_*(\mu^{\otimes m})$

$\Lambda_{\mu,m}(t) = \mathbb{E}_{P_\mu}[\lambda(t)]$

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

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

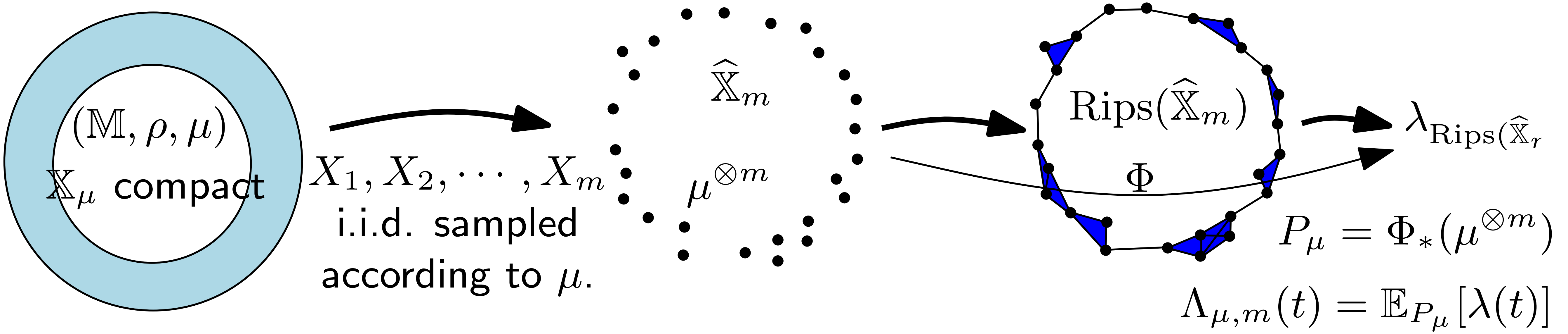
where W_p denotes the Wasserstein distance with cost function $\rho(x,y)^p$.

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]



$P_\mu = \Phi_*(\mu^{\otimes m})$

$\Lambda_{\mu,m}(t) = \mathbb{E}_{P_\mu}[\lambda(t)]$

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 $\rho(x,y)^p$.

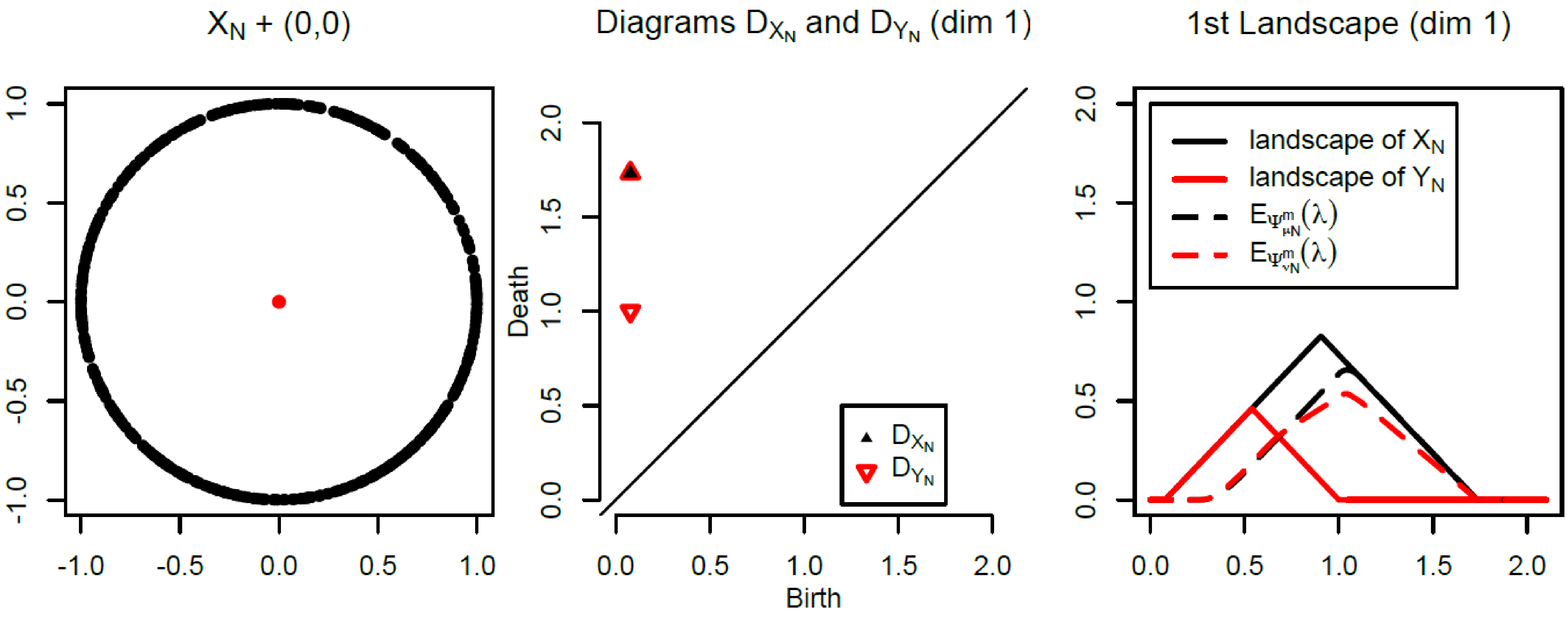
Proof:

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]

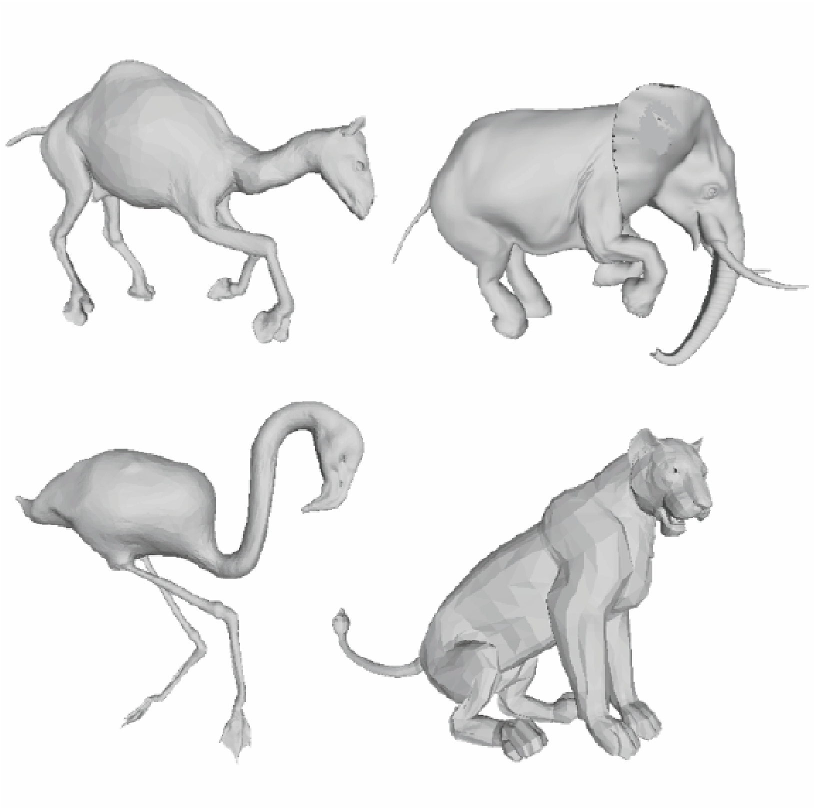
Example: Circle with one outlier.

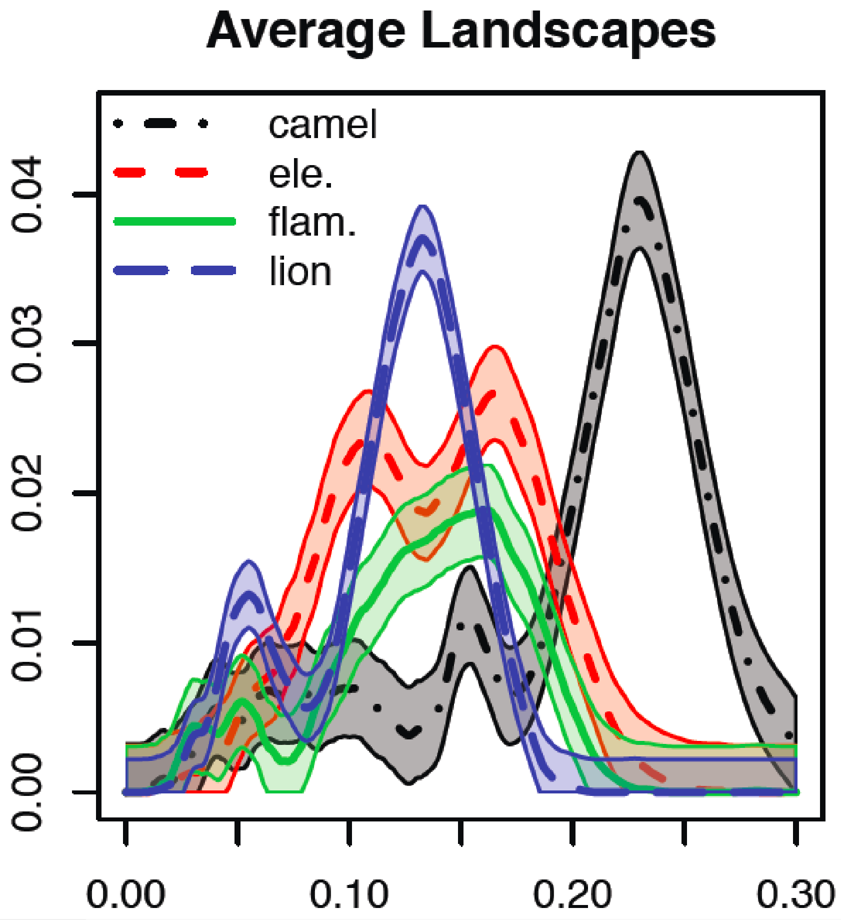


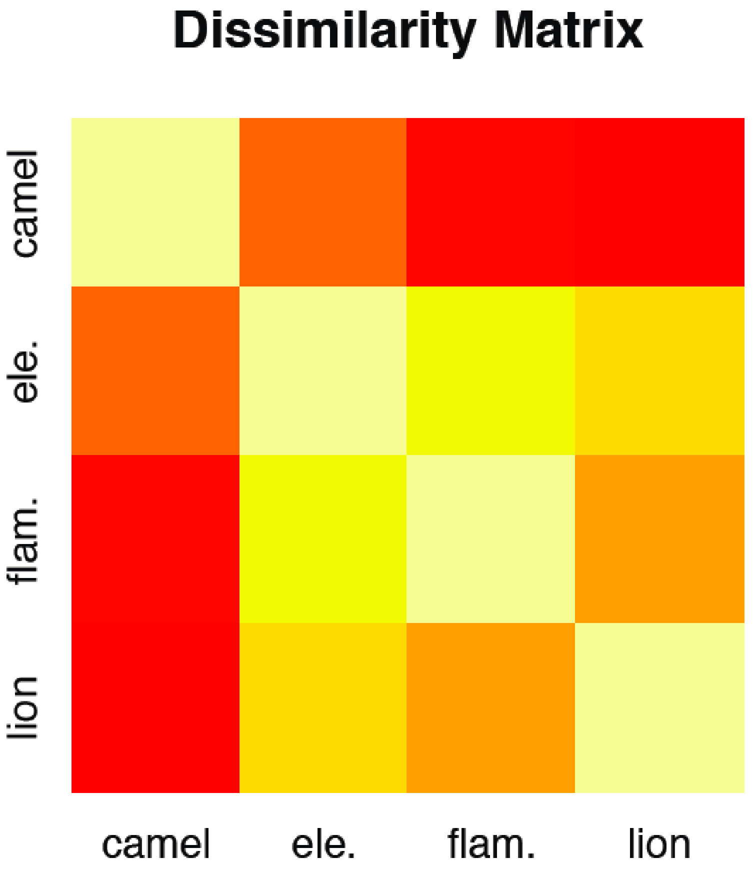
(Sub)sampling and stability of expected landscapes

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

Example: 3D shapes





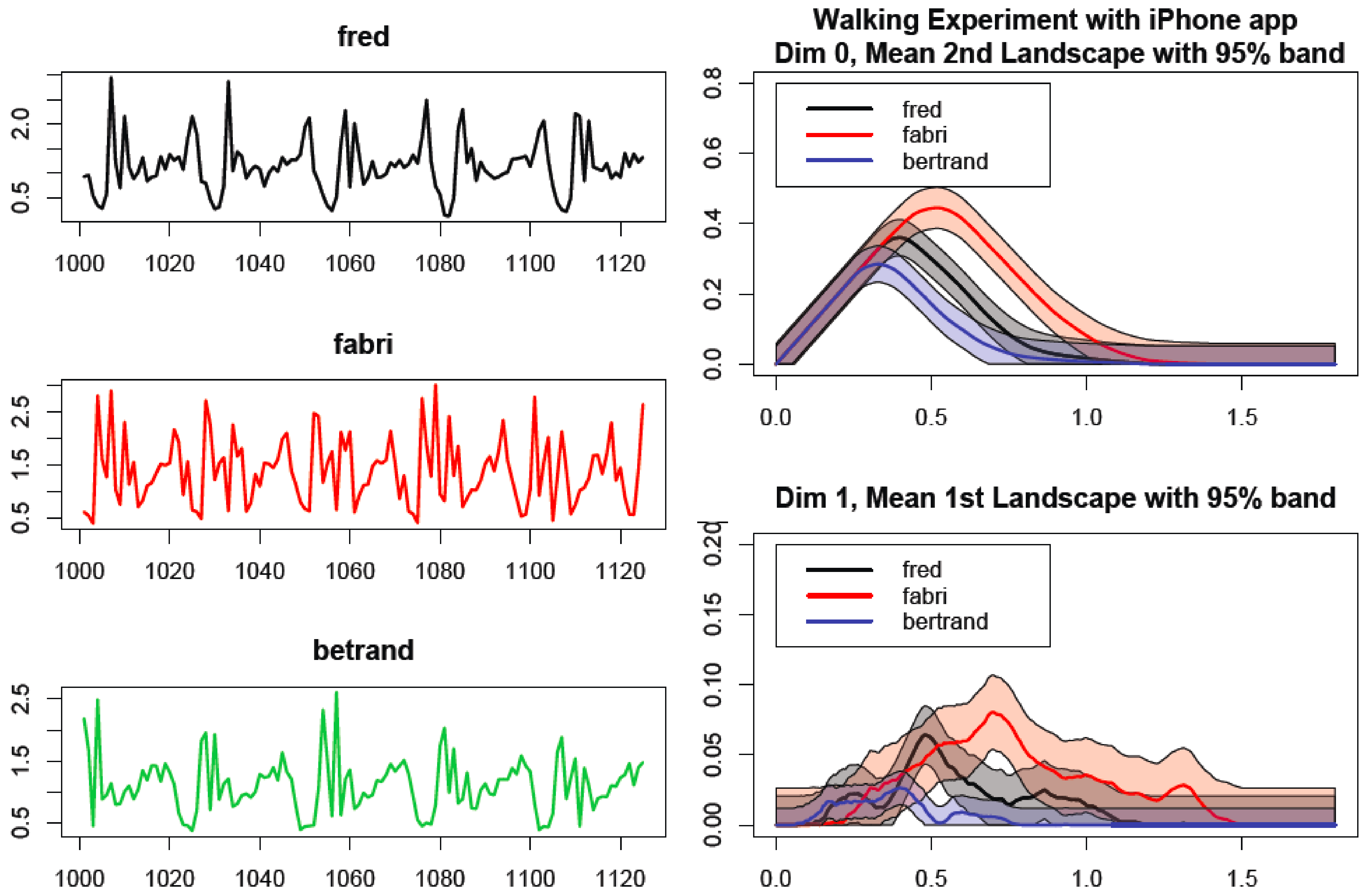


From $n = 100$ subsamples of size $m = 300$

(Sub)sampling and stability of expected landscapes

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

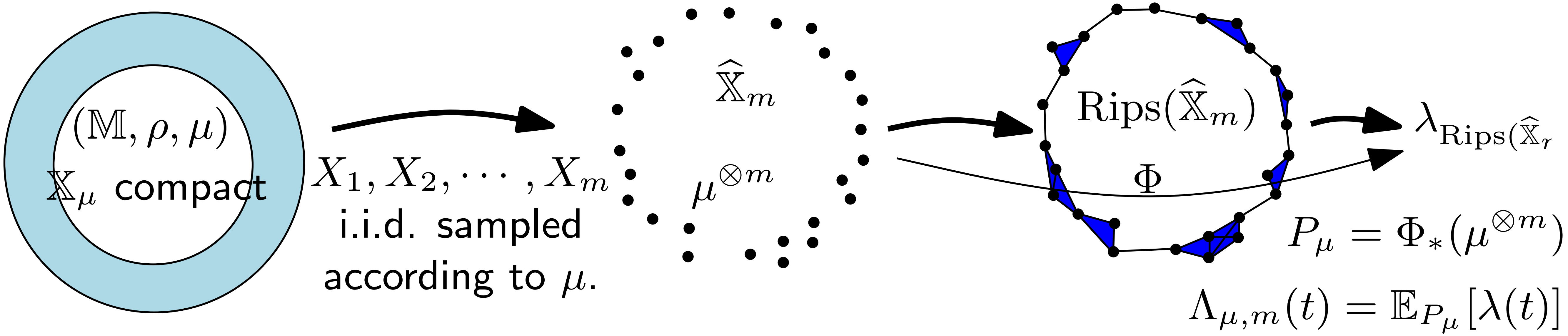
(Toy) Example: Accelerometer data from smartphone.



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

(Sub)sampling and stability of expected landscapes

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



$(\mathbb{M}, \rho, \mu)$

$\mathbb{X}_\mu$ compact

$X_1, X_2, \cdots, X_m$
i.i.d. sampled
according to μ.

$\widehat{\mathbb{X}}_m$

$\mu^{\otimes m}$

$\mathrm{Rips}(\widehat{\mathbb{X}}_m)$

Φ

$\lambda_{\mathrm{Rips}(\widehat{\mathbb{X}}_r}$

$$P_\mu = \Phi_*(\mu^{\otimes m})$$

$$\Lambda_{\mu,m}(t) = \mathbb{E}_{P_\mu}[\lambda(t)]$$

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)

Proof

Lemma 1: For any $\mu, \nu \in \mathcal{P}(\mathbb{M})$,

$$W_p(\mu^{\otimes m}, \nu^{\otimes m}) \le m^{\frac{1}{p}} W_p(\mu, \nu)$$

where the metric ρ_m in $\mathbb{M}^m$ is any metric satisfying for any $X = (x_1, \cdots x_m)$, $Y = (y_1, \cdots y_m)$,

$$\rho_m(X, Y) \le \left(\sum_{i=1}^{m} \rho(x_i, y_i)^p \right)^{\frac{1}{p}}$$

Proof: If $\Pi \in \mathcal{P}(\mathbb{M} \times \mathbb{M})$ is a transport plan between μ and ν, then $\Pi^{\otimes m}$ is a transport plan between $\mu^{\otimes m}$ and $\nu^{\otimes m}$ (up to reordering the comp. of $\mathbb{M}^{2m}$ and

$$\begin{aligned} \int_{\mathbb{M}^{2m}} \rho_m(X, Y)^p d\Pi^{\otimes m}(X, Y) &\le \int_{\mathbb{M}^m \times \mathbb{M}^m} \sum_{i=1}^{m} \rho(x_i, y_i)^p \, d\Pi(x_1, y_1) \cdots d\Pi(x_m, y_m) \\ &= m \int_{\mathbb{M} \times \mathbb{M}} \rho(x_1, y_1)^p d\Pi(x_1, y_1). \end{aligned}$$

Proof

Lemma 2:

$$W_p(\phi_\mu^m, \phi_\nu^m) \leq W_p(\mu^{\otimes m}, \nu^{\otimes m})$$

where $\phi^m : \mathbb{M}^m \to \mathcal{D}$, $\phi^m(X) = \text{dgm}(\text{Filt}(X))$, $\mathcal{D}$ is the space of persistence diagrams endowed with the bottleneck distance and $\phi_\mu^m = (\phi^m)_*\mu$, $\phi_\nu^m = (\phi^m)_*\nu$.

Notations:

- $\Lambda_m : \mathbb{M}^m \times \mathbb{M}^m \to \mathcal{D} \times \mathcal{D}$, $\Lambda_m(X, Y) = (\psi(\phi^m(X)), \psi(\phi^m(X)))$.

Proof: if $\Pi \in \mathcal{P}(\mathbb{M}^m \times \mathbb{M}^m)$ is a transport plan between $\mu^{\otimes m}$ and $\nu^{\otimes m}$ then $\Lambda_{m,*}\Pi$ is a transport plan between Φ_μ^m and Φ_ν^m and

$$\begin{aligned}
\int_{\mathcal{D}_T^2} d_b(D_X, D_Y)^p d\Lambda_{m,*}\Pi(D_X, D_Y) &= \int_{\mathbb{M}^{2m}} d_b(\phi_m(X), \phi_m(Y))^p d\Pi(X, Y) \\
&\leq \int_{\mathbb{M}^{2m}} d_H(X, Y)^p d\Pi(X, Y) \quad \text{(stab.thm)} \\
&\leq \int_{\mathbb{M}^{2m}} \rho_m(X, Y)^p d\Pi(X, Y).
\end{aligned}$$

Proof

Notations:

- $\mathcal{L}$: space of landscapes (with sup. norm)
- $\psi : \mathcal{D} \to \mathcal{L}$, $\psi(D) = \lambda_D$
- $\Psi_\mu^m = \psi_* \phi_\mu^m$, $\Psi_\nu^m = \psi_* \phi_\nu^m$

Lemma 3: Let $\lambda_X \sim \Psi_\mu^m$ and $\lambda_Y \sim \Psi_\nu^m$. Then

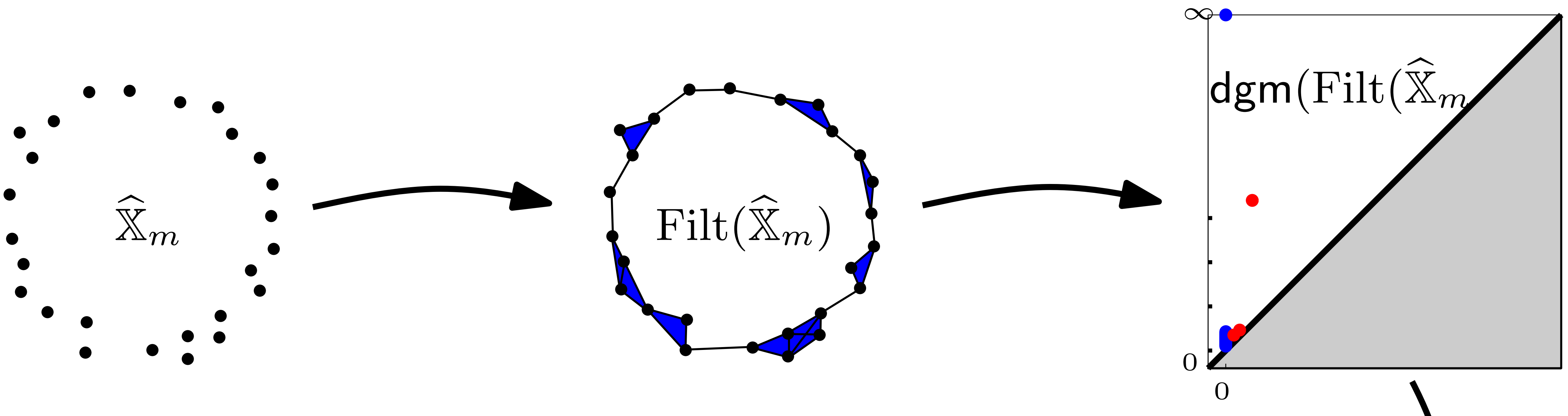
$$\left\| \mathbb{E}_{\Psi_\mu^m}[\lambda_X] - \mathbb{E}_{\Psi_\nu^m}[\lambda_Y] \right\|_\infty \leq W_{d_b,p}\left(\Phi_\mu^m,\ \Phi_\nu^m\right).$$

Proof: Let Π be a transport plan between Φ_μ^m and Φ_ν^m. For any $t \in \mathbb{R}$ we have

$$\begin{aligned}
\left| \mathbb{E}_{\Psi_\mu^m}[\lambda_X](t) - \mathbb{E}_{\Psi_\nu^m}[\lambda_Y](t) \right|^p &= |\mathbb{E}[\lambda_X(t) - \lambda_Y(t)]|^p \\
&\leq \mathbb{E}\left[|\lambda_X(t) - \lambda_Y(t)|^p\right] \quad \text{(Jensen inequality)} \\
&\leq \mathbb{E}\left[d_b(D_X, D_Y)^p\right] \quad \text{(Stability of landscapes)} \\
&= \int_{\mathcal{D}_T \times \mathcal{D}_T} d_b(D_X, D_Y)^p d\Pi(D_X, D_Y)
\end{aligned}$$

TDA and Machine Learning
The problem of representation of persistent homology

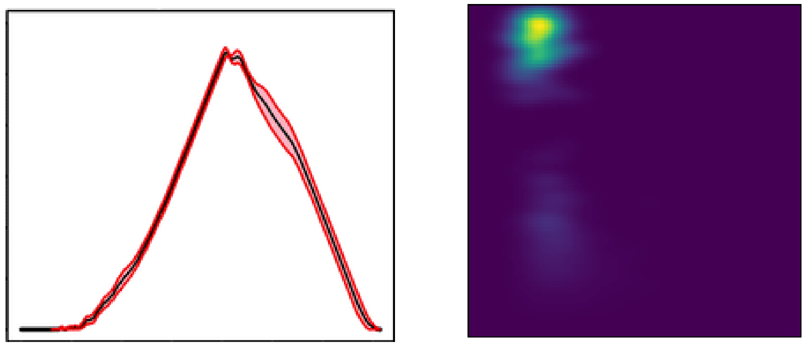
The problem of representation of persistence



Persistence diagrams are not well-suited for classical ML algorithms (the space of PD is highly non linear)

Not always clear which part of the diagrams carries the relevant information.

Machine Learning / AI



Representations of persistence

A zoo of representations of persistence

(non exhaustive list - see also Gudhi representations)

- Collections of 1D functions

  → landscapes [Bubenik 2012]

  → Betti curves [Umeda 2017]

- discrete measures: (interesting statistical properties [Chazal, Divol 2018])

  → persistence images [Adams et al 2017]

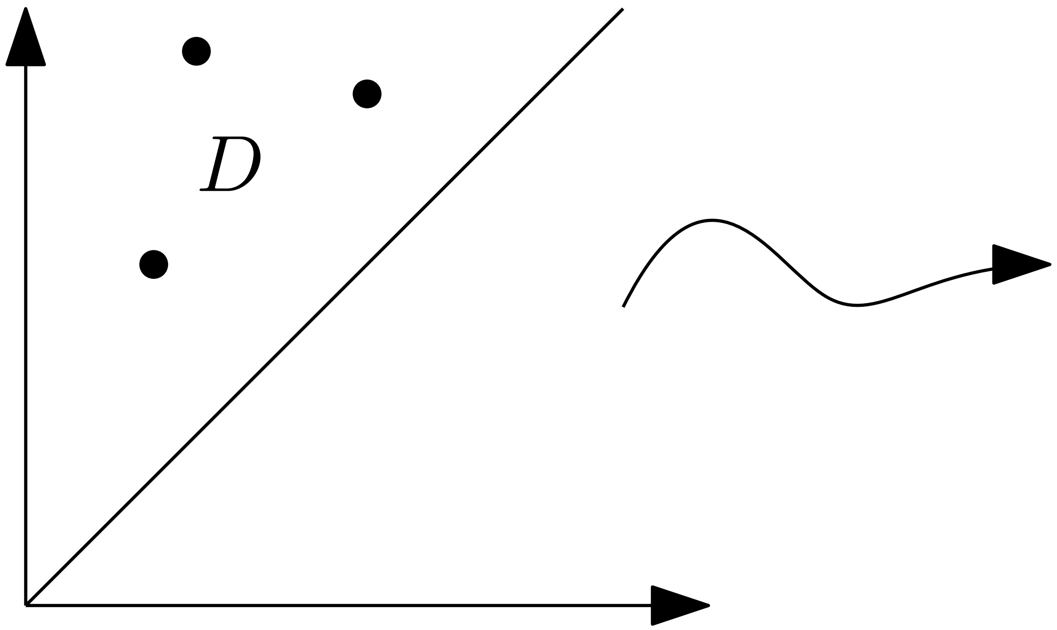
  → convolution with Gaussian kernel [Reininghaus et al. 2015] [Chepushtanova et al. 2015] [Kusano Fukumisu Hiraoka 2016-17] [Le Yamada 2018]

  → sliced on lines [Carrière Oudot Cuturi 2017]

- finite metric spaces [Carrière Oudot Ovsjanikov 2015]

- polynomial roots or evaluations [Di Fabio Ferri 2015] [Kališnik 2016]

Persistence diagrams as discrete measures



$$D := \sum_{p \in D} \delta_p$$

Motivations:

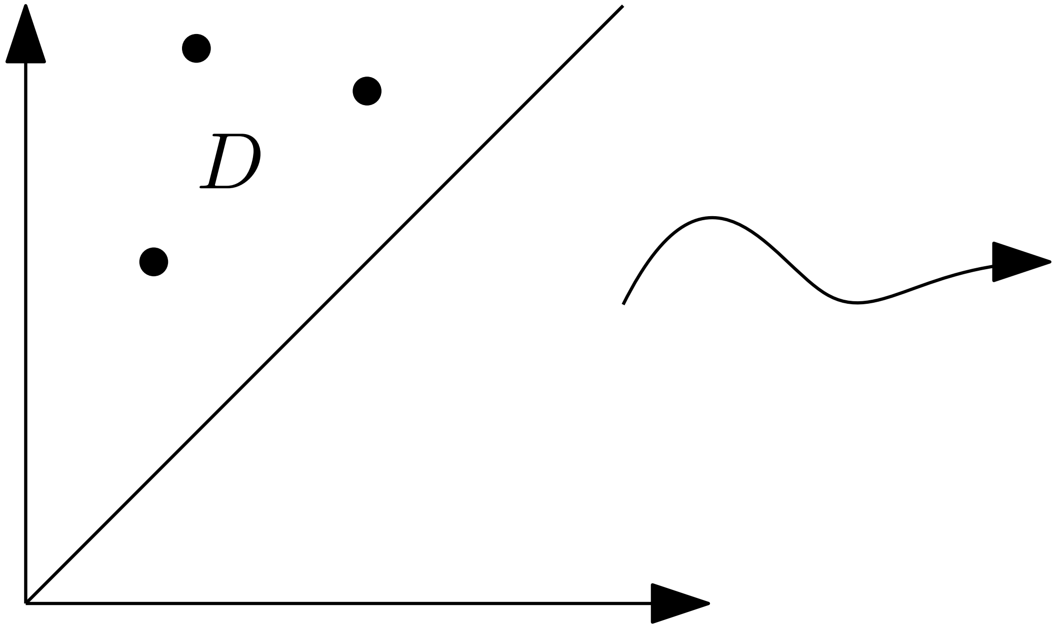
- The space of measures is much nicer that the space of P. D. !
- In the general algebraic persistence theory, persistence diagrams naturally appears as discrete measures in the plane.

[C., de Silva, Glisse, Oudot 16]

- Many persistence representations can be expressed as

$$D(f) = \sum_{p \in D} f(p) = \int f dD$$

Persistence diagrams as discrete measures



$$D := \sum_{p \in D} \delta_p$$

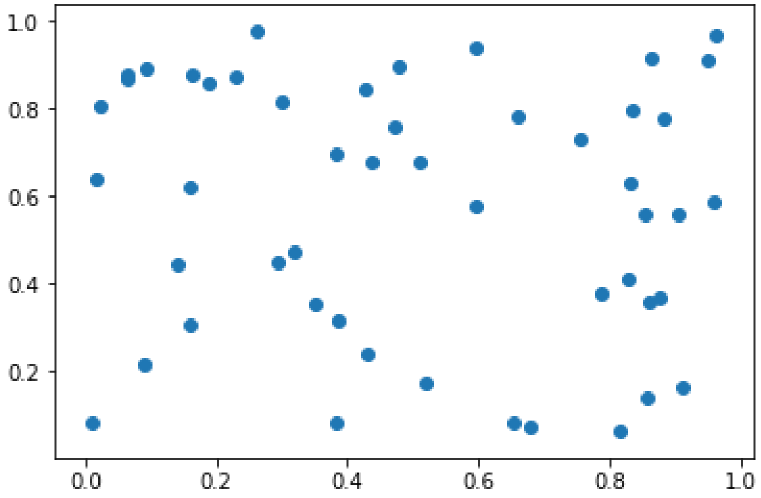
Benefits:

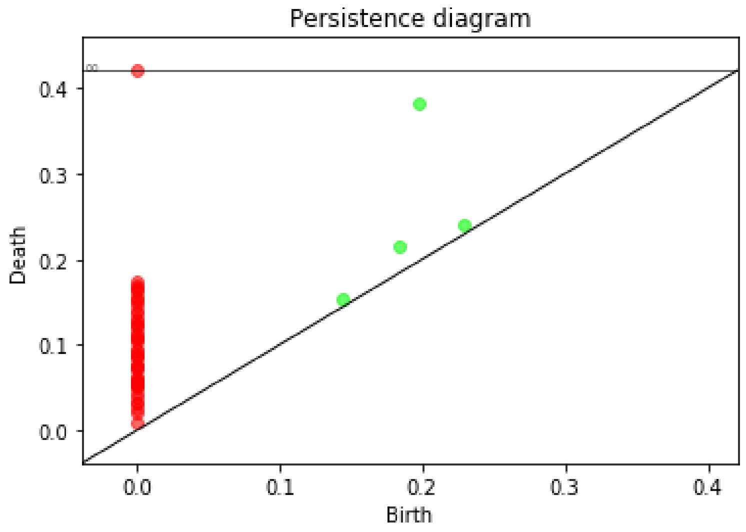
- Interesting statistical properties
- Data-driven selection of well-adapted representations (supervised and unsupervised, coming with guarantees)
- Optimisation of persistence-based functions

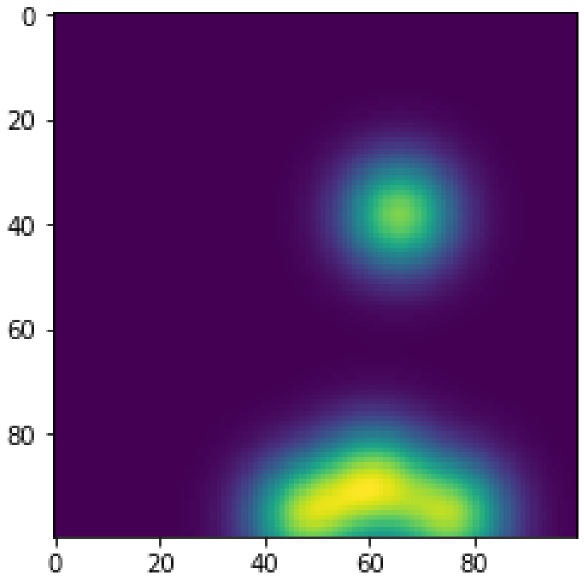
Many tools available and implemented in the GUDHI library

Persistence images

[Adams et al, JMLR 2017]







For $K : \mathbb{R}^2 \to \mathbb{R}$ a kernel and H a bandwidth matrix (e.g. a symmetric positive definite matrix), pose for $u \in \mathbb{R}^2$, $K_H(u) = |H|^{-1/2} K(H^{-1/2} \cdot u)$

For $D = \sum_i \delta_{p_i}$ a diagram, $K : \mathbb{R}^2 \to \mathbb{R}$ a kernel, H a bandwidth matrix and $w : \mathbb{R}^2 \to \mathbb{R}_+$ a weight function, one defines the persistence surface of D with kernel K and weight function w by:

$$\forall u \in \mathbb{R}^2,\ \rho(D)(u) = \sum_i w(p_i) K_H(u - p_i) = D(w K_H(u - \cdot))$$

A zoo of representations of persistence

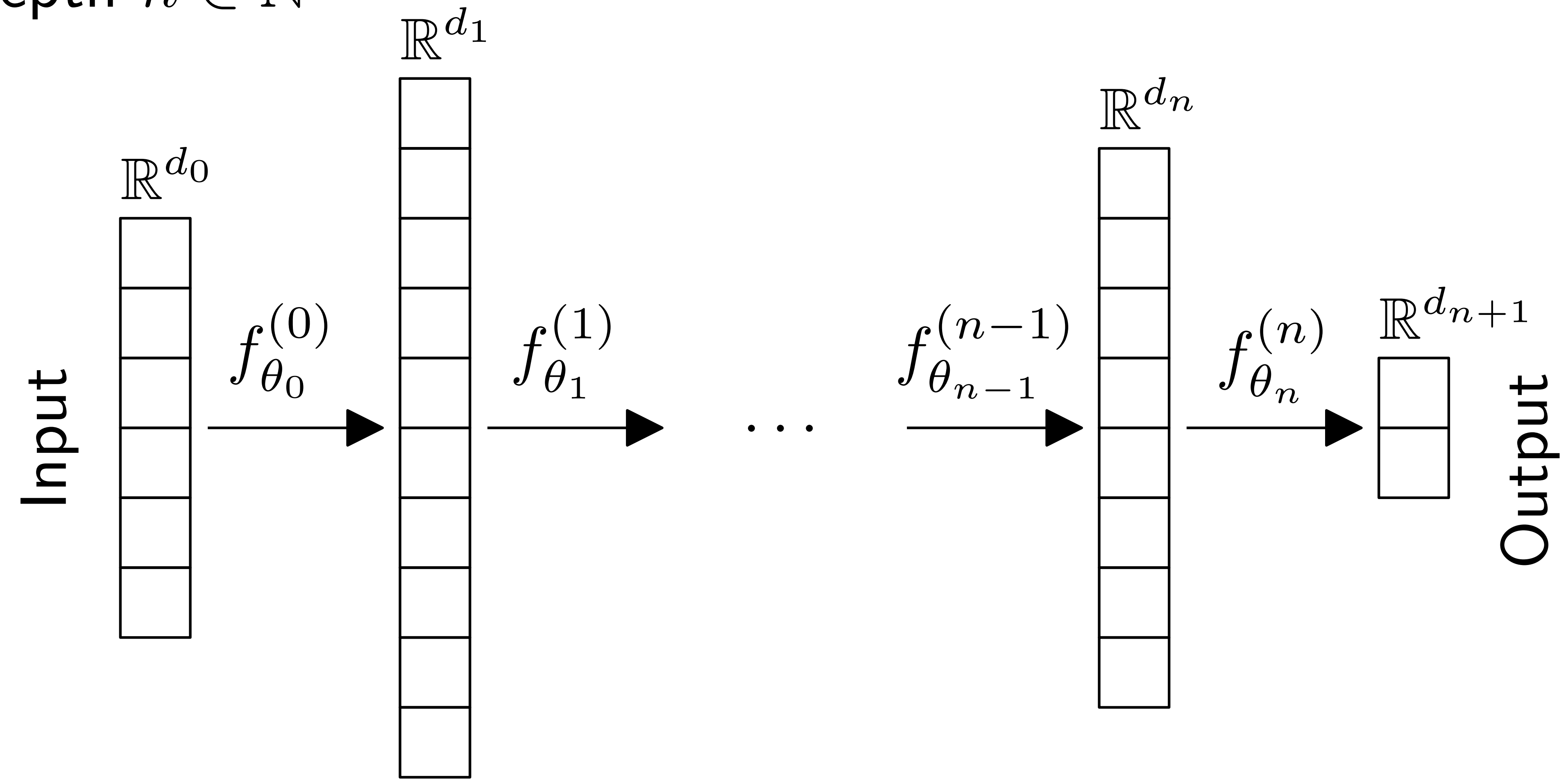
(non exhaustive list - see also Gudhi representations)

- Collections of 1D functions
  - → landscapes [Bubenik 2012]
  - → Betti curves [Umeda 2017]
- discrete measures: (interesting statistical properties [Chazal, Divol 2018])
  - → persistence images [Adams et al 2017]
  - → convolution with Gaussian kernel [Reininghaus et al. 2015] [Chepushtanova et al. 2015] [Kusano Fukumisu Hiraoka 2016-17] [Le Yamada 2018]
  - → sliced on lines [Carrière Oudot Cuturi 2017]
- finite metric spaces [Carrière Oudot Ovsjanikov 2015]
- polynomial roots or evaluations [Di Fabio Ferri 2015] [Kališnik 2016]

Problem: How to chose the right representation?

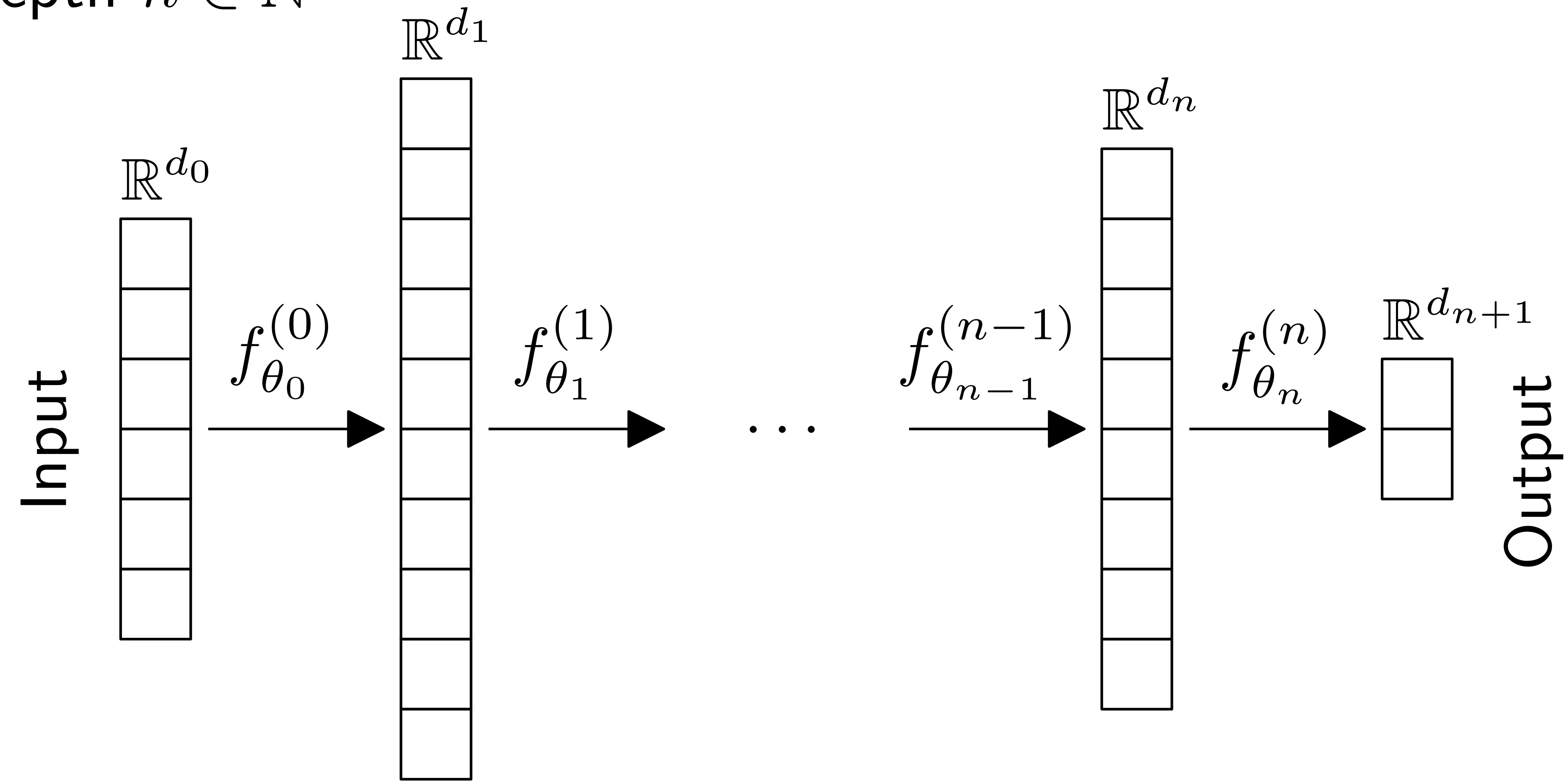
Reminder: Neural Net

NN with depth $n \in \mathbb{N}^*$



Reminder: Neural Net

NN with depth $n \in \mathbb{N}^*$



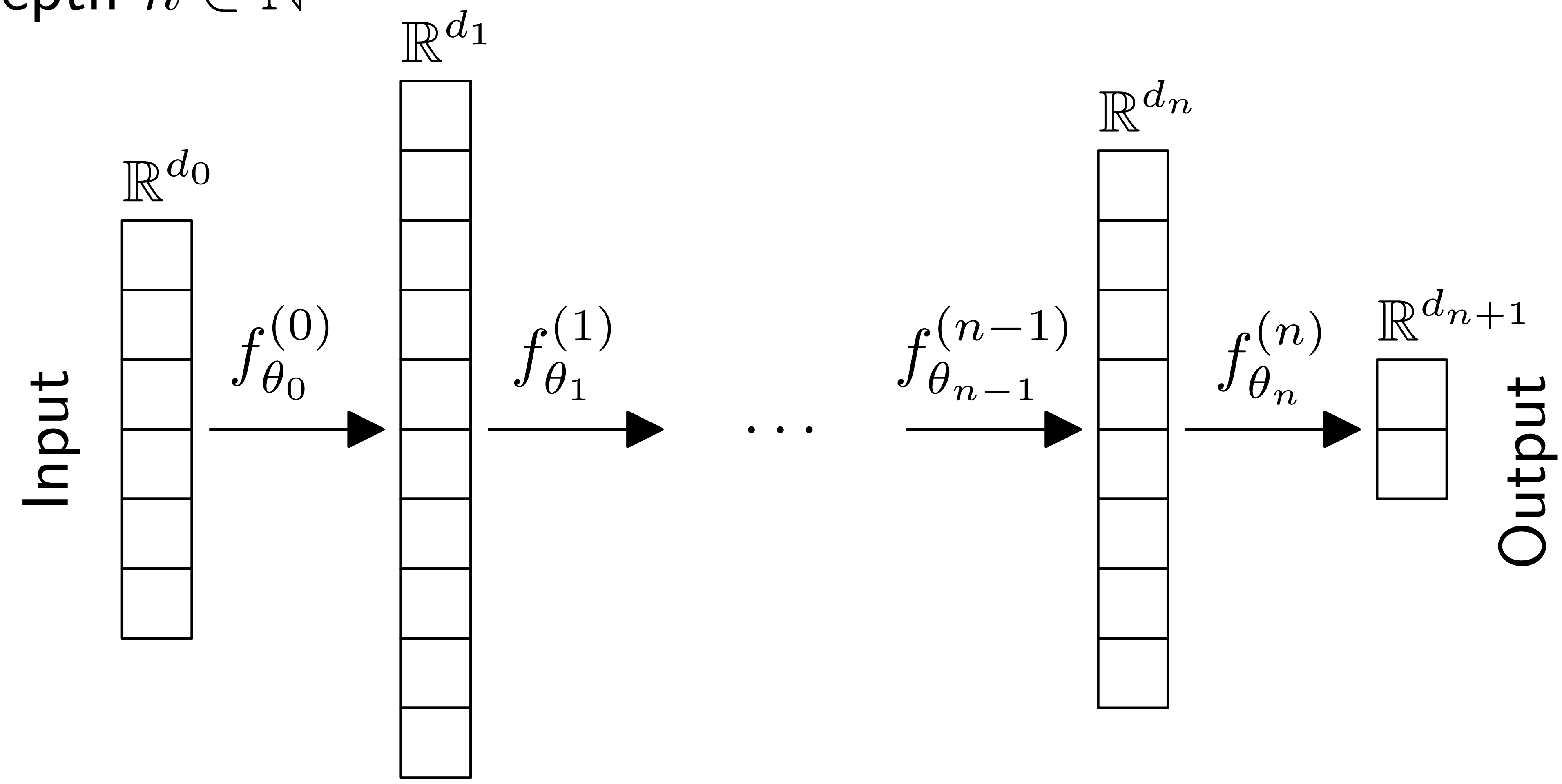
$\theta_k = (W_k \in \mathbb{R}^{d_{k+1} \times d_k},\ b_k \in \mathbb{R}^{d_{k+1}})$, $\quad \sigma : x \mapsto \max(0, x)$ or $(1 + \mathrm{e}^{-x})^{-1}$

$$f_{\theta_k}^{(k)} : x \in \mathbb{R}^{d_k} \mapsto \sigma(W_k \cdot x + b_k) \in \mathbb{R}^{d_{k+1}}$$

Final classifier: $F_\theta = f_{\theta_n}^{(n)} \circ \cdots \circ f_{\theta_0}^{(0)}$

Reminder: Neural Net

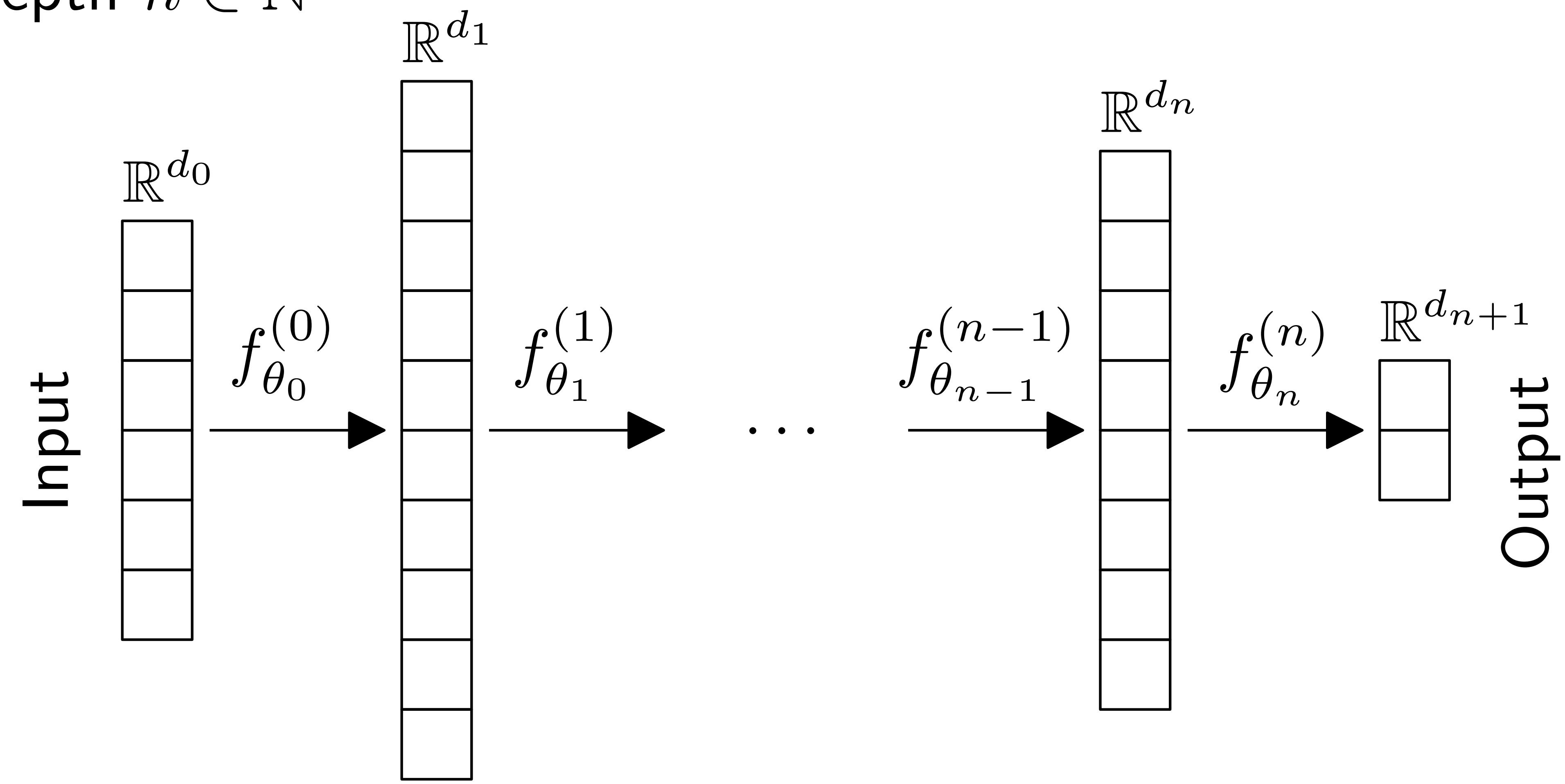
NN with depth $n \in \mathbb{N}^*$



Goal: Minimize $\ell(\theta) = \sum_i \|f_\theta(x_i) - y_i\|_2^2$ w.r.t. θ

Reminder: Neural Net

NN with depth $n \in \mathbb{N}^*$



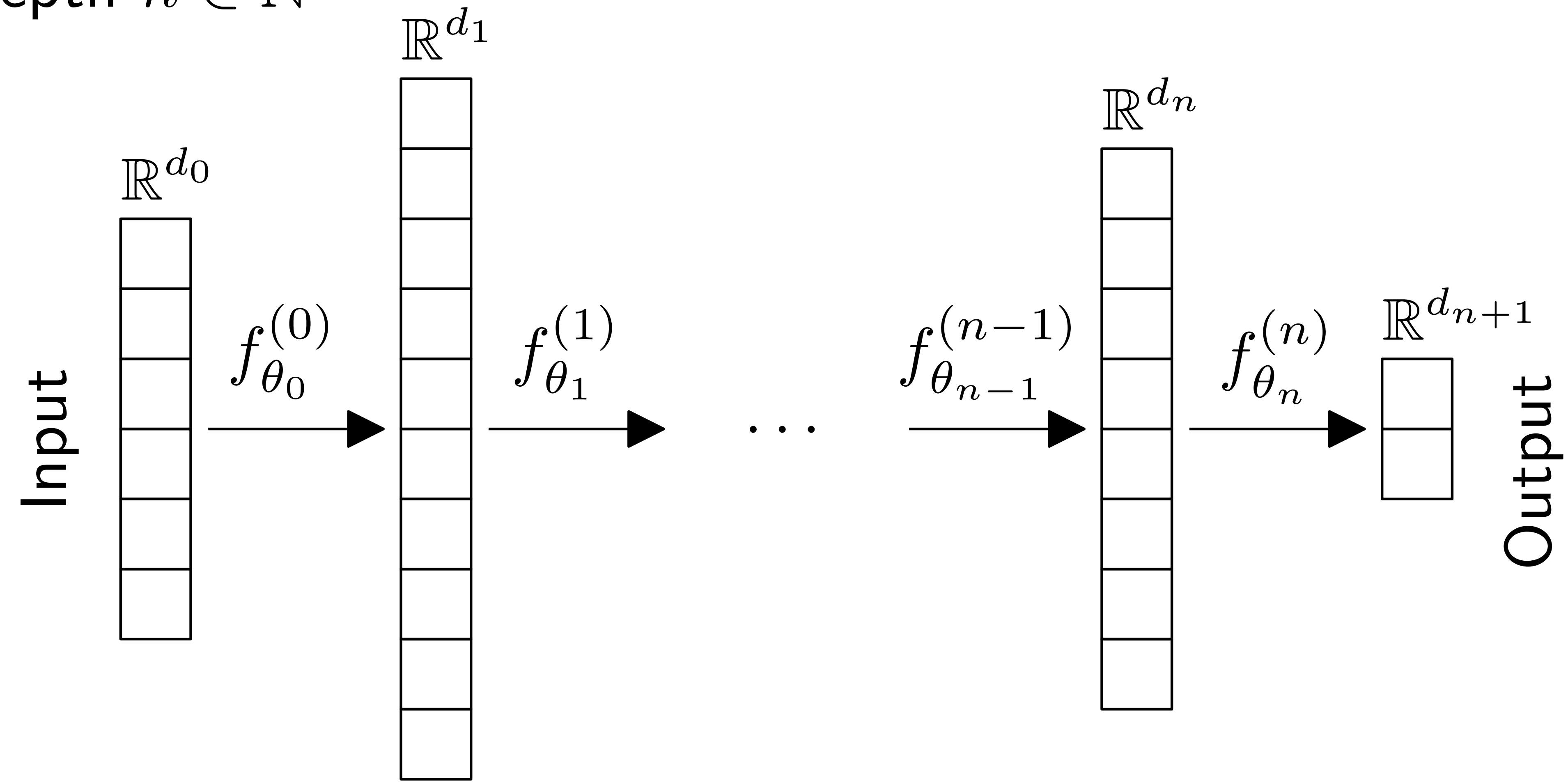
Goal: Minimize $\ell(\theta) = \sum_i \|f_\theta(x_i) - y_i\|_2^2$ w.r.t. θ

Backpropagation: for each k:

1. compute $\nabla\ell(\theta_k)$ with chain rule
2. update $\theta_k := \theta_k - \eta\nabla\ell(\theta_k)$

Reminder: Neural Net

NN with depth $n \in \mathbb{N}^*$



Goal: Minimize $\ell(\theta) = \sum_i \|f_\theta(x_i) - y_i\|_2^2$ w.r.t. θ

Backpropagation: for each k:

1. compute $\nabla\ell(\theta_k)$ with chain rule
2. update $\theta_k := \theta_k - \eta\nabla\ell(\theta_k)$

Requirement: $f^{(k)}_{\theta_k}$ needs to be differentiable w.r.t. θ_k and x

Deep Set Architecture

Originally defined in [Zaheer et al. 2017]

Tailored to handle sets instead of finite dimensional vectors

Input: $\{x_1, ..., x_n\} \subset \mathbb{R}^d$ instead of $x \in \mathbb{R}^d$

Deep Set Architecture

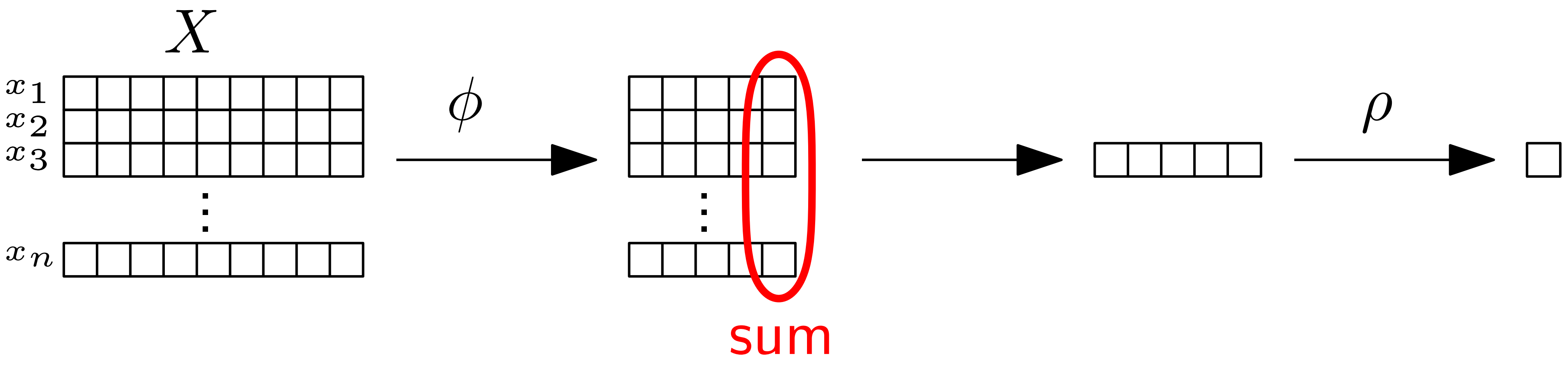
Originally defined in [Zaheer et al. 2017]

Tailored to handle sets instead of finite dimensional vectors

Input: $\{x_1, ..., x_n\} \subset \mathbb{R}^d$ instead of $x \in \mathbb{R}^d$

Network is *permutation invariant*: $F(X) = \rho\left(\sum_i \phi(x_i)\right)$

$$\Rightarrow F(\{x_1, ..., x_n\}) = F(\{x_{\sigma(1)}, ..., x_{\sigma(n)}\}),\ \forall \sigma$$



In practice: $\phi(x_i) = W \cdot x_i + b$

Deep Set Architecture

Originally defined in [Zaheer et al. 2017]

Tailored to handle sets instead of finite dimensional vectors

Input: $\{x_1, ..., x_n\} \subset \mathbb{R}^d$ instead of $x \in \mathbb{R}^d$

Network is *permutation invariant*: $F(X) = \rho\left(\sum_i \phi(x_i)\right)$

Universality theorem

Th: [Zaheer et al. 2017]

A function f is permutation invariant iif $f(X) = \rho\left(\sum_i \phi(x_i)\right)$ for some ρ and ϕ, whenever X is included in a *countable* space

Adaptation to persistence diagrams

[Carrière et al 2019]

Permutation invariant layers generalize several TDA approaches

→ persistence images → silhouettes → Betti curves

But not all of them since $\mathbb{R}^2$ is not countable

Using any permutation invariant operation (such as max, min, kth largest value) allows to generalize other TDA approaches

$$\mathrm{PersLay}(\mathsf{dgm}) = \rho\left(\mathsf{op}\{w(p)\cdot\phi(p)\}_{p\in\mathsf{dgm}}\right)$$

Permutation-invariant operation

Weight function

Point transformation $\phi : \mathbb{R}^2 \to \mathbb{R}^k$

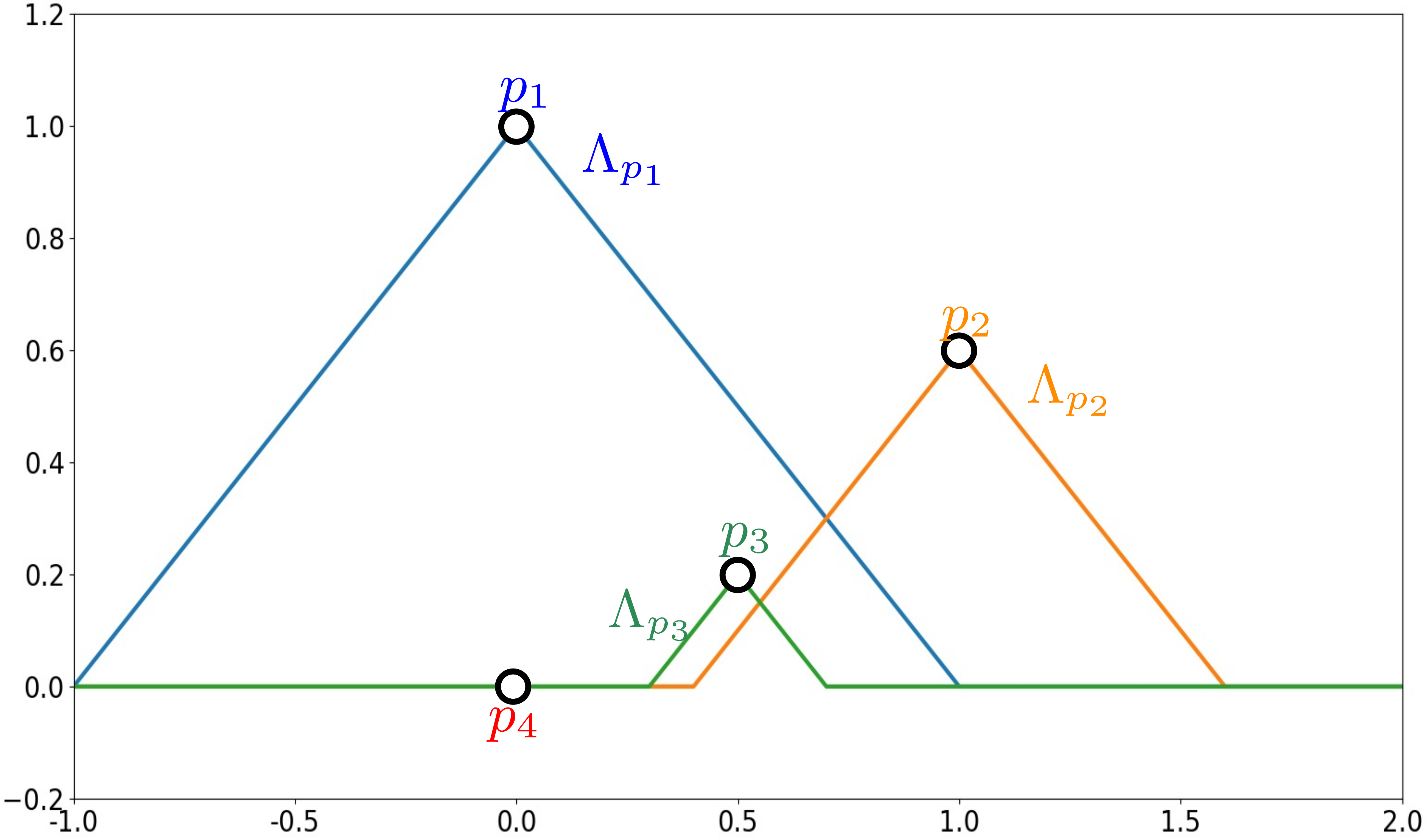
https://github.com/MathieuCarriere/perslay (will be released in gudhi in a near future)

Adaptation to persistence diagrams

[Carrière et al 2019]

Parameters $t_1, \cdots, t_q \in \mathbb{R}$

$$w(p) = 1 \qquad \phi_\Lambda : p \mapsto \begin{bmatrix} \Lambda_p(t_1) \\ \Lambda_p(t_2) \\ \vdots \\ \Lambda_p(t_q) \end{bmatrix} \qquad \text{op} = \text{top-}k$$



Persistence landscape

Adaptation to persistence diagrams

[Carrière et al 2019]

Parameters $t_1, \cdots, t_q \in \mathbb{R}^2$

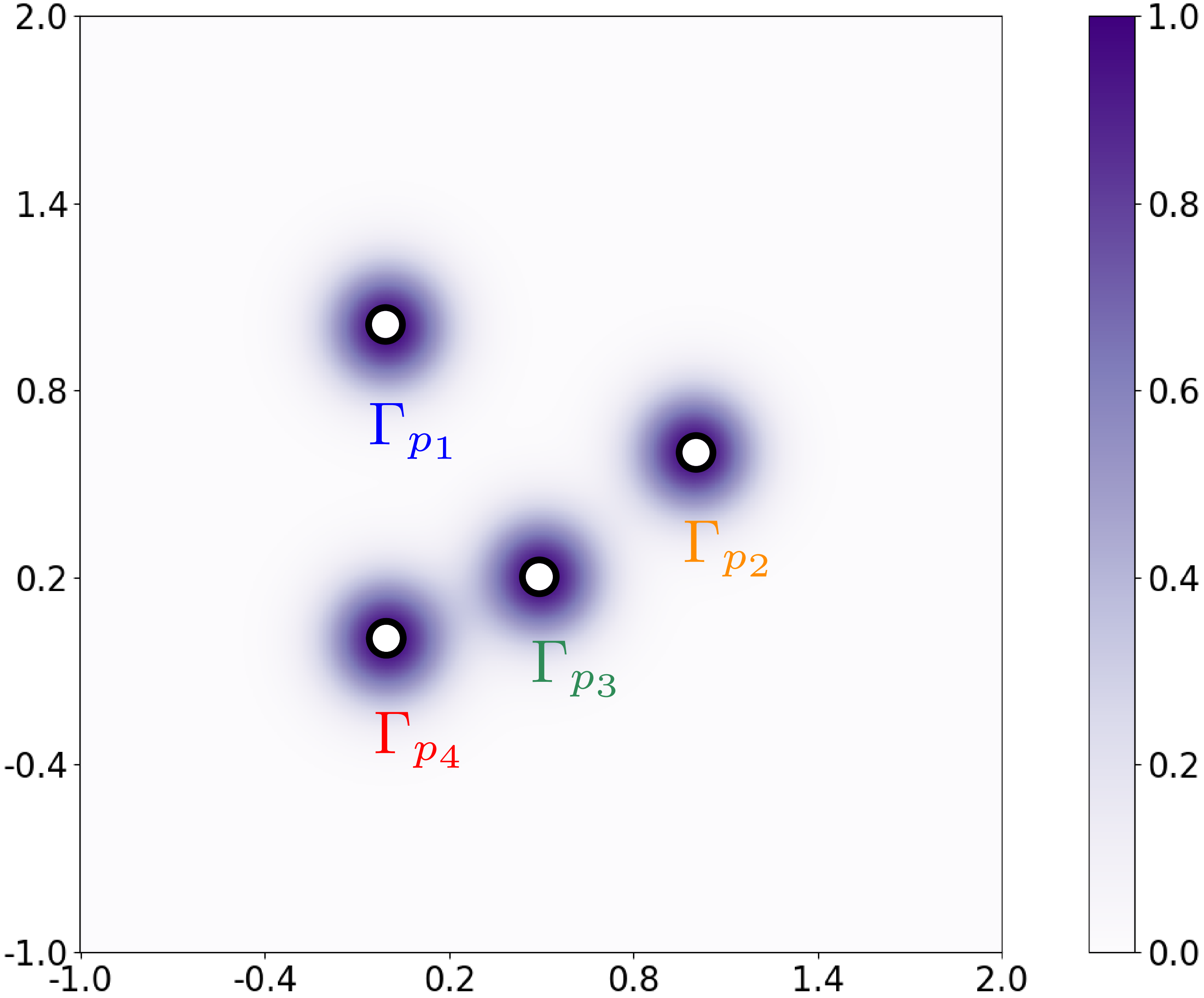
$$w(p) = w_t((x, y))$$

$$\phi_\Gamma : p \mapsto \begin{bmatrix} \Gamma_p(t_1) \\ \Gamma_p(t_2) \\ \vdots \\ \Gamma_p(t_q) \end{bmatrix}$$

op = sum

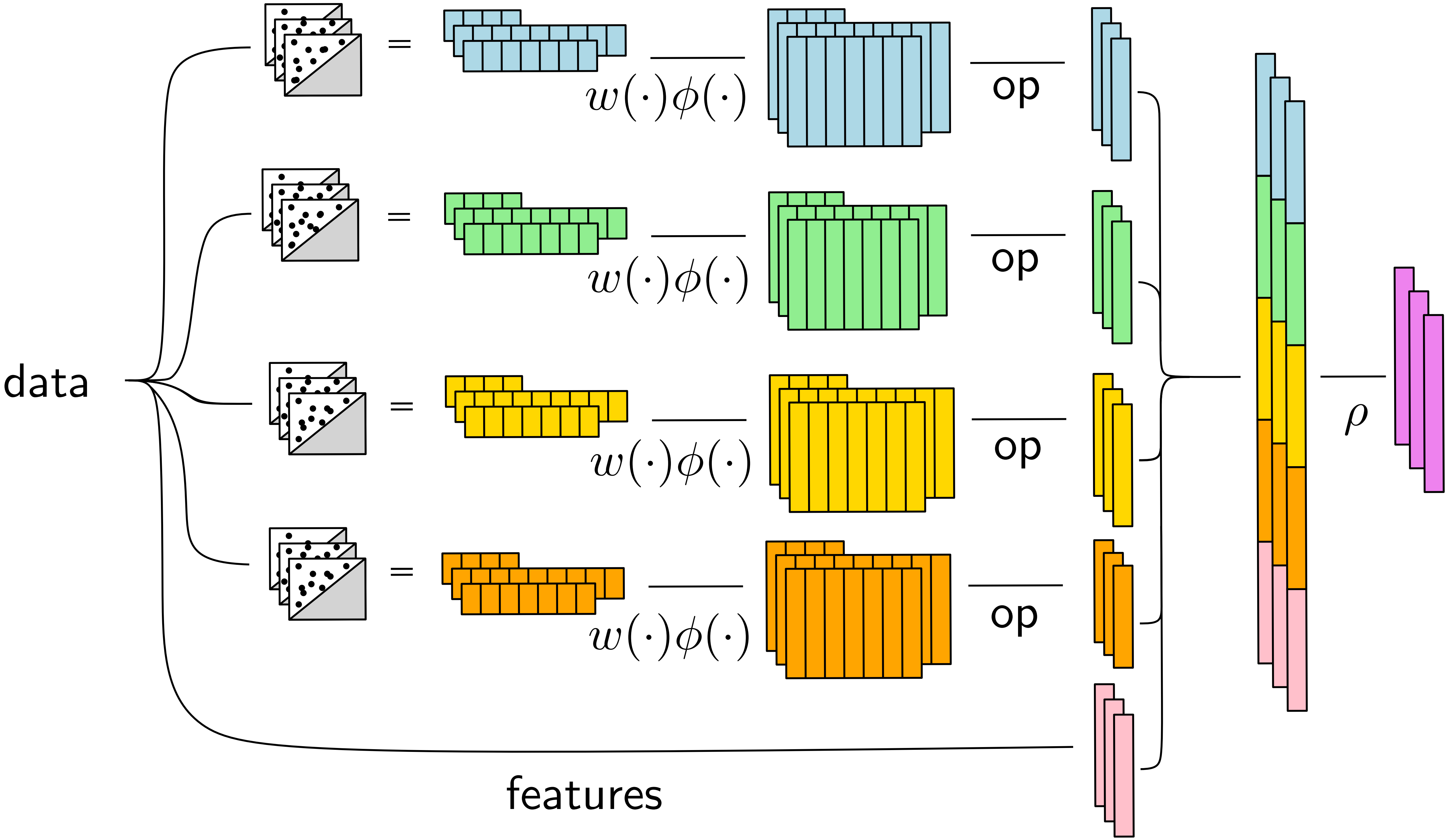
$$\Gamma_p : t \mapsto \exp\left(-\|p - t\|_2^2/(2\sigma^2)\right)$$

Persistence surface



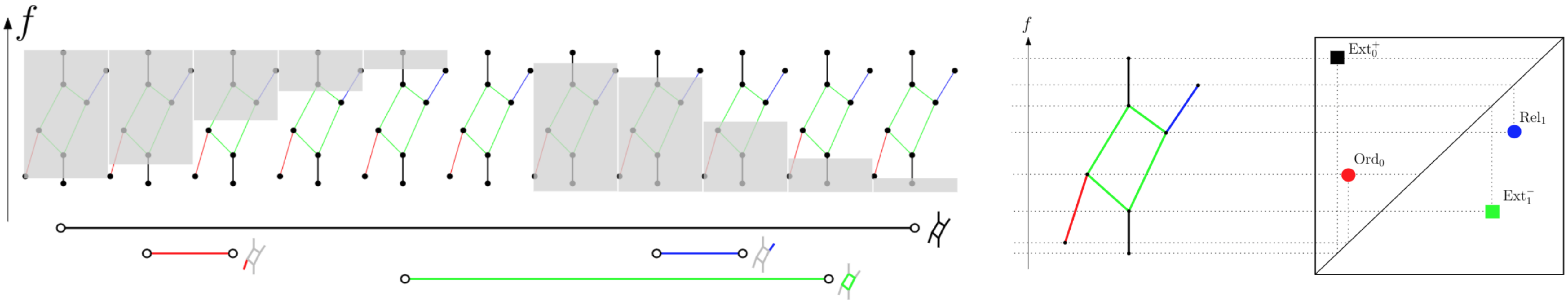
Adaptation to persistence diagrams

[Carrière et al 2019]



Adaptation to persistence diagrams

[Carrière et al 2019]



Dataset	ScaleVariant[1]	RetGK1 * [2]	RetGK11 * [2]	FGSD [3]	GCNN [4]	Spectral + HKS [5]	PersLay
REDDIT5K	—	56.1(±0.5)	55.3(±0.3)	47.8	52.9	49.7(±0.3)	**56.6**(±0.3)
REDDIT12K	—	**48.7**(±0.2)	47.1(±0.3)	—	46.6	39.7(±0.1)	47.7(±0.2)
COLLAB	—	**81.0(±0.3)**	80.6(±0.3)	80.0	79.6	67.8(±0.2)	76.4(±0.4)
IMDB-B	72.9	71.9(±1.0)	72.3(±0.6)	**73.6**	73.1	67.6(±0.6)	70.9(±0.7)
IMDB-M	50.3	47.7(±0.3)	48.7(±0.6)	**52.4**	50.3	44.5(±0.4)	48.7(±0.6)
BZR *	86.6	—	—	—	—	80.8(±0.8)	**87.2(±0.7)**
COX2 *	78.4	80.1(±0.9)	81.4(±0.6)	—	—	78.2(±1.3)	**81.6(±1.0)**
DHFR *	78.4	81.5(±0.9)	**82.5(±0.8)**	—	—	69.5(±1.0)	81.8(±0.8)
MUTAG *	88.3	90.3(±1.1)	90.1(±1.0)	**92.1**	86.7	85.8(±1.3)	89.8(±0.9)
PROTEINS *	72.6	75.8(±0.6)	75.2(±0.3)	73.4	**76.3**	73.5(±0.3)	74.8(±0.3)
NCI1 *	71.6	**84.5(±0.2)**	83.5(±0.2)	79.8	78.4	65.3(±0.2)	72.8(±0.3)
NCI109 *	70.5	—	—	**78.8**	—	64.9(±0.2)	71.7(±0.3)
FRANKENSTEIN	69.4	—	—	—	—	62.9(±0.1)	**70.7(±0.4)**

Average scores from 10 times 10-folds cross-validation

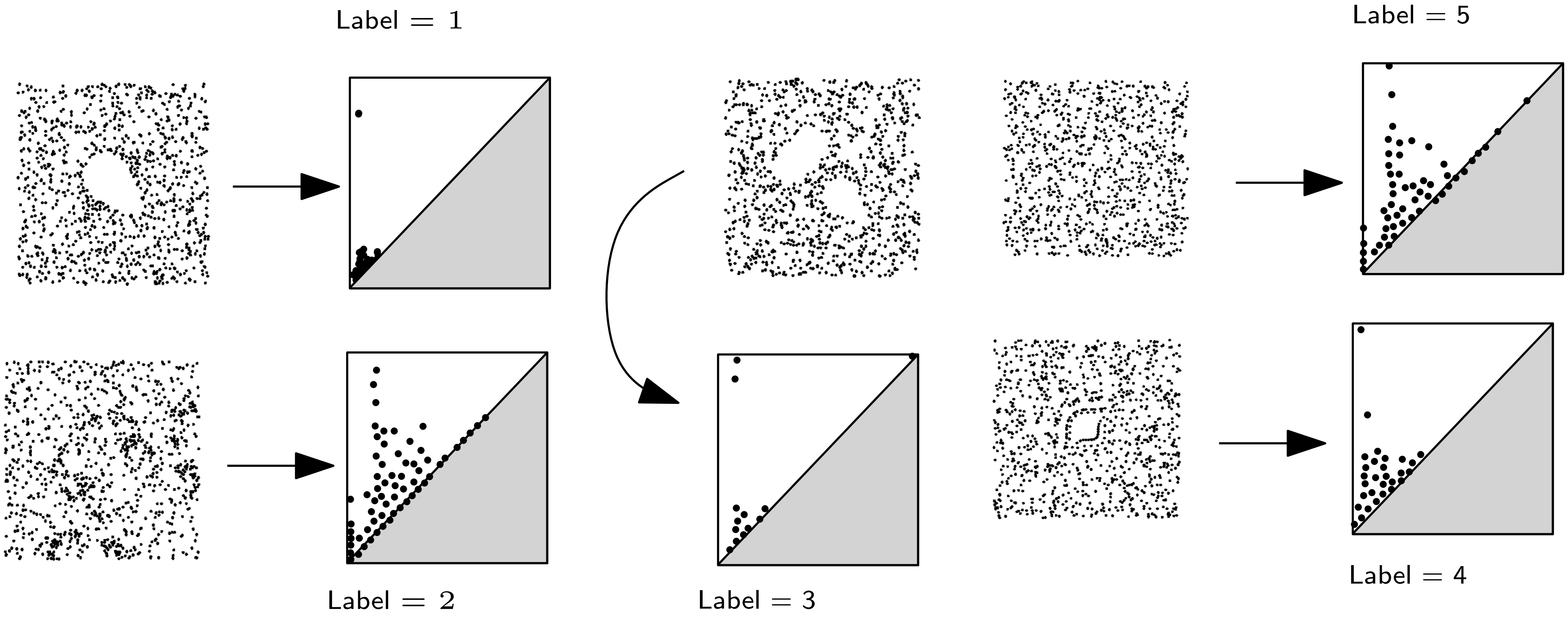
Adaptation to persistence diagrams

[Carrière et al 2019]

Goal: classify orbits of *linked twisted map*

Orbits described by (depending on parameter r):

$$\begin{cases} x_{n+1} = x_n + r\, y_n(1 - y_n) \mod 1 \\ y_{n+1} = y_n + r\, x_{n+1}(1 - x_{n+1}) \mod 1 \end{cases}$$

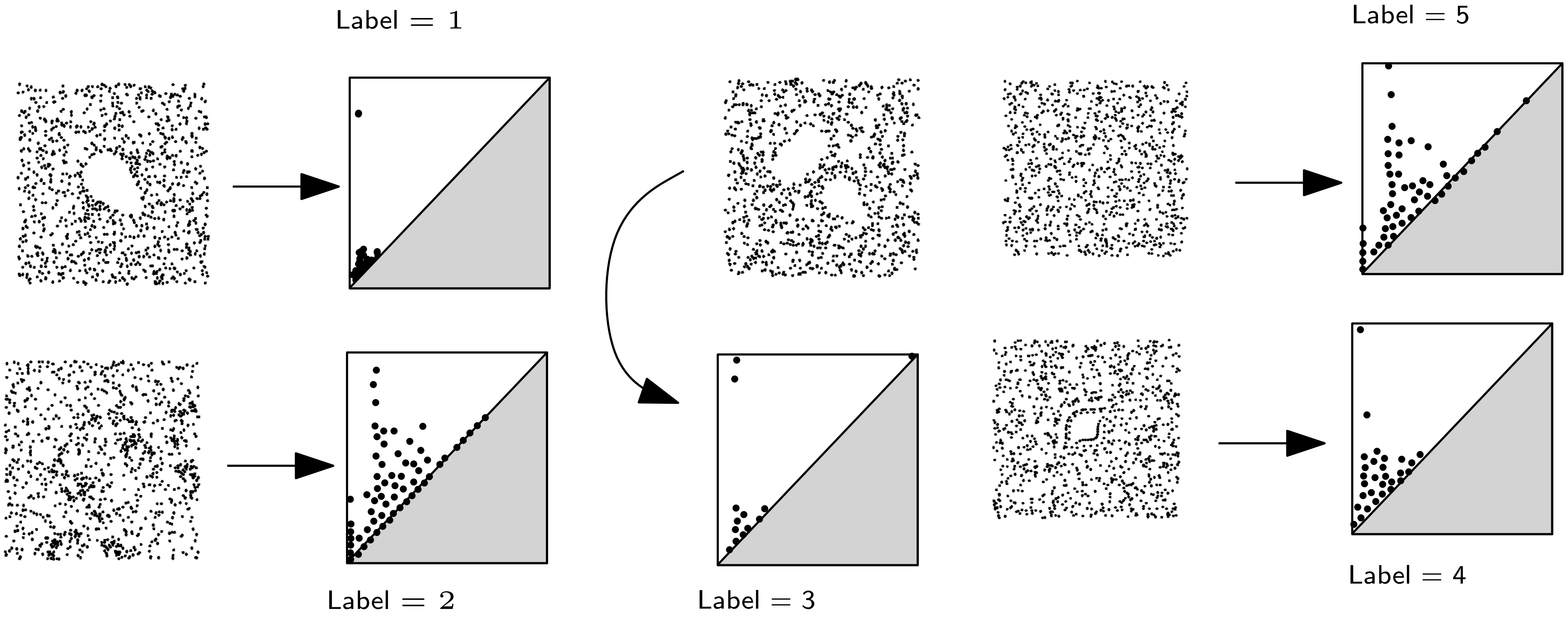


Adaptation to persistence diagrams

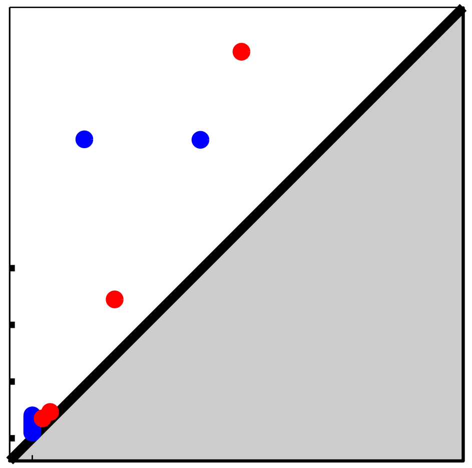
[Carrière et al 2019]

Goal: classify orbits of *linked twisted map*

Dataset	PSS-K	PWG-K	SW-K	PF-K	PersLay
ORBIT5K	72.38(±2.4)	76.63(±0.7)	83.6(±0.9)	85.9(±0.8)	**87.7(±1.0)**
ORBIT100K	—	—	—	—	**89.2(±0.3)**

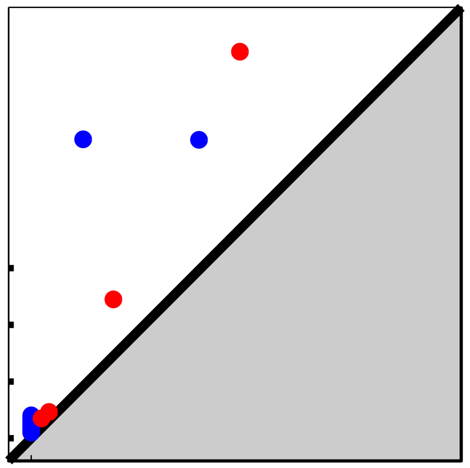


A general question

$\arg\min f($  $)?$

How to minimize functions depending of persistence diagrams (e.g. total persistence)?

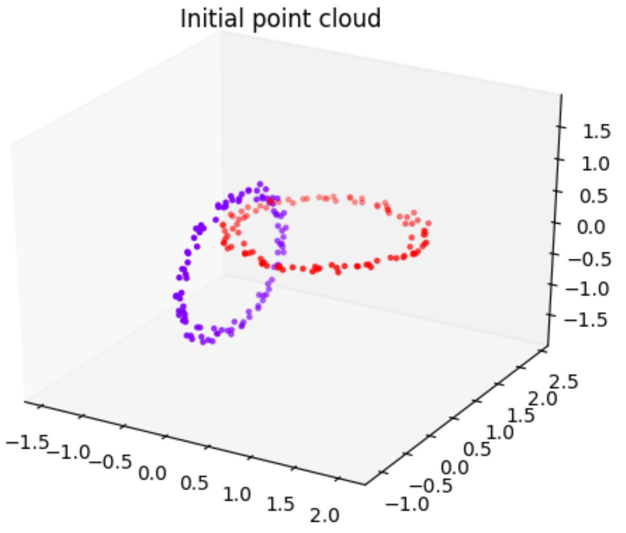
A general question

$\arg\min f($  $)?$

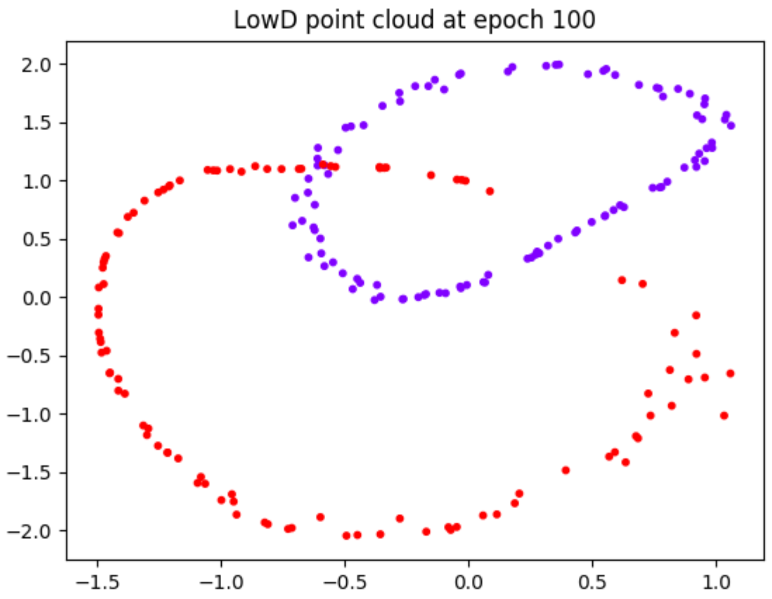
How to minimize functions depending of persistence diagrams (e.g. total persistence)?

→ Need to understand the “differentiability of persistence”

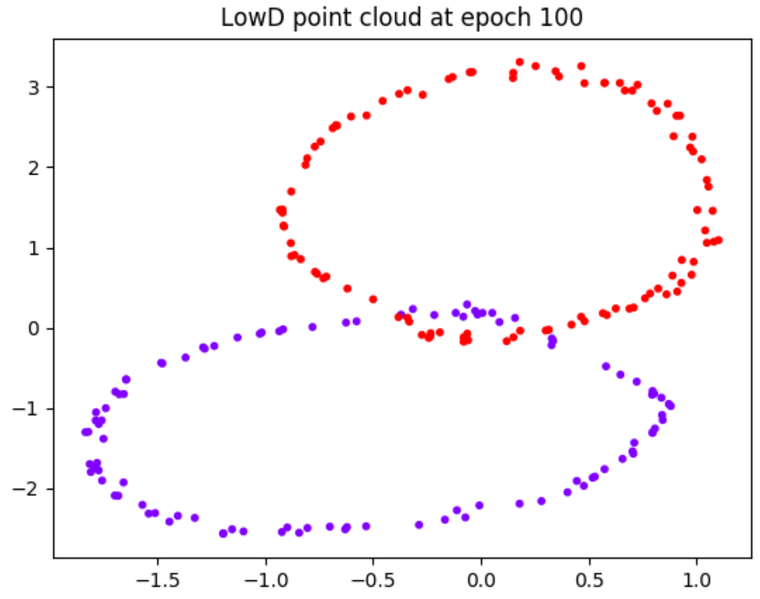
Example: dimensionality reduction



Input: 2 sampled circles in $\mathbb{R}^9$ (3D view)

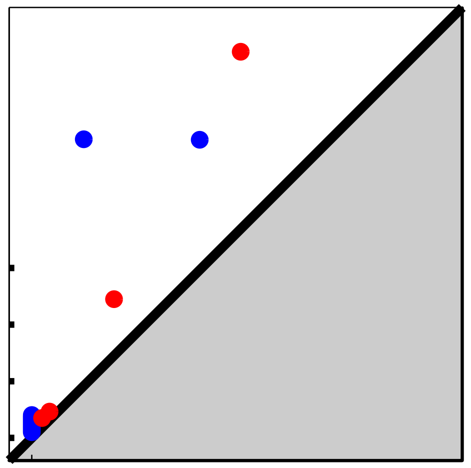


Dim reduction in $\mathbb{R}^2$ without topol. constraint



Dim reduction in $\mathbb{R}^2$ with topol. constraint

The minimization problem

$\arg\min f($  $)?$

A "long-standing" question in Topological Data Analysis

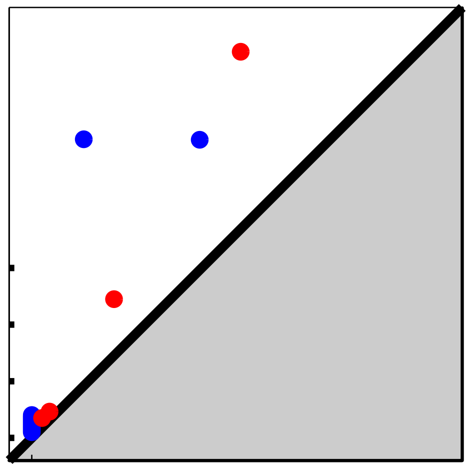
[*Continuation of point clouds via persistence diagrams*, Gameiro, Hiraoka, Obayashi, Physica D, 2015]

[*Topological Function Optimization for Continuous Shape Matching*, Poulenard, Skraba, Ovsjanikov, SGP, 2018]

[*A topology layer for machine learning*, Brüel-Gabrielsson et al., AISTATS, 2020]

[*Topological Autoencoders*, Moor et al., ICML, 2020]

The minimization problem

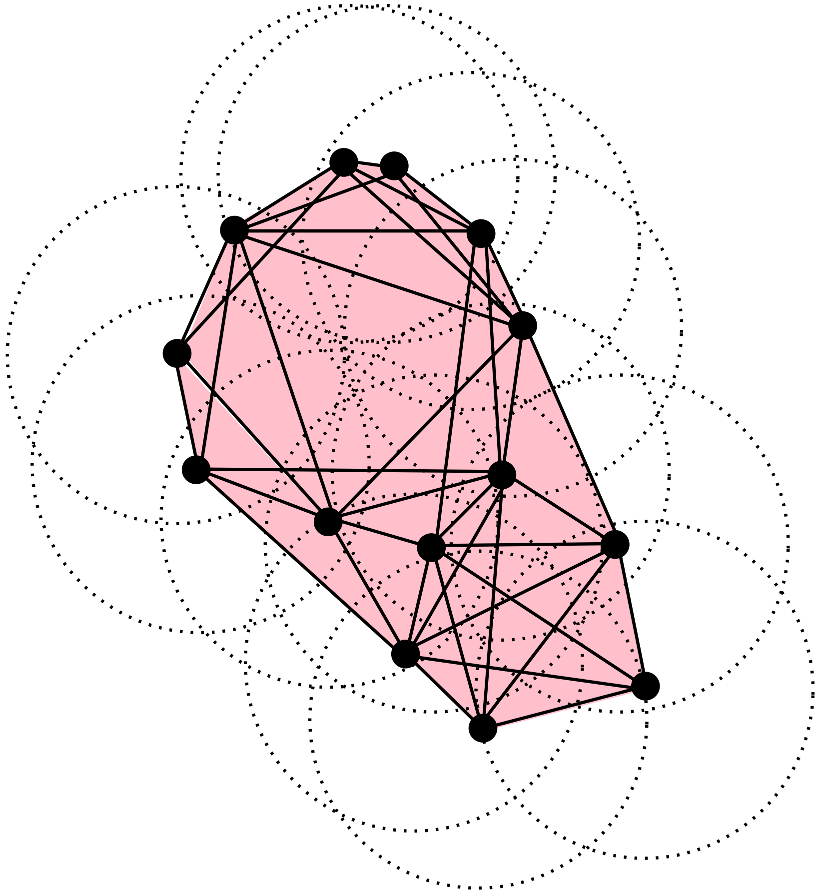
$\arg\min f($  $)?$

A “long-standing” question in Topological Data Analysis

[*Continuation of point clouds via persistence diagrams*, Gameiro, Hiraoka, Obayashi, Physica D, 2015]	Point cloud data
[*Topological Function Optimization for Continuous Shape Matching*, Poulenard, Skraba, Ovsjanikov, SGP, 2018]	Bottleneck distance loss
[*A topology layer for machine learning*, Brüel-Gabrielsson et al., AISTATS, 2020]	Total persistence loss
[*Topological Autoencoders*, Moor et al., ICML, 2020]	Vietoris-Rips filtration

All restricted to specific data type / loss function / filtration!

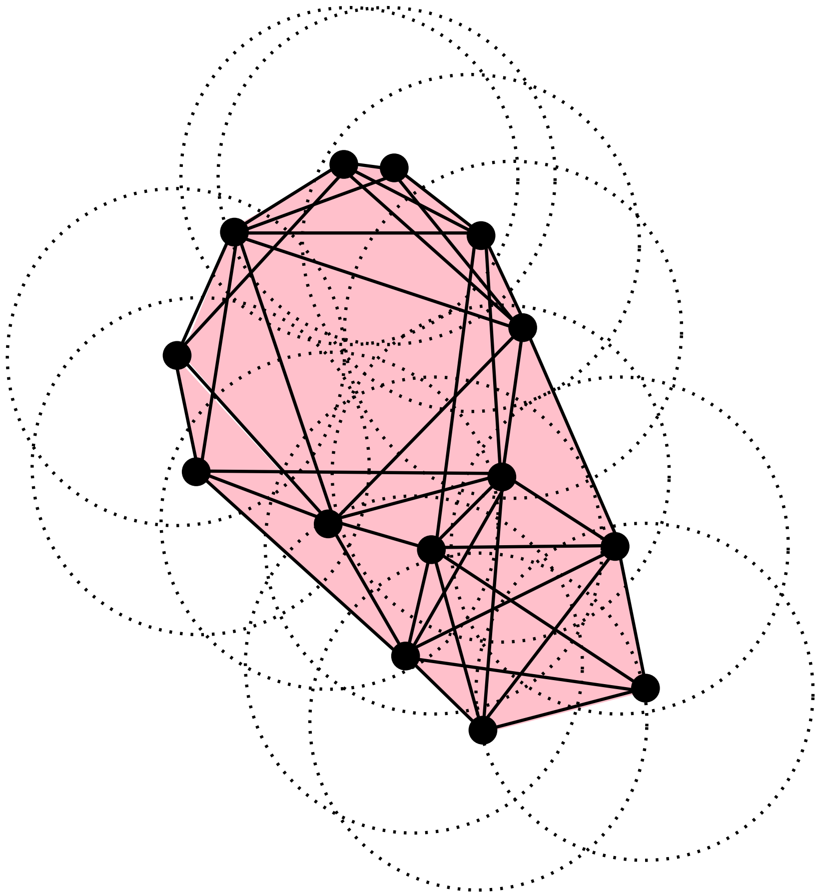
Simplicial complexes and filtrations



Given a set V, a simplicial complex K is a collection of finite subsets of V s. t.
- $\{v\} \in K$ for any $v \in V$,
- if $\sigma \in K$ and $\tau \subseteq \sigma$ then $\tau \in K$.

Given K and $R \subseteq \mathbb{R}$, a filtration of K is an increasing sequence $(K_r)_{r \in R}$ of subcomplexes of K with respect to the inclusion such that $\bigcup_{r \in R} K_r = K$.

Simplicial complexes and filtrations



Given a set V, a simplicial complex K is a collection of finite subsets of V s. t.
- $\{v\} \in K$ for any $v \in V$,
- if $\sigma \in K$ and $\tau \subseteq \sigma$ then $\tau \in K$.

Given K and $R \subseteq \mathbb{R}$, a filtration of K is an increasing sequence $(K_r)_{r \in R}$ of subcomplexes of K with respect to the inclusion such that $\bigcup_{r \in R} K_r = K$.

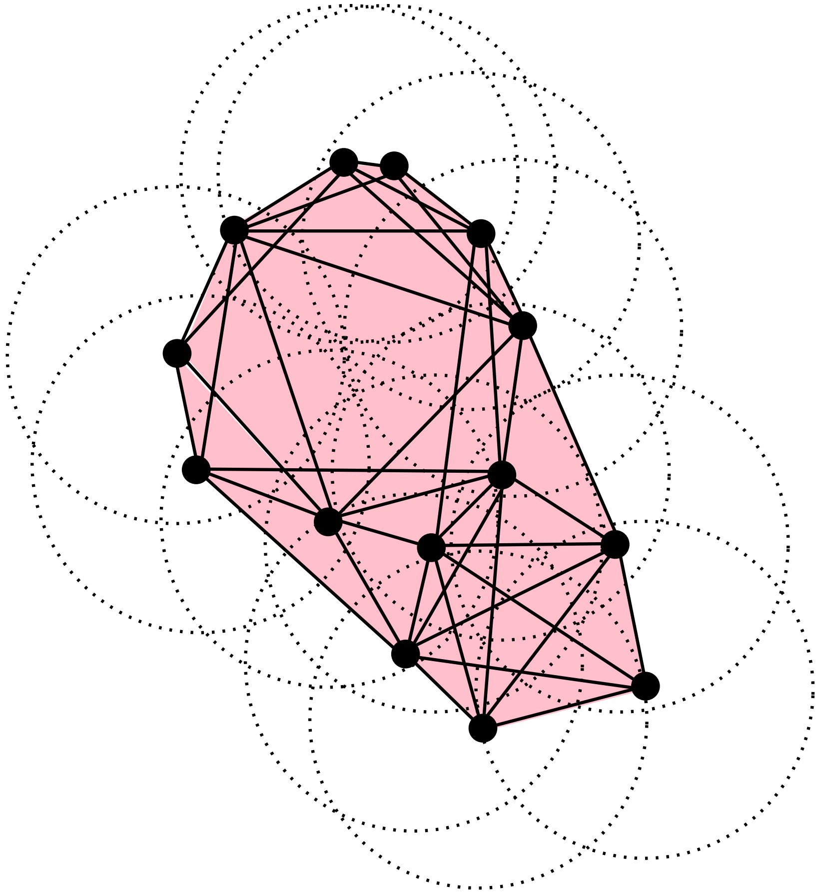
To $\sigma \in K$, one can associate $\Phi_\sigma = \inf\{r \in R : \sigma \in K_r\}$

$\Rightarrow$ A filtration of K is a $|K|$-dimensional vector

$$\Phi = (\Phi_\sigma)_{\sigma \in K} \in \mathbb{R}^{|K|} \text{ s. t. } \tau \subseteq \sigma \Rightarrow \Phi_\tau \leq \Phi_\sigma$$

The set $\mathrm{Filt}_K \subset \mathbb{R}^{|K|}$ of the vectors in $\mathbb{R}^{|K|}$ defining a filtration on K is semi-algebraic.

Simplicial complexes and filtrations



Given a set V, a simplicial complex K is a collection of finite subsets of V s. t.
- $\{v\} \in K$ for any $v \in V$,
- if $\sigma \in K$ and $\tau \subseteq \sigma$ then $\tau \in K$.

Given K and $R \subseteq \mathbb{R}$, a filtration of K is an increasing sequence $(K_r)_{r \in R}$ of subcomplexes of K with respect to the inclusion such that $\bigcup_{r \in R} K_r = K$.

Definition: Let K be a simplicial complex and A a set. A map $\Phi \colon A \to \mathbb{R}^{|K|}$ is said to be a parametrized family of filtrations if for any $x \in A$ and $\sigma, \tau \in K$ with $\tau \subseteq \sigma$, one has $\Phi_\tau(x) \leq \Phi_\sigma(x)$.

Persistent homology computation

Let K be a finite filtered simplicial complex and let $\sigma_1 \preceq \cdots \preceq \sigma_{|K|}$ the simplices of K ordered according the increasing entries of $\Phi = (\Phi_\sigma)_{\sigma \in K} \in \mathbb{R}^{|K|}$

Persistent homology computation

Let K be a finite filtered simplicial complex and let $\sigma_1 \preceq \cdots \preceq \sigma_{|K|}$ the simplices of K ordered according the increasing entries of $\Phi = (\Phi_\sigma)_{\sigma \in K} \in \mathbb{R}^{|K|}$

Process the simplices according to their order of entrance in the filtration:

Let $k = \dim \sigma_i$ and denote $K_{i-1} = \cup_{l=1}^{i-1} \sigma_l$

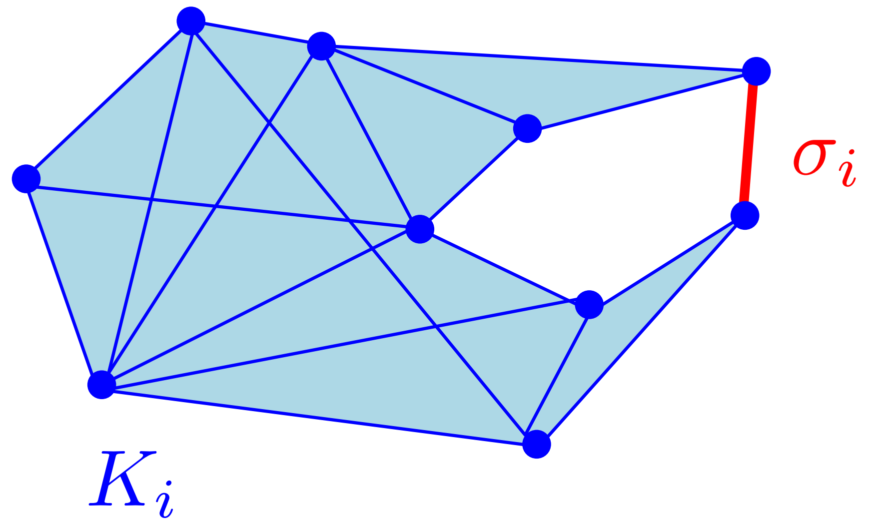
Persistent homology computation

Let K be a finite filtered simplicial complex and let $\sigma_1 \preceq \cdots \preceq \sigma_{|K|}$ the simplices of K ordered according the increasing entries of $\Phi = (\Phi_\sigma)_{\sigma \in K} \in \mathbb{R}^{|K|}$

Process the simplices according to their order of entrance in the filtration:

Let $k = \dim \sigma_i$ and denote $K_{i-1} = \cup_{l=1}^{i-1} \sigma_l$

Case 1: adding σ_i to K_{i-1} creates a new k-dimensional topological feature in K_i (new homology class in H_k).



$\Rightarrow$ the birth of a k-dim feature is registered.

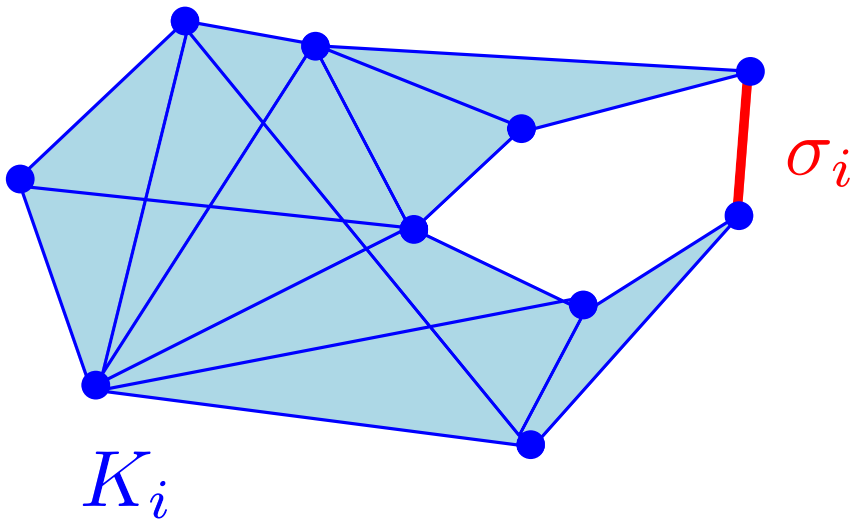
Persistent homology computation

Let K be a finite filtered simplicial complex and let $\sigma_1 \preceq \cdots \preceq \sigma_{|K|}$ the simplices of K ordered according the increasing entries of $\Phi = (\Phi_\sigma)_{\sigma \in K} \in \mathbb{R}^{|K|}$

Process the simplices according to their order of entrance in the filtration:

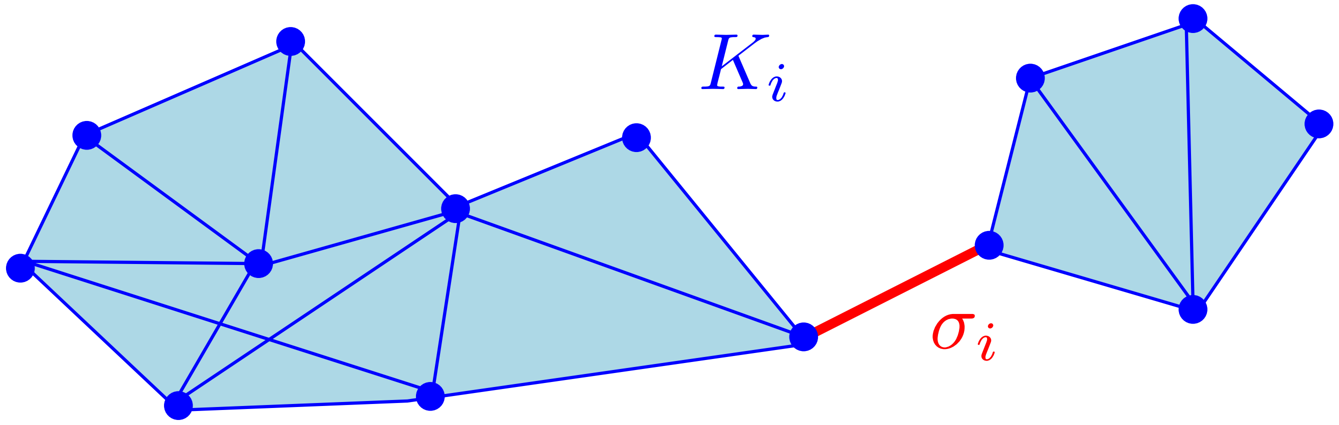
Let $k = \dim \sigma_i$ and denote $K_{i-1} = \cup_{l=1}^{i-1} \sigma_l$

Case 1: adding σ_i to K_{i-1} creates a new k-dimensional topological feature in K_i (new homology class in H_k).



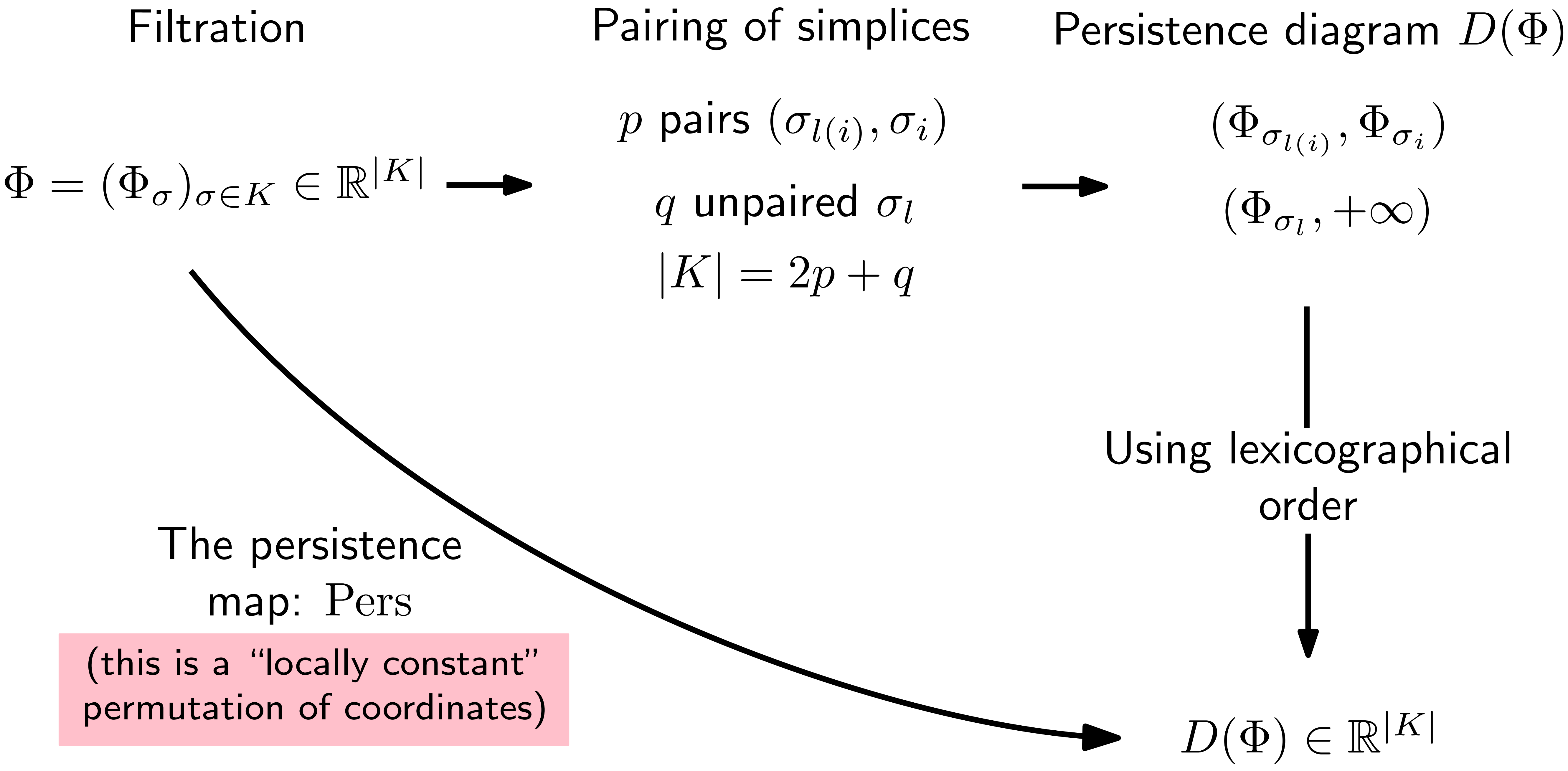
$\Rightarrow$ the birth of a k-dim feature is registered.

Case 2: adding σ_i to K_{i-1} kills a $(k-1)$-dimensional topological feature in K_i (homology class in H_{k-1}).



$\Rightarrow$ persistence algo. pairs the simplex σ_i to the simplex $\sigma_{l(i)}$ that gave birth to the killed feature.

Persistent homology computation



Filtration

$\Phi = (\Phi_\sigma)_{\sigma \in K} \in \mathbb{R}^{|K|}$

Pairing of simplices

p pairs $(\sigma_{l(i)}, \sigma_i)$

q unpaired σ_l

$|K| = 2p + q$

Persistence diagram $D(\Phi)$

$(\Phi_{\sigma_{l(i)}}, \Phi_{\sigma_i})$

$(\Phi_{\sigma_l}, +\infty)$

Using lexicographical order

$D(\Phi) \in \mathbb{R}^{|K|}$

The persistence map: Pers

(this is a "locally constant" permutation of coordinates)

The persistence map is semi-algebraic

Proposition: Given a simplicial complex K, the map

$$\mathrm{Pers}\colon \mathrm{Filt}_K \subseteq \mathbb{R}^{|K|} \to \mathbb{R}^{|K|}$$

is semi-algebraic, (and thus definable in any o-minimal structure). Moreover, there exists a semi-algebraic partition of Filt_K such that the restriction of Pers to each element of this partition is a Lipschitz map.

Corollary: Let K be a simplicial complex and $\Phi\colon A \to \mathbb{R}^{|K|}$ be a semi-algebraic (or definable in a given o-minimal structure) parametrized family of filtrations. The map $\mathrm{Pers} \circ \Phi\colon A \to \mathbb{R}^{|K|}$ is semi-algebraic (definable).

The persistence map is semi-algebraic

Proposition: Let K be a simplicial complex and $\Phi\colon A \to \mathbb{R}^{|K|}$ a definable parametrized family of filtrations, where $\dim A = m$. Then there exists a finite definable partition of A, $A = S \sqcup O_1 \sqcup \cdots \sqcup O_k$ such that $\dim S < \dim A := m$ and, for any $i = 1, \ldots, k$, O_i is a definable manifold of dimension m and $\mathrm{Pers} \circ \Phi\colon O_i \to \mathbb{R}^{|K|}$ is differentiable.

This is an immediate consequence of finiteness and stratifiability properties of definable sets

Semi-algebraic sets and maps

A semialgebraic subset of $\mathbb{R}^n$ is a subset defined as a finite unions and intersections of polynomial equations and inequations with real coefficients.
In other words, the set of semialgebraic subsets of $\mathbb{R}^n$ is the smallest class $\mathcal{SA}_n$ of subsets of $\mathbb{R}^n$ satisfaying:

1. if $P \in \mathbb{R}[X_1, \cdots, X_n]$ is a polynomial, $\{x \in \mathbb{R}^n : P(x) = 0\} \in \mathcal{SA}_n$ and $\{x \in \mathbb{R}^n : P(x) > 0\} \in \mathcal{SA}_n$.
2. If $A, B \in \mathcal{SA}_n$, then $A \cup B$, $A \cap B$ and $\mathbb{R}^n \setminus A$ belong to $\mathcal{SA}_n$.

Given $A \subseteq \mathbb{R}^n$ and $A \subseteq \mathbb{R}^m$ two semialgebraic sets, a map $\Phi : A \to B$, where A and B is a semialgebraic map if its graph

$$G_\Phi = \{(x, \Phi(x)) : x \in A\} \subseteq A \times B$$

is a semialgebraic subset of $\mathbb{R}^n \times \mathbb{R}^m$.

o-minimal structures

An o-minimal structure on the field of real numbers $\mathbb{R}$ is a collection $(S_n)_{n\in\mathbb{N}}$, where each S_n is a set of subsets of $\mathbb{R}^n$ such that:

1. S_1 is exactly the collection of finite unions of points and intervals;
2. all algebraic subsets of $\mathbb{R}^n$ are in S_n;
3. S_n is a Boolean subalgebra of $\mathbb{R}^n$ for any $n \in \mathbb{N}$;
4. if $A \in S_n$ and $B \in S_m$, then $A \times B \in S_{n+m}$;
5. if $\pi\colon \mathbb{R}^{n+1} \to \mathbb{R}^n$ is the linear projection onto the first n coordinates and $A \in S_{n+1}$, then $\pi(A) \in S_n$.

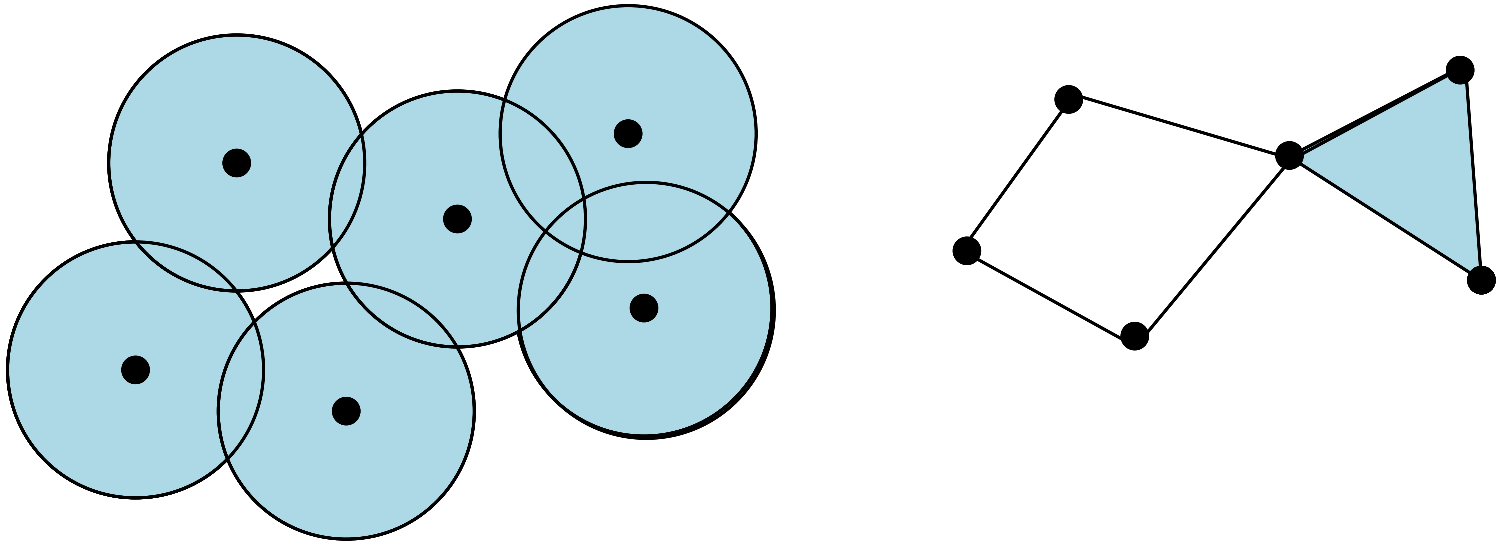
$A \in S_n$ is called a definable set in the o-minimal structure.

For $A \subseteq \mathbb{R}^n$, a map $f\colon A \to \mathbb{R}^m$ is a definable map if its graph is a definable set in $\mathbb{R}^{n+m}$.

Example: Semi-algebraic sets define an o-minimal structure.

Important property: Definable sets admit finite (Whitney) stratification.

Example: the Vietoris-Rips filtration

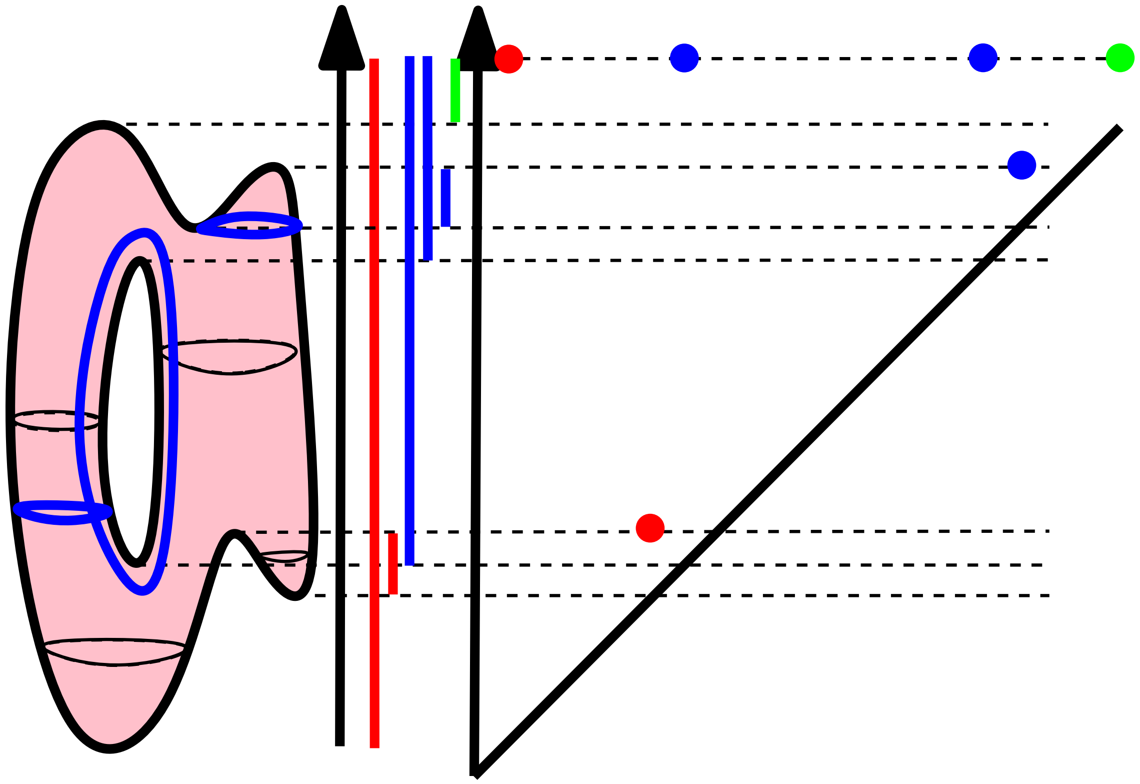


$$\Phi\colon A = (\mathbb{R}^d)^n \to \mathbb{R}^{|\Delta_n|} = \mathbb{R}^{2^n-1}$$

where Δ_n is the simplicial complex made of all the faces of the $(n-1)$-dimensional simplex and, for any $x = (x_1, \ldots, x_n) \in A$ and any simplex $\sigma \subseteq \{1, \ldots, n\}$,

$$\Phi_\sigma(x) = \max_{i,j\in\sigma} \|x_i - x_j\|.$$

Example: sublevel sets filtrations



K a simplicial complex with n vertices $v_1, \dots, v_n$.

Any real-valued function f defined on the vertices of K can be represented as a vector $(f(v_1), \dots, f(v_n)) \in \mathbb{R}^n$.

$$\Phi \colon A = \mathbb{R}^n \to \mathbb{R}^{|K|}$$

where for any $f = (f_1, \dots, f_n) \in A$ and any simplex $\sigma \subseteq \{1, \dots, n\}$,

$$\Phi_\sigma(f) = \max_{i \in \sigma} f_i$$

Functions of persistence

Definition: A function

$$E\colon \mathbb{R}^{|K|} = (\mathbb{R}^2)^p \times \mathbb{R}^q \to \mathbb{R}$$

is a function of persistence if it is invariant to permutations of the points of the persistence diagram: for any $(p_1, \ldots, p_p, e_1, \ldots, e_q) \in (\mathbb{R}^2)^p \times \mathbb{R}^q$ and any permutations α, β of the sets $\{1, \ldots, p\}$ and $\{1, \ldots, q\}$, respectively, one has

$$E(p_{\alpha(1)}, \ldots, p_{\alpha(p)}, e_{\beta(1)}, \ldots, e_{\beta(q)}) = E(p_1, \ldots, p_p, e_1, \ldots, e_q).$$

Functions of persistence

Definition: A function

$$E\colon \mathbb{R}^{|K|} = (\mathbb{R}^2)^p \times \mathbb{R}^q \to \mathbb{R}$$

is a function of persistence if it is invariant to permutations of the points of the persistence diagram: for any $(p_1, \dots, p_p, e_1, \dots, e_q) \in (\mathbb{R}^2)^p \times \mathbb{R}^q$ and any permutations α, β of the sets $\{1, \dots, p\}$ and $\{1, \dots, q\}$, respectively, one has

$$E(p_{\alpha(1)}, \dots, p_{\alpha(p)}, e_{\beta(1)}, \dots, e_{\beta(q)}) = E(p_1, \dots, p_p, e_1, \dots, e_q).$$

Properties:

If E is locally Lipschitz, then the composition $E \circ \mathrm{Pers}$ is also locally Lipschitz.

If E and $\Phi\colon A \subseteq \mathbb{R}^d \to \mathbb{R}^{|K|}$ are semi-algebraic (or definable), then $\mathcal{L} = E \circ \mathrm{Pers} \circ \Phi\colon A \to \mathbb{R}$ has a well-defined Clarke subdifferential $\partial\mathcal{L}(z) := \mathrm{Conv}\{\lim_{z_i \to z} \nabla\mathcal{L}(z_i) \ : \ \mathcal{L}$ is differentiable at $z_i\}$.

Examples

Total persistence.

$$E(D) = \sum_{i=1}^{p} |d_i - b_i|, \quad \text{for} \quad D = ((b_1, d_1), \ldots, (b_p, d_p), e_1, \ldots, e_q).$$

E is semi-algebraic and Lipschitz.

Examples

Total persistence.

$$E(D) = \sum_{i=1}^{p} |d_i - b_i|, \quad \text{for} \quad D = ((b_1, d_1), \dots, (b_p, d_p), e_1, \dots, e_q).$$

E is semi-algebraic and Lipschitz.

Bottleneck distance.

$$E(D) = d_B(D, D^*) = \min_{m} \quad \max_{(p,p^*)\in m} ||p - p^*||_\infty$$

where denoting $\Delta = \{(x,x) : x \in \mathbb{R}\}$ the diagonal in $\mathbb{R}^2$, m is a partial matching between D and D^*, i.e., a subset of $(D \cup \Delta) \times (D^* \cup \Delta)$ such that every point of $D \setminus \Delta$ and $D^* \setminus \Delta$, appears exactly once in m.
E is semi-algebraic and Lipschitz.

Minimization via stochastic (sub-)gradient descent

If E and $\Phi \colon A \subseteq \mathbb{R}^d \to \mathbb{R}^{|K|}$ are semi-algbraic (or definable), then $\mathcal{L} = E \circ \mathrm{Pers} \circ \Phi \colon A \to \mathbb{R}$ has a well-defined Clarke subdifferential $\partial\mathcal{L}(z) := \mathrm{Conv}\{\lim_{z_i \to z} \nabla\mathcal{L}(z_i) \ : \ \mathcal{L}$ is differentiable at $z_i\}$.

Minimization of $\mathcal{L}$ through the differential inclusion

$$\frac{dz}{dt} \in -\partial\mathcal{L}(z(t)) \quad \text{for almost every } t.$$

Standard stochastic subgradient algorithm

$$x_{k+1} = x_k - \alpha_k(y_k + \zeta_k), \ y_k \in \partial\mathcal{L}(x_k),$$

where the sequence $(\alpha_k)_k$ is the learning rate and $(\zeta_k)_k$ is a sequence of random variables.

Minimization via stochastic (sub-)gradient descent

If E and $\Phi\colon A \subseteq \mathbb{R}^d \to \mathbb{R}^{|K|}$ are semi-algbraic (or definable), then $\mathcal{L} = E \circ \mathrm{Pers} \circ \Phi\colon A \to \mathbb{R}$ has a well-defined Clarke subdifferential $\partial\mathcal{L}(z) := \mathrm{Conv}\{\lim_{z_i \to z} \nabla\mathcal{L}(z_i)\ :\ \mathcal{L}$ is differentiable at $z_i\}$.

Minimization of $\mathcal{L}$ through the differential inclusion

$$\frac{dz}{dt} \in -\partial\mathcal{L}(z(t)) \quad \text{for almost every } t.$$

Standard stochastic subgradient algorithm

$$x_{k+1} = x_k - \alpha_k(y_k + \zeta_k),\ y_k \in \partial\mathcal{L}(x_k),$$

where the sequence $(\alpha_k)_k$ is the learning rate and $(\zeta_k)_k$ is a sequence of random variables.

Question: convergence of the algorithm?

Convergence

Convergence follows from [Davis et al, Stochastic subgradient method converges on tame functions. Found. Comp. Math. 2020].

Standard stochastic subgradient algorithm

$$x_{k+1} = x_k - \alpha_k(y_k + \zeta_k), \; y_k \in \partial \mathcal{L}(x_k),$$

where the sequence $(\alpha_k)_k$ is the learning rate and $(\zeta_k)_k$ is a sequence of random variables.

Technical (but classical) assumptions:

1. for any k, $\alpha_k \geq 0$, $\sum_{k=1}^{\infty} \alpha_k = +\infty$ and, $\sum_{k=1}^{\infty} \alpha_k^2 < +\infty$;
2. $\sup_k \|x_k\| < +\infty$, almost surely;
3. denoting by $\mathcal{F}_k$ the increasing sequence of σ-algebras $\mathcal{F}_k = \sigma(x_j, y_j, \zeta_j, j < k)$, there exists a function $p\colon \mathbb{R}^d \to \mathbb{R}$ which is bounded on bounded sets such that almost surely, for any k,

$$\mathbb{E}[\zeta_k | \mathcal{F}_k] = 0 \;\text{ and }\; \mathbb{E}[\|\zeta_k\|^2 | \mathcal{F}_k] < p(x_k).$$

Convergence

Convergence follows from [Davis et al, Stochastic subgradient method converges on tame functions. Found. Comp. Math. 2020].

Standard stochastic subgradient algorithm

$$x_{k+1} = x_k - \alpha_k(y_k + \zeta_k),\ y_k \in \partial\mathcal{L}(x_k),$$

where the sequence $(\alpha_k)_k$ is the learning rate and $(\zeta_k)_k$ is a sequence of random variables.

Theorem:
Let K be a simplicial complex, $A \subseteq \mathbb{R}^d$, and $\Phi\colon A \to \mathbb{R}^{|K|}$ a parametrized family of filtrations of K that is definable in an o-minimal structure. Let $E\colon \mathbb{R}^{|K|} \to \mathbb{R}$ be a definable function of persistence such that $\mathcal{L} = E \circ \mathrm{Pers} \circ \Phi$ is locally Lipschitz. Then, under the above assumptions 1, 2, and 3, almost surely the limit points of the sequence $(x_k)_k$ obtained from the iterations of the algo. are critical points of $\mathcal{L}$ and the sequence $(\mathcal{L}(x_k))_k$ converges.

Numerical illustration

The differential of persistence map is obvious to compute → easy implementation (soon available in GUDHI)

Point cloud optimization

Input: a point cloud X sampled uniformly from the unit square $S = [0,1]^2$

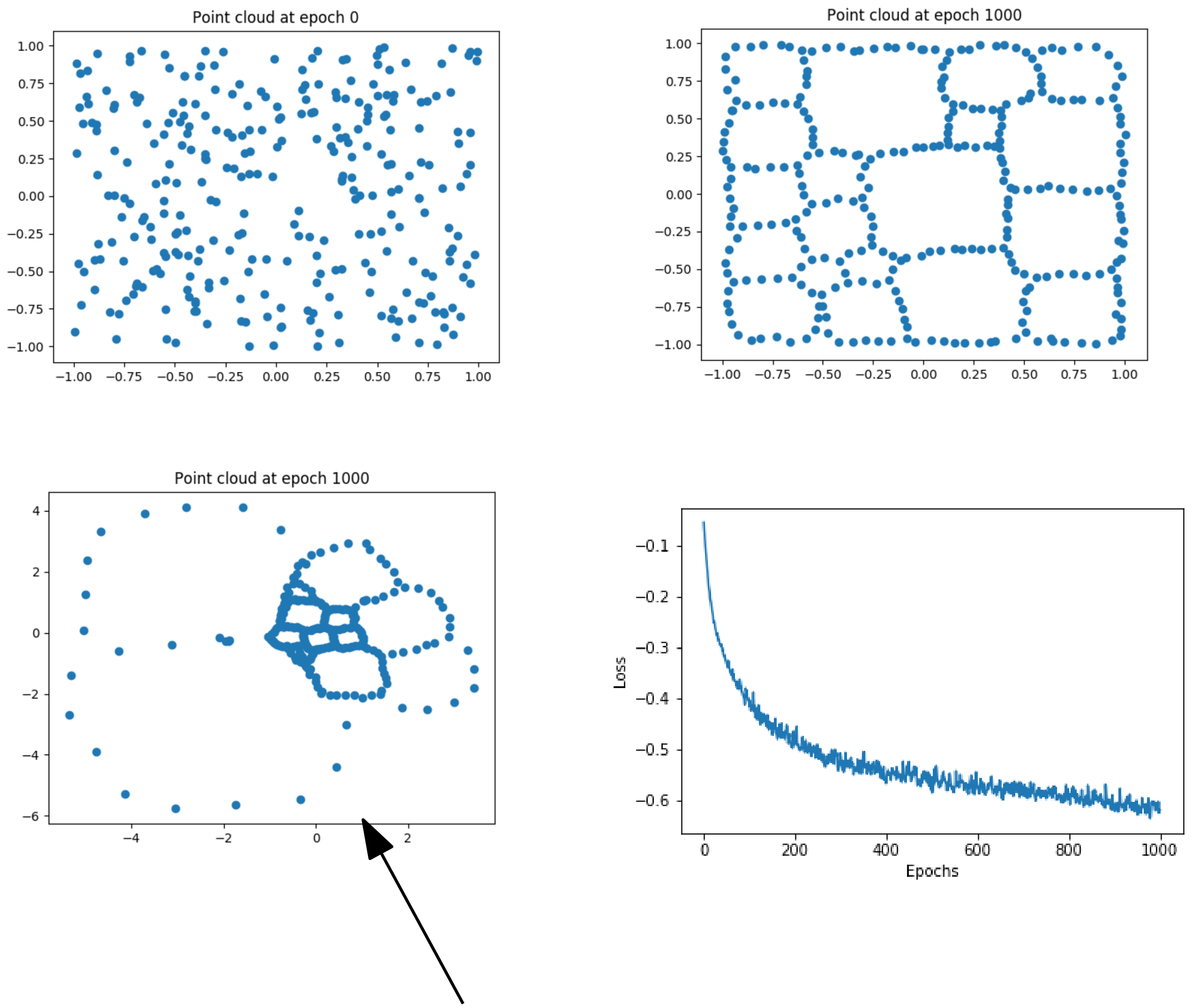
Loss: $\mathcal{L}(X) = P(X) + T(X)$ where

$$T(X) := -\sum_{p \in D} \|p - \pi_\Delta(p)\|_\infty^2$$

with D is the 1-dimensional persistence diagram associated to the Vietoris-Rips filtration of X, π_Δ stands for the projection onto the diagonal Δ, and

$$P(X) := \sum_{x \in X} d(x, S)$$

is a penalty term ensuring that the point coordinates stay in the unit square.



Point cloud at epoch 0

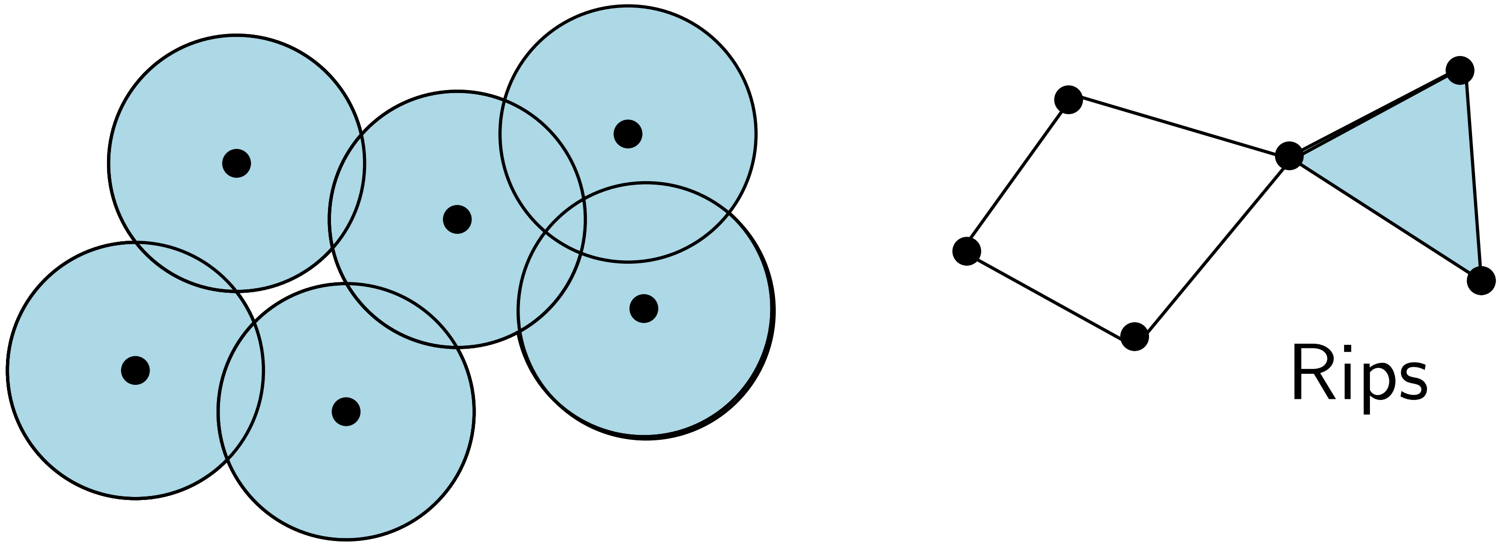
Point cloud at epoch 1000

Point cloud at epoch 1000

With $T(X)$ only

Another example: The density of expected persistence diagrams

Reminder: the Vietoris-Rips filtration



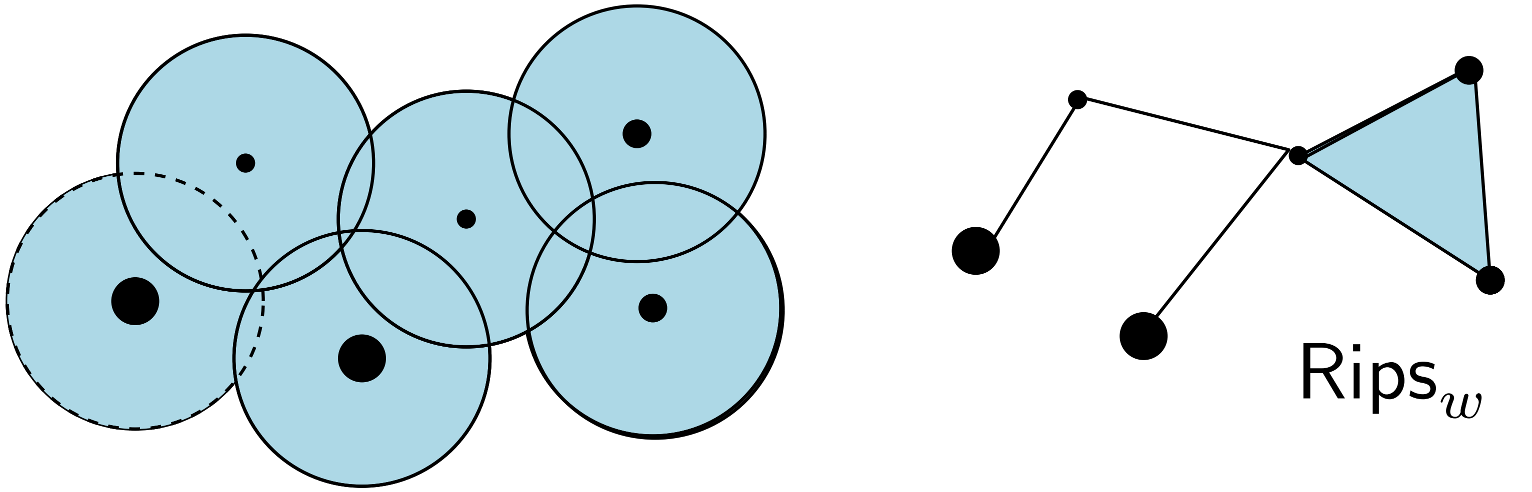
Let V be a point cloud (in a metric space (X, d)).

The Vietoris-Rips complex $\mathrm{Rips}(V)$ is the filtered simplicial complex indexed by $\mathbb{R}$ whose vertex set is V and defined by:

$$\sigma = [p_0 p_1 \cdots p_k] \in \mathrm{Rips}(V, \alpha) \;\text{ iff }\; \forall i, j \in \{0, \cdots, k\},\; d(p_i, p_j) \leq \alpha$$

Easy to compute and fully determined by its 1-skeleton

The weighted Vietoris-Rips filtration



Let V be a weigthed point cloud (in a metric space (X, d)): $V \subset \mathbb{X}$ and $w : V \to \mathbb{R}$.

The weighted Vietoris-Rips complex Rips$_w(V)$ is the filtered simplicial complex indexed by $\mathbb{R}$ whose vertex set is V and defined by:

$$\sigma = [p_0 p_1 \cdots p_k] \in \text{Rips}_w(V, \alpha)$$

iff

$$\forall i, j \in \{0, \cdots, k\},\ d(p_i, p_j) \leq \alpha \text{ and } \forall i \in \{0, \cdots, k\},\ w(p_i) \leq \alpha$$

Persistent homology computation (reminder)

Let $\mathbb{S} = (\mathbb{S}_a \mid a \in \mathbf{R})$ be a finite filtered simplicial complex with N simplices and let $\mathbb{S}_{a_1} \subset \mathbb{S}_{a_2} \subset \cdots \subset \mathbb{S}_{a_N}$ be the discrete filtration induced by the entering times of the simplices: $\mathbb{S}_{a_i} \setminus \mathbb{S}_{a_{i-1}} = \sigma_{a_i}$.

Persistent homology computation (reminder)

Let $\mathbb{S} = (\mathbb{S}_a \mid a \in \mathbf{R})$ be a finite filtered simplicial complex with N simplices and let $\mathbb{S}_{a_1} \subset \mathbb{S}_{a_2} \subset \cdots \subset \mathbb{S}_{a_N}$ be the discrete filtration induced by the entering times of the simplices: $\mathbb{S}_{a_i} \setminus \mathbb{S}_{a_{i-1}} = \sigma_{a_i}$.

Process the simplices according to their order of entrance in the filtration:

Let $k = \dim \sigma_{a_i}$ (ie. $\sigma_{a_i} = [v_0, \cdots, v_k]$)

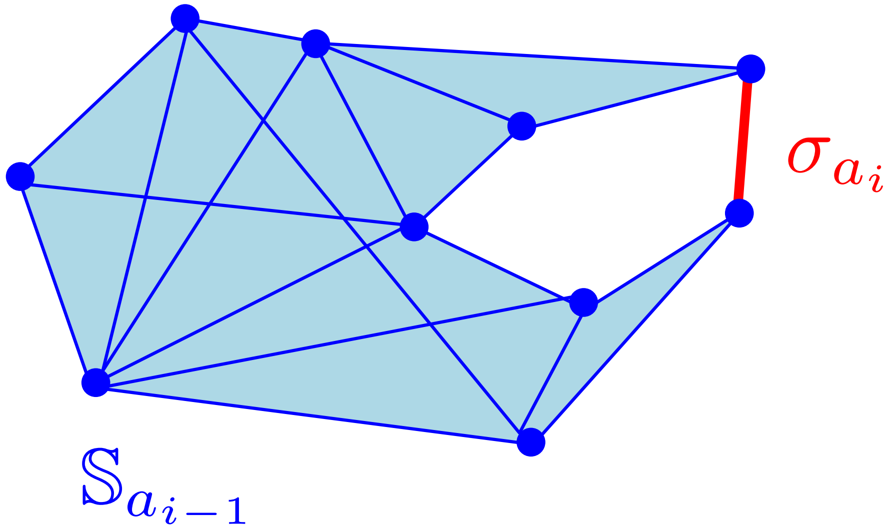
Persistent homology computation (reminder)

Let $\mathbb{S} = (\mathbb{S}_a \mid a \in \mathbf{R})$ be a finite filtered simplicial complex with N simplices and let $\mathbb{S}_{a_1} \subset \mathbb{S}_{a_2} \subset \cdots \subset \mathbb{S}_{a_N}$ be the discrete filtration induced by the entering times of the simplices: $\mathbb{S}_{a_i} \setminus \mathbb{S}_{a_{i-1}} = \sigma_{a_i}$.

Process the simplices according to their order of entrance in the filtration:

Let $k = \dim \sigma_{a_i}$ (ie. $\sigma_{a_i} = [v_0, \cdots, v_k]$)

Case 1: adding σ_{a_i} to $\mathbb{S}_{a_{i-1}}$ creates a new k-dimensional topological feature in $\mathbb{S}_{a_i}$ (new homology class in H_k).



$\Rightarrow$ the birth of a k-dim feature is registered.

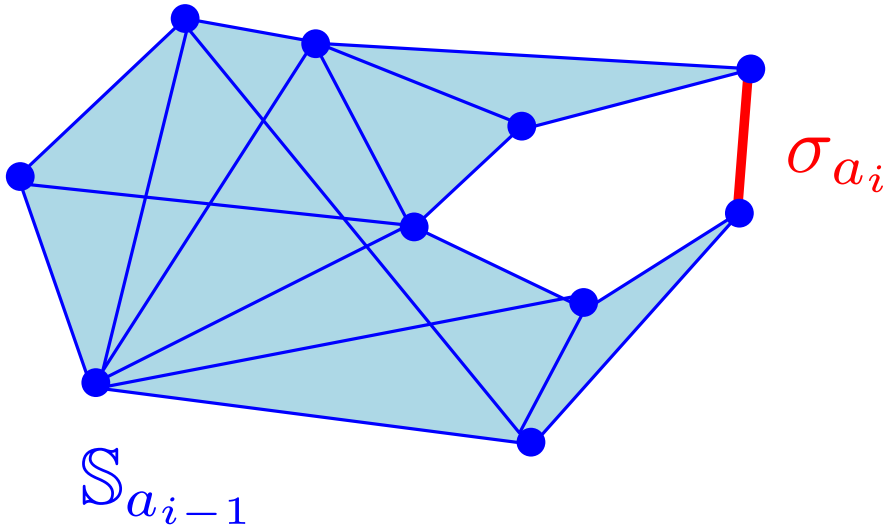
Persistent homology computation (reminder)

Let $\mathbb{S} = (\mathbb{S}_a \mid a \in \mathbf{R})$ be a finite filtered simplicial complex with N simplices and let $\mathbb{S}_{a_1} \subset \mathbb{S}_{a_2} \subset \cdots \subset \mathbb{S}_{a_N}$ be the discrete filtration induced by the entering times of the simplices: $\mathbb{S}_{a_i} \setminus \mathbb{S}_{a_{i-1}} = \sigma_{a_i}$.

Process the simplices according to their order of entrance in the filtration:

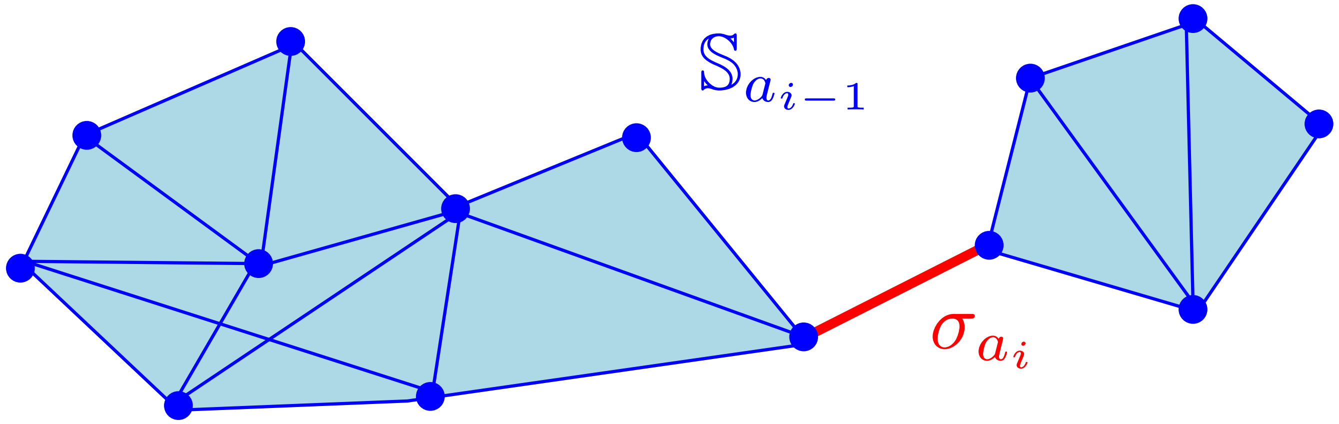
Let $k = \dim \sigma_{a_i}$ (ie. $\sigma_{a_i} = [v_0, \cdots, v_k]$)

Case 1: adding σ_{a_i} to $\mathbb{S}_{a_{i-1}}$ creates a new k-dimensional topological feature in $\mathbb{S}_{a_i}$ (new homology class in H_k).



$\Rightarrow$ the birth of a k-dim feature is registered.

Case 2: adding σ_{a_i} to $\mathbb{S}_{a_{i-1}}$ kills a $(k-1)$-dimensional topological feature in $\mathbb{S}_{a_i}$ (homology class in H_{k-1}).



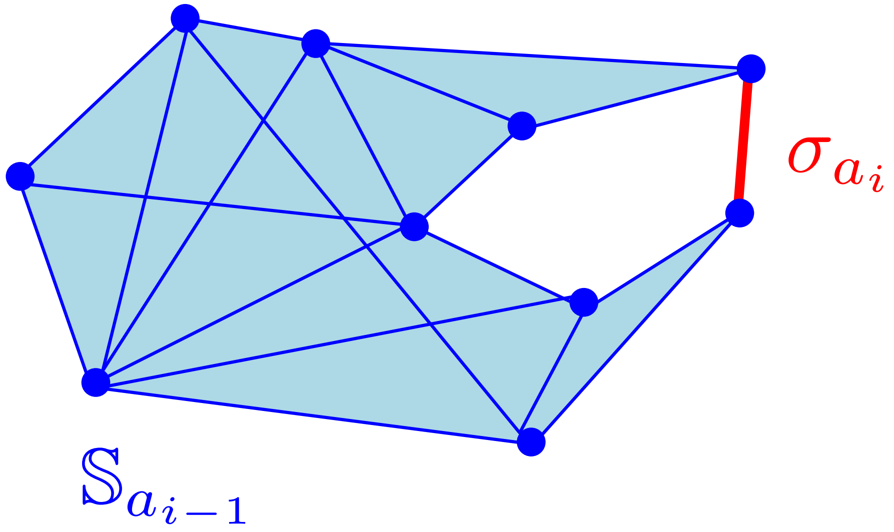
$\Rightarrow$ persistence algo. pairs the simplex σ_{a_i} to the simplex σ_{a_j} that gave birth to the killed feature.

Persistent homology computation (reminder)

Process the simplices according to their order of entrance in the filtration:

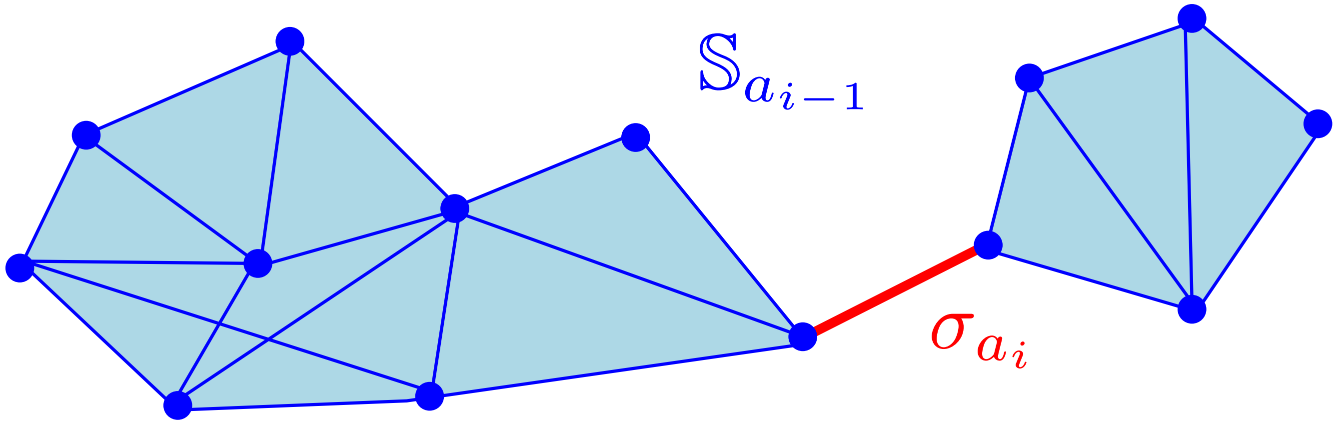
Let $k = \dim \sigma_{a_i}$ (ie. $\sigma_{a_i} = [v_0, \cdots, v_k]$)

Case 1: adding σ_{a_i} to $\mathbb{S}_{a_{i-1}}$ creates a new k-dimensional topological feature in $\mathbb{S}_{a_i}$ (new homology class in H_k).



$\Rightarrow$ the birth of a k-dim feature is registered.

Case 2: adding σ_{a_i} to $\mathbb{S}_{a_{i-1}}$ kills a $(k-1)$-dimensional topological feature in $\mathbb{S}_{a_i}$ (homology class in H_{k-1}).



$\Rightarrow$ persistence algo. pairs the simplex σ_{a_i} to the simplex σ_{a_j} that gave birth to the killed feature.

→ $(\sigma_{a_j}, \sigma_{a_i})$: persistence pair

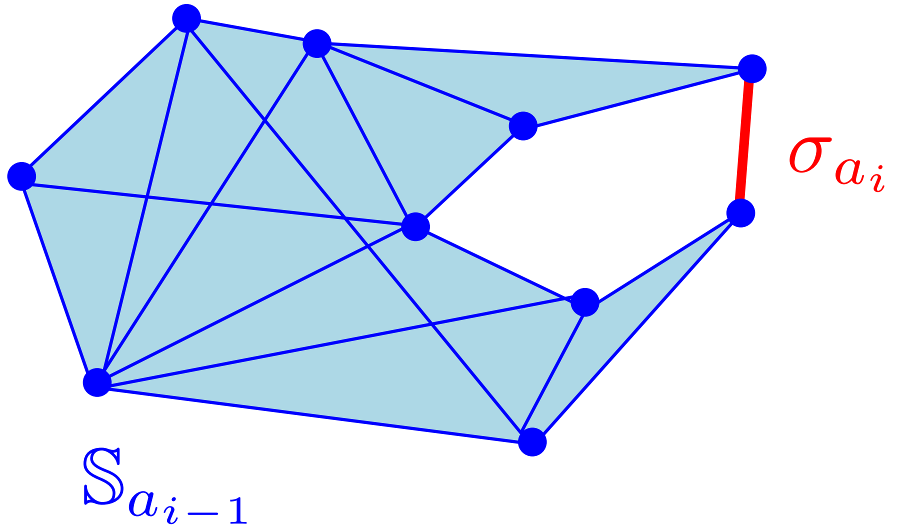
→ $(a_j, a_i) \in \mathbb{R}^2$: point in the persistence diagram

Persistent homology computation (reminder)

Process the simplices according to their order of entrance in the filtration:

Let $k = \dim \sigma_{a_i}$ (ie. $\sigma_{a_i} = [v_0, \cdots, v_k]$)

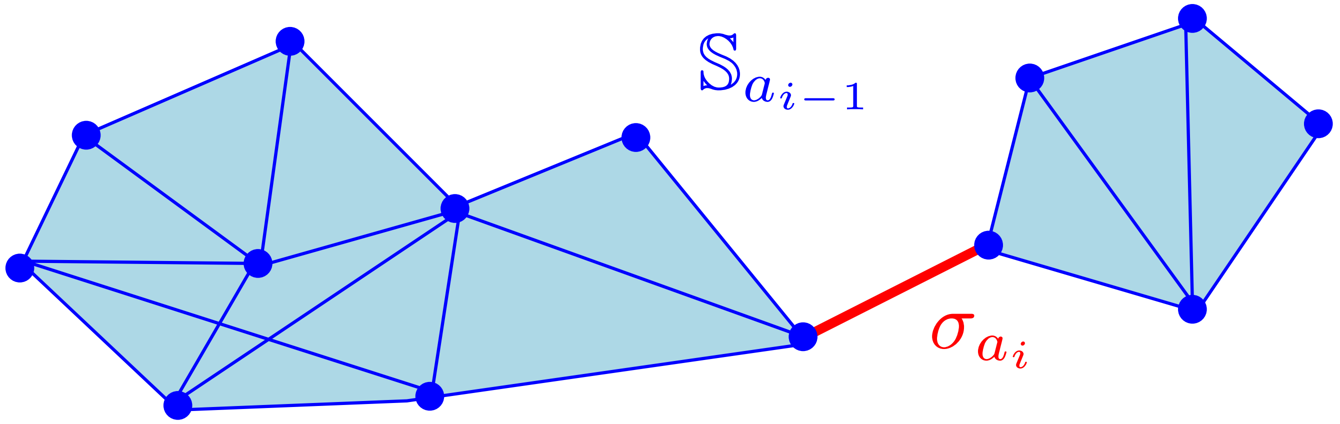
Case 1: adding σ_{a_i} to $\mathbb{S}_{a_{i-1}}$ creates a new k-dimensional topological feature in $\mathbb{S}_{a_i}$ (new homology class in H_k).



$\Rightarrow$ the birth of a k-dim feature is registered.

Important to remember: the persistence pairs are determined by the order on the simplices; the corresponding points in the diagrams are determined by the indices.

Case 2: adding σ_{a_i} to $\mathbb{S}_{a_{i-1}}$ kills a $(k-1)$-dimensional topological feature in $\mathbb{S}_{a_i}$ (homology class in H_{k-1}).



$\Rightarrow$ persistence algo. pairs the simplex σ_{a_i} to the simplex σ_{a_j} that gave birth to the killed feature.

$\rightarrow$ $(\sigma_{a_j}, \sigma_{a_i})$: persistence pair

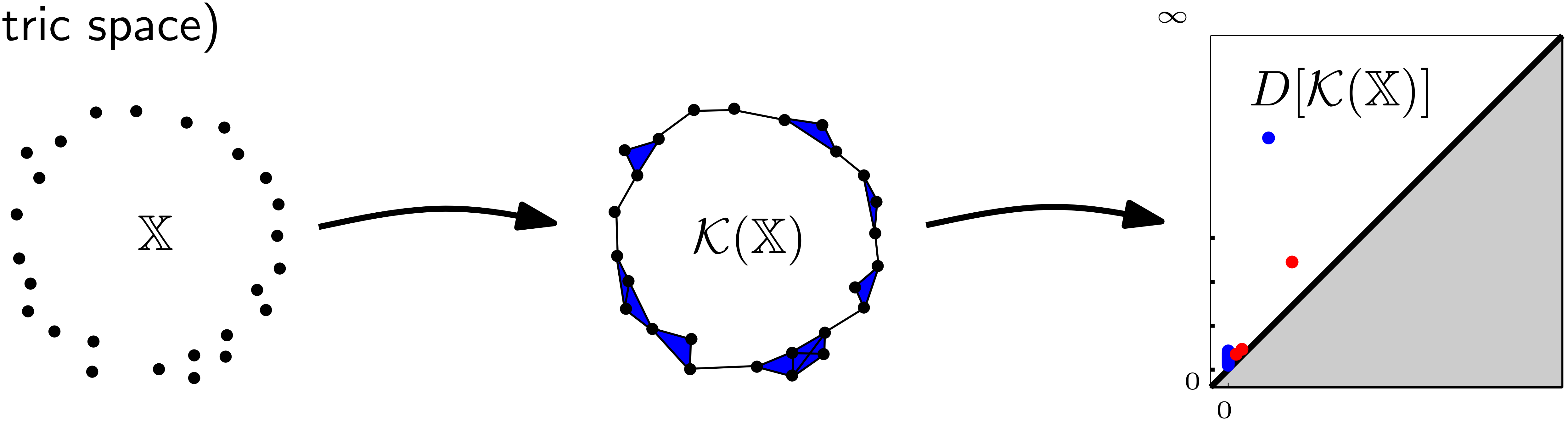
$\rightarrow$ $(a_j, a_i) \in \mathbb{R}^2$: point in the persistence diagram

Statistical setting

$\mathbb{X}$ is now a random point coud (in some metric space)

$\mathcal{K}$ is a deterministic filtration (e.g. Rips)

$D[\mathcal{K}(\mathbb{X})]$ becomes random

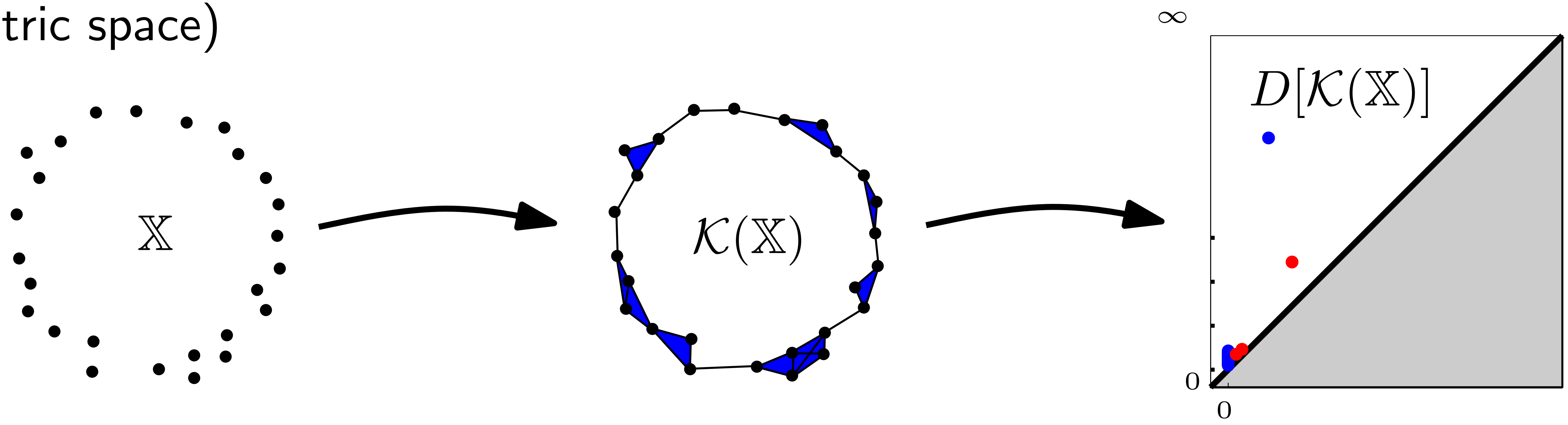


Statistical setting

$\mathbb{X}$ is now a random point coud (in some metric space)

$\mathcal{K}$ is a deterministic filtration (e.g. Rips)

$D[\mathcal{K}(\mathbb{X})]$ becomes random



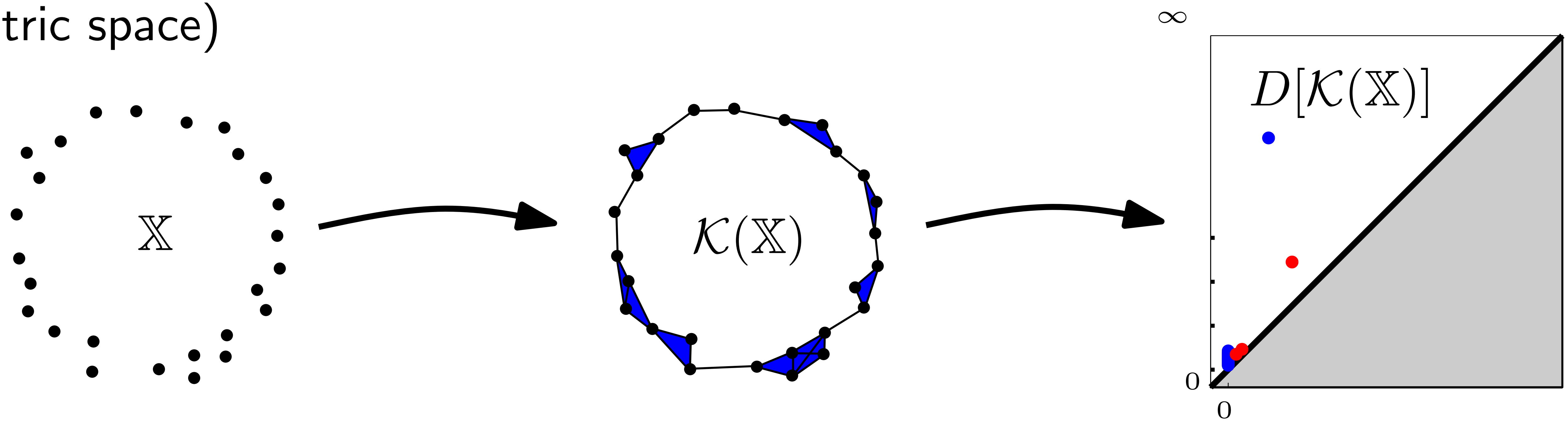
What can be said about the distribution of diagrams $D[\mathcal{K}(\mathbb{X})]$?

Statistical setting

$\mathbb{X}$ is now a random point coud (in some metric space)

$\mathcal{K}$ is a deterministic filtration (e.g. Rips)

$D[\mathcal{K}(\mathbb{X})]$ becomes random



What can be said about the distribution of diagrams $D[\mathcal{K}(\mathbb{X})]$?

- Stability properties $\Rightarrow$ asymptotic properties, confidence bands, Wasserstein stability,...
- Other representation of persistence (landscapes, Betti curves, pers. images, kernels,...)

Goal: understand the structure of $E[D[\mathcal{K}(\mathbb{X})]]$ in the non asymptotic setting ($|\mathbb{X}| = n$ is fixed, or bounded)

Filtrations revisited

Let $n > 0$ be an integer,
$\mathcal{F}_n$: the collection of non-empty subsets of $\{1, \dots, n\}$,
M: a real analytic compact d-dim. connected manifold (poss. with boundary).

Filtering function:

$$\varphi = (\varphi[J])_{J \in \mathcal{F}_n} : M^n \to \mathbb{R}^{|\mathcal{F}_n|}$$

satisfiying the following conditions:

(K2) *Invariance by permutation:* For $J \in \mathcal{F}_n$ and for $(x_1, \dots, x_n) \in M^n$, if τ is a permutation of the entries having support included in J, then $\varphi[J](x_{\tau(1)}, \dots, x_{\tau(n)}) = \varphi[J](x_1, \dots, x_n)$.

(K3) *Monotony:* For $J \subset J' \in \mathcal{F}_n$, $\varphi[J] \leq \varphi[J']$.

Given $x = (x_1, \cdots, x_n)$, $\varphi(x)$ induces an order on the faces of the simplex with n vertices that is a filtration $\mathcal{K}(x)$:

$$\forall J \in \mathcal{F}_n, \; J \in \mathcal{K}(x, r) \Longleftrightarrow \varphi[J](x) \leq r.$$

Filtrations revisited

Not: for $x = (x_1, \ldots, x_n) \in M^n$ and for J a simplex, $x(J) := (x_j)_{j \in J}$

(K1) *Absence of interaction:* For $J \in \mathcal{F}_n$, $\varphi[J](x)$ only depends on $x(J)$.

(K2) *Invariance by permutation:* For $J \in \mathcal{F}_n$ and for $(x_1, \ldots, x_n) \in M^n$, if τ is a permutation of the entries having support included in J, then $\varphi[J](x_{\tau(1)}, \ldots, x_{\tau(n)}) = \varphi[J](x_1, \ldots, x_n)$.

(K3) *Monotony:* For $J \subset J' \in \mathcal{F}_n$, $\varphi[J] \leq \varphi[J']$.

(K4) *Compatibility:* For a simplex $J \in \mathcal{F}_n$ and for $j \in J$, if $\varphi[J](x_1, \ldots, x_n)$ is not a function of x_j on some open set U of M^n, then $\varphi[J] \equiv \varphi[J \backslash \{j\}]$ on U.

(K5) *Smoothness:* The function φ is subanalytic and the gradient of each of its entries (which is defined a.s.e.) is non vanishing a.s.e..

The example of the Vietoris-Rips filtration

$$\varphi[J](x) = \max_{i,j \in J} d(x_i, x_j)$$

- (K1) *Absence of interaction:* For $J \in \mathcal{F}_n$, $\varphi[J](x)$ only depends on $x(J)$.
- (K2) *Invariance by permutation:* For $J \in \mathcal{F}_n$ and for $(x_1, \ldots, x_n) \in M^n$, if τ is a permutation of the entries having support included in J, then $\varphi[J](x_{\tau(1)}, \ldots, x_{\tau(n)}) = \varphi[J](x_1, \ldots, x_n)$.
- (K3) *Monotony:* For $J \subset J' \in \mathcal{F}_n$, $\varphi[J] \leq \varphi[J']$.
- (K4) *Compatibility:* For a simplex $J \in \mathcal{F}_n$ and for $j \in J$, if $\varphi[J](x_1, \ldots, x_n)$ is not a function of x_j on some open set U of M^n, then $\varphi[J] \equiv \varphi[J \backslash \{j\}]$ on U.
- (K5') *Smoothness:* The function φ is subanalytic and the gradient of each of its entries J of size larger than 1 is non vanishing a.e. and for $J = \{j\}$, $\varphi[\{j\}] \equiv 0$.

The example of the Vietoris-Rips filtration

$$\varphi[J](x) = \max_{i,j \in J} d(x_i, x_j)$$

- (K1) *Absence of interaction:* For $J \in \mathcal{F}_n$, $\varphi[J](x)$ only depends on $x(J)$. ✓
- (K2) *Invariance by permutation:* For $J \in \mathcal{F}_n$ and for $(x_1, \dots, x_n) \in M^n$, if τ is a permutation of the entries having support included in J, then $\varphi[J](x_{\tau(1)}, \dots, x_{\tau(n)}) = \varphi[J](x_1, \dots, x_n)$.
- (K3) *Monotony:* For $J \subset J' \in \mathcal{F}_n$, $\varphi[J] \leq \varphi[J']$.
- (K4) *Compatibility:* For a simplex $J \in \mathcal{F}_n$ and for $j \in J$, if $\varphi[J](x_1, \dots, x_n)$ is not a function of x_j on some open set U of M^n, then $\varphi[J] \equiv \varphi[J \backslash \{j\}]$ on U.
- (K5') *Smoothness:* The function φ is subanalytic and the gradient of each of its entries J of size larger than 1 is non vanishing a.e. and for $J = \{j\}$, $\varphi[\{j\}] \equiv 0$.

The example of the Vietoris-Rips filtration

$$\varphi[J](x) = \max_{i,j \in J} d(x_i, x_j)$$

(K1) *Absence of interaction:* For $J \in \mathcal{F}_n$, $\varphi[J](x)$ only depends on $x(J)$. ✓

(K2) *Invariance by permutation:* For $J \in \mathcal{F}_n$ and for $(x_1, \ldots, x_n) \in M^n$, if τ is a permutation of the entries having support included in J, then $\varphi[J](x_{\tau(1)}, \ldots, x_{\tau(n)}) = \varphi[J](x_1, \ldots, x_n)$. ✓

(K3) *Monotony:* For $J \subset J' \in \mathcal{F}_n$, $\varphi[J] \leq \varphi[J']$.

(K4) *Compatibility:* For a simplex $J \in \mathcal{F}_n$ and for $j \in J$, if $\varphi[J](x_1, \ldots, x_n)$ is not a function of x_j on some open set U of M^n, then $\varphi[J] \equiv \varphi[J \backslash \{j\}]$ on U.

(K5') *Smoothness:* The function φ is subanalytic and the gradient of each of its entries J of size larger than 1 is non vanishing a.e. and for $J = \{j\}$, $\varphi[\{j\}] \equiv 0$.

The example of the Vietoris-Rips filtration

$$\varphi[J](x) = \max_{i,j \in J} d(x_i, x_j)$$

(K1) *Absence of interaction:* For $J \in \mathcal{F}_n$, $\varphi[J](x)$ only depends on $x(J)$. ✓

(K2) *Invariance by permutation:* For $J \in \mathcal{F}_n$ and for $(x_1, \dots, x_n) \in M^n$, if τ is a permutation of the entries having support included in J, then $\varphi[J](x_{\tau(1)}, \dots, x_{\tau(n)}) = \varphi[J](x_1, \dots, x_n)$. ✓

(K3) *Monotony:* For $J \subset J' \in \mathcal{F}_n$, $\varphi[J] \leq \varphi[J']$. ✓

(K4) *Compatibility:* For a simplex $J \in \mathcal{F}_n$ and for $j \in J$, if $\varphi[J](x_1, \dots, x_n)$ is not a function of x_j on some open set U of M^n, then $\varphi[J] \equiv \varphi[J \backslash \{j\}]$ on U.

(K5') *Smoothness:* The function φ is subanalytic and the gradient of each of its entries J of size larger than 1 is non vanishing a.e. and for $J = \{j\}$, $\varphi[\{j\}] \equiv 0$.

The example of the Vietoris-Rips filtration

$$\varphi[J](x) = \max_{i,j \in J} d(x_i, x_j)$$

(K1) *Absence of interaction:* For $J \in \mathcal{F}_n$, $\varphi[J](x)$ only depends on $x(J)$. ✓

(K2) *Invariance by permutation:* For $J \in \mathcal{F}_n$ and for $(x_1, \ldots, x_n) \in M^n$, if τ is a permutation of the entries having support included in J, then $\varphi[J](x_{\tau(1)}, \ldots, x_{\tau(n)}) = \varphi[J](x_1, \ldots, x_n)$. ✓

(K3) *Monotony:* For $J \subset J' \in \mathcal{F}_n$, $\varphi[J] \leq \varphi[J']$. ✓

(K4) *Compatibility:* For a simplex $J \in \mathcal{F}_n$ and for $j \in J$, if $\varphi[J](x_1, \ldots, x_n)$ is not a function of x_j on some open set U of M^n, then $\varphi[J] \equiv \varphi[J \backslash \{j\}]$ on U. ✓

(K5') *Smoothness:* The function φ is subanalytic and the gradient of each of its entries J of size larger than 1 is non vanishing a.e. and for $J = \{j\}$, $\varphi[\{j\}] \equiv 0$.

The example of the Vietoris-Rips filtration

$$\varphi[J](x) = \max_{i,j \in J} d(x_i, x_j)$$

(K1) *Absence of interaction:* For $J \in \mathcal{F}_n$, $\varphi[J](x)$ only depends on $x(J)$. ✓

(K2) *Invariance by permutation:* For $J \in \mathcal{F}_n$ and for $(x_1, \ldots, x_n) \in M^n$, if τ is a permutation of the entries having support included in J, then $\varphi[J](x_{\tau(1)}, \ldots, x_{\tau(n)}) = \varphi[J](x_1, \ldots, x_n)$. ✓

(K3) *Monotony:* For $J \subset J' \in \mathcal{F}_n$, $\varphi[J] \leq \varphi[J']$. ✓

(K4) *Compatibility:* For a simplex $J \in \mathcal{F}_n$ and for $j \in J$, if $\varphi[J](x_1, \ldots, x_n)$ is not a function of x_j on some open set U of M^n, then $\varphi[J] \equiv \varphi[J \backslash \{j\}]$ on U. ✓

(K5') *Smoothness:* The function φ is subanalytic and the gradient of each of its entries J of size larger than 1 is non vanishing a.e. and for $J = \{j\}$, $\varphi[\{j\}] \equiv 0$. ✓

The density of expected persistence diagrams

Theorem: Fix $n \geq 1$. Assume that:

- M is a real analytic compact d-dimensional connected submanifold possibly with boundary,
- $\mathbb{X}$ is a random variable on M^n having a density with respect to the Haussdorf measure $\mathcal{H}_{dn}$,
- $\mathcal{K}$ satisfies the assumptions (K1)-(K5).

Then, for $s \geq 0$, $E[D_s[\mathcal{K}(\mathbb{X})]]$ has a density with respect to the Lebesgue measure on the half plane $\Delta = \{(b, d) \in \mathbb{R}^2 : b \leq d\}$.

The density of expected persistence diagrams

Theorem: Fix $n \geq 1$. Assume that:

- M is a real analytic compact d-dimensional connected riemannian manifold possibly with boundary,
- $\mathbb{X}$ is a random variable on M^n having a density with respect to the Haussdorf measure $\mathcal{H}_{dn}$,
- $\mathcal{K}$ satisfies the assumptions (K1)-(K4) and (K5').

Then, for $s \geq 1$, $E[D_s[\mathcal{K}(\mathbb{X})]]$ has a density with respect to the Lebesgue measure on Δ. Moreover, $E[D_0[\mathcal{K}(\mathbb{X})]]$ has a density with respect to the Lebesgue measure on the vertical line $\{0\} \times [0, \infty)$.

The density of expected persistence diagrams

Technical assumption (related to finiteness properties of subanalytic sets) that can be discarded in most cases.

Theorem: Fix $n \geq 1$. Assume that:

- M is a real analytic compact d-dimensional connected riemannian manifold possibly with boundary,
- $\mathbb{X}$ is a random variable on M^n having a density with respect to the Haussdorf measure $\mathcal{H}_{dn}$,
- $\mathcal{K}$ satisfies the assumptions (K1)-(K4) and (K5').

Then, for $s \geq 1$, $E[D_s[\mathcal{K}(\mathbb{X})]]$ has a density with respect to the Lebesgue measure on Δ. Moreover, $E[D_0[\mathcal{K}(\mathbb{X})]]$ has a density with respect to the Lebesgue measure on the vertical line $\{0\} \times [0, \infty)$.

The density of expected persistence diagrams

Technical assumption (related to finiteness properties of subanalytic sets) that can be discarded in most cases.

Theorem: Fix $n \geq 1$. Assume that:

- M is a real analytic compact d-dimensional connected riemannian manifold possibly with boundary,
- $\mathbb{X}$ is a random variable on M^n having a density with respect to the Haussdorf measure $\mathcal{H}_{dn}$,
- $\mathcal{K}$ satisfies the assumptions (K1)-(K4) and (K5').

Then, for $s \geq 1$, $E[D_s[\mathcal{K}(\mathbb{X})]]$ has a density with respect to the Lebesgue measure on Δ. Moreover, $E[D_0[\mathcal{K}(\mathbb{X})]]$ has a density with respect to the Lebesgue measure on the vertical line $\{0\} \times [0, \infty)$.

Theorem [smoothness]: Under the assumption of previous theorem, if moreover $\mathbb{X} \in M^n$ has a density of class C^k with respect to $\mathcal{H}_{nd}$. Then, for $s \geq 0$, the density of $E[D_s[\mathcal{K}(\mathbb{X})]]$ is of class C^k.

The Hausdorff measure and the co-area formula

Definition: Let k be a non-negative number. For $A \subset \mathbb{R}^D$, and $\delta > 0$, consider

$$\mathcal{H}_k^\delta(A) := \inf \left\{ \sum_i \text{diam}(U_i)^k, A \subset \bigcup_i U_i \text{ and } \text{diam}(U_i) < \delta \right\}.$$

The *k-dimensional Haussdorf measure* on $\mathbb{R}^D$ of A is defined by $\mathcal{H}_k(A) := \lim_{\delta \to 0} \mathcal{H}_k^\delta(A)$.

Theorem [Co-area formula]: Let M (resp. N) be a smooth Riemannian manifold of dimension m (resp n). Assume that $m \geq n$ and let $\Phi : M \to N$ be a differentiable map. Denote by $D\Phi$ the differential of Φ. The Jacobian of Φ is defined by $J\Phi = \sqrt{\det((D\Phi) \times (D\Phi)^t)}$. For $f : M \to \mathbb{R}_+$ a positive measurable function, the following equality holds:

$$\int_M f(x) J\Phi(x) d\mathcal{H}_m(x) = \int_N \left(\int_{x \in \Phi^{-1}(\{y\})} f(x) d\mathcal{H}_{m-n}(x) \right) d\mathcal{H}_n(y).$$

Background on subanalytic sets

Let $M \subset \mathbb{R}^D$ be a connected real analytic submanifold (poss. with boundary), of dim. d.

- $X \subset M$ is *semianalytic* if any $p \in M$ has a neighbourhood U_p such that

$$X \cap U_p = \bigcup_{i=1}^{p} \bigcap_{j=1}^{q} X_{ij},$$

where X_{ij} is either $f_{ij}^{-1}(\{0\})$ or $f_{ij}^{-1}((0,\infty))$ for some analytic functions $f_{ij} : U \to \mathbb{R}$.

- $X \subseteq M$ is *subanalytic* if for each point of M, there exists a neighborhood U of this point, a real analytic manifold N and A, a relatively compact semianalytic set of $N \times M$, such that $X \cap U$ is the projection of A on M.

- $f : X \to \mathbb{R}$ is subanalytic if its graph is subanalytic in $M \times \mathbb{R}$. The set of real-valued subanalytic functions on X is denoted by $\mathcal{S}(X)$.

Background on subanalytic sets

Let $M \subset \mathbb{R}^D$ be a connected real analytic submanifold (poss. with boundary), of dim. d.

- $x \in X \subseteq M$ is smooth of dimension k if, in some neighbourhood of x in M, X is an analytic submanifold (of dimension k).
- The dimension of X is the maximal dimension of a smooth point of X.
- $\mathrm{Reg}(X)$: regular points of X, i.e. smooth points of X of dimension d.
- $\mathrm{Sing}(X)$: sigular points of X, i.e. the non-regular points.
- $\mathrm{Reg}(X)$ is an open subset of M, possibly empty.

Background on subanalytic sets

Let $M \subset \mathbb{R}^D$ be a connected real analytic submanifold (poss. with boundary), of dim. d.

Lemma: For $f \in \mathcal{S}(M)$, the set $A(f)$ on which f is analytic is an open subanalytic set of M. Its complement is a subanalytic set of dimension smaller than d.

Lemma: Let X be a subanalytic subset of M and let $f, g : X \to \mathbb{R}$ be subanalytic such that the image of a bounded set is bounded. Then
- fg and $f + g$ are subanalytic,
- the sets $f^{-1}(\{0\})$ and $f^{-1}((0, \infty))$ are subanalytic in M.

Lemma: Let X be a subanalytic subset of M. If the dimension of X is smaller than d, then $\mathcal{H}_d(X) = 0$.

Background on subanalytic sets

Let $M \subset \mathbb{R}^D$ be a connected real analytic submanifold (poss. with boundary), of dim. d.

Lemma: For $f \in \mathcal{S}(M)$, the set $A(f)$ on which f is analytic is an open subanalytic set of M. Its complement is a subanalytic set of dimension smaller than d.

Lemma: Let X be a subanalytic subset of M and let $f, g : X \to \mathbb{R}$ be subanalytic such that the image of a bounded set is bounded. Then
- fg and $f + g$ are subanalytic,
- the sets $f^{-1}(\{0\})$ and $f^{-1}((0, \infty))$ are subanalytic in M.

Lemma: Let X be a subanalytic subset of M. If the dimension of X is smaller than d, then $\mathcal{H}_d(X) = 0$.

Consequences:
- $\mathcal{H}_d(X) = \mathcal{H}_d(\text{Reg}(X))$,
- for any $f \in \mathcal{S}(M)$, the gradient ∇f is defined everywhere but on some subanalytic set of dimension smaller than d (of zero Hausdorff measure).

Sketch of proof

1. There exists a partition of the complement of a (subanalytic) set of measure 0 in M^n by open sets $V_1, \cdots, V_R$ such that :

- the order of the simplices of $\mathcal{K}(x)$ is constant on each V_r,
- for any $r = 1, \cdots, R$, and any $x \in V_r$,

$$D_s[\mathcal{K}(x)] = \sum_{i=1}^{N_r} \delta_{\mathbf{r}_i}$$

with $\mathbf{r}_i = (\varphi[J_{i_1}](x), \varphi[J_{i_2}](x))$ where N_r, J_{i_1}, J_{i_2} only depends on V_r.

- J_{i_1}, J_{i_2} can be chosen so that the differential of

$$\Phi_{ir} : x \in V_r \to \mathbf{r}_i = (\varphi[J_{i_1}](x), \varphi[J_{i_2}](x))$$

has maximal rank (2).

Sketch of proof

2. The expected diagram can be written as

$$\begin{aligned} E[D_s[\mathcal{K}(\mathbb{X})]] &= \sum_{r=1}^{R} E\left[\mathbb{1}\{\mathbb{X} \in V_r\} D_s[\mathcal{K}(\mathbb{X})]\right] = \sum_{r=1}^{R} E\left[\mathbb{1}\{\mathbb{X} \in V_r\} \sum_{i=1}^{N_r} \delta_{\mathbf{r}_i}\right] \\ &= \sum_{r=1}^{R} \sum_{i=1}^{N_r} E\left[\mathbb{1}\{\mathbb{X} \in V_r\} \delta_{\mathbf{r}_i}\right] \end{aligned}$$

Sketch of proof

2. The expected diagram can be written as

$$E[D_s[\mathcal{K}(\mathbb{X})]] = \sum_{r=1}^{R} E\left[\mathbb{1}\{\mathbb{X}\in V_r\}D_s[\mathcal{K}(\mathbb{X})]\right] = \sum_{r=1}^{R} E\left[\mathbb{1}\{\mathbb{X}\in V_r\}\sum_{i=1}^{N_r}\delta_{\mathbf{r}_i}\right]$$

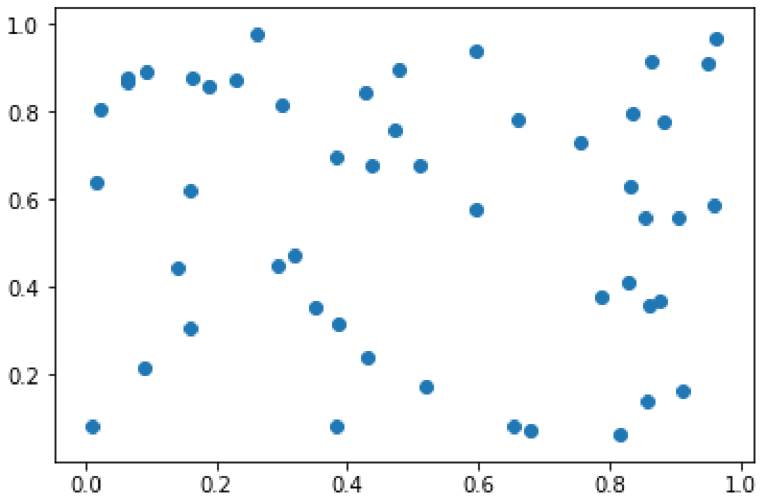
$$= \sum_{r=1}^{R}\sum_{i=1}^{N_r} \underbrace{E\left[\mathbb{1}\{\mathbb{X}\in V_r\}\delta_{\mathbf{r}_i}\right]}_{\mu_{ir}}$$

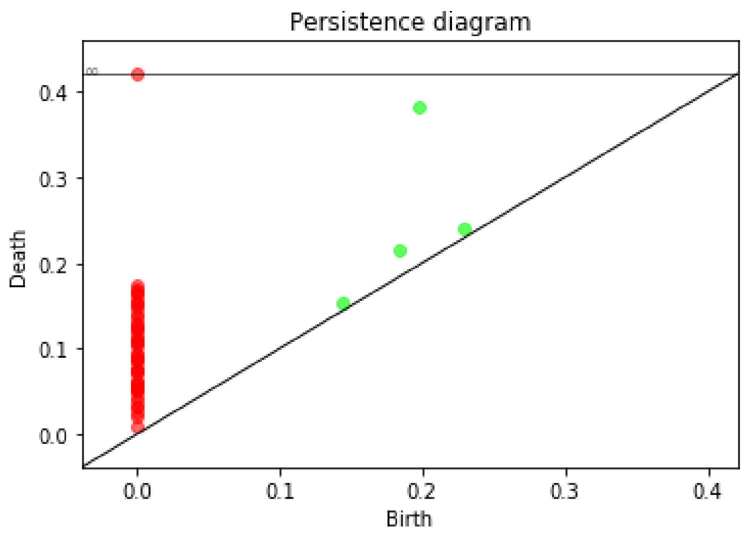
3. Use the co-area formula:

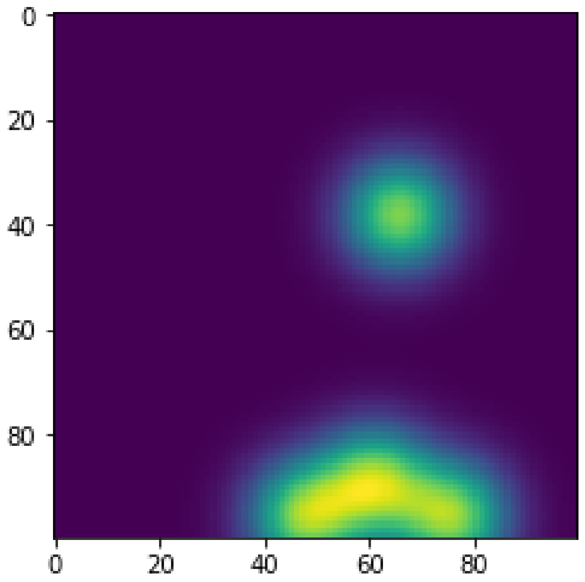
$$\begin{aligned}
\mu_{ir}(B) &= P(\Phi_{ir}(\mathbb{X})\in B, \mathbb{X}\in V_r)\\
&= \int_{V_r} \mathbb{1}\{\Phi_{ir}(x)\in B\}\overbrace{\kappa(x)}^{\text{Density of }\mathbb{X}}d\mathcal{H}_{nd}(x)\\
&= \int_{u\in B}\underbrace{\int_{x\in\Phi_{ir}^{-1}(u)}(J\Phi_{ir}(x))^{-1}\kappa(x)d\mathcal{H}_{nd-2}(x)}_{\text{Density of }\mu_{ir}}du.
\end{aligned}$$

Persistence images

[Adams et al, JMLR 2017]







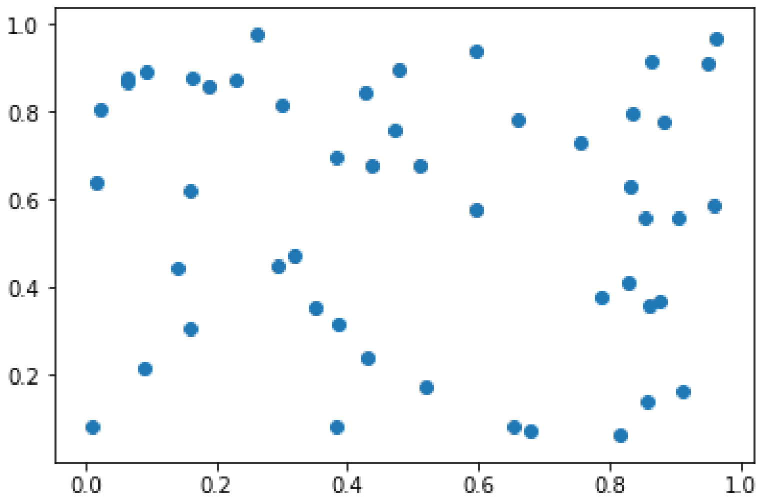
For $K : \mathbb{R}^2 \to \mathbb{R}$ a kernel and H a bandwidth matrix (e.g. a symmetric positive definite matrix), pose for $u \in \mathbb{R}^2$, $K_H(z) = |H|^{-1/2} K(H^{-1/2} \cdot u)$

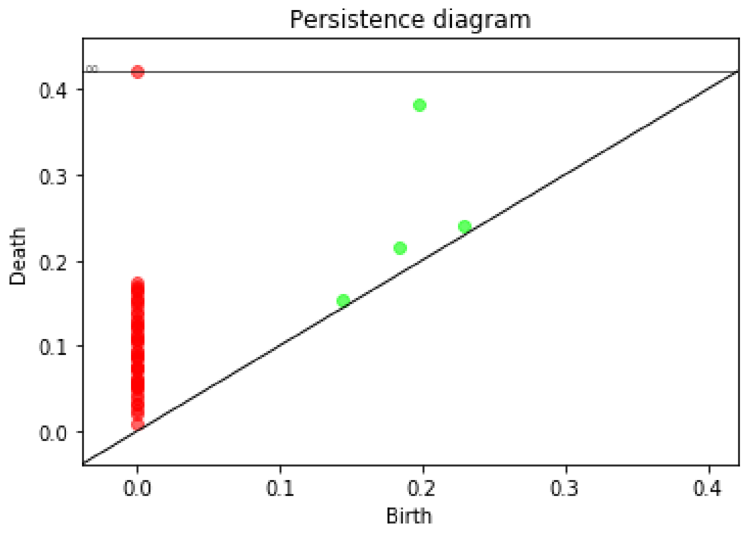
For $D = \sum_i \delta_{\mathbf{r}_i}$ a diagram, $K : \mathbb{R}^2 \to \mathbb{R}$ a kernel, H a bandwidth matrix and $w : \mathbb{R}^2 \to \mathbb{R}_+$ a weight function, one defines the persistence surface of D with kernel K and weight function w by:

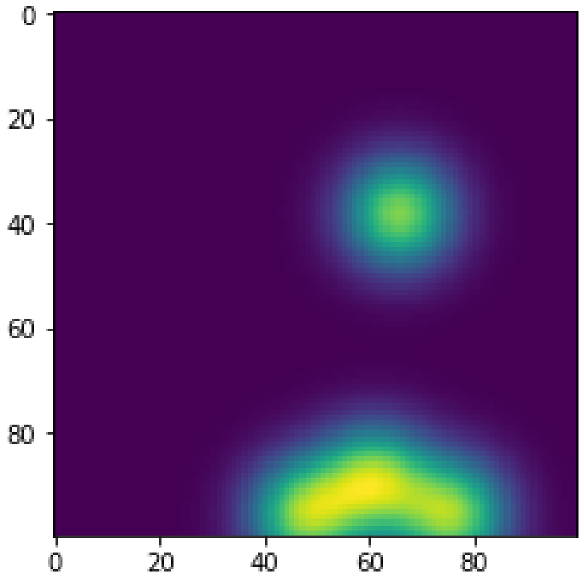
$$\forall z \in \mathbb{R}^2,\ \rho(D)(u) = \sum_i w(\mathbf{r}_i) K_H(u - \mathbf{r}_i) = D(wK_H(u - \cdot))$$

Persistence images

[Adams et al, JMLR 2017]







For $K : \mathbb{R}^2 \to \mathbb{R}$ a kernel and H a bandwidth matrix (e.g. a symmetric positive definite matrix), pose for $u \in \mathbb{R}^2$, $K_H(z) = |H|^{-1/2} K(H^{-1/2} \cdot u)$

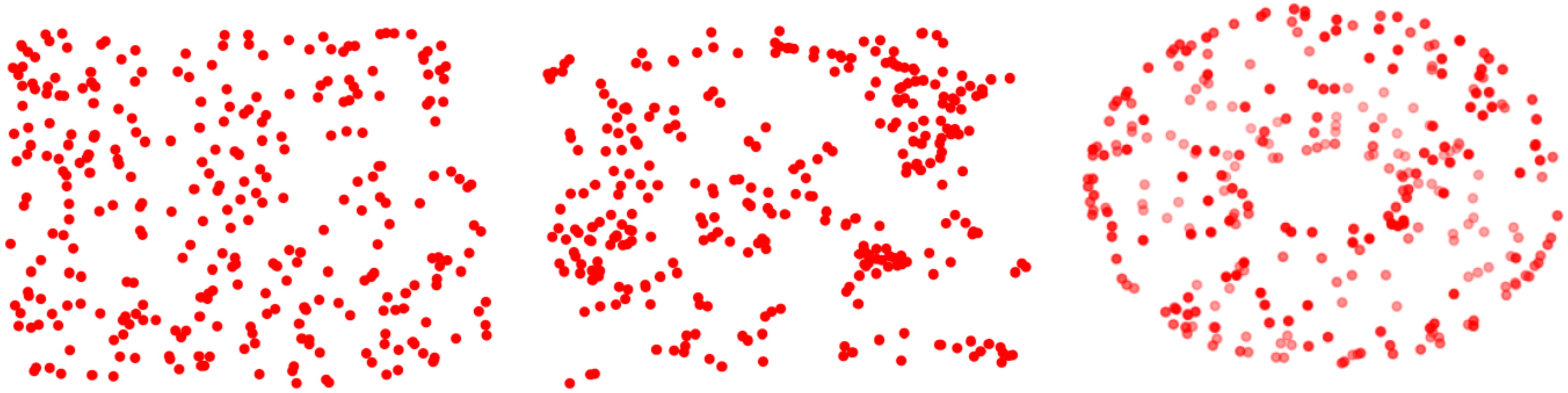
For $D = \sum_i \delta_{\mathbf{r}_i}$ a diagram, $K : \mathbb{R}^2 \to \mathbb{R}$ a kernel, H a bandwidth matrix and $w : \mathbb{R}^2 \to \mathbb{R}_+$ a weight function, one defines the persistence surface of D with kernel K and weight function w by:

$$\forall z \in \mathbb{R}^2, \ \rho(D)(u) = \sum_i w(\mathbf{r}_i) K_H(u - \mathbf{r}_i) = D(w K_H(u - \cdot))$$

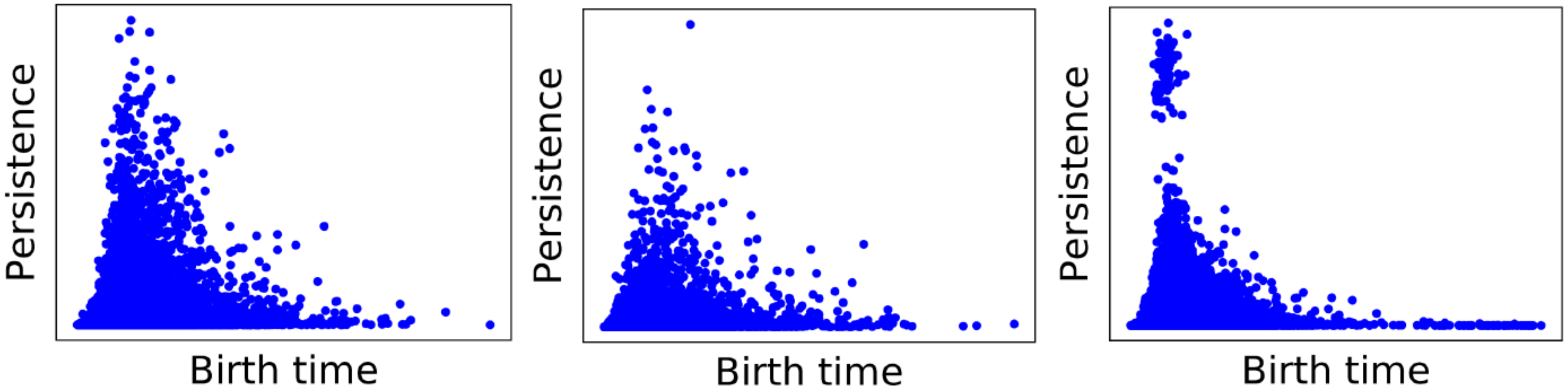
$\Rightarrow$ persistence surfaces can be seen as kernel based estimators of $E[D_s[\mathcal{K}(\mathbb{X})]]$.

Persistence images

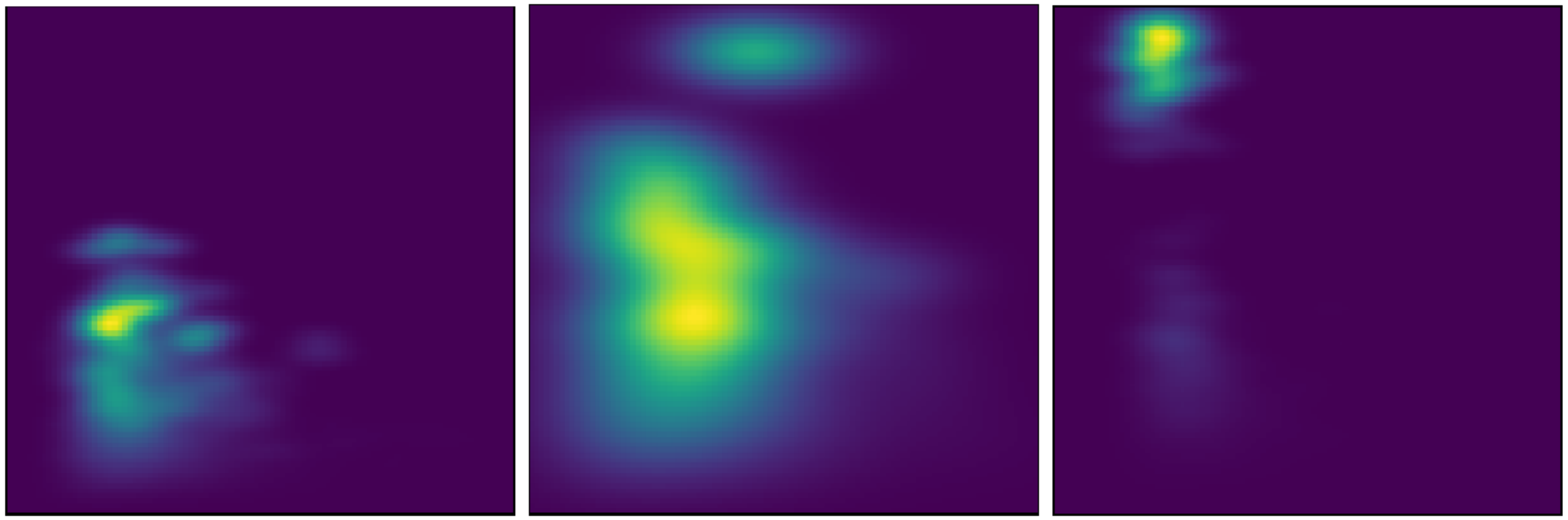
The realization of 3 different processes



The overlay of 40 different persistence diagrams



The persistence images with weight function $w(\mathbf{r}) = (r_2 - r_1)^3$ and bandwith selected using cross-validation.

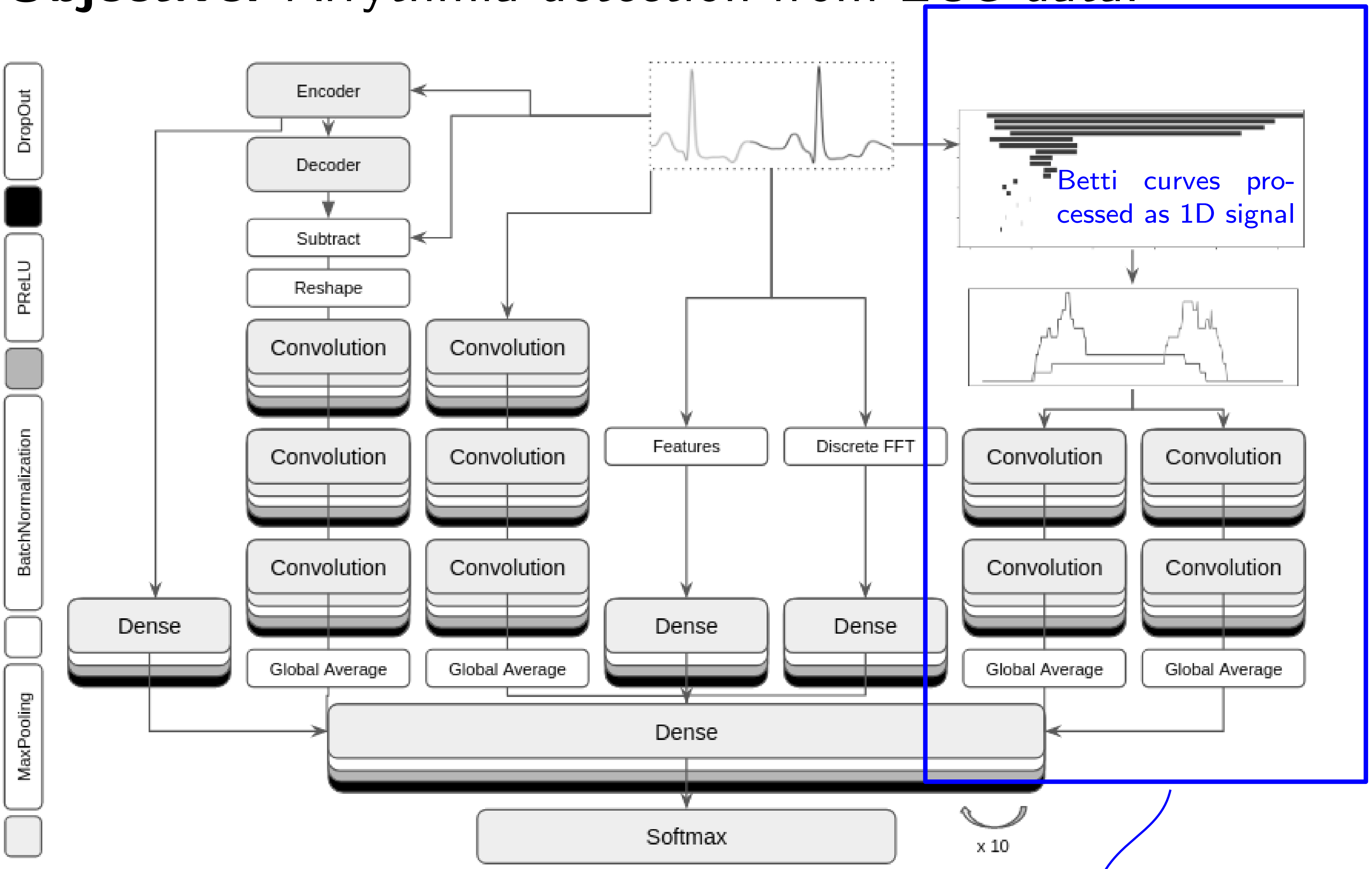


A few illustrative applications

Example of application: arrhythmia detection

[Dindin, Umeda, C. Can. Conf. AI 2020]

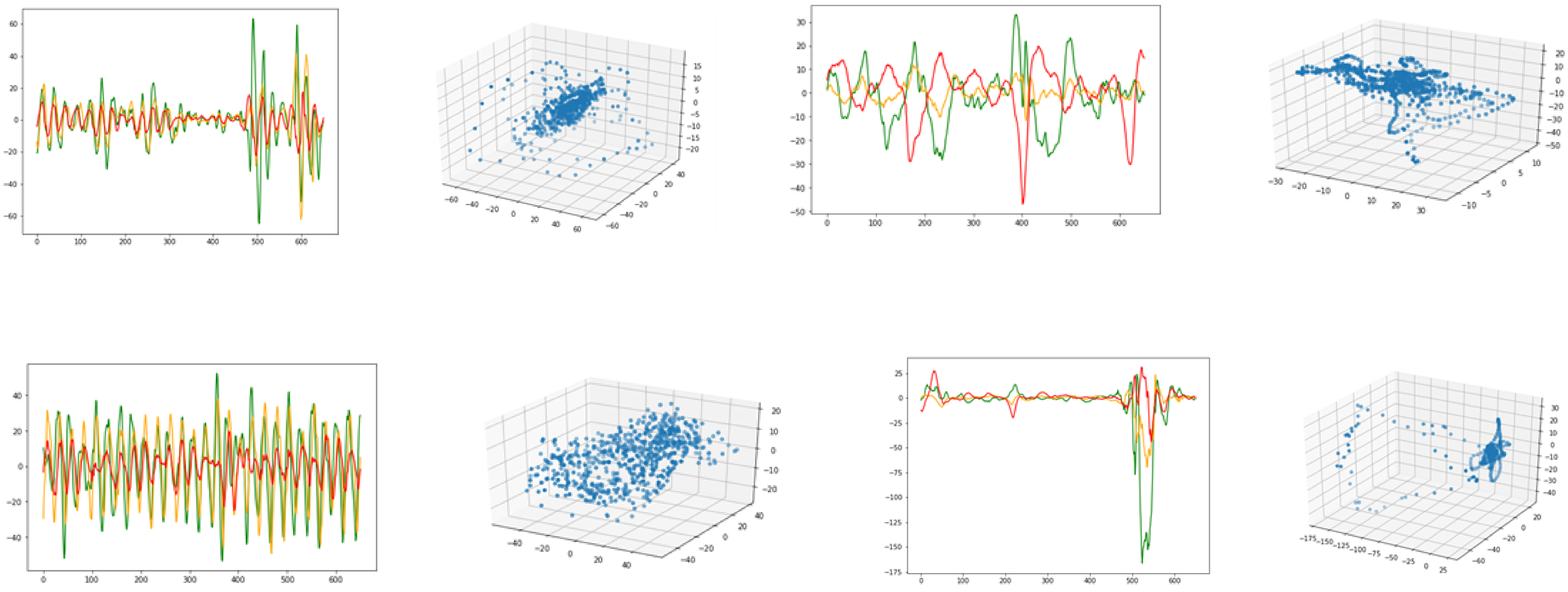
Objective: Arrythmia detection from ECG data.



- Improvement over state-of-the-art.
- Better generalization.

	Accuracy[%]
UCLA (2018)	93.4
Li et al. (2016)	94.6
Inria-Fujitsu (2018)*	**98.6**

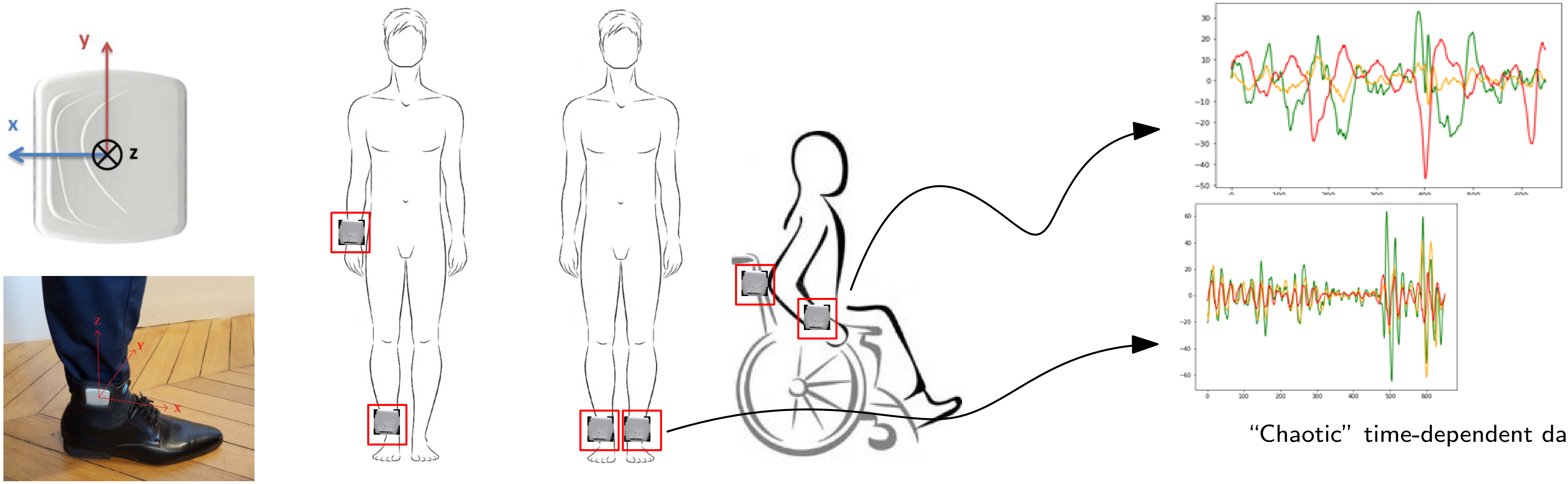
TDA and Machine Learning for sensor data



(Multivariate) time-dependent data can be converted into point clouds: sliding window, time-delay embedding,...

With landscapes: patient monitoring

[Beaufils, C., Dindin, Grelet, Michel 2018]



"Chaotic" time-dependent data

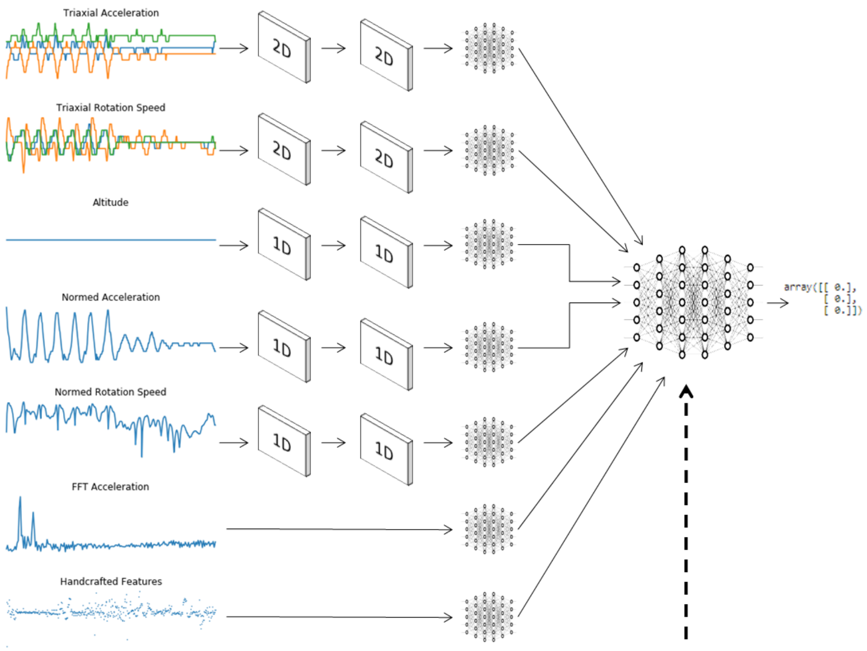
Objective: precise analysis of movements and activities of pedestrians.

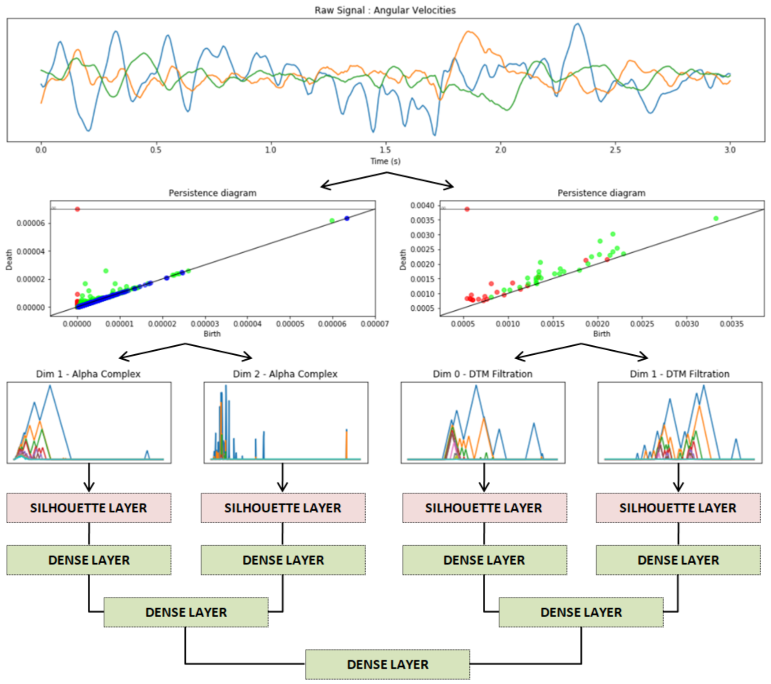
Applications: personal healthcare; medical studies; defense.

With landscapes: patient monitoring

[Beaufils, C., Dindin, Grelet, Michel 2018]

Example: Dyskinesia crisis detection and activity recognition:





Class	Naive	Multi	FEA	QUA	TDA
Walking	97.6	98.4	99.3	99.0	**99.5**
Upstairs	97.2	**99.8**	97.8	98.0	97.7
Downstairs	99.6	**99.7**	99.0	98.4	98.3
Sitting	87.1	93.1	89.7	91.8	**96.5**
Standing	87.0	97.7	97.2	97.2	**98.1**
Laying	92.4	**100.**	99.8	99.9	**100.**
Stand-Sit	90.8	**95.6**	89.1	91.3	93.4
Sit-Stand	**100.**	99.9	**100.**	**100.**	**100.**
Sit-Lie	87.1	81.1	84.2	90.0	**95.1**
Lie-Sit	81.4	81.8	85.9	**91.8**	87.9
Stand-Lie	74.2	**87.6**	86.5	87.4	81.5
Lie-Stand	80.4	72.1	**83.2**	77.7	**83.2**

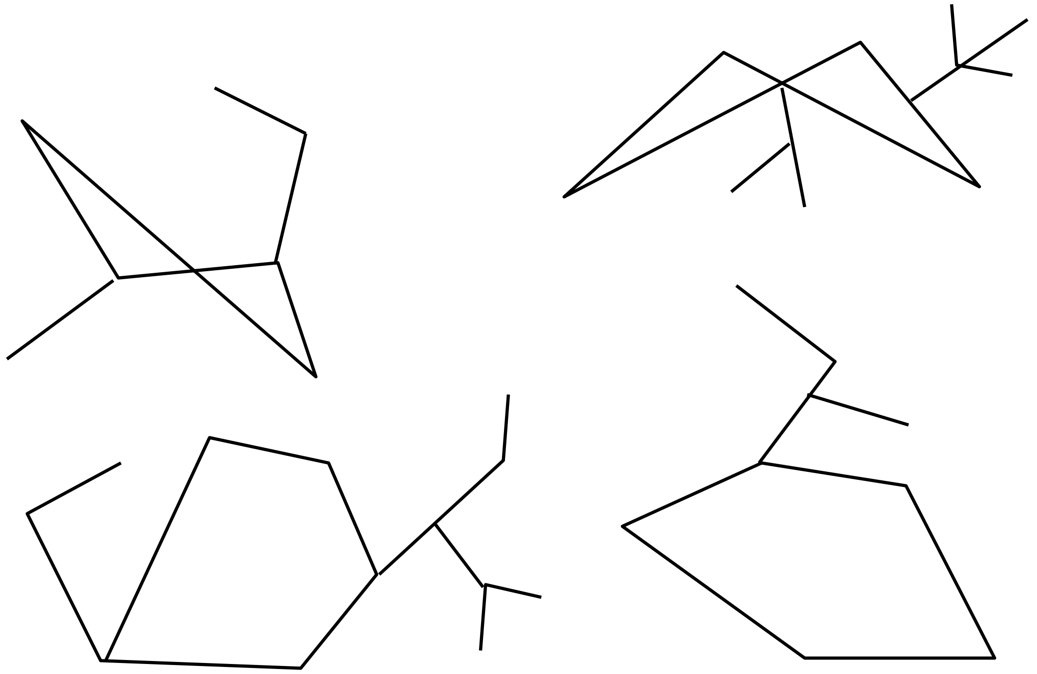
Multi-channels CNN + TDA neural network

Results on publicly available data set (HAPT) - improve the state-of-the-art.

- Data collected in non controlled environments (home) are very chaotic.
- Data registration (uncertainty in sensors orientation/position).
- Reliable and robust information is mandatory.
- Events of interest are often rare and difficult to characterize.

Extra slides

Graph classification using persistent homology



Input: A collection of graphs $G_1, \cdots, G_n$ belonging to different classes $y_1, \cdots, y_k$.

Goal: Recover the classes, i.e. build, from the input data, a function

$$f : \mathcal{G} \to Y = \{y_1, \cdots, y_k\}$$

that assigns each graph in $\mathcal{G}$ to its expected class.

Unformal assumption (hope): the class of a graph is determined by its geometric structure.

Simple idea: Build functions encoding the structure of the graphs at different scales and use their persistence diagrams as features.

Heat Kernel Signature on Graphs

Let $G = (V, E)$ be a non oriented graph with vertex set $V = \{v_1, \cdots, v_n\}$ and adjacency matrix $W = (w_{i,j})$.

The degree matrix D is the diagonal matrix defined by $D_{i,i} = \sum_j w_{i,j}$.

The normalized graph Laplacian is defined by $L_w = I - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$.

Let $\Psi = \{\psi_1, \ldots, \psi_n\}$ be an orthonormal basis of eigenfunctions of L_w with corresponding eigenvalues $0 \leq \lambda_1 \leq \cdots \leq \lambda_n \leq 2$.

Definition:
Given $t \geq 0$, the heat kernel signature at time t is defined by

$$\mathrm{hks}_{G,t} \colon v \mapsto \sum_{k=1}^{n} \exp(-t\lambda_k)\psi_k(v)^2.$$

[Sun et al 2009].

Heat Kernel Signature on Graphs

Definition:
Given $t \geq 0$, the heat kernel signature at time t is defined by

$$\mathrm{hks}_{G,t} \colon v \mapsto \sum_{k=1}^{n} \exp(-t\lambda_k)\psi_k(v)^2.$$

[Sun et al 2009].

Theorem: [Stability] [Hu et al 2014, Carriere et al 2019].

Let $t \geq 0$ and let L_w be the Laplacian matrix of a graph G with n vertices. Let G' be another graph with n vertices and Laplacian matrix $\tilde{L}_w = L_w + E$. Then there exists a constant $C(G,t) > 0$ only depending on t and the spectrum of L_w such that, for small enough $\|E\|$:

$$d_B(Dg(G, \mathrm{hks}_{G,t}), Dg(G, \mathrm{hks}_{G',t})) \leq C(G,t)\|W\|.$$

Rmk: Here Dg stands for sub-level sets, upper-level sets or extended persistence.