Sophia-Antipolis - January 2016

Homology inference

F. Chazal
Geometrica Group
INRIA Saclay

If you have any question:
frederic.chazal@inria.fr

Weak feature size and stability

The *weak feature size* of a compact $K \subset \mathbb{R}^d$:

$$\mathsf{wfs}(K) = \inf\{c > 0 : c \text{ is a critical value of } d_K\}$$

Proposition: [C-Lieutier'05] Let $K, K' \subset \mathbb{R}^d$ be such that

$$d_H(K, K') < \varepsilon := \frac{1}{2}\min(\mathsf{wfs}(K), \mathsf{wfs}(K'))$$

Then for all $0 < r \leq 2\varepsilon$, K^r and K'^r are homotopy equivalent.

Weak feature size and stability

The *weak feature size* of a compact $K \subset \mathbb{R}^d$:

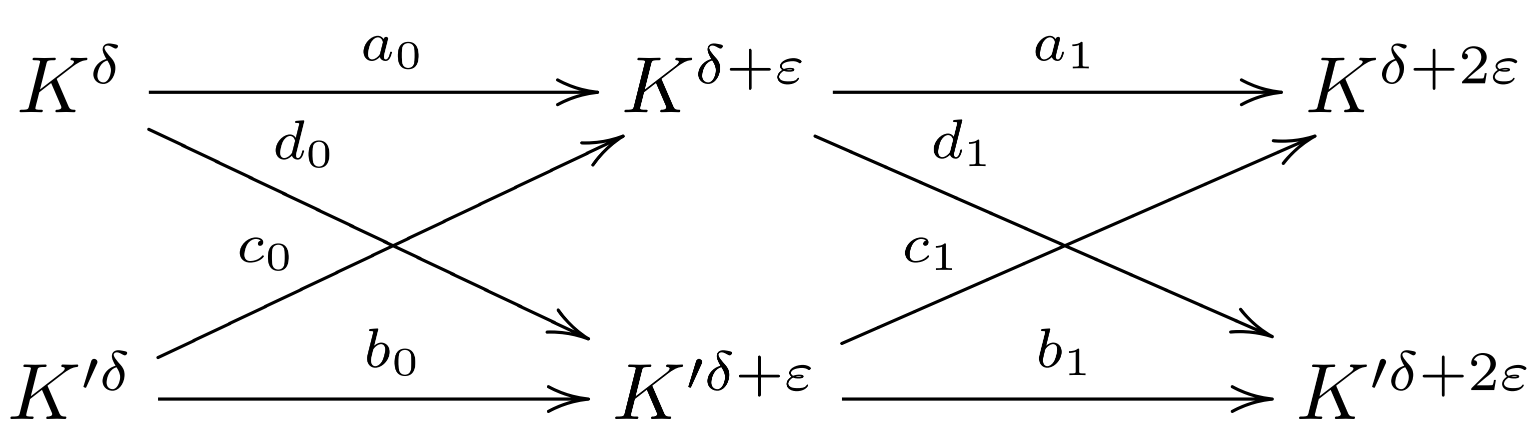
$$\mathsf{wfs}(K) = \inf\{c > 0 : c \text{ is a critical value of } d_K\}$$

Proposition: [C-Lieutier'05] Let $K, K' \subset \mathbb{R}^d$ be such that

$$d_H(K, K') < \varepsilon := \frac{1}{2}\min(\mathsf{wfs}(K), \mathsf{wfs}(K'))$$

Then for all $0 < r \leq 2\varepsilon$, K^r and K'^r are homotopy equivalent.

Proof: let $\delta > 0$ be s.t. $\delta + 2\varepsilon < \min(\mathsf{wfs}(K), \mathsf{wfs}(K'))$.



Weak feature size and stability

The *weak feature size* of a compact $K \subset \mathbb{R}^d$:

$$\text{wfs}(K) = \inf\{c > 0 : c \ \text{ is a critical value of } \ d_K\}$$

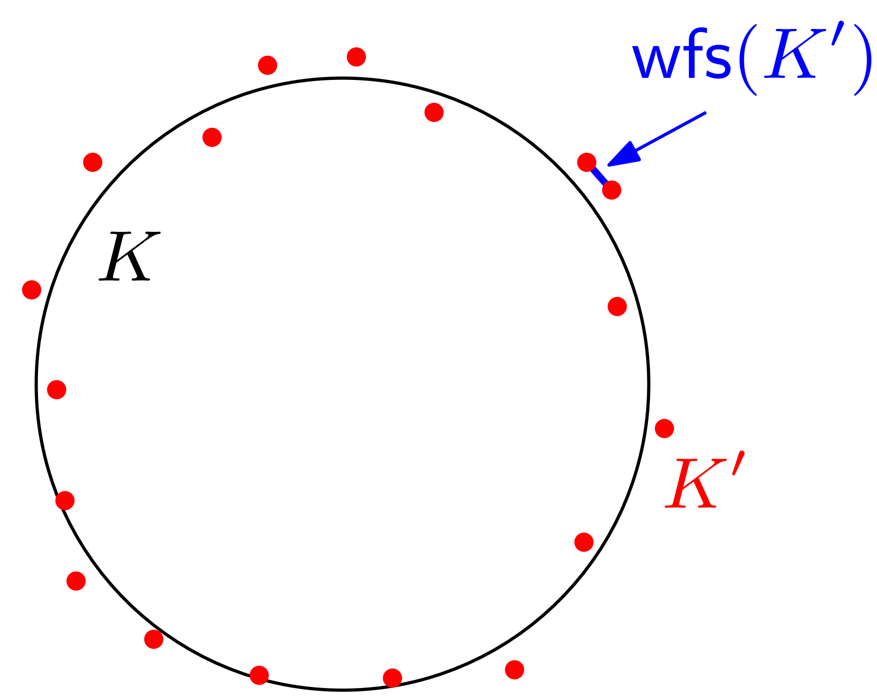
Proposition: [C-Lieutier'05] Let $K, K' \subset \mathbb{R}^d$ be such that

$$d_H(K, K') < \varepsilon := \frac{1}{2} \min(\text{wfs}(K), \text{wfs}(K'))$$

Then for all $0 < r \leq 2\varepsilon$, K^r and K'^r are homotopy equivalent.

Compact set with positive wfs:

- Stability properties
- Large class of compact sets (including sub-analytic sets)
- $K \to \text{wfs}(K)$ is not continuous (unstability of critical points).



Overcoming the discontinuity of wfs

Proposition: [C-Lieutier'05] Let $K, K' \subset \mathbb{R}^d$ be such that

$$d_H(K, K') < \varepsilon := \frac{1}{2}\min(\mathsf{wfs}(K), \mathsf{wfs}(K'))$$

Then for all $0 < r \leq 2\varepsilon$, K^r and K'^r are homotopy equivalent.

$K \to \mathsf{wfs}(K)$ is not continuous (unstability of critical points).

Option 1:
Restrict to a smaller class of compact sets with some stability properties of the critical points.

Option 2:
Try to get topological information about K without any assumption on $\mathsf{wfs}(K')$.

Overcoming the discontinuity of wfs

Proposition: [C-Lieutier'05] Let $K, K' \subset \mathbb{R}^d$ be such that

$$d_H(K, K') < \varepsilon := \frac{1}{2} \min(\mathsf{wfs}(K), \mathsf{wfs}(K'))$$

Then for all $0 < r \leq 2\varepsilon$, K^r and K'^r are homotopy equivalent.

$K \to \mathsf{wfs}(K)$ is not continuous (unstability of critical points).

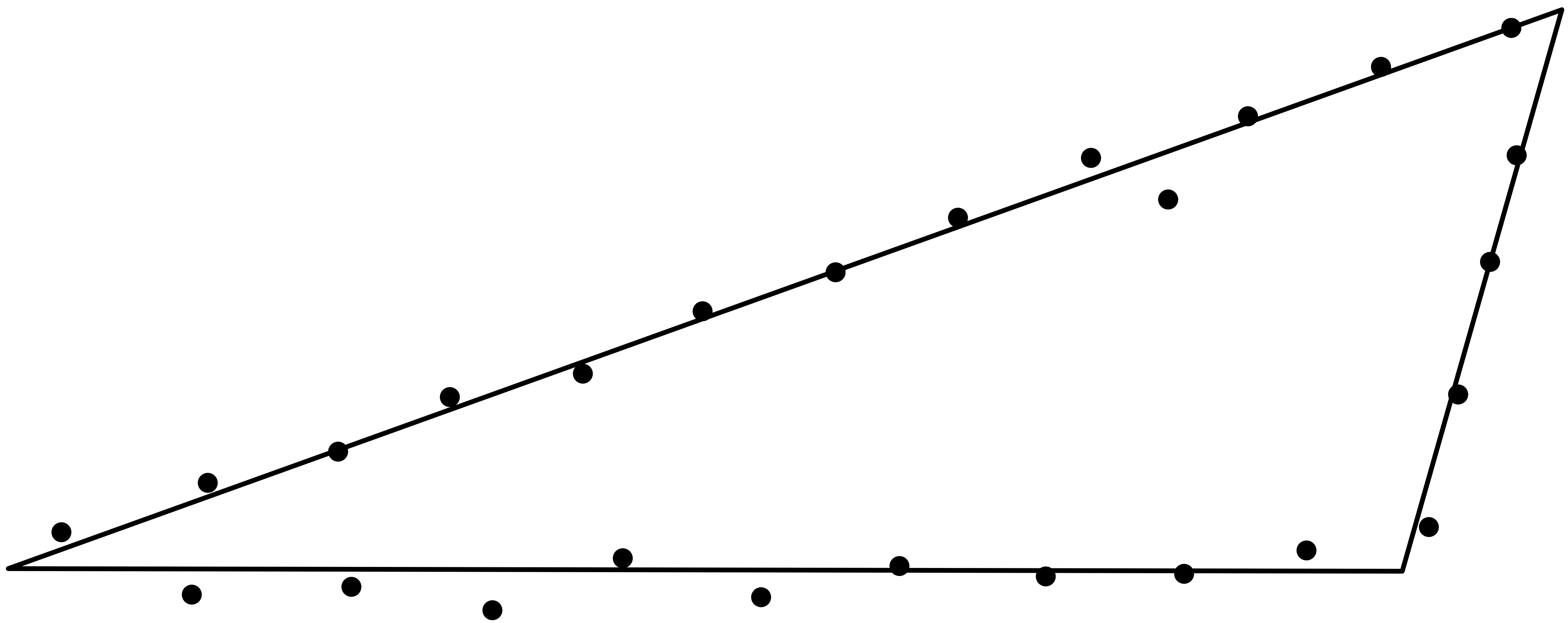
Option 1:
Restrict to a smaller class of compact sets with some stability properties of the critical points.

Notion of μ-critical points. Strong reconstruction results. (not in this course)

Option 2:
Try to get topological information about K without any assumption on $\mathsf{wfs}(K')$.

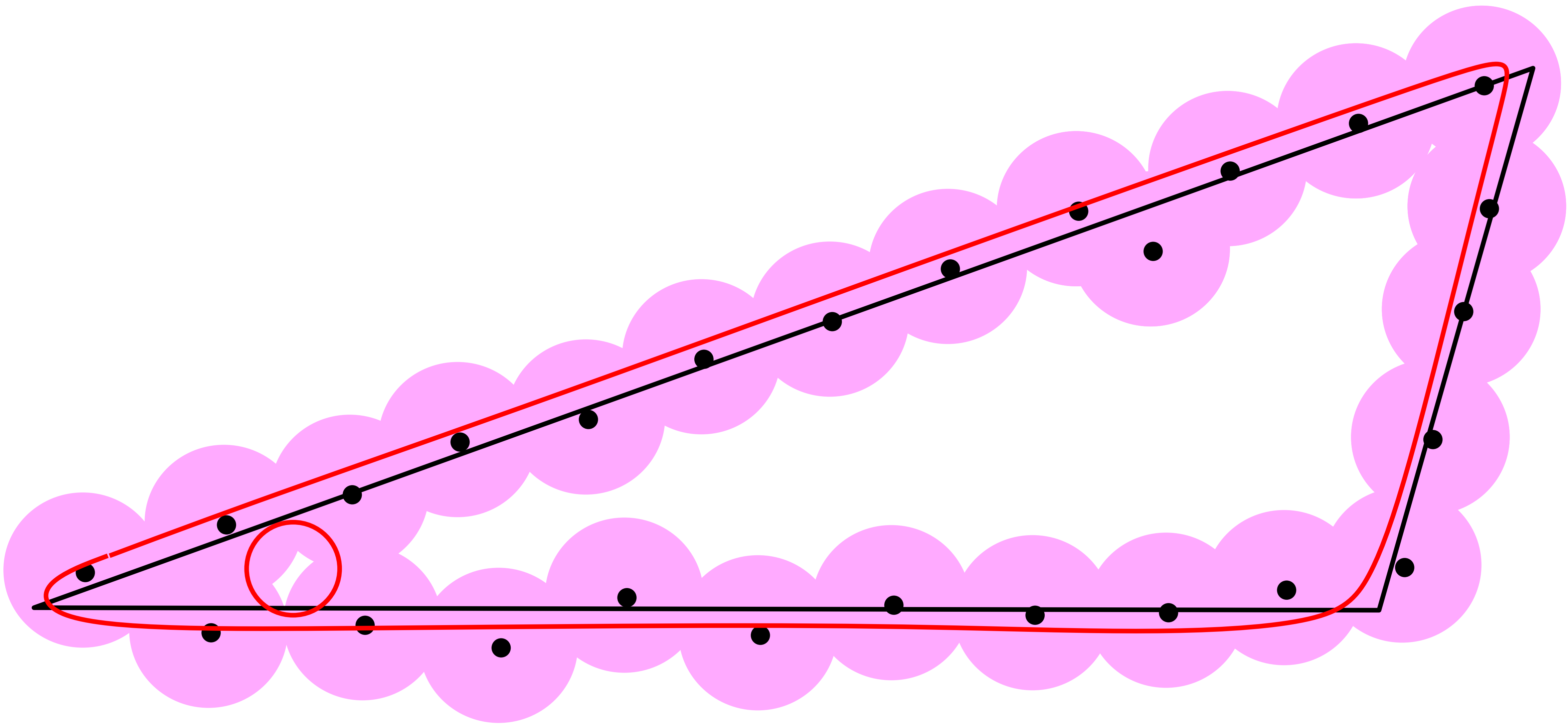
Persistence-based inference

Motivation: getting topological information without reconstructing



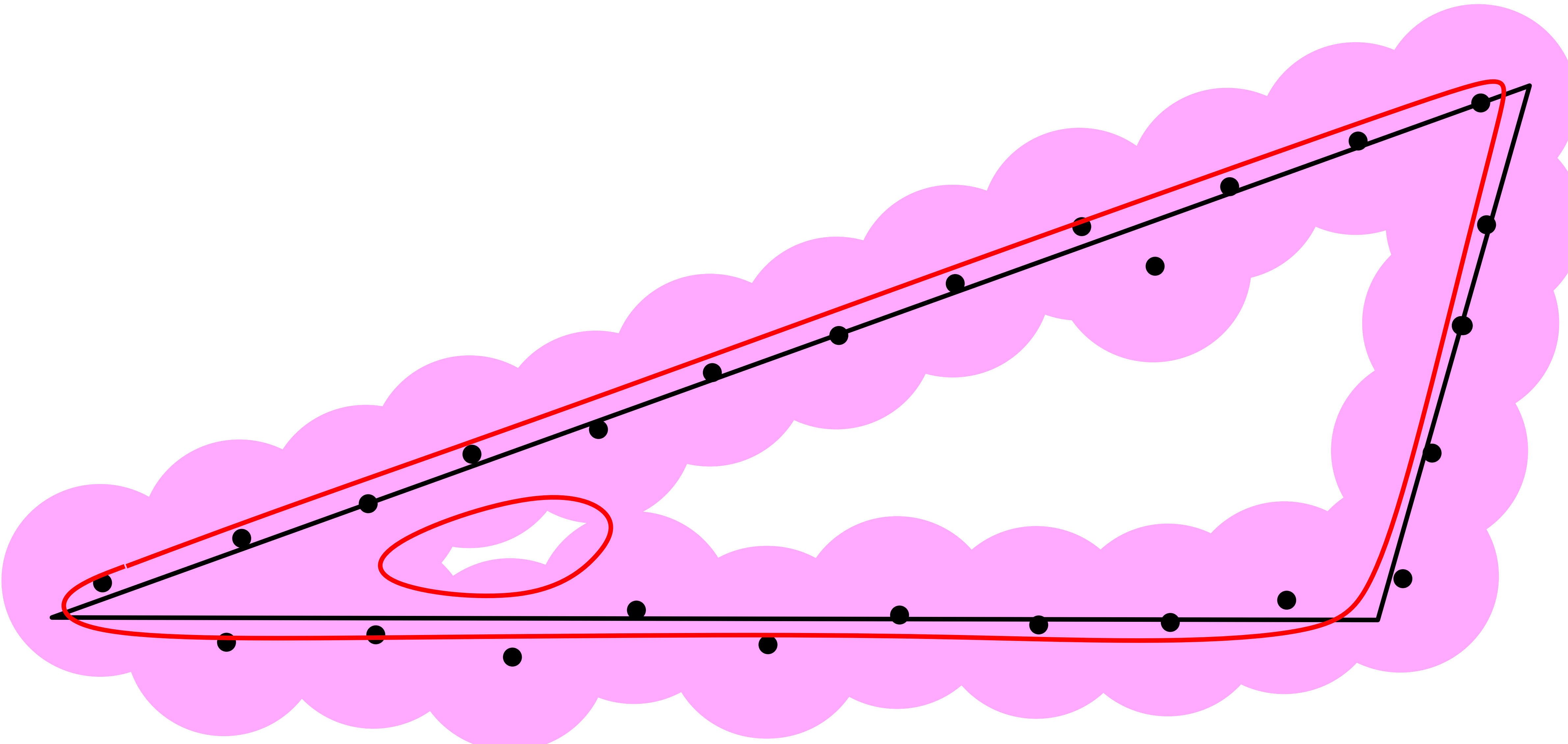
How to determine the number of "cycles" of the underlying shape from the point cloud approximation?

Motivation: getting topological information without reconstructing



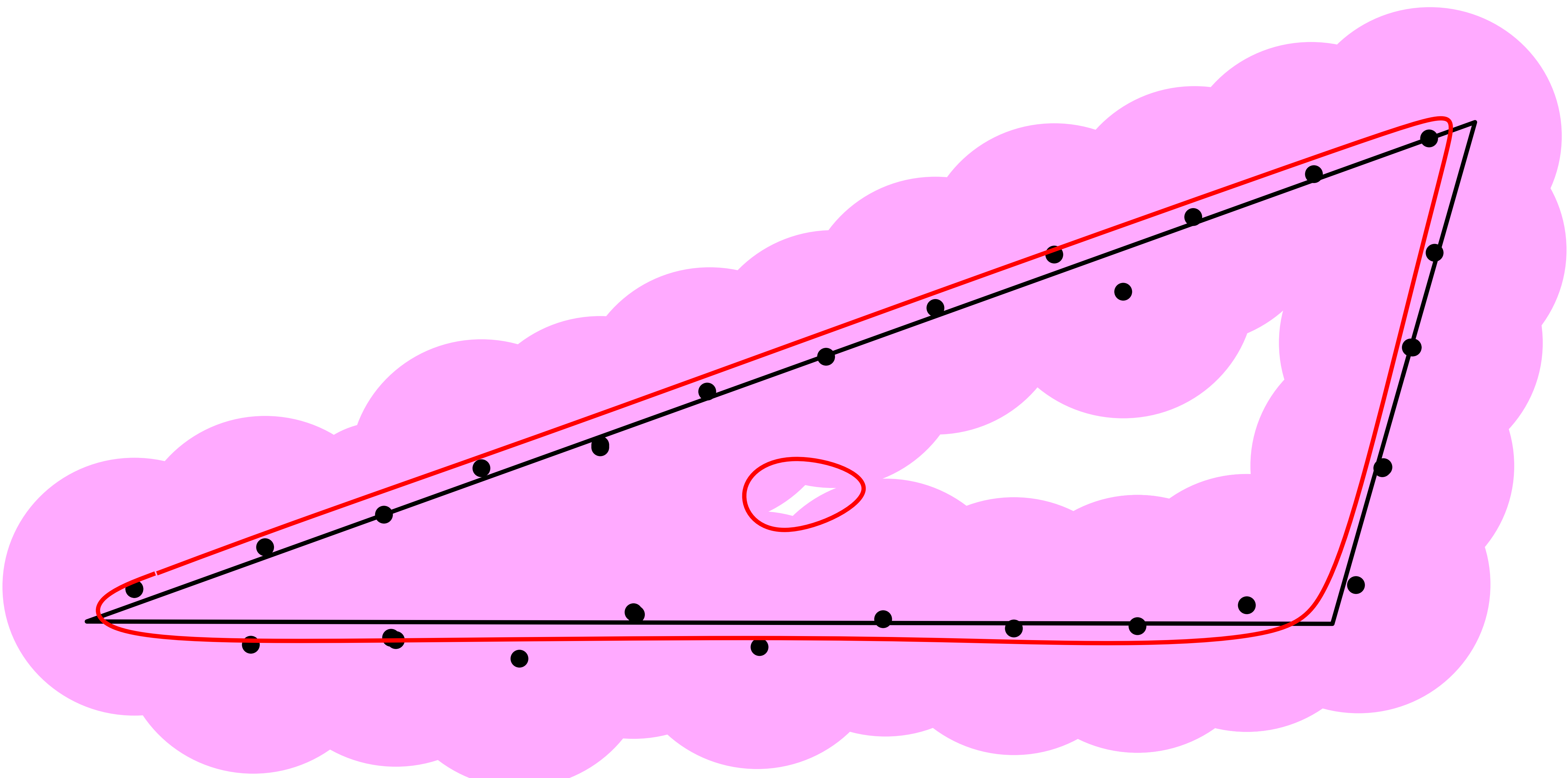
How to determine the number of “cycles” of the underlying shape from the point cloud approximation?

Motivation: getting topological information without reconstructing



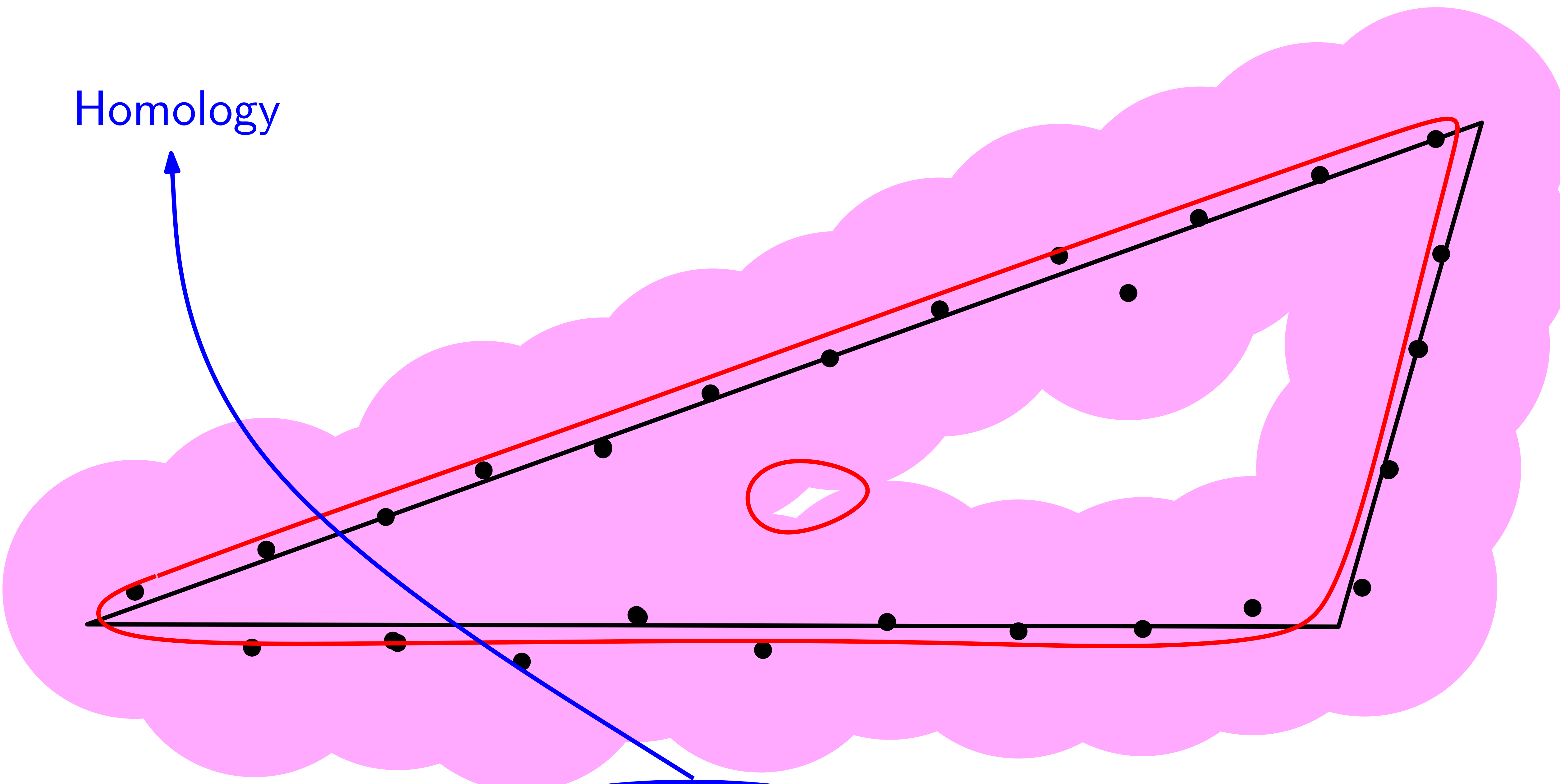
How to determine the number of “cycles” of the underlying shape from the point cloud approximation?

Motivation: getting topological information without reconstructing



How to determine the number of “cycles” of the underlying shape from the point cloud approximation?

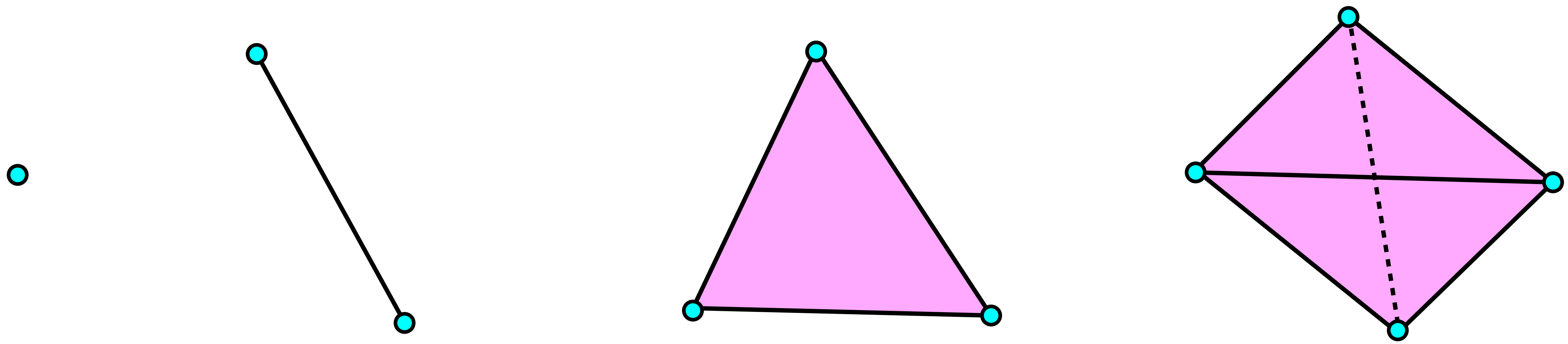
Motivation: getting topological information without reconstructing



How to determine the number of "cycles" of the underlying shape from the point cloud approximation?

Persistent homology

Simplices



0-simplex: vertex

1-simplex: edge

2-simplex: triangle

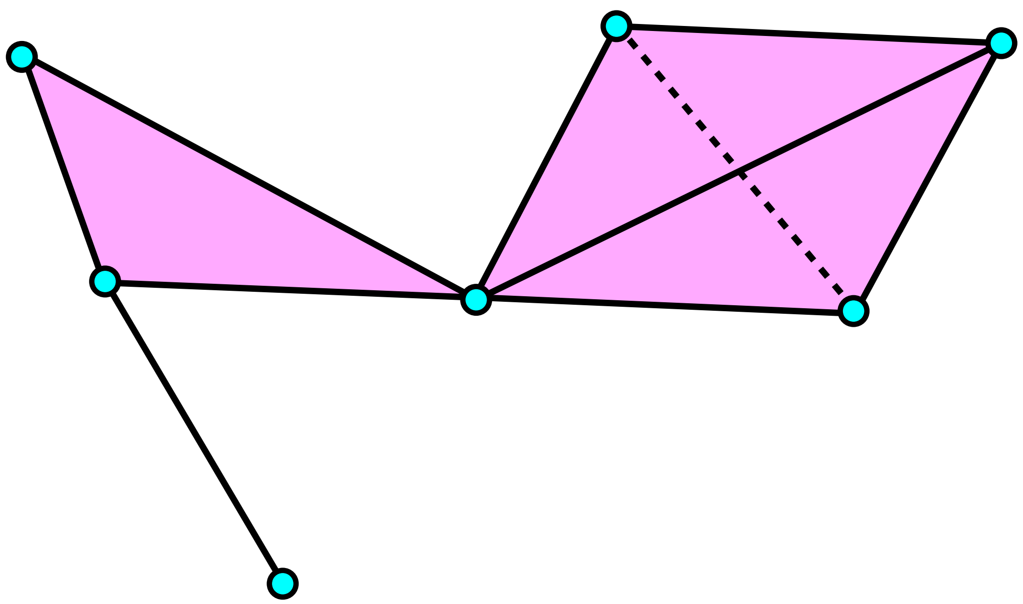
3-simplex: tetraedron

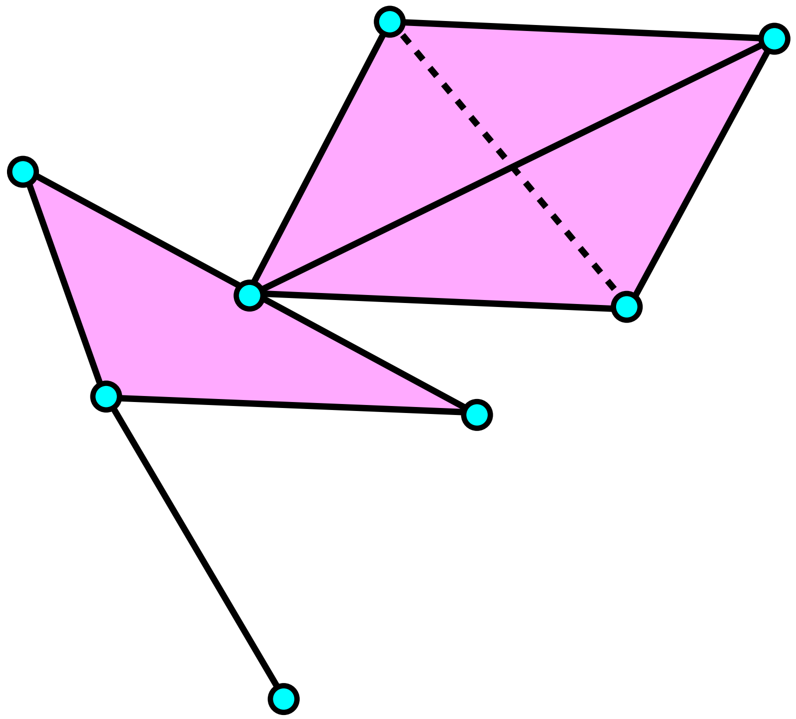
$v_0, v_1, \cdots, v_k \in \mathbb{R}^d$ are affinely independant if

$$\left(\sum_{i=0}^{k} t_i v_i = 0 \text{ and } \sum_{i=0}^{k} t_i = 0 \right) \Rightarrow t_0 = t_1 = \cdots = t_k = 0$$

In this case $\sigma = [v_0, v_1, \cdots, v_k]$ is a simplex of dimension d. A simplex generated by a subset of the vertices $v_0, v_1, \cdots, v_k$ of σ is a face of σ.

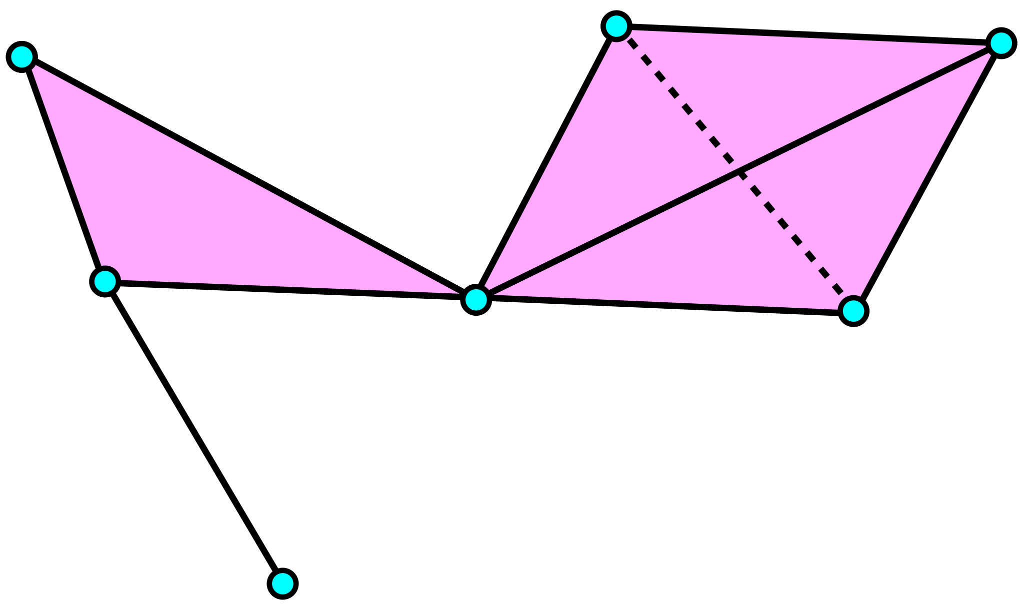
Simplicial complexes

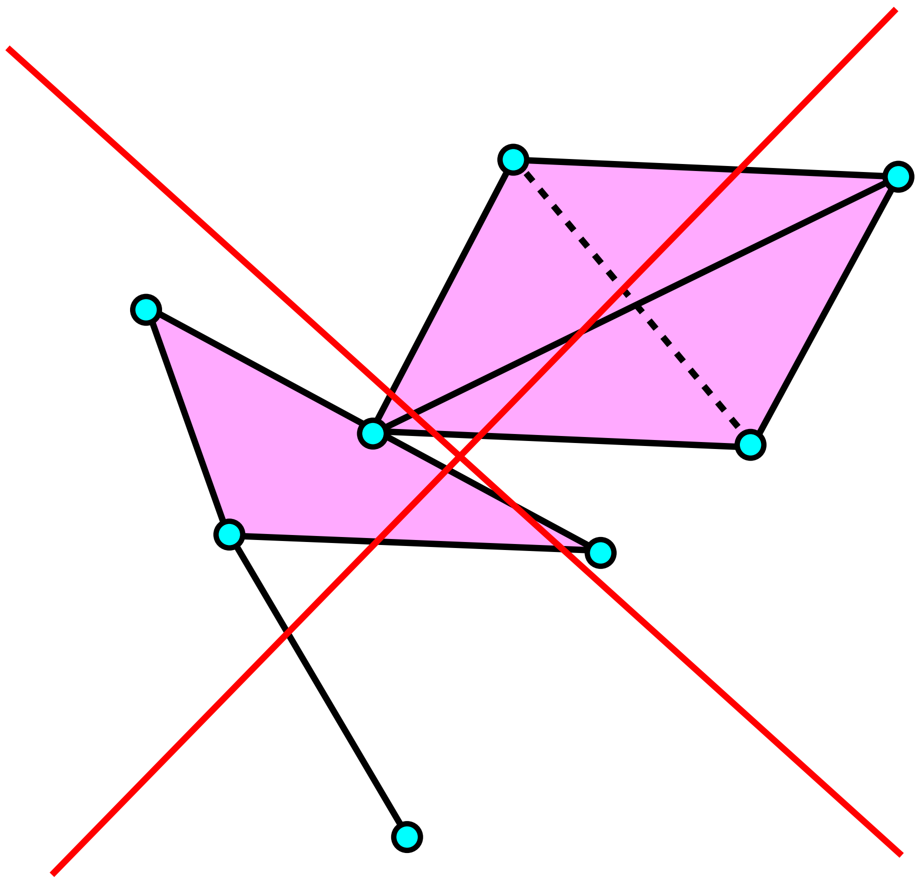




A (finite) simplicial complex C is a (finite) union of simplices s.t.
i) for any $\sigma \in C$, all the faces of σ are in C,
ii) the intersection of any two simplices of C is either empty or a simplex which is their common face of highest dimension.

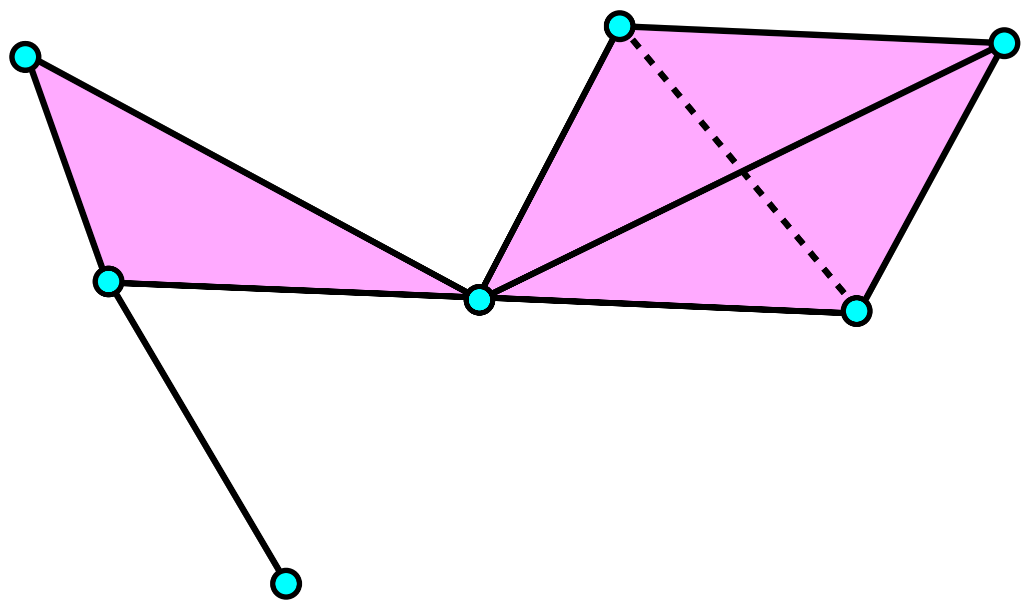
Simplicial complexes

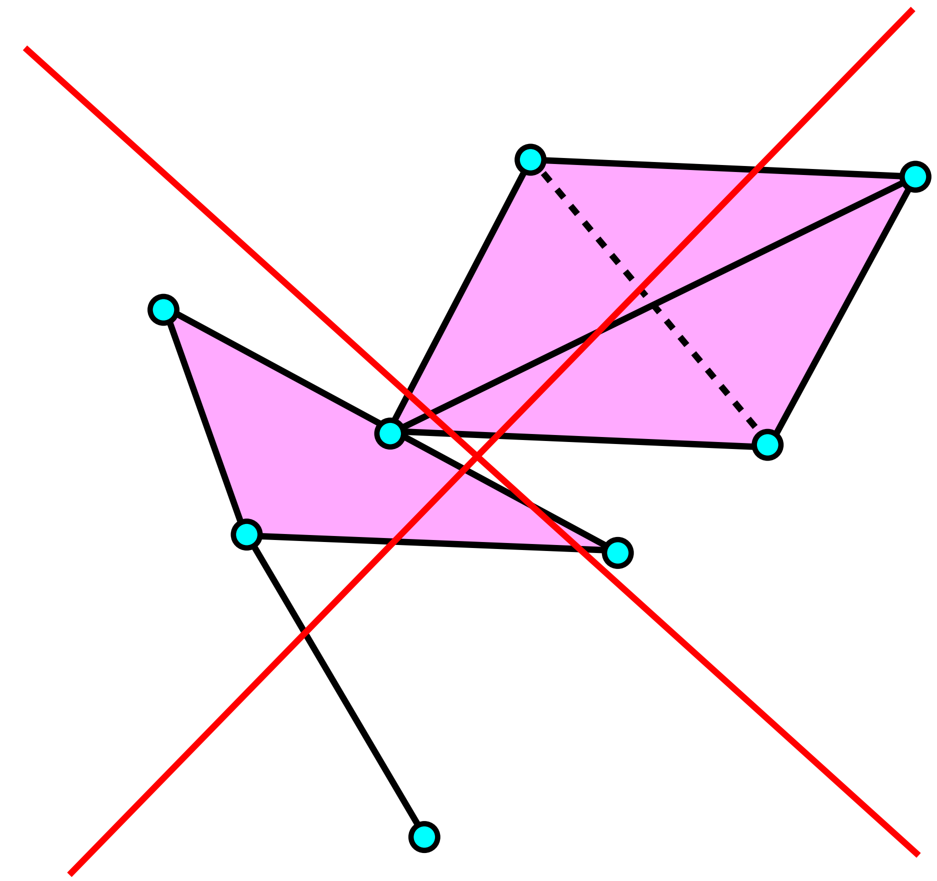




A (finite) simplicial complex C is a (finite) union of simplices s.t.
i) for any $\sigma \in C$, all the faces of σ are in C,
ii) the intersection of any two simplices of C is either empty or a simplex which is their common face of highest dimension.

Simplicial complexes





A (finite) simplicial complex C is a (finite) union of simplices s.t.
i) for any $\sigma \in C$, all the faces of σ are in C,
ii) the intersection of any two simplices of C is either empty or a simplex which is their common face of highest dimension.

Faces: the simplices of C.

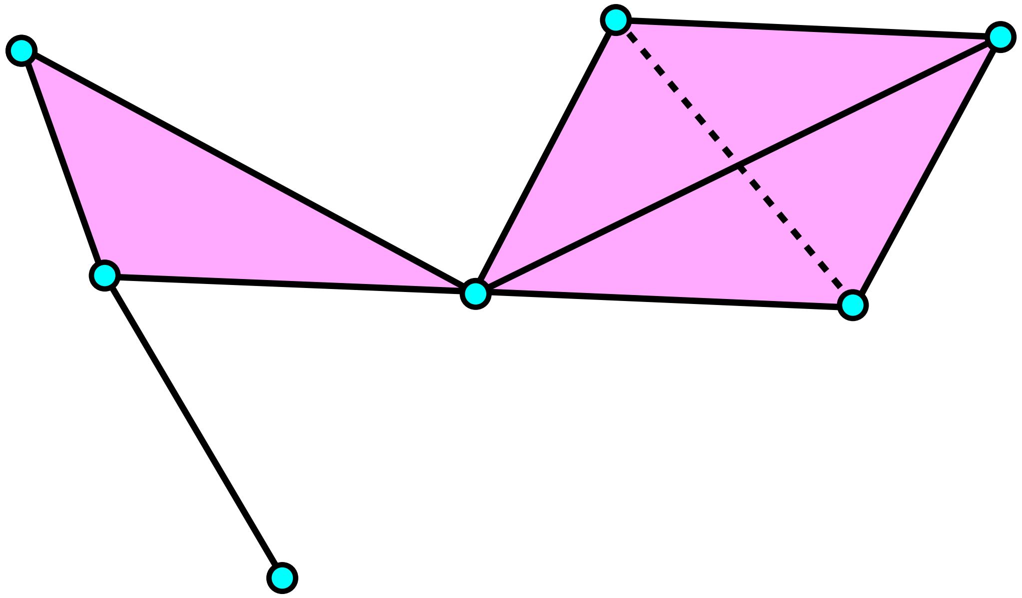
j-skeleton: the subcomplex made of the simplices of dimension at most j.

Dimension of C: the maximum of the dimensions of the faces. C is homogenousof dimension n if any of its faces is a face of a n-dimensional simplex.

Abstract simplicial complexes

Let $P = \{p_1, \cdots p_n\}$ be a (finite) set. An abstract simplicial complex K with vertex set P is a set of 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.

Let $\{e_1, \cdots e_n\}$ a basis of $\mathbb{R}^n$ }. The geometric realization of K is the (geometric) subcomplex $|K|$ of the simplex spanned by $e_1, \cdots e_n$ such that:

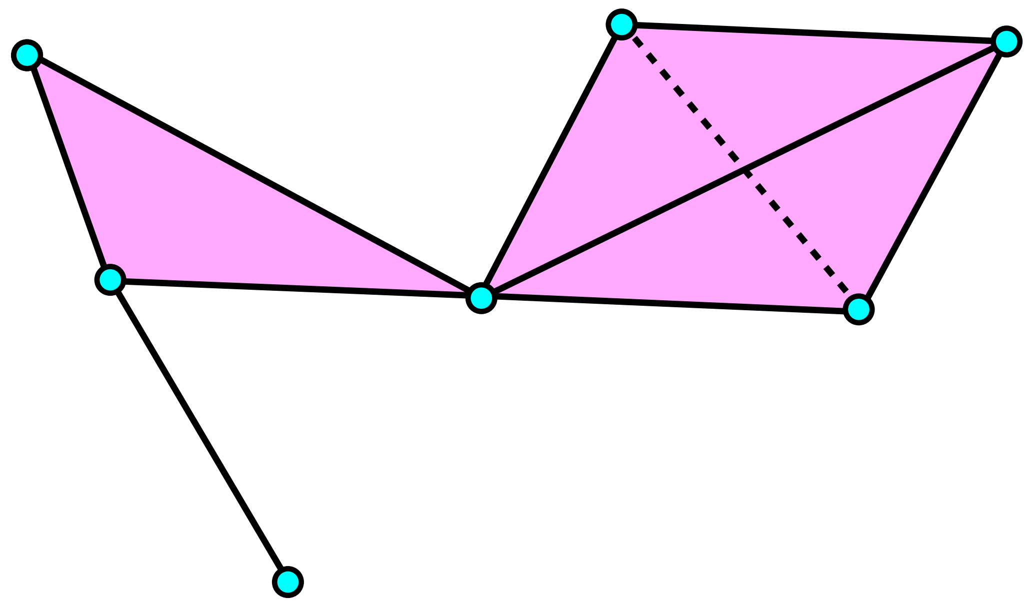
$$[e_{i_0} \cdots e_{i_k}] \in |K| \text{ iff } \{p_{i_0}, \cdots, p_{i_k}\} \in K$$

$|K|$ is a topological space (subspace of an Euclidean space)!

Abstract simplicial complexes

Let $P = \{p_1, \cdots p_n\}$ be a (finite) set. An abstract simplicial complex K with vertex set P is a set of 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 $\sigma \in K$.

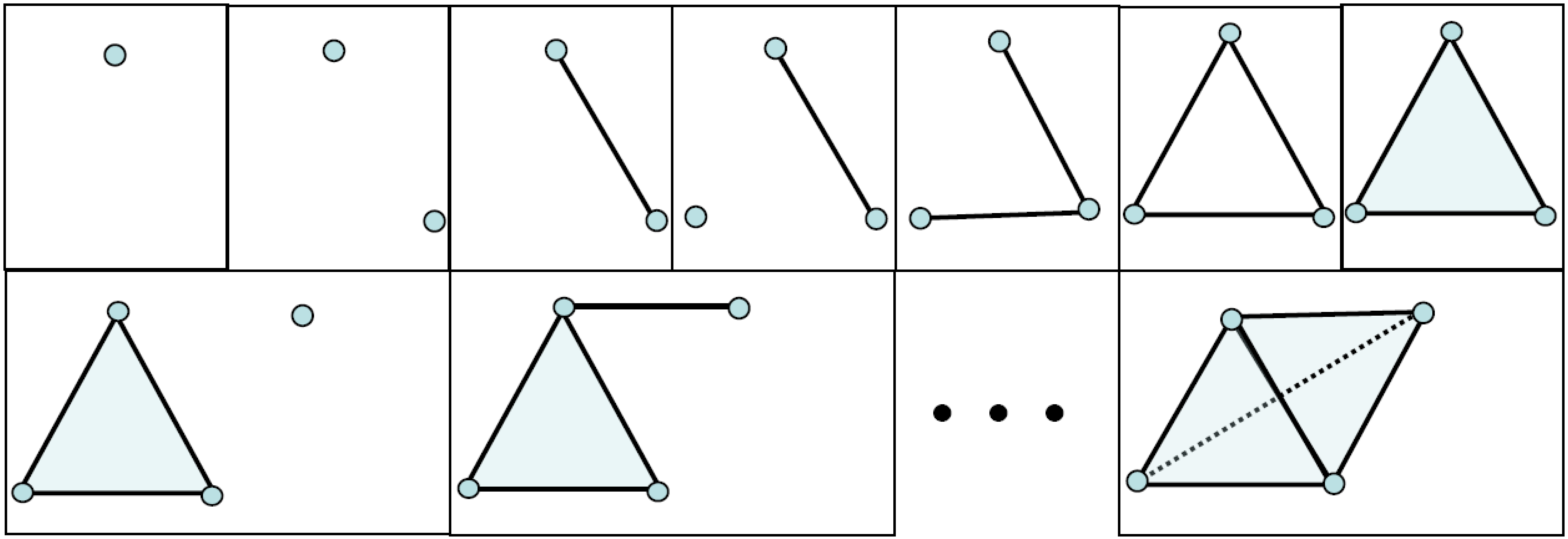


The elements of K are the simplices.

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).

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.

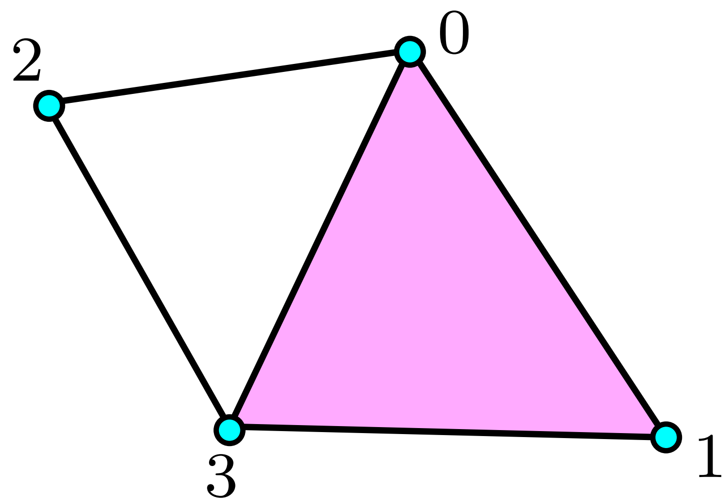
Example: filtration associated to a function

- 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).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

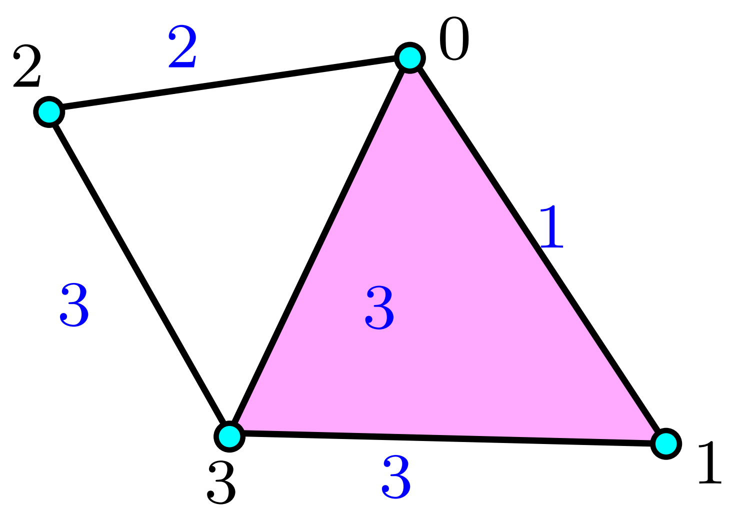


- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

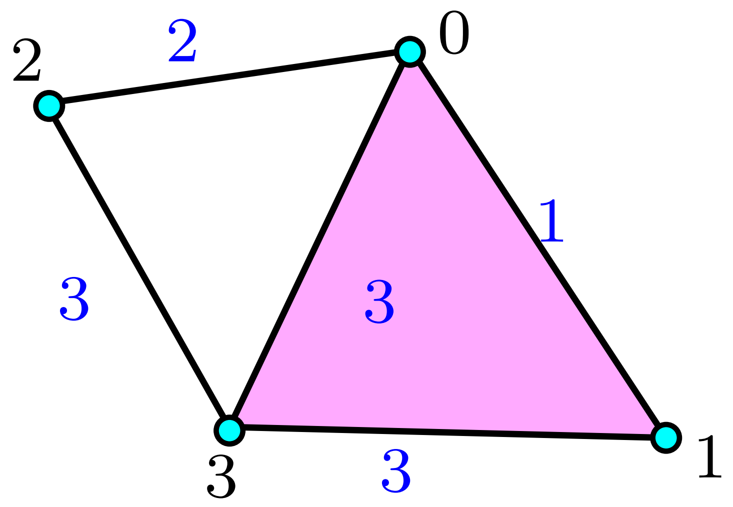


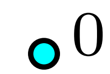
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



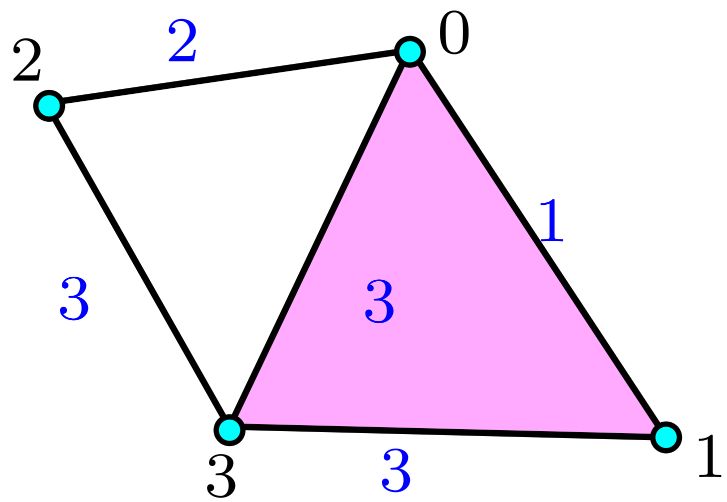


- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

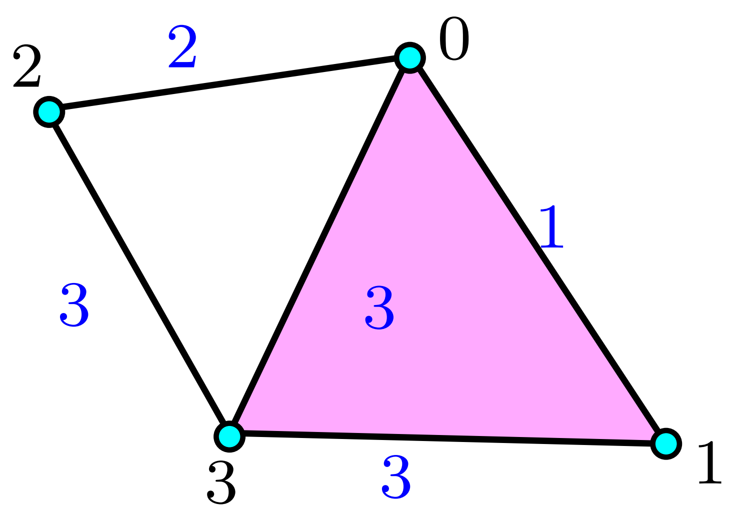


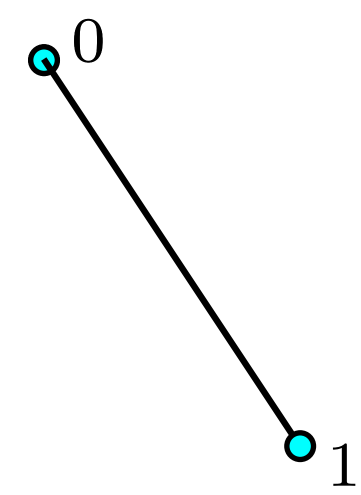
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



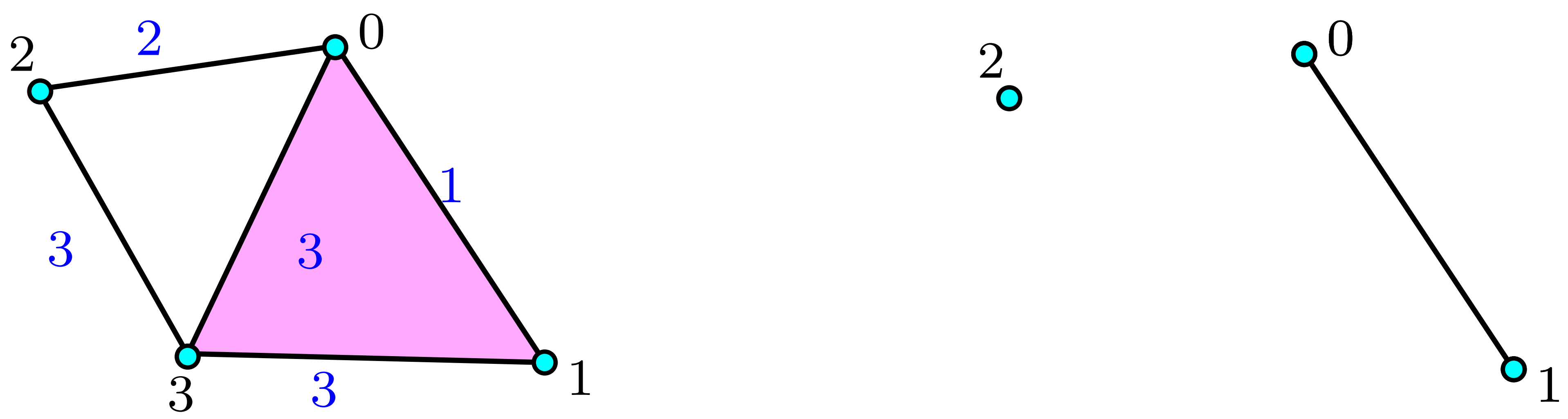


- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function

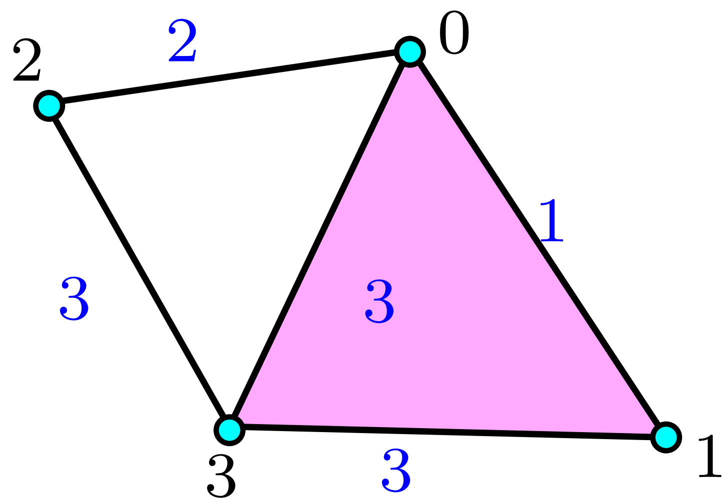


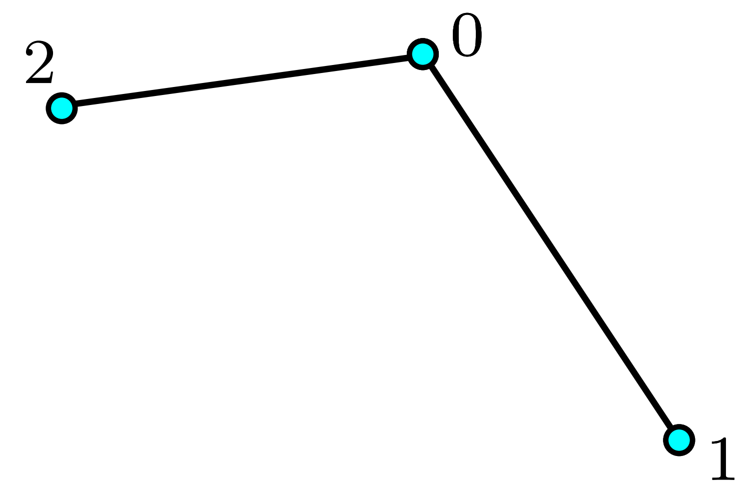
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



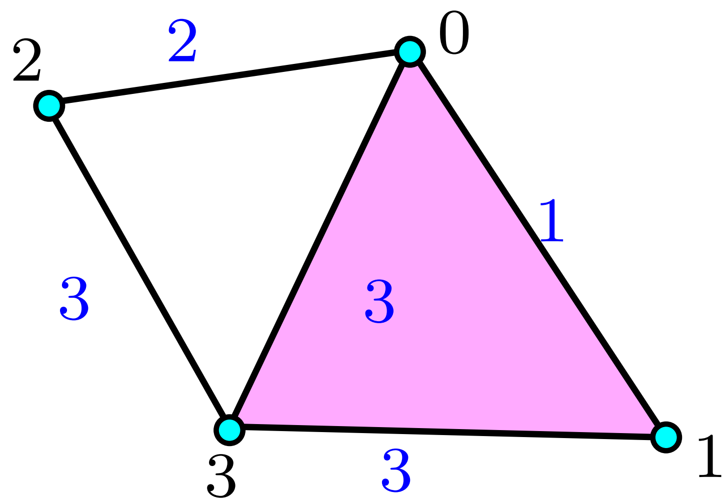


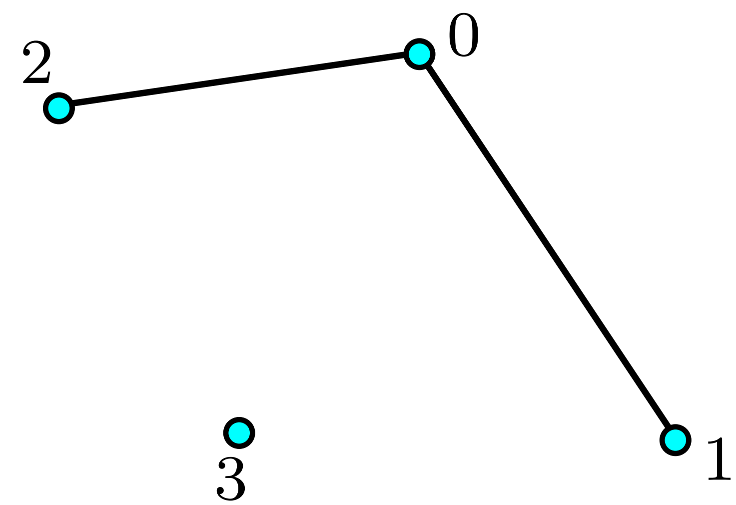
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



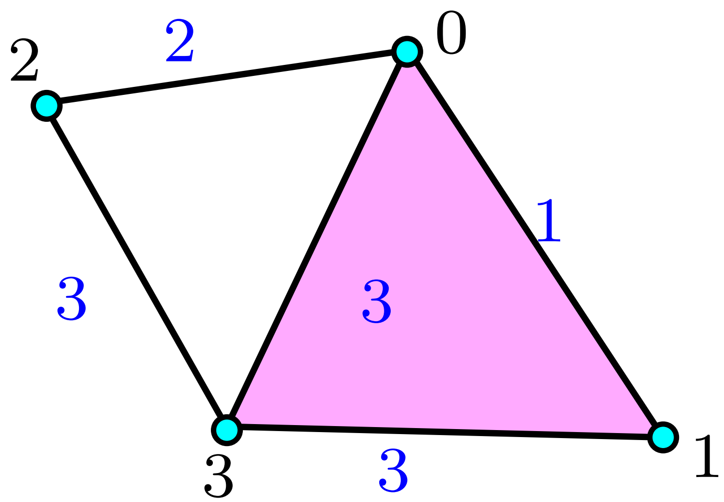


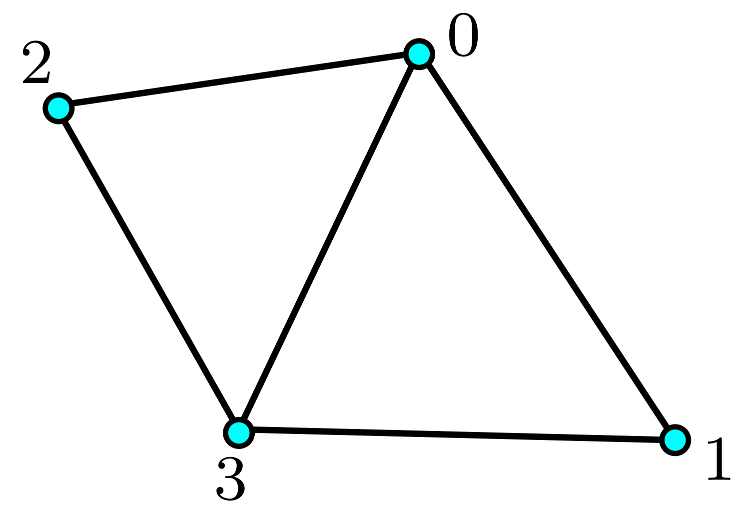
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



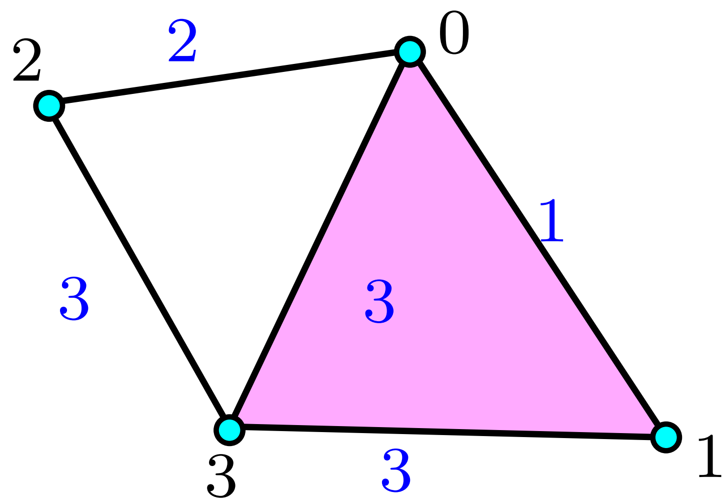


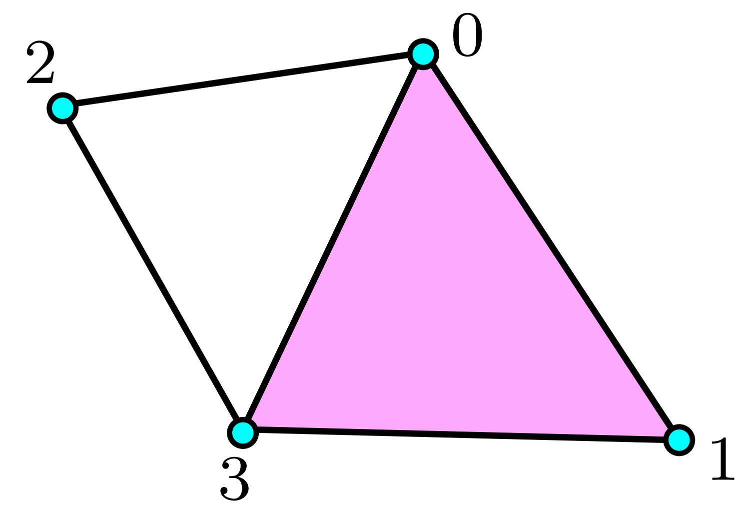
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: filtration associated to a function



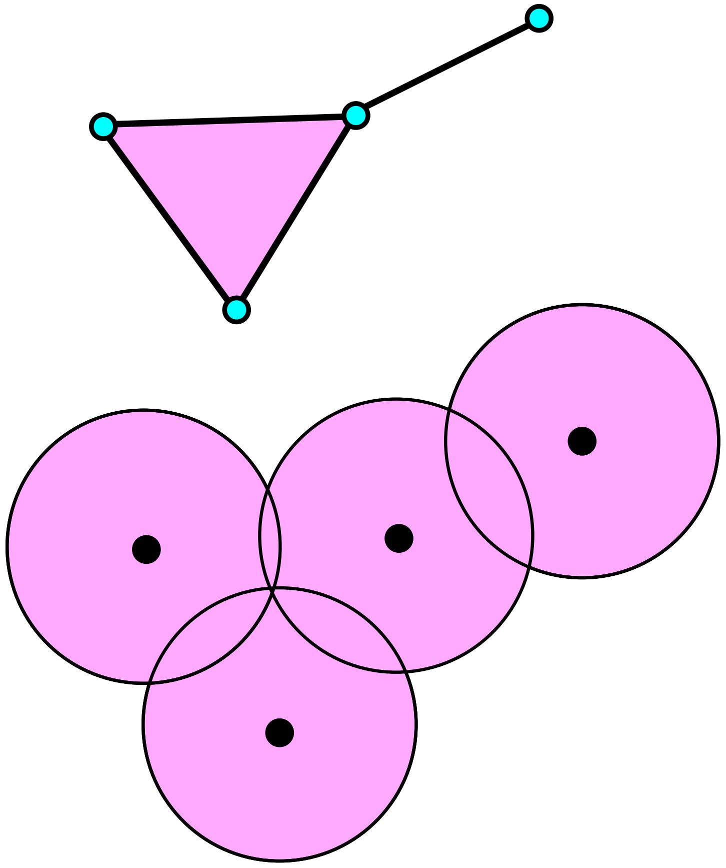


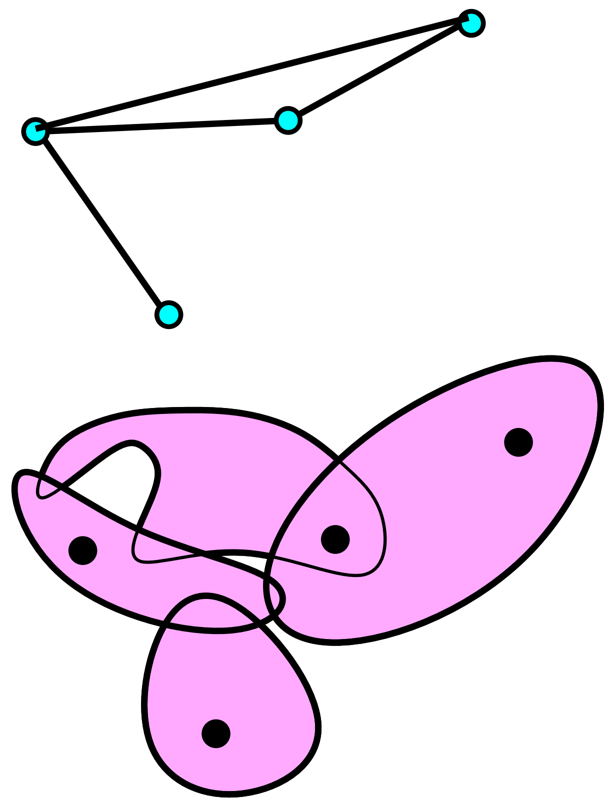
- f a real valued function defined on the vertices of K
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

$\Rightarrow$ The sublevel sets filtration.

Exercise: show that this is a filtration.

Example: The Cěch complex

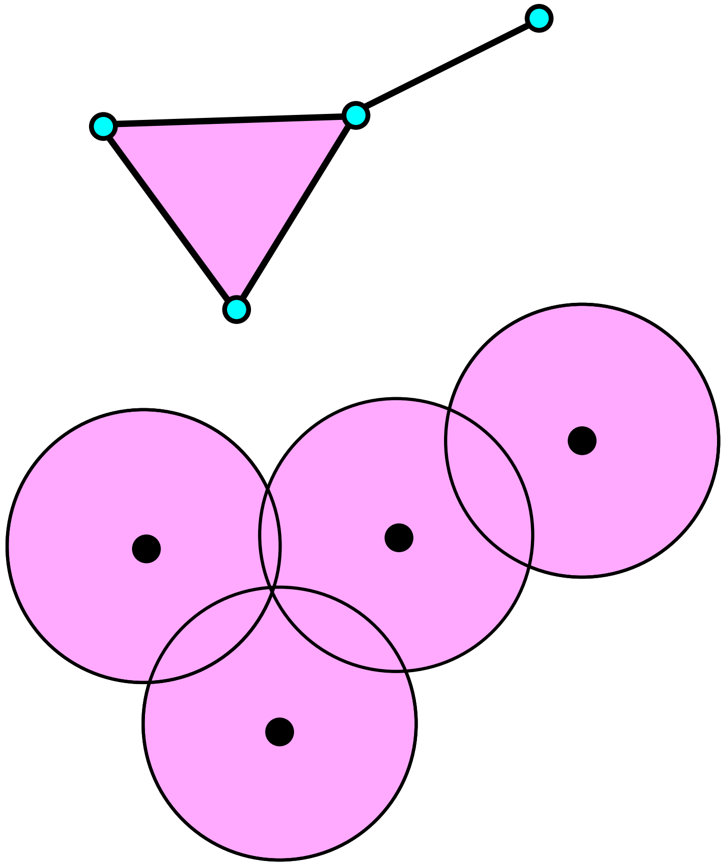


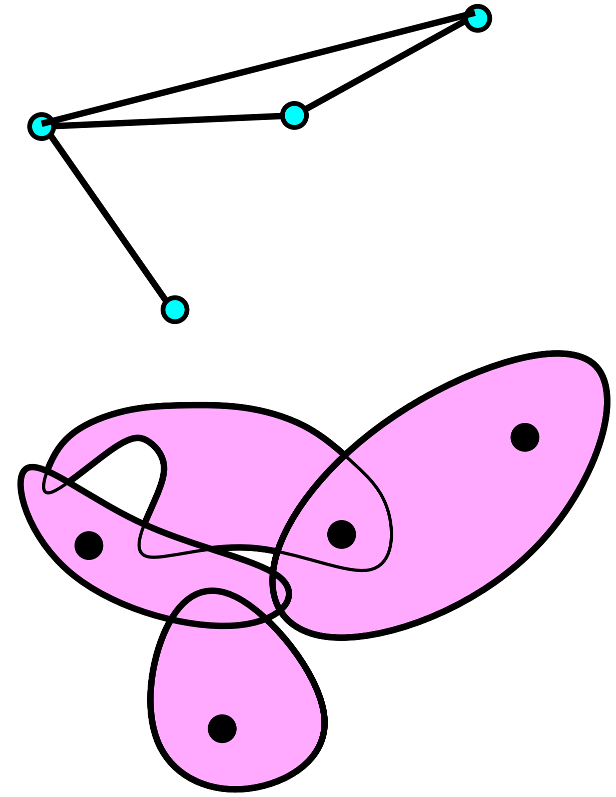


- Let $\mathcal{U} = (U_i)_{i \in I}$ be a covering of a topological space X by open sets: $X = \cup_{i \in I} U_i$.

- 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 $\cap_{j=0}^{k} U_{i_j} \neq \emptyset$.

Example: The Cěch complex

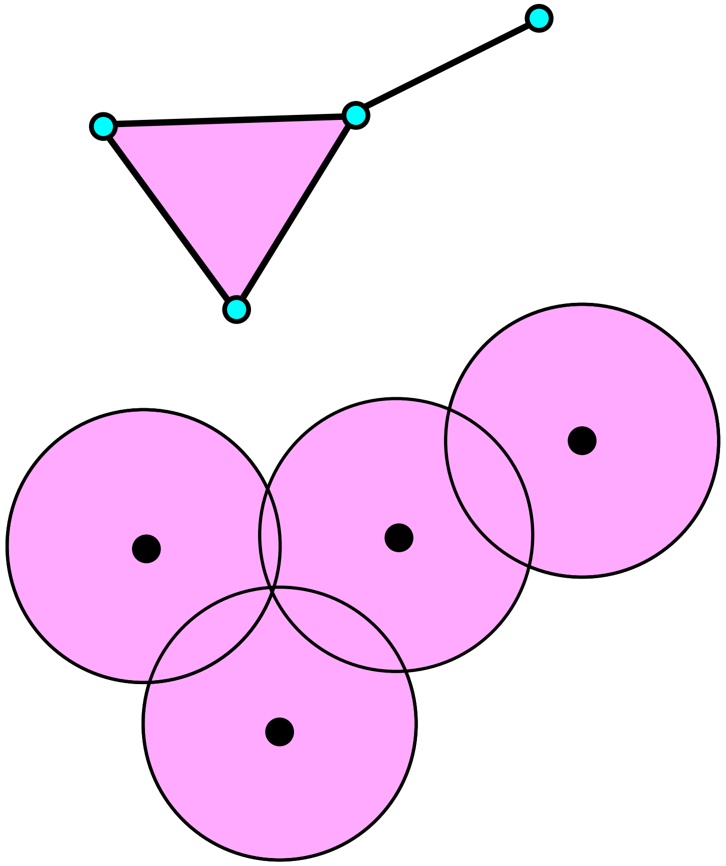


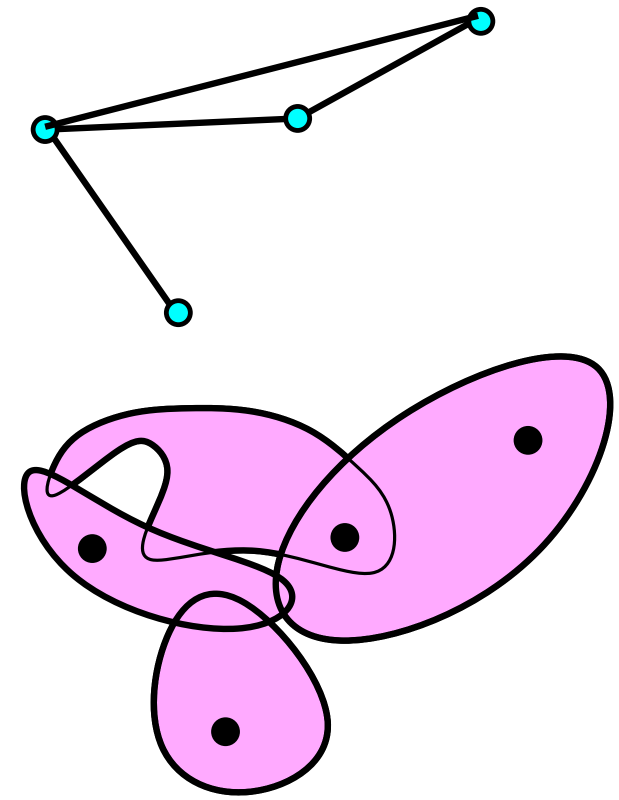


Nerve theorem (Leray): If all the intersections between opens in $\mathcal{U}$ are either empty or contractible then $C(\mathcal{U})$ and $X = \cup_{i \in I} U_i$ are homotopy equivalent.

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

Example: The Cěch complex



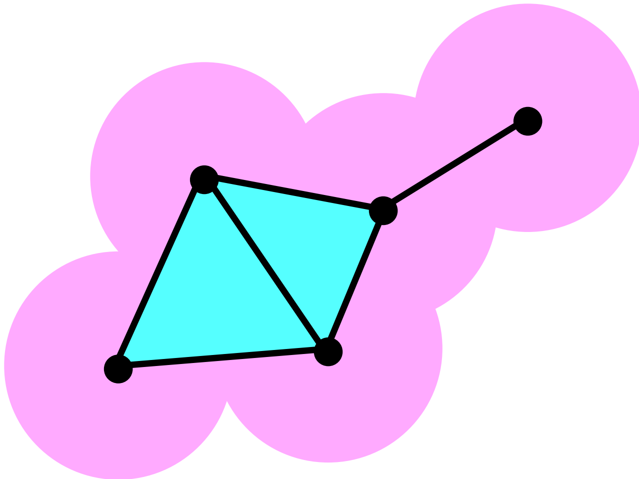


Nerve theorem (Leray): If all the intersections between opens in $\mathcal{U}$ are either empty or contractible then $C(\mathcal{U})$ and $X = \cup_{i \in I} U_i$ are homotopy equivalent.

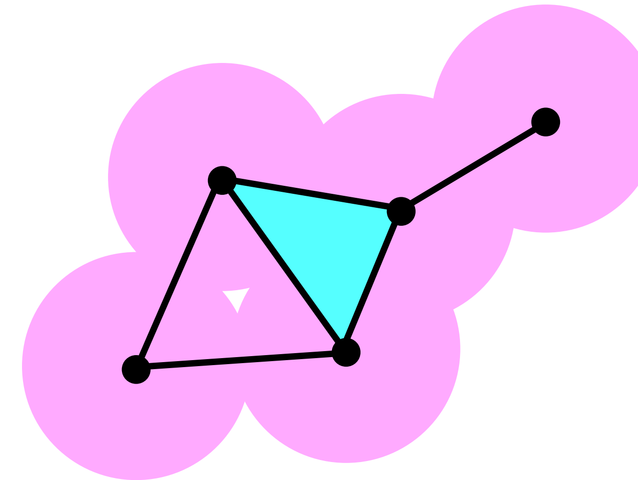
$\Rightarrow$ The combinatorics of the covering (a simplicial complex) carries the topology of the space.

Warning: even when the open sets are euclidean balls, the computation of the Cěch complex is a very difficult task!

Example: the Rips complex



Rips vs Čech



Let $L = \{p_0, \cdots p_n\}$ be a (finite) point cloud (in a metric space).

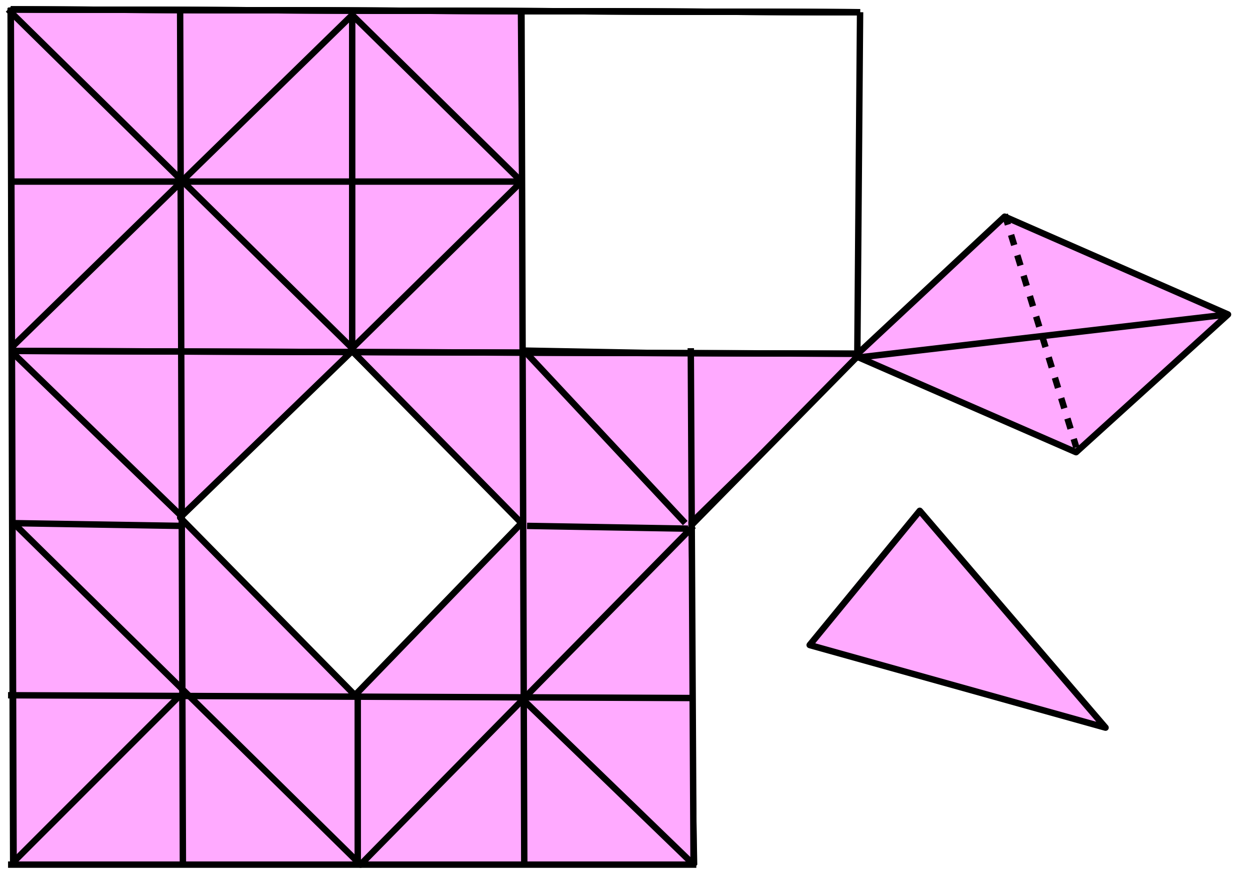
The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \cdots p_k \in L$,

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

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

$$\mathcal{C}^{\frac{\alpha}{2}}(L) \subseteq \mathcal{R}^\alpha(L) \subseteq \mathcal{C}^\alpha(L) \subseteq \mathcal{R}^{2\alpha}(L) \subseteq \cdots$$

Homology of simplicial complexes



- 2 connected components
- Intuitively: 2 cycles

Topological invariants:
- Number of connected components
- Number of cycles: how to define a cycle?
- Number of voids: how to define a void?
- ...

(Simplicial) homology and Betti numbers

In the following: homology with coefficient in $\mathbb{Z}/2$

Refs: J.R. Munkres, *Elements of Algebraic Topology*, Addison-Wesley, 1984.
A. Hatcher, *Algebraic Topology*, Cambridge University Press 2002.

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.

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:

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

The space $\mathcal{C}_k(K)$ of k-chains is a $\mathbb{Z}/2$-vector space

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

Geometric interpretation:
k-chain $=$ union of k-simplices
sum $c + c'$ $=$ symmetric difference

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:

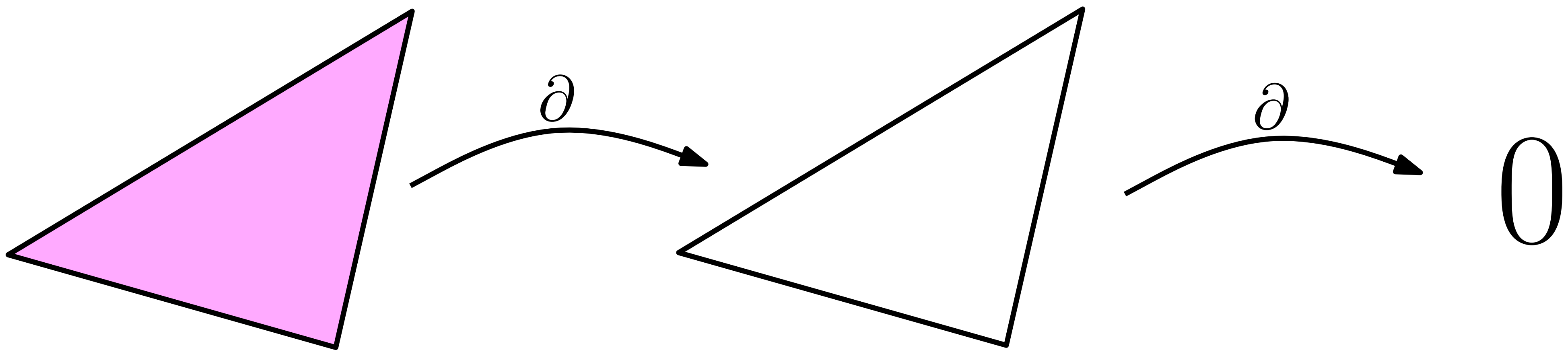
$$c + c' = \sum_{i=1}^{p} (\varepsilon_i + \varepsilon'_i) \sigma_i \quad \text{and} \quad \lambda.c = \sum_{i=1}^{p} (\lambda \varepsilon'_i) \sigma_i$$

The space $\mathcal{C}_k(K)$ of k-chains is a $\mathbb{Z}/2$-vector space

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

Geometric interpretation:
k-chain $=$ union of k-simplices
sum $c + c'$ $=$ symmetric difference

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.

$$If \sigma = [v_0, \cdots, v_k] \;\; \text{then} \;\; \partial\sigma = \sum_{i=0}^{k} [v_0 \cdots \hat{v}_i \cdots v_k]$$

The boundary operator is the linear map defined by

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

Fundamental property of the boundary operator

$$\partial\partial := \partial \circ \partial = 0$$

Proof: by linearity it is just necessary to prove it for a simplex.

$$\begin{aligned}
\partial\partial\sigma &= \partial\left(\sum_{i=0}^{k}[v_0\cdots\hat{v}_i\cdots v_k]\right) \\
&= \sum_{i=0}^{k}\partial[v_0\cdots\hat{v}_i\cdots v_k] \\
&= \sum_{j<i}[v_0\cdots\hat{v}_j\cdots\hat{v}_i\cdots v_k] + \sum_{j>i}[v_0\cdots\hat{v}_i\cdots\hat{v}_j\cdots v_k] \\
&= 0
\end{aligned}$$

Cycles and boundaries

The chain complex associated to a complex K of dimension d

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

k-cycles:

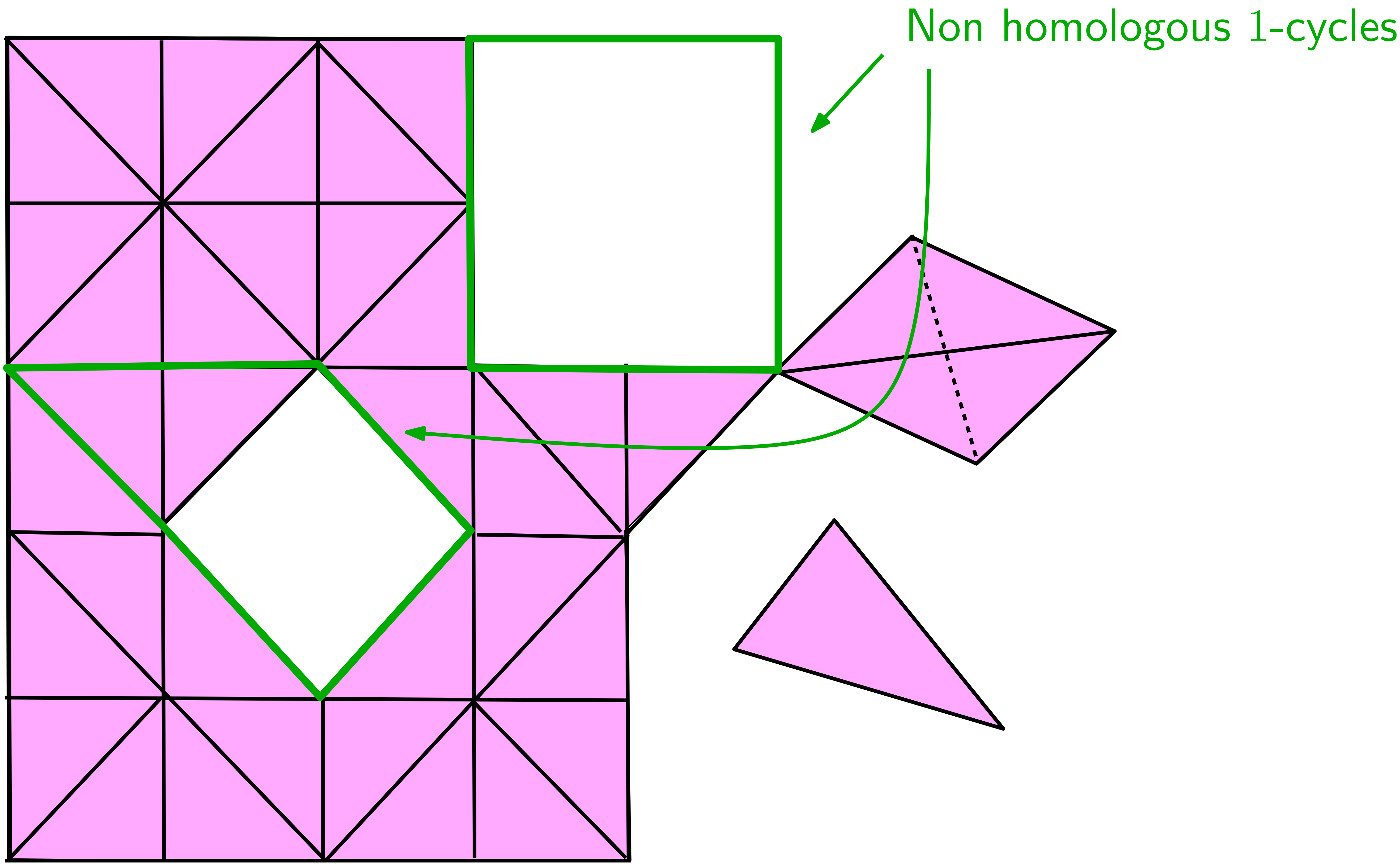
$$Z_k(K) := \ker(\partial : \mathcal{C}_k \to \mathcal{C}_{k-1}) = \{c \in \mathcal{C}_k : \partial c = \emptyset\}$$

k-boundaries:

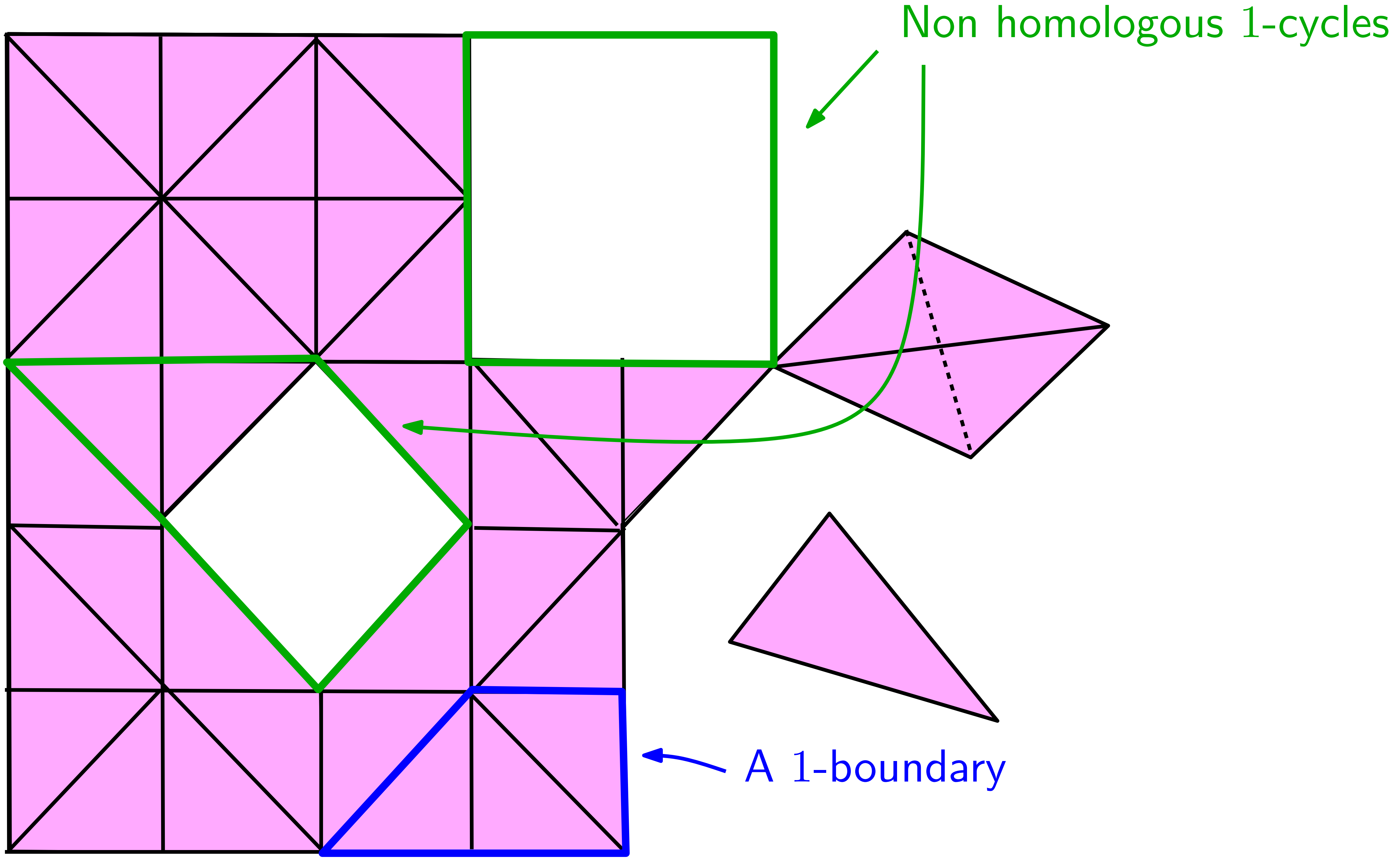
$$B_k(K) := im(\partial : \mathcal{C}_{k+1} \to \mathcal{C}_k) = \{c \in \mathcal{C}_k : \exists c' \in \mathcal{C}_{k+1}, c = \partial c'\}$$

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

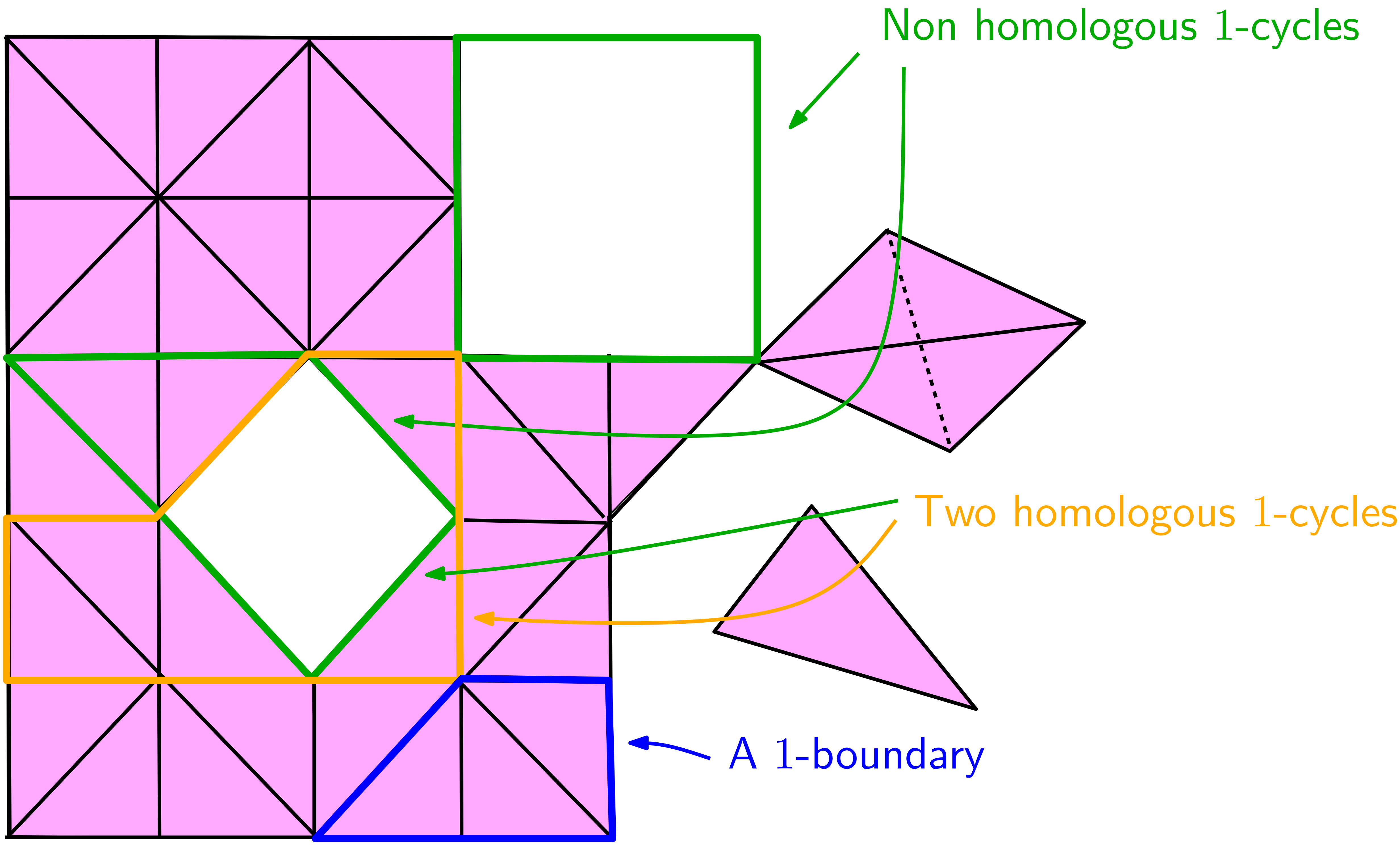
Cycles and boundaries



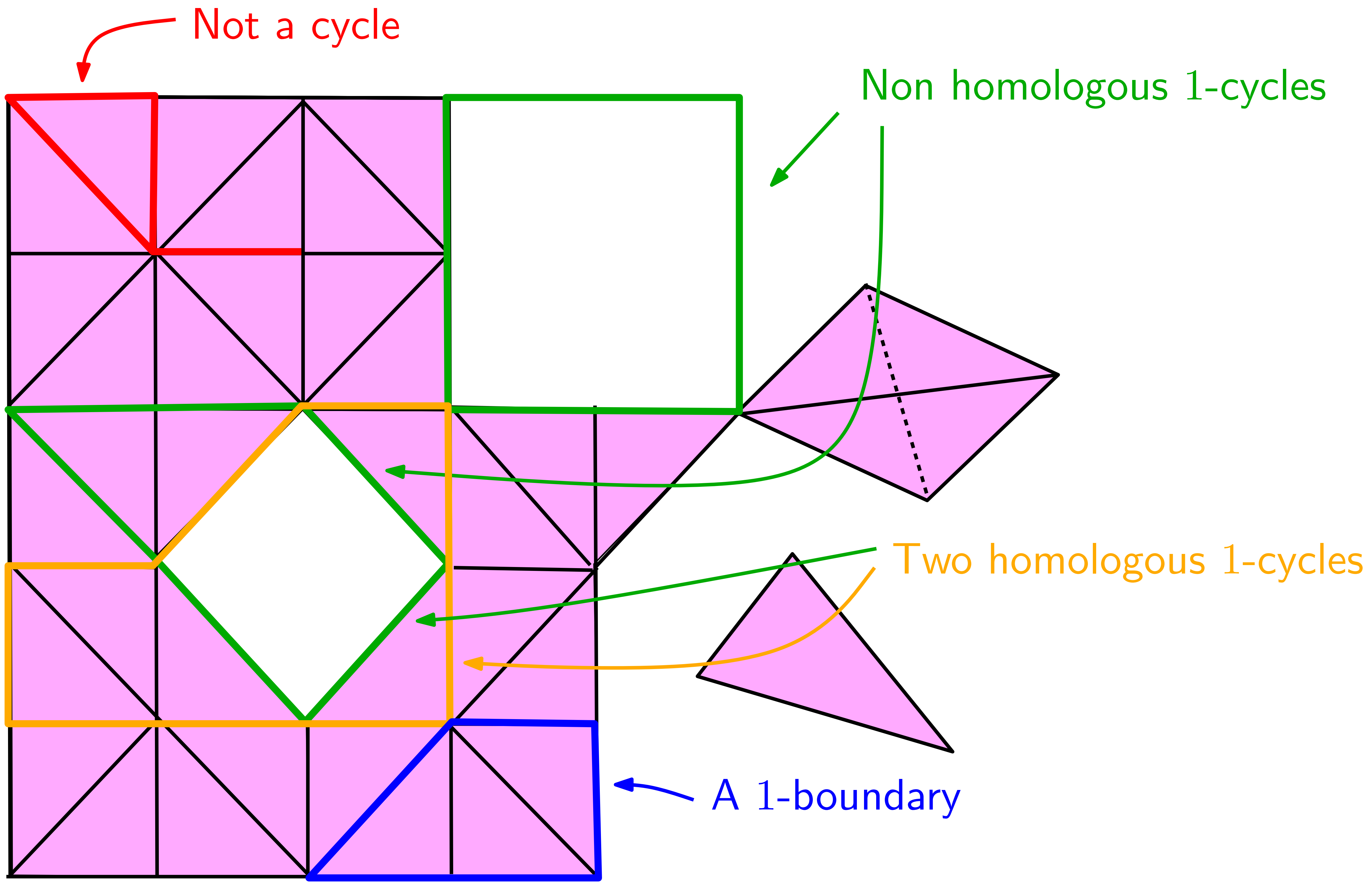
Cycles and boundaries



Cycles and boundaries



Cycles and boundaries



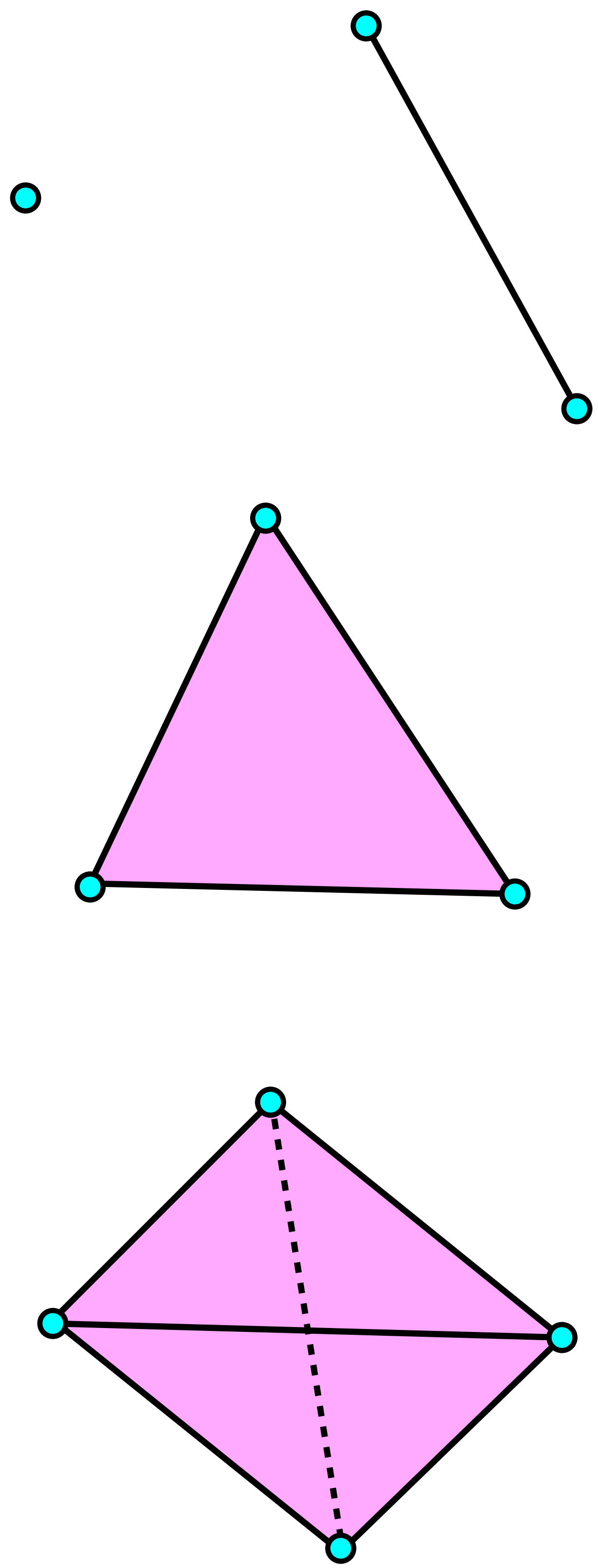
Homology groups and Betti numbers

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

- 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: $\beta_k(K) = \dim(H_k(K))$.

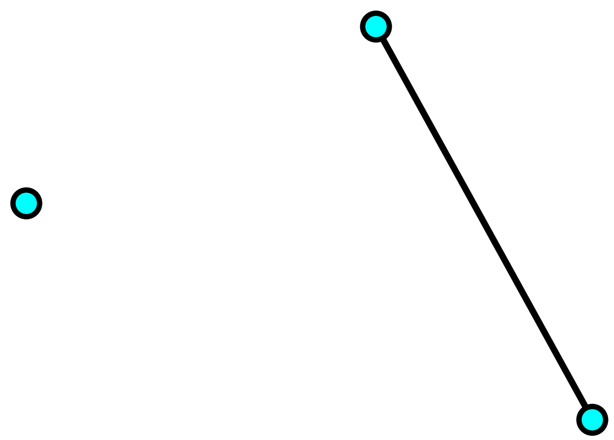
Elementary examples

Remark: $\beta_0 =$ number of connected components of K

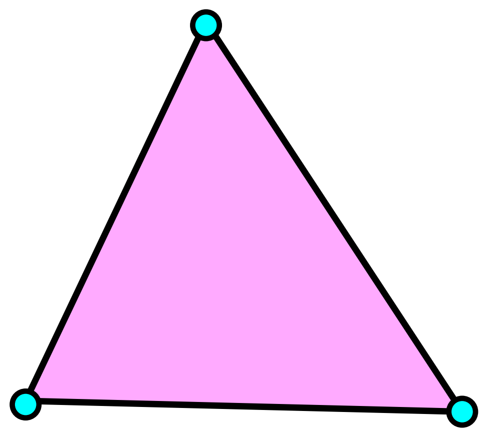


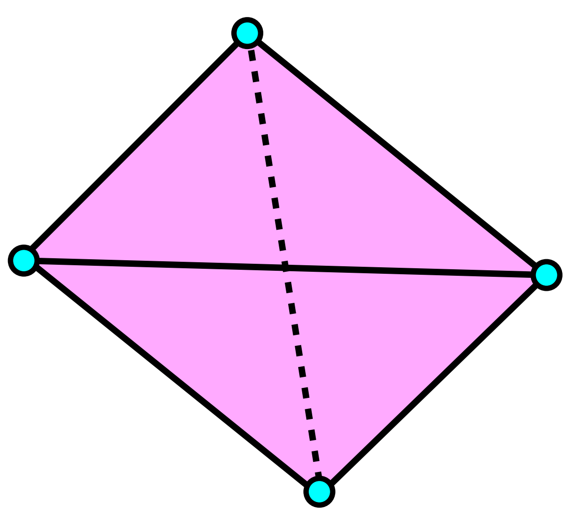
Elementary examples

Remark: $\beta_0 =$ number of connected components of K



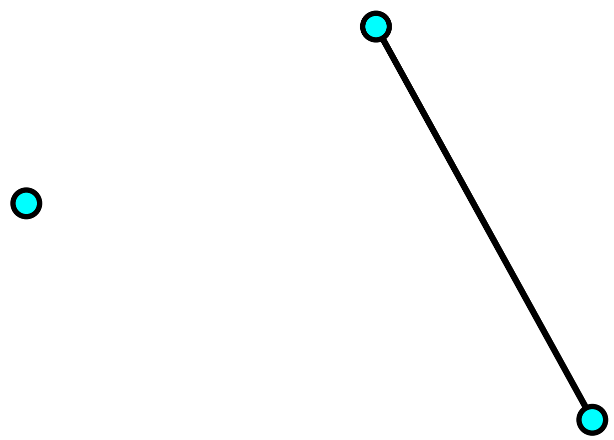
$\beta_0 = 2$
$\beta_1 = 0$
$\beta_2 = 0$



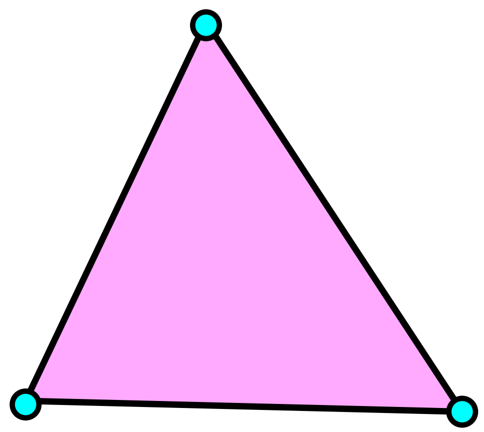


Elementary examples

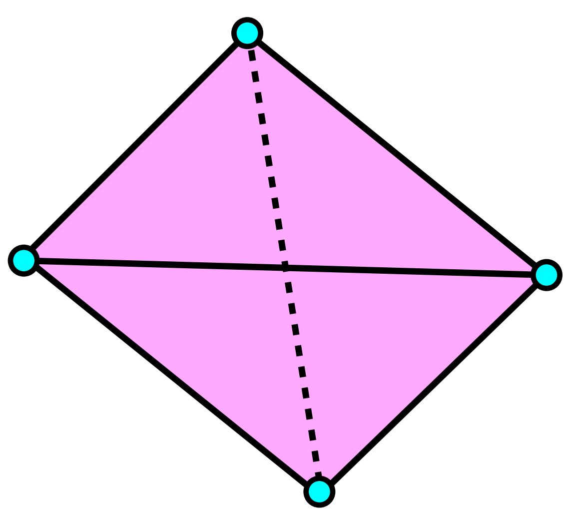
Remark: $\beta_0 =$ number of connected components of K



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

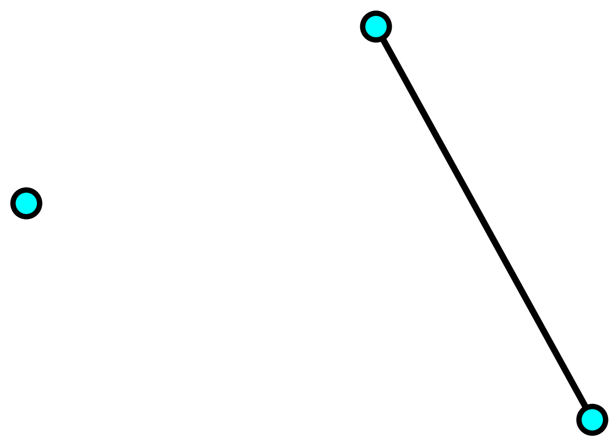


$\beta_0 = 1$
$\beta_1 = 0$
$\beta_2 = 0$

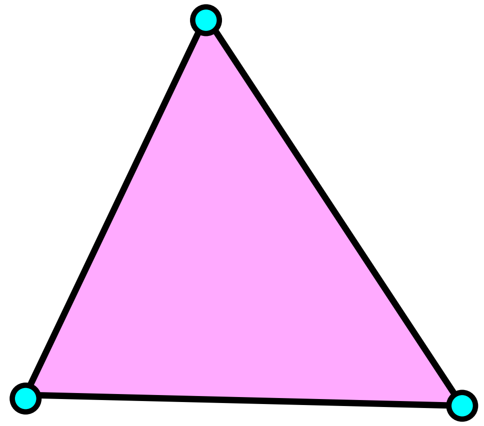


Elementary examples

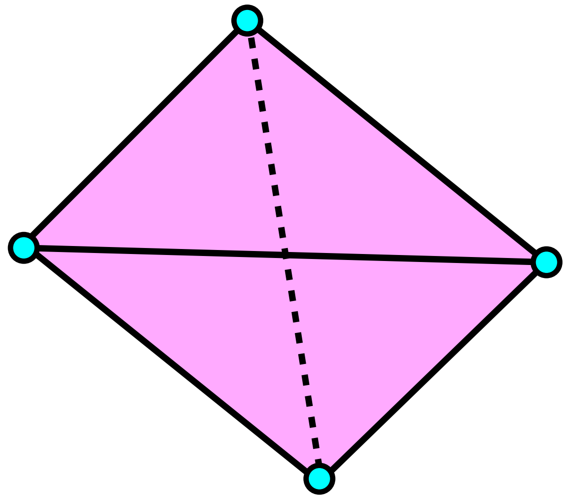
Remark: β_0 = number of connected components of K



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

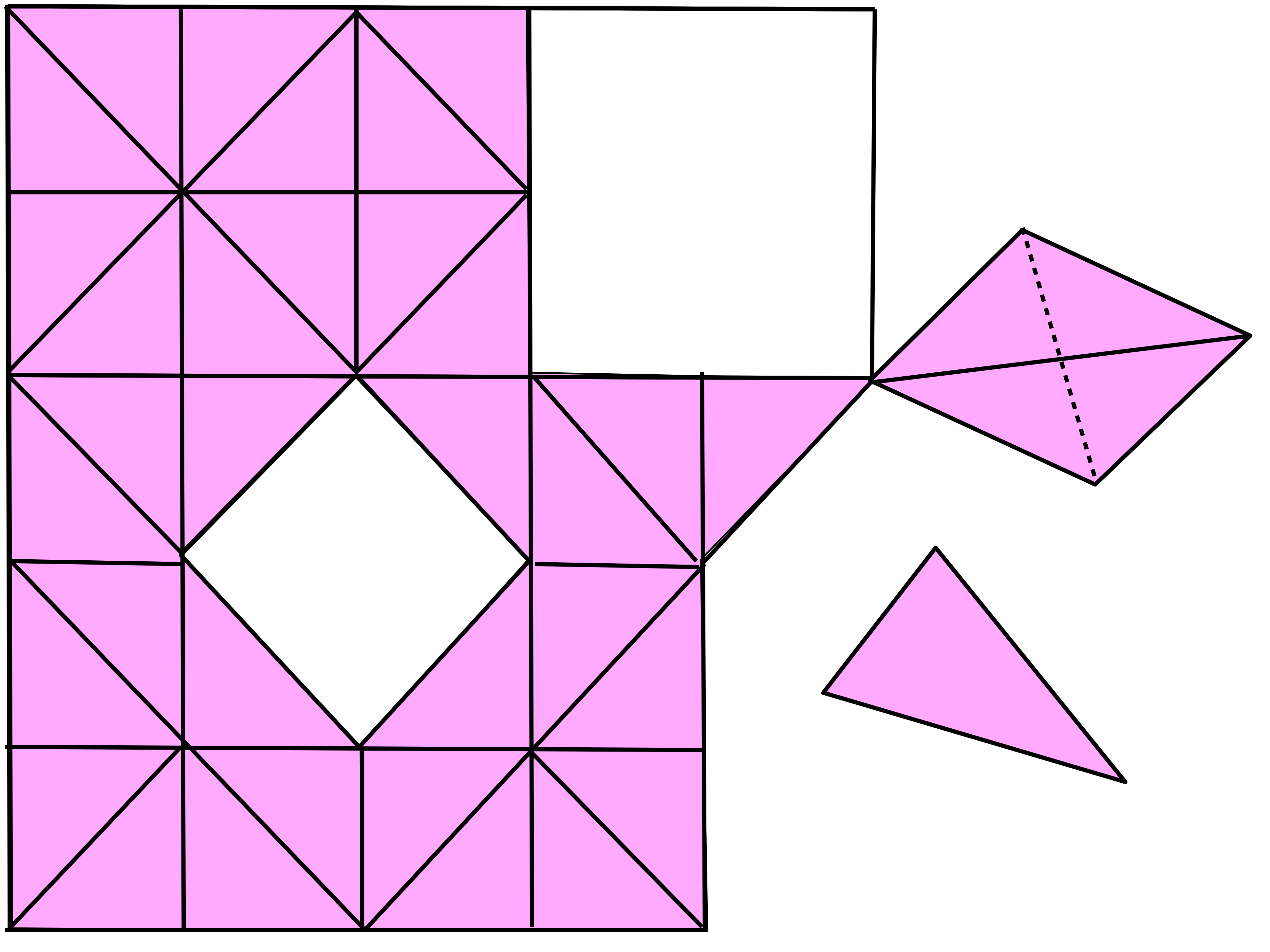


$\beta_0 = 1$
$\beta_1 = 0$
$\beta_2 = 0$

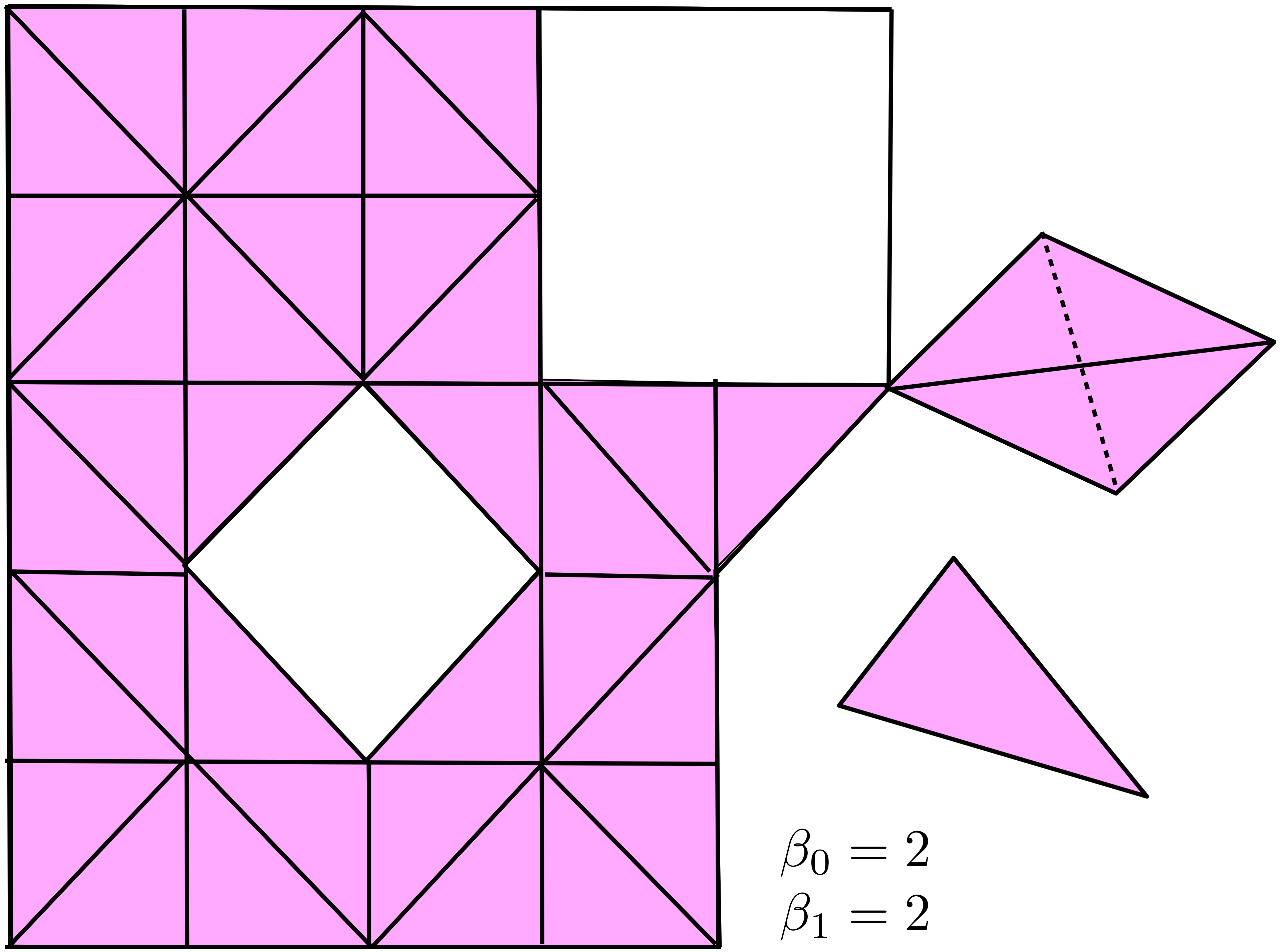


$\beta_0 = 1$
$\beta_1 = 0$
$\beta_2 = 1$ if empty and $\beta_2 = 0$ if filled
$\beta_3 = 0$

Elementary examples

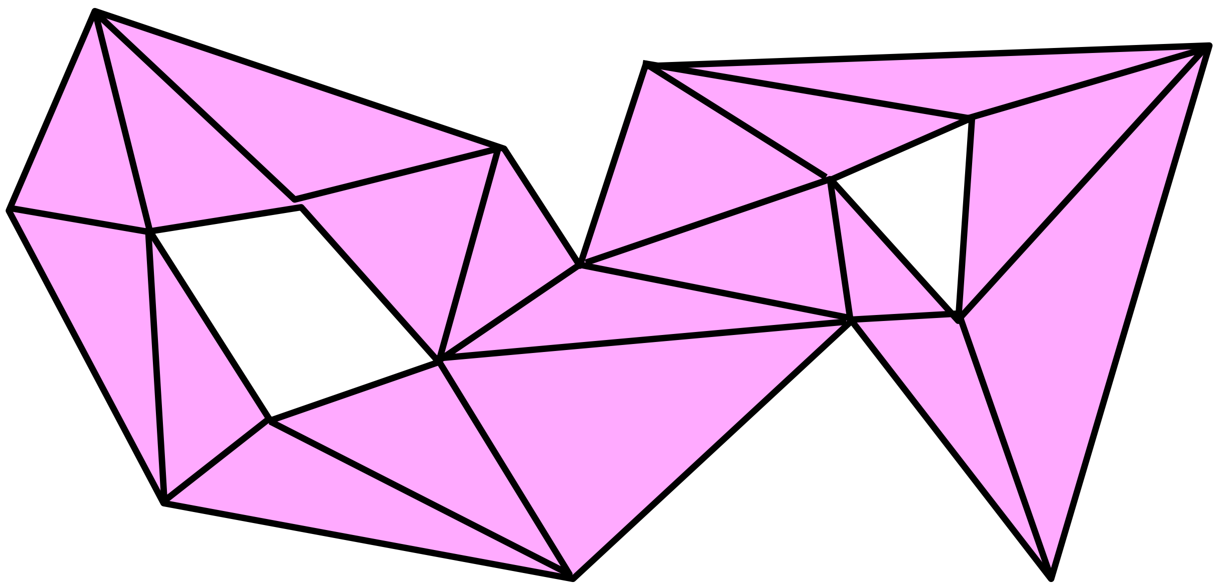


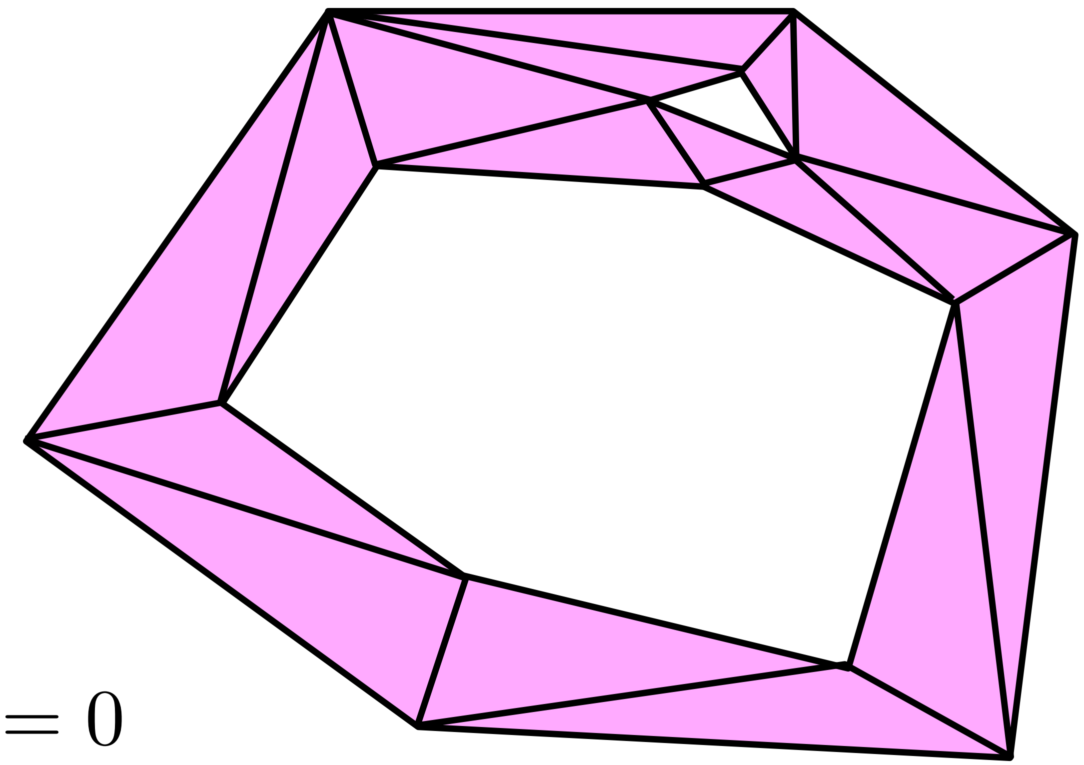
Elementary examples



$\beta_0 = 2$
$\beta_1 = 2$
$\beta_2 = 1$ if empty and $\beta_2 = 0$ if filled
$\beta_3 = 0$

Topological invariance and singular homology



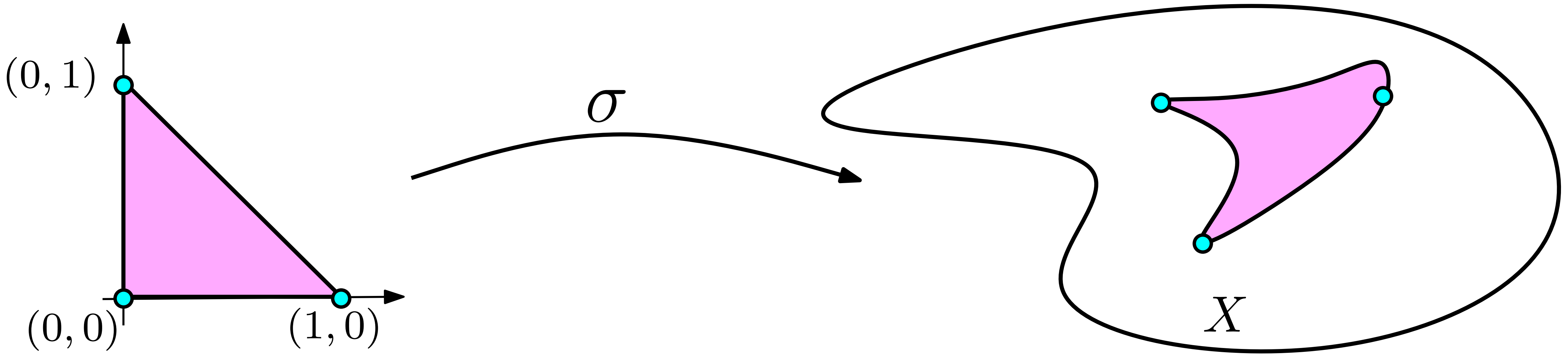


$$\beta_0 = 1,\ \beta_1 = 2, \beta_2 = 0$$

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 $\rightarrow$ defined for any topological space.

Topological invariance and singular homology



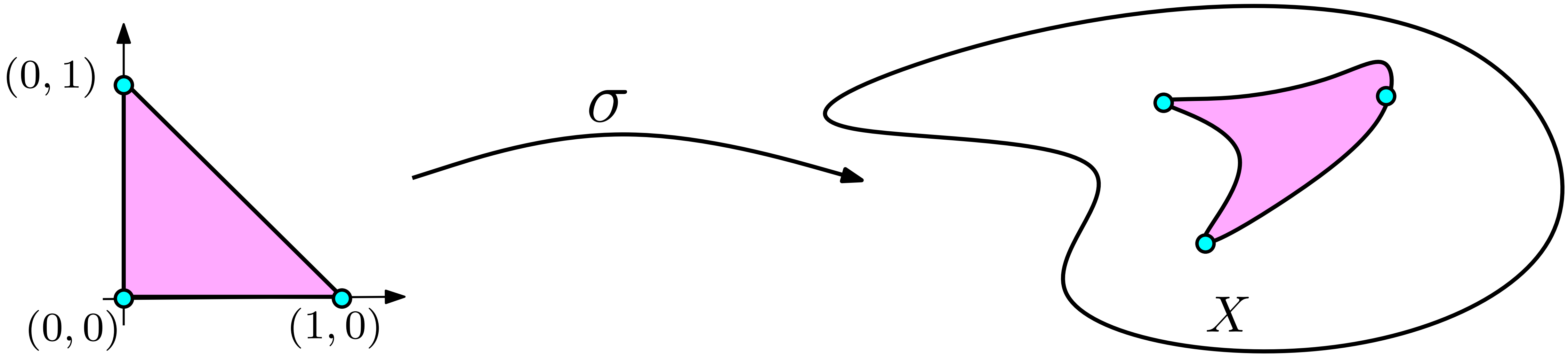
Let Δ_k be the standard simplex in $\mathbb{R}^k$. A singular k-simplex in a topological space X is a continuous map $\sigma : \Delta_k \to X$.

The same construction as for simplicial homology can be done with singular complexes $\to$ 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



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

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.

An algorithm for geometric inference

- $X \subset \mathbb{R}^d$ be a compact set such that $\text{wfs}(X) > 0$.
- $L \subset \mathbb{R}^d$ be a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon > 0$.

An algorithm for geometric inference

- $X \subset \mathbb{R}^d$ be a compact set such that wfs$(X) > 0$.
- $L \subset \mathbb{R}^d$ be a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon > 0$.

Goal: Compute the Betti numbers of X^r for $0 < r < \mathrm{wfs}(X)$ from L.

An algorithm for geometric inference

- $X \subset \mathbb{R}^d$ be a compact set such that wfs$(X) > 0$.
- $L \subset \mathbb{R}^d$ be a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon > 0$.

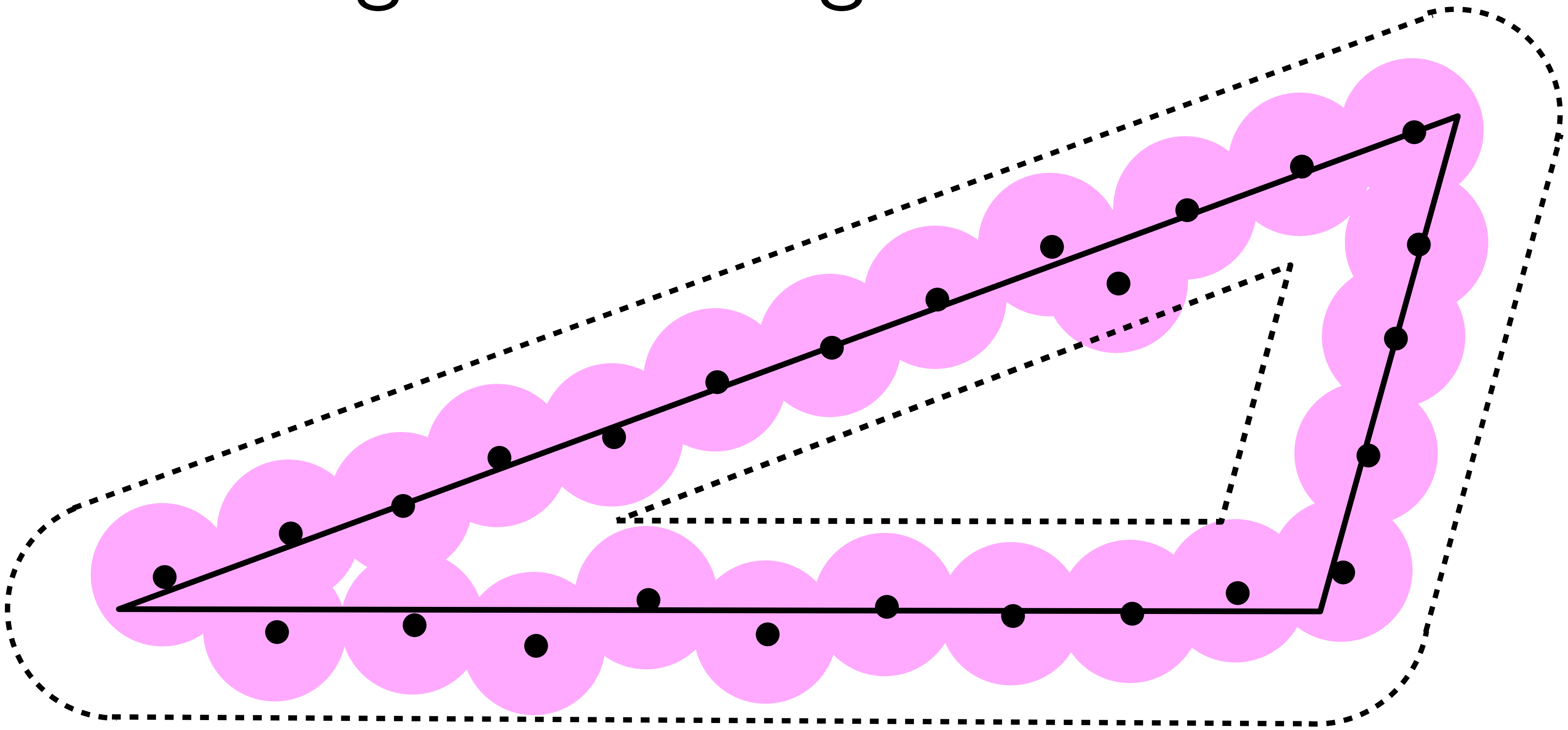
Goal: Compute the Betti numbers of X^r for $0 < r < \mathrm{wfs}(X)$ from L.

Theorem: [CL'05 - CSEH'05]
Assume that wfs$(X) > 4\varepsilon$. For $\alpha > 0$ s.t. $\alpha + 4\varepsilon <$ wfs(X), let $i : L^{\alpha+\varepsilon} \hookrightarrow L^{\alpha+3\varepsilon}$ be the canonical inclusion.For any $0 < r < \mathrm{wfs}(X)$,

$$H_k(X^r) \cong im\left(i_* : H_k(L^{\alpha+\varepsilon}) \to H_k(L^{\alpha+3\varepsilon})\right)$$

An algorithm for geometric inference

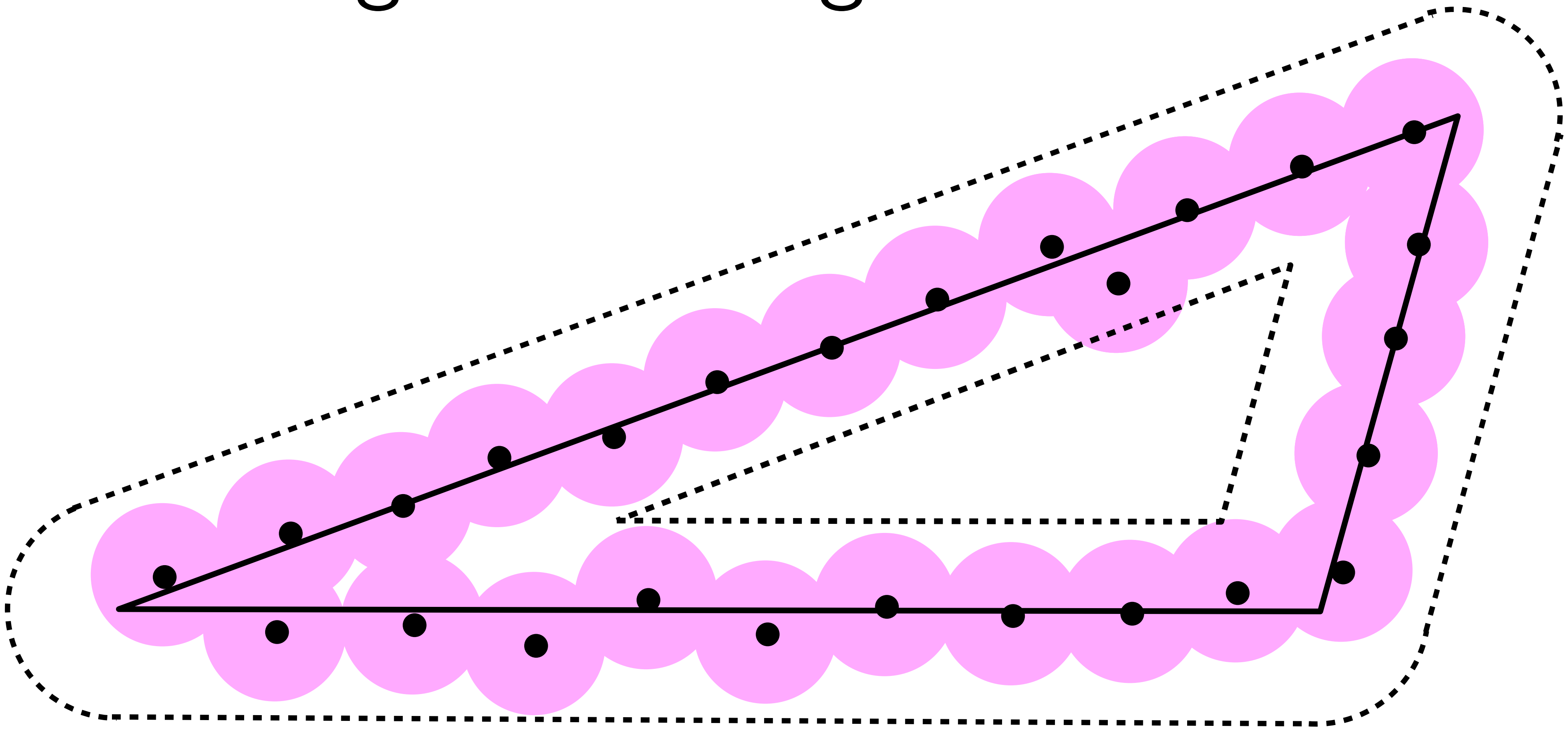
Proof:



For any $\alpha > 0$, $\quad X^{\alpha} \subseteq L^{\alpha+\varepsilon} \subseteq X^{\alpha+2\varepsilon} \subseteq L^{\alpha+3\varepsilon} \subseteq X^{\alpha+4\varepsilon} \subseteq \cdots$

An algorithm for geometric inference

Proof:



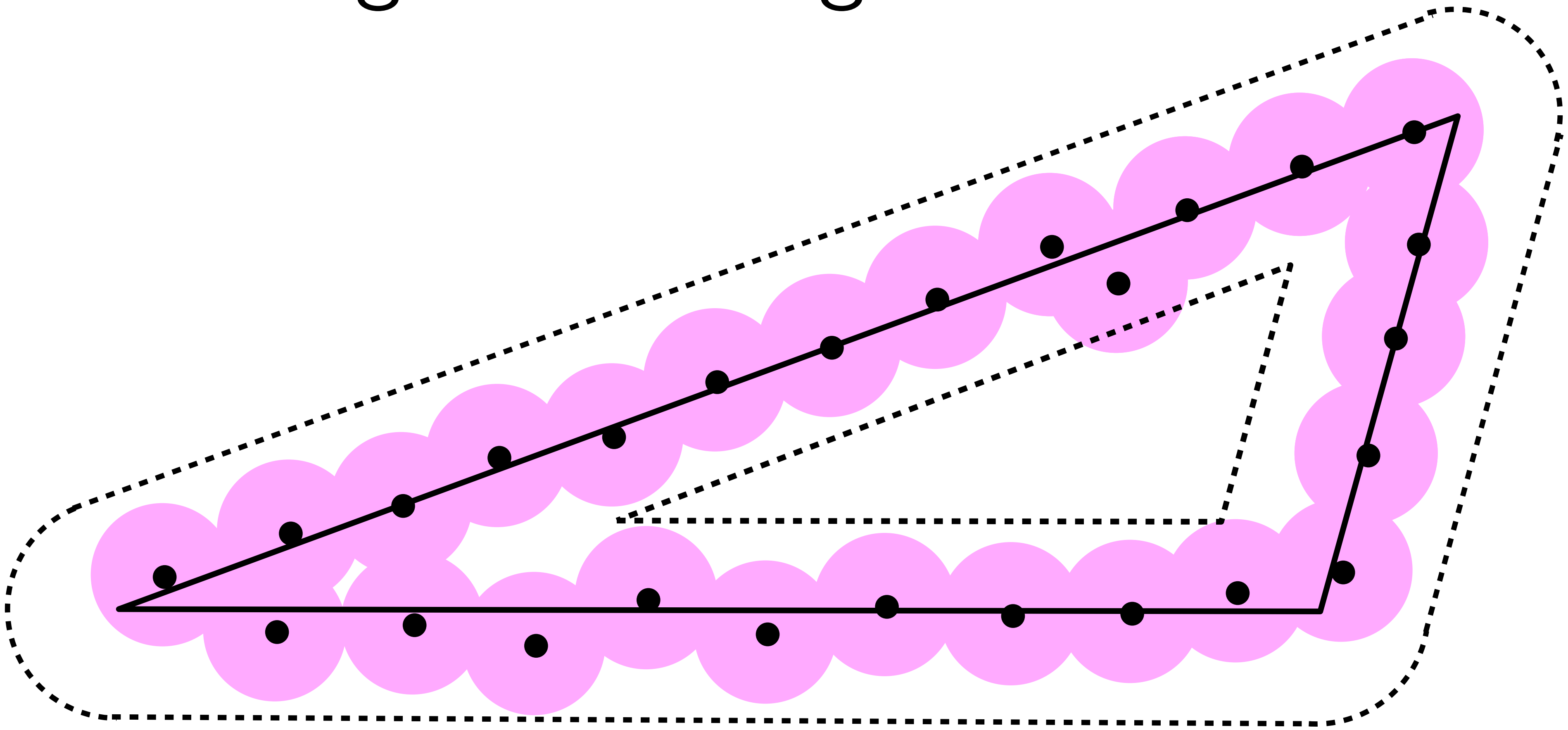
For any $\alpha > 0$, $\quad X^{\alpha} \subseteq L^{\alpha+\varepsilon} \subseteq X^{\alpha+2\varepsilon} \subseteq L^{\alpha+3\varepsilon} \subseteq X^{\alpha+4\varepsilon} \subseteq \cdots$

At homology level:

$$H_k(X^{\alpha}) \to H_k(L^{\alpha+\varepsilon}) \to H_k(X^{\alpha+2\varepsilon}) \to H_k(L^{\alpha+3\varepsilon}) \to H_k(X^{\alpha+4\varepsilon}) \to \cdots$$

An algorithm for geometric inference

Proof:



For any $\alpha > 0$, $X^\alpha \subseteq L^{\alpha+\varepsilon} \subseteq X^{\alpha+2\varepsilon} \subseteq L^{\alpha+3\varepsilon} \subseteq X^{\alpha+4\varepsilon} \subseteq \cdots$

At homology level:

rank $= \dim H_k(X^\alpha)$

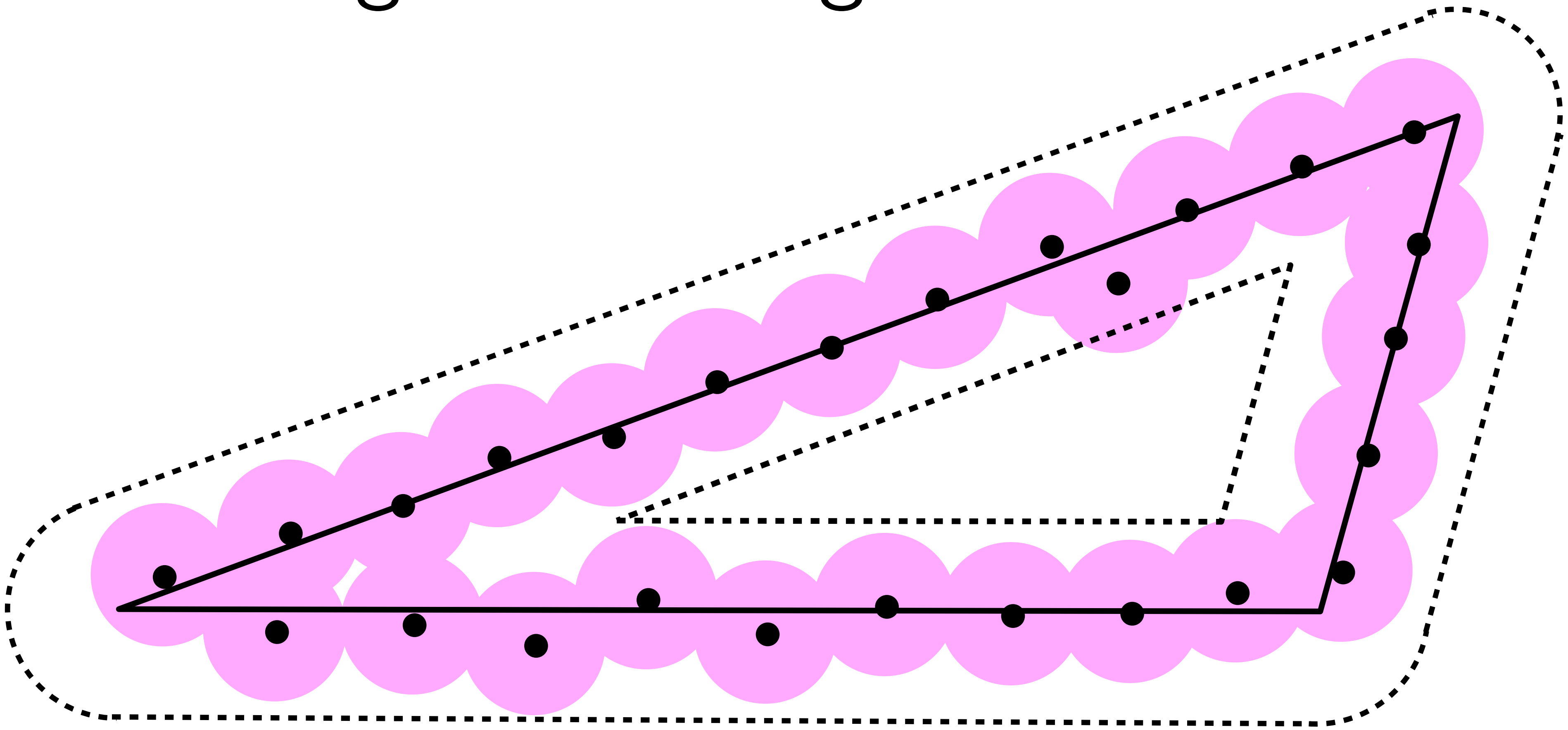
$$H_k(X^\alpha) \to H_k(L^{\alpha+\varepsilon}) \to H_k(X^{\alpha+2\varepsilon}) \to H_k(L^{\alpha+3\varepsilon}) \to H_k(X^{\alpha+4\varepsilon}) \to \cdots$$

isomorphism

isomorphism

An algorithm for geometric inference

Proof:



For any $\alpha > 0$, $\quad X^{\alpha} \subseteq L^{\alpha+\varepsilon} \subseteq X^{\alpha+2\varepsilon} \subseteq L^{\alpha+3\varepsilon} \subseteq X^{\alpha+4\varepsilon} \subseteq \cdots$

At homology level:

Cannot be directly computed !

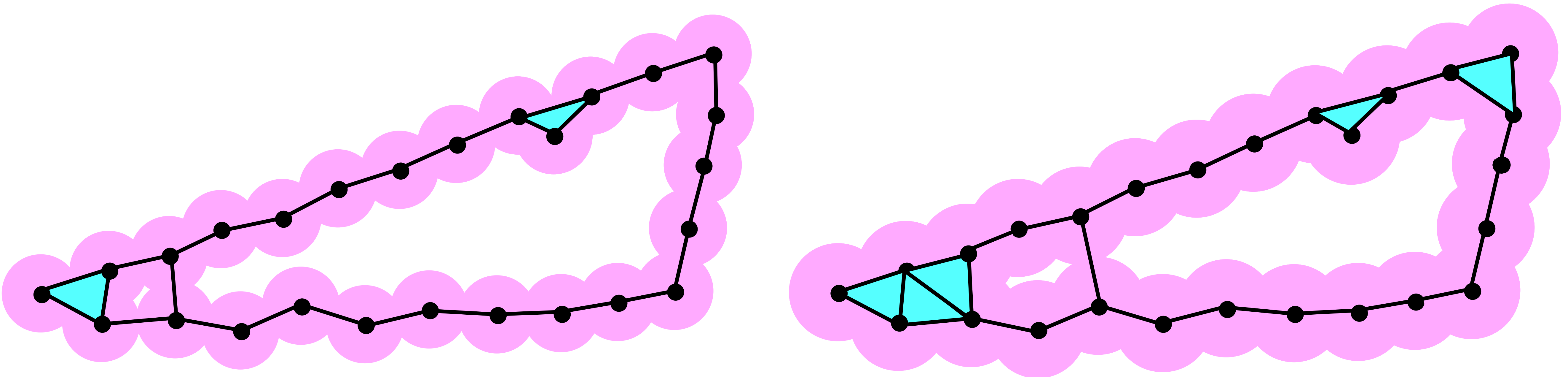
rank $= \dim H_k(X^{\alpha})$

$$H_k(X^{\alpha}) \to H_k(L^{\alpha+\varepsilon}) \to H_k(X^{\alpha+2\varepsilon}) \to H_k(L^{\alpha+3\varepsilon}) \to H_k(X^{\alpha+4\varepsilon}) \to \cdots$$

isomorphism

isomorphism

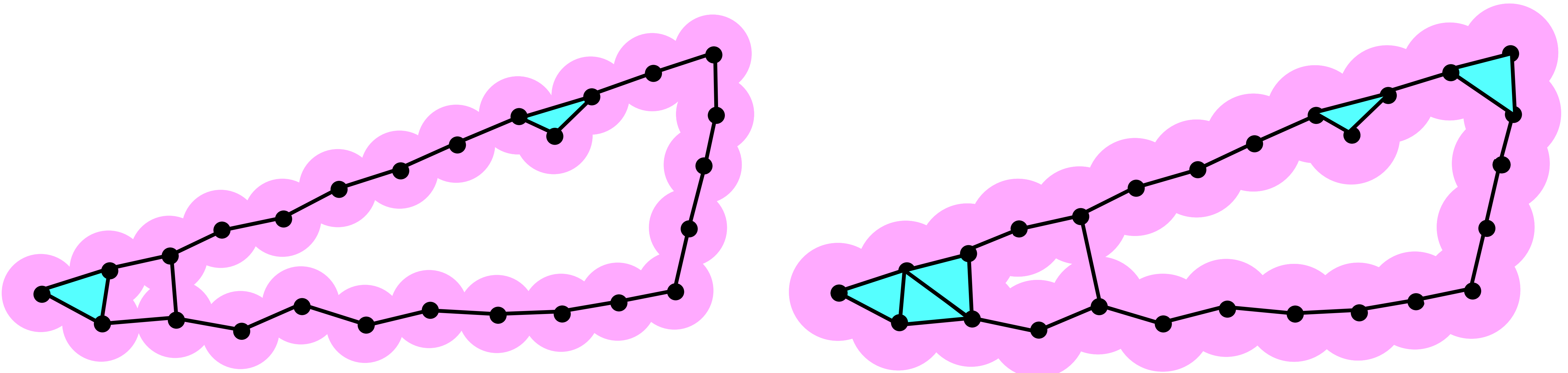
Using the Čech complex



The Čech complex $\mathcal{C}^\alpha(L)$:

$$\text{for } p_0, \cdots p_k \in L, \quad \sigma = [p_0 p_1 \cdots p_k] \in \mathcal{C}^\alpha(L) \quad \text{iff} \quad \bigcap_{i=0}^{k} B(p_i, \alpha) \neq \emptyset$$

Using the Čech complex

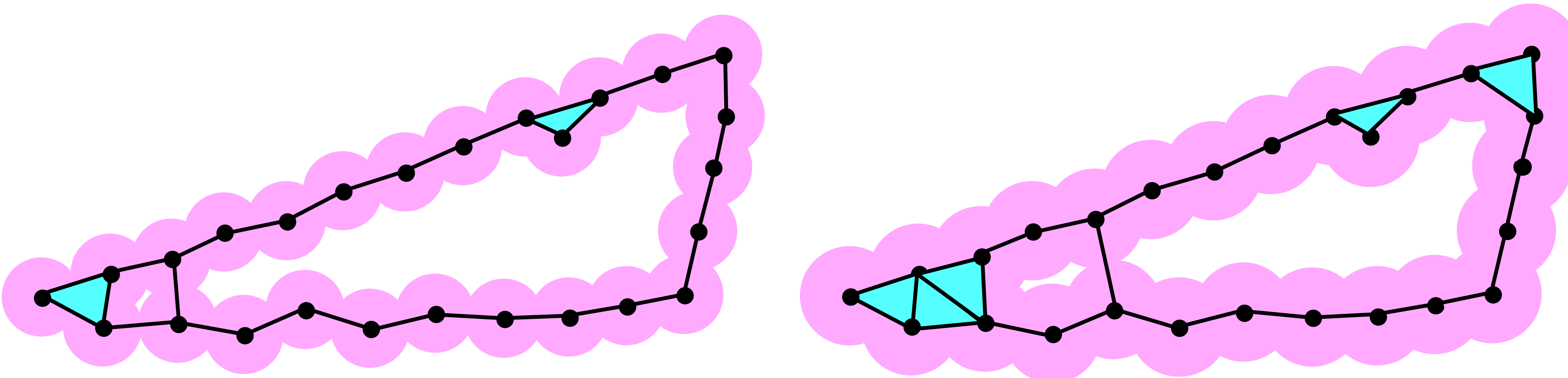


The Čech complex $\mathcal{C}^\alpha(L)$:

$$\text{for } p_0, \cdots p_k \in L, \quad \sigma = [p_0 p_1 \cdots p_k] \in \mathcal{C}^\alpha(L) \;\text{ iff }\; \bigcap_{i=0}^{k} B(p_i, \alpha) \neq \emptyset$$

Nerve theorem: For any $\alpha > 0$, L^α and $\mathcal{C}^\alpha(L)$ are homotopy equivalent and the homotopy equivalences can be chosen to commute with inclusions.

Using the Čech complex



The Čech complex $\mathcal{C}^{\alpha}(L)$:

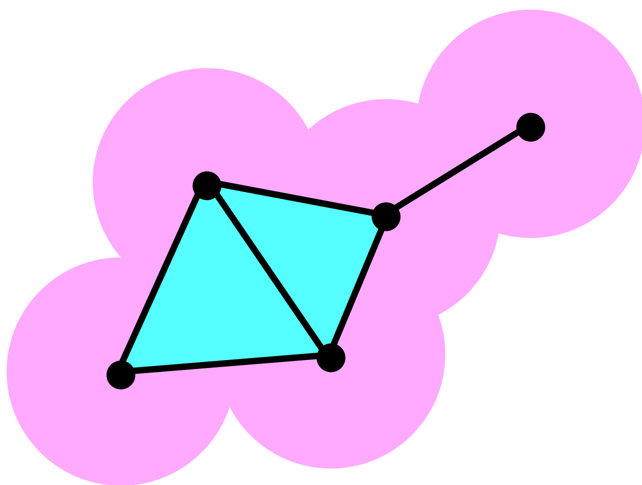
$$\text{for } p_0, \cdots p_k \in L, \quad \sigma = [p_0 p_1 \cdots p_k] \in \mathcal{C}^{\alpha}(L) \;\text{ iff }\; \bigcap_{i=0}^{k} B(p_i, \alpha) \neq \emptyset$$

Nerve theorem: For any $\alpha > 0$, L^{α} and $\mathcal{C}^{\alpha}(L)$ are homotopy equivalent and the homotopy equivalences can be chosen to commute with inclusions.

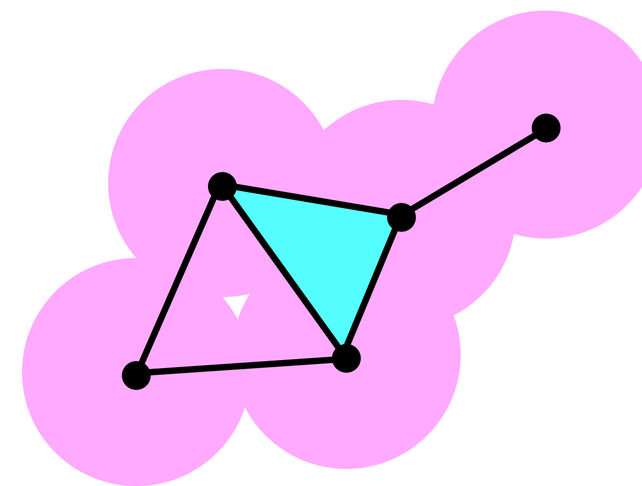
$$\begin{array}{ccccccc} \cdots & \to & H_k(L^{\alpha+\varepsilon}) & \to & H_k(L^{\alpha+3\varepsilon}) & \to & \cdots \\ & & \downarrow & & \downarrow & & \\ \cdots & \to & H_k(\mathcal{C}^{\alpha+\varepsilon}(L)) & \to & H_k(\mathcal{C}^{\alpha+3\varepsilon}(L)) & \to & \cdots \end{array}$$

Allow to work with simplicial complexes but... still too difficult to compute

Using the Rips complex



Rips vs Čech



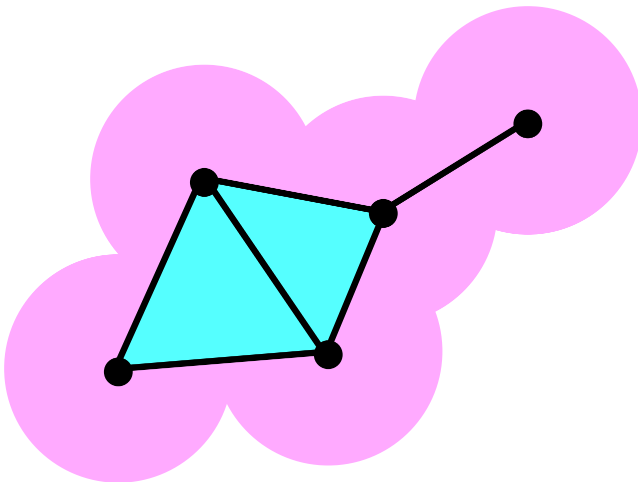
The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \cdots p_k \in L$,

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

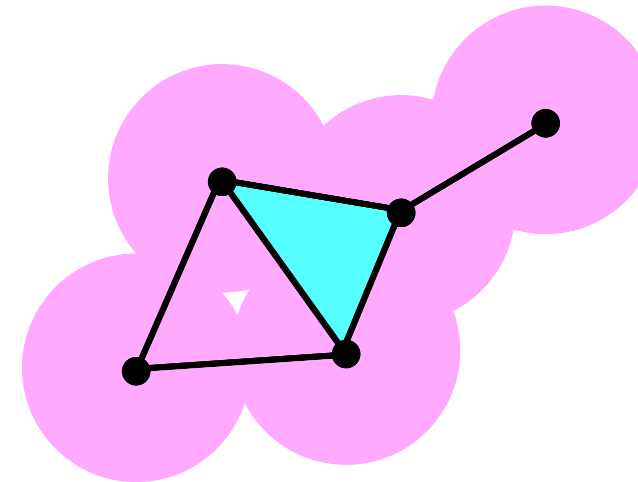
- Easy to compute and fully determined by its 1-skeleton
- Rips-Čech interleaving: for any $\alpha > 0$,

$$\mathcal{C}^{\frac{\alpha}{2}}(L) \subseteq \mathcal{R}^\alpha(L) \subseteq \mathcal{C}^\alpha(L) \subseteq \mathcal{R}^{2\alpha}(L) \subseteq \cdots$$

Using the Rips complex



Rips vs Čech



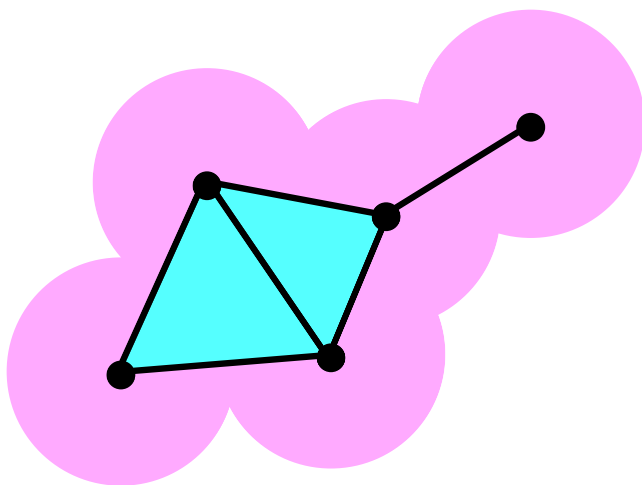
The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \cdots p_k \in L$,

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

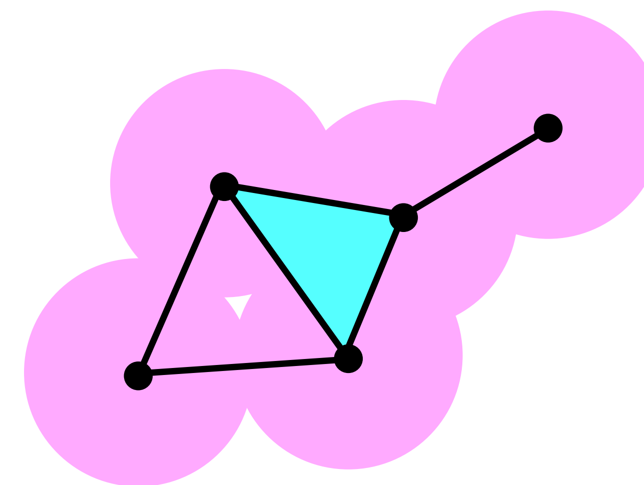
Theorem: [C-Oudot'08]
Let $X \subset \mathbb{R}^d$ be a compact set and $L \subset \mathbb{R}^d$ a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{9}$ wfs(X). Then for all $\alpha \in [2\varepsilon, \frac{1}{4}(\mathrm{wfs}(X) - \varepsilon)]$ and all $\lambda \in (0, \mathrm{wfs}(X))$, one has: $\forall k \in \mathbb{N}$

$$\beta_k(X^\lambda) = \dim(H_k(X^\lambda)) = \mathrm{rk}(\mathcal{R}^\alpha(L) \to \mathcal{R}^{4\alpha}(L))$$

Using the Rips complex



Rips vs Čech



The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \cdots p_k \in L$,

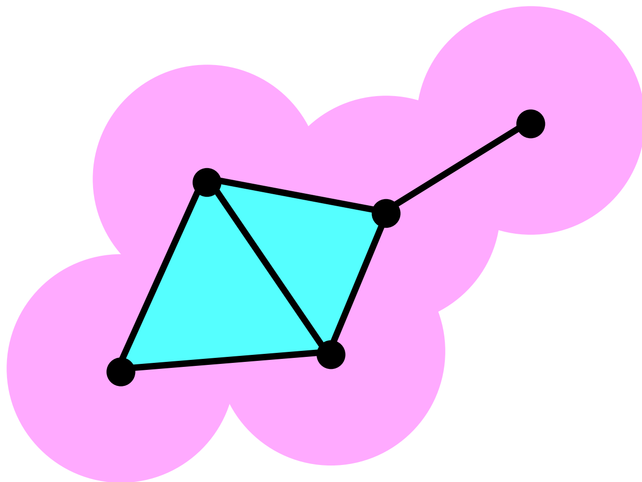
$$\sigma = [p_0 p_1 \cdots p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i,j \in \{0, \cdots k\},\ d(p_i, p_j) \leq \alpha$$

Theorem: [C-Oudot'08]
Let $X \subset \mathbb{R}^d$ be a compact set and $L \subset \mathbb{R}^d$ a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{9}$ wfs(X). Then for all $\alpha \in [2\varepsilon, \frac{1}{4}(\mathrm{wfs}(X) - \varepsilon)]$ and all $\lambda \in (0, \mathrm{wfs}(X))$, one has: $\forall k \in \mathbb{N}$

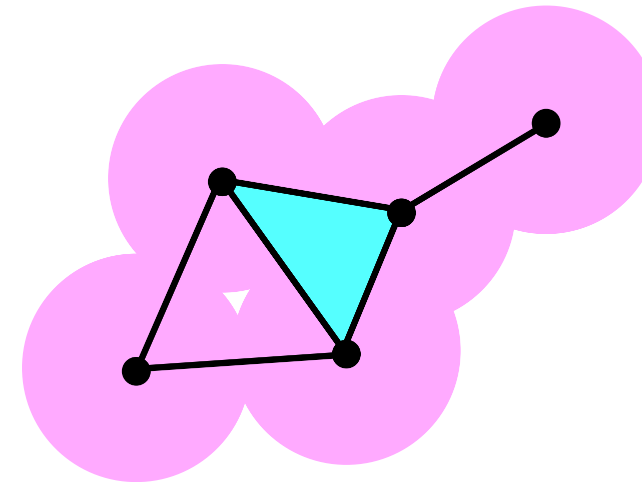
$$\beta_k(X^\lambda) = \dim(H_k(X^\lambda)) = \boxed{\mathrm{rk}(\mathcal{R}^\alpha(L) \to \mathcal{R}^{4\alpha}(L))}$$

Easy to compute using persistence algo.

Using the Rips complex



Rips vs Čech



The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \cdots p_k \in L$,

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

Theorem: [C-Oudot'08]
Let $X \subset \mathbb{R}^d$ be a compact set and $L \subset \mathbb{R}^d$ a finite set such that $d_H(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{9}$ wfs(X). Then for all $\alpha \in [2\varepsilon, \frac{1}{4}(\mathrm{wfs}(X) - \varepsilon)]$ and all $\lambda \in (0, \mathrm{wfs}(X))$, one has: $\forall k \in \mathbb{N}$

$$\beta_k(X^\lambda) = \dim(H_k(X^\lambda)) = \mathrm{rk}(\mathcal{R}^\alpha(L) \to \mathcal{R}^{4\alpha}(L))$$

→ **Pb:** Choice of α when wfs(X) is unknown?

Multiscale inference

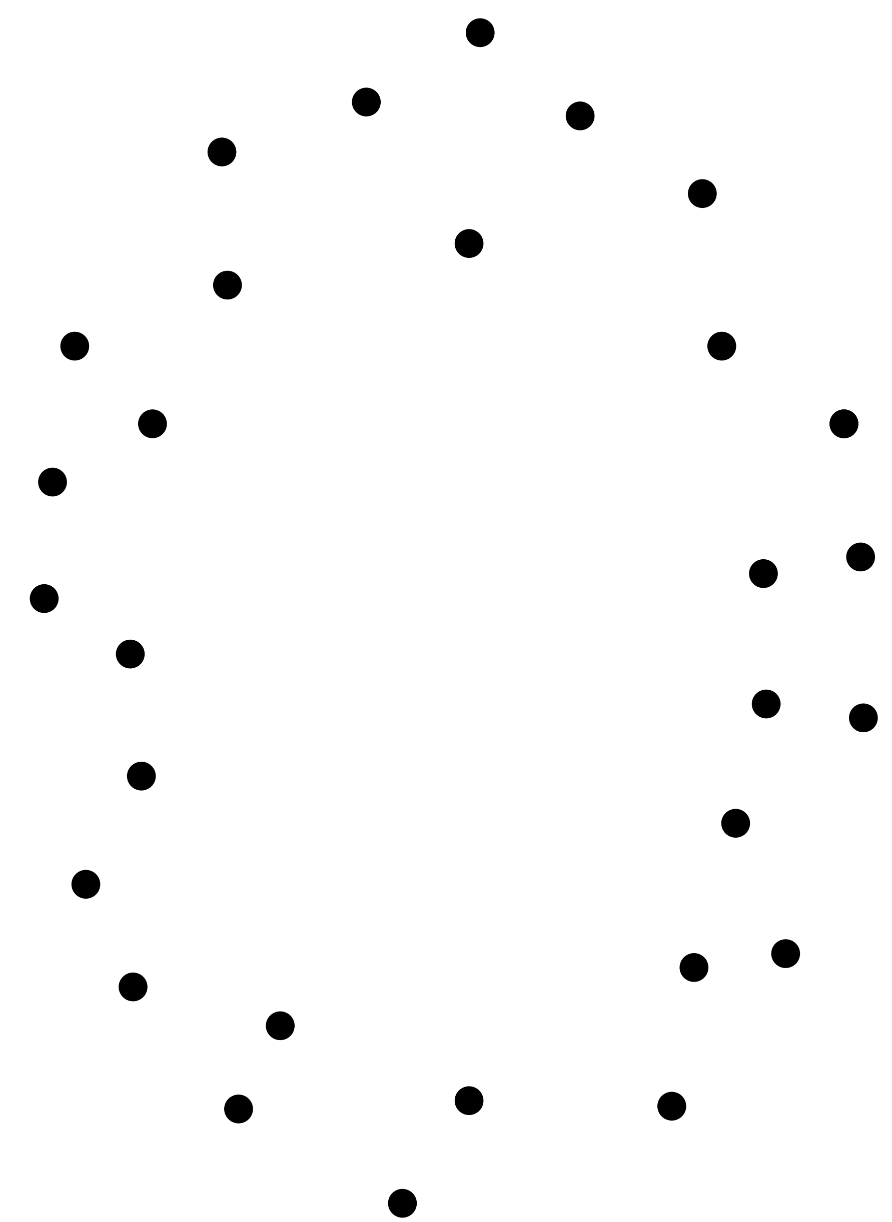
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w, L)$ in L
update $\varepsilon = \max_{w\in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers
of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

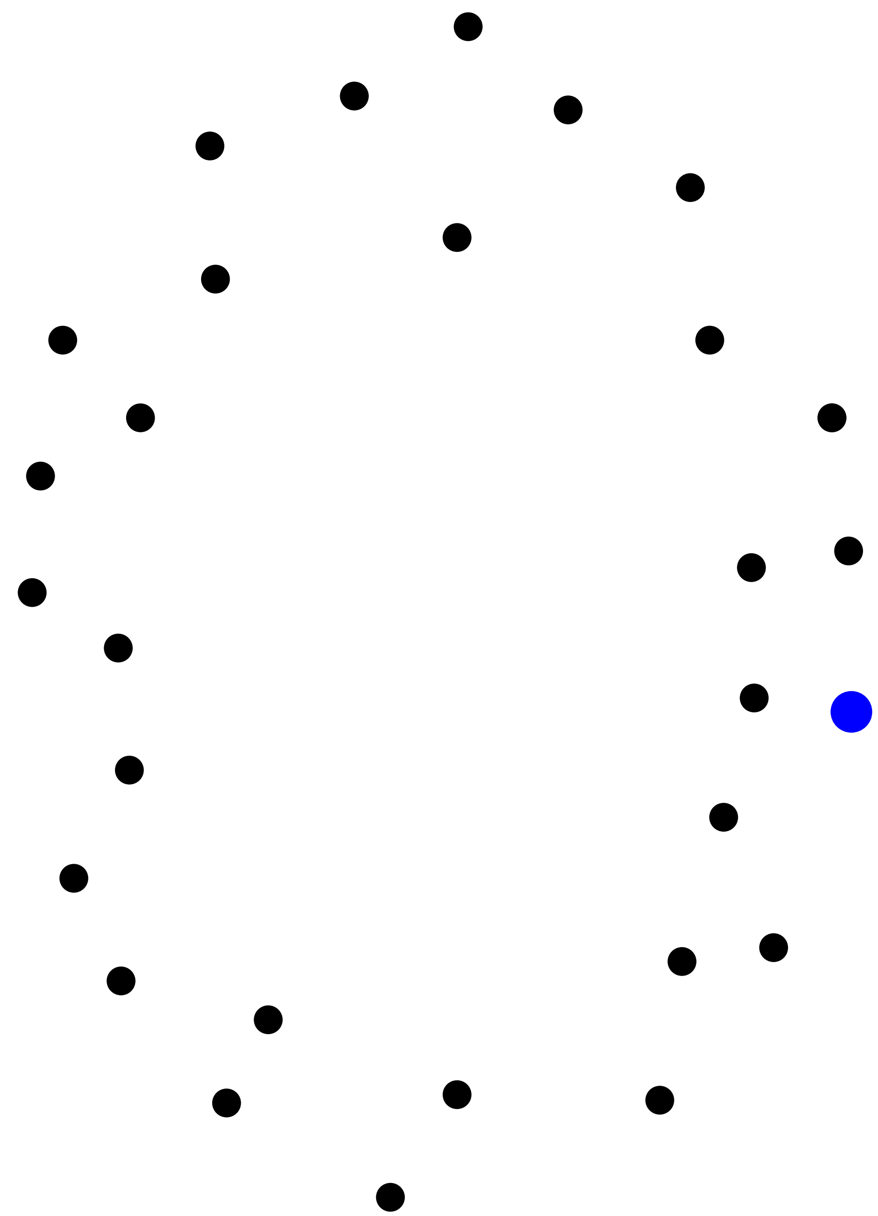
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w, L)$ in L
update $\varepsilon = \max_{w\in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers
of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

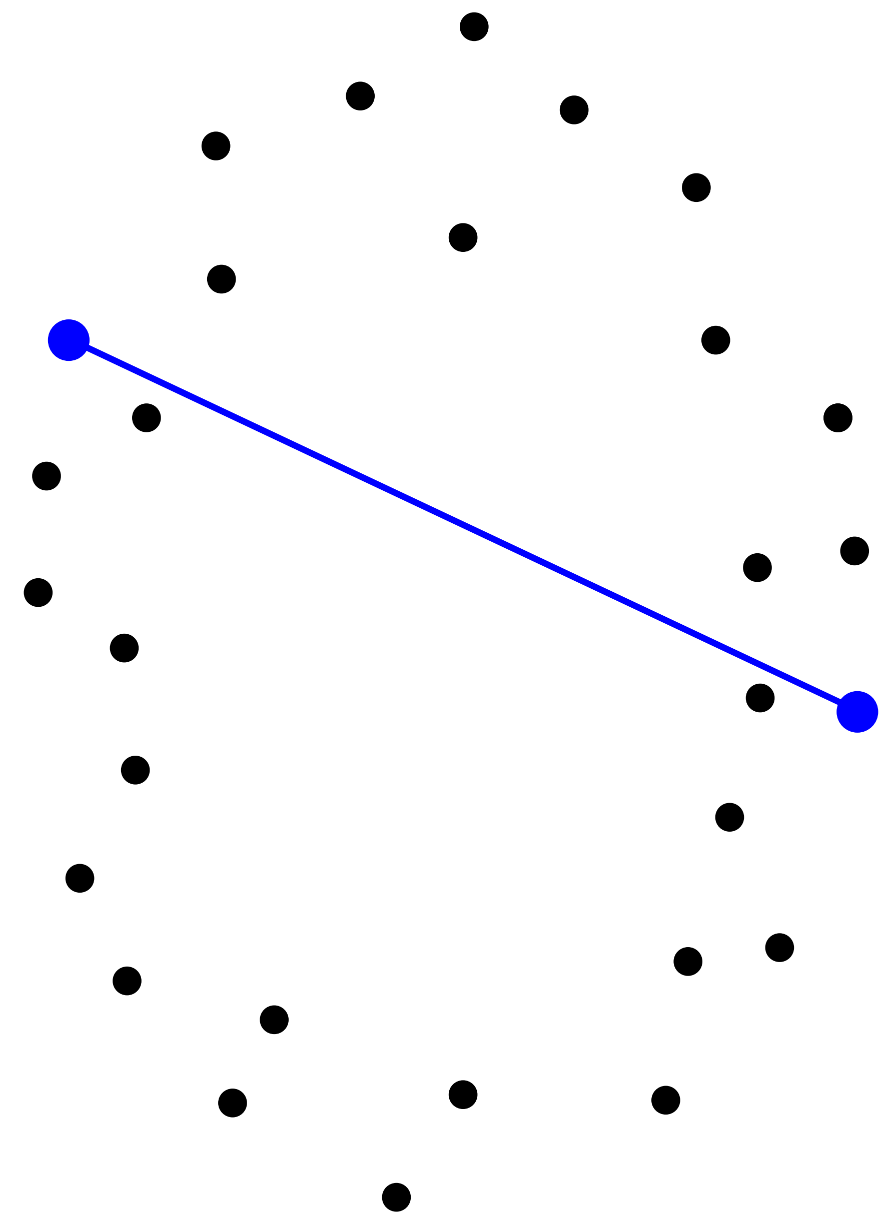
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w' \in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w \in W} d(w, L)$ in L
update $\varepsilon = \max_{w \in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

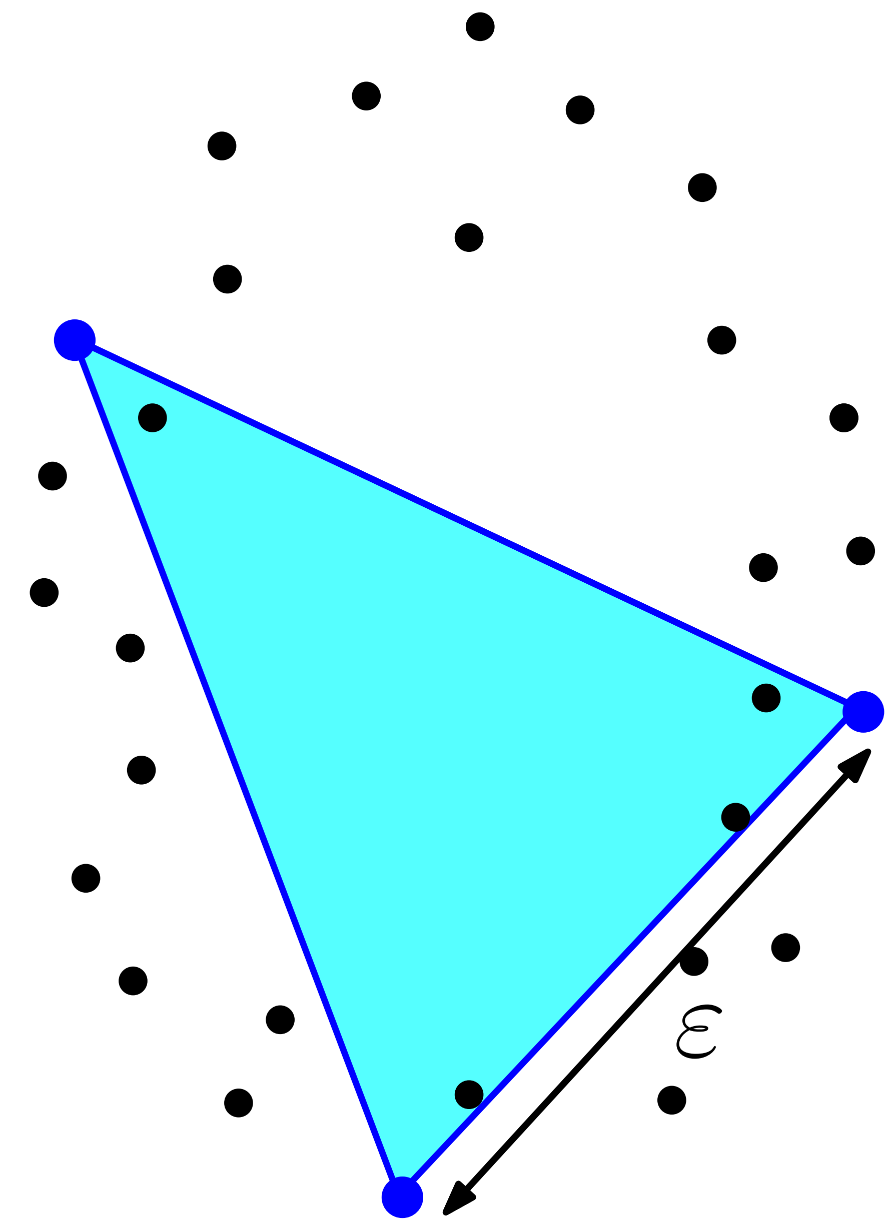
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w, L)$ in L
update $\varepsilon = \max_{w\in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



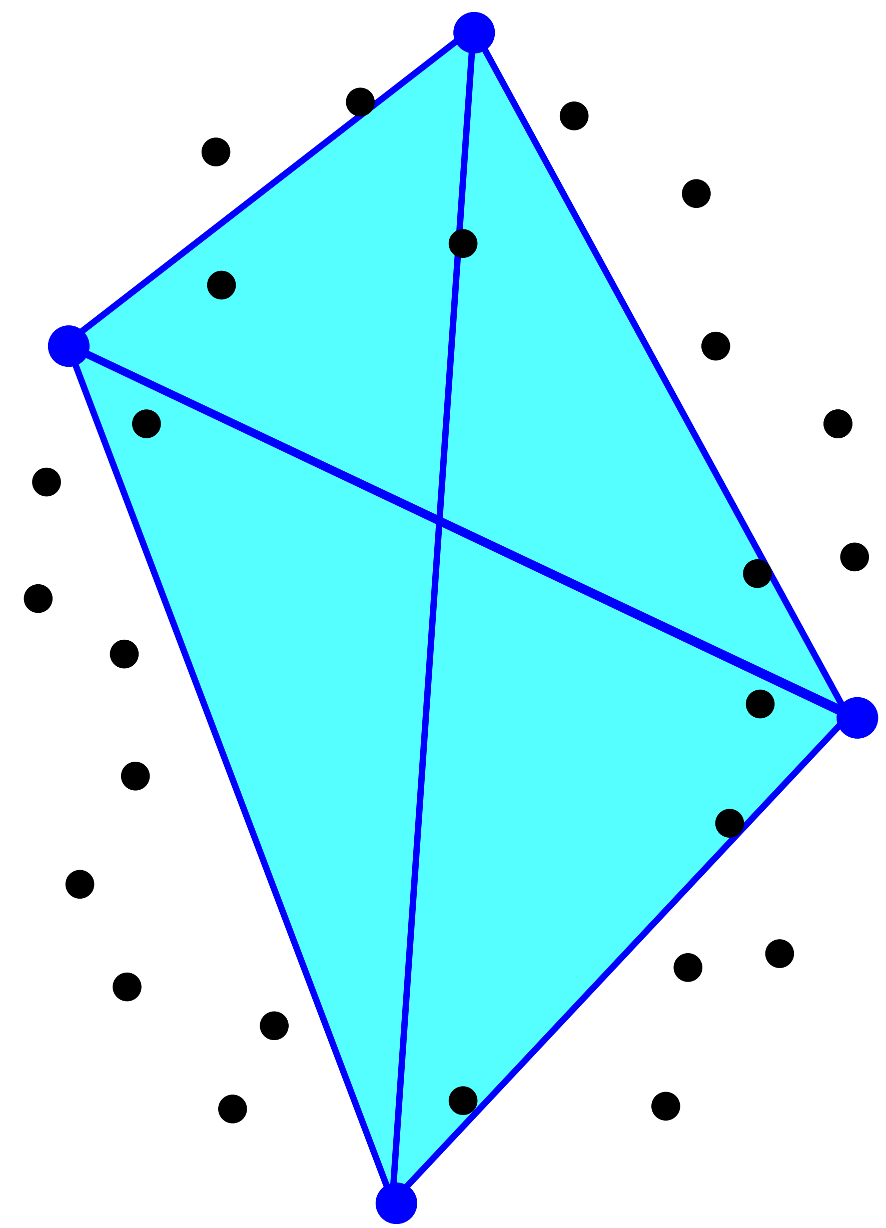
Multiscale inference

Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w' \in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w \in W} d(w, L)$ in L
update $\varepsilon = \max_{w \in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$) ← Rank of the map induced at homology level
END_WHILE

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

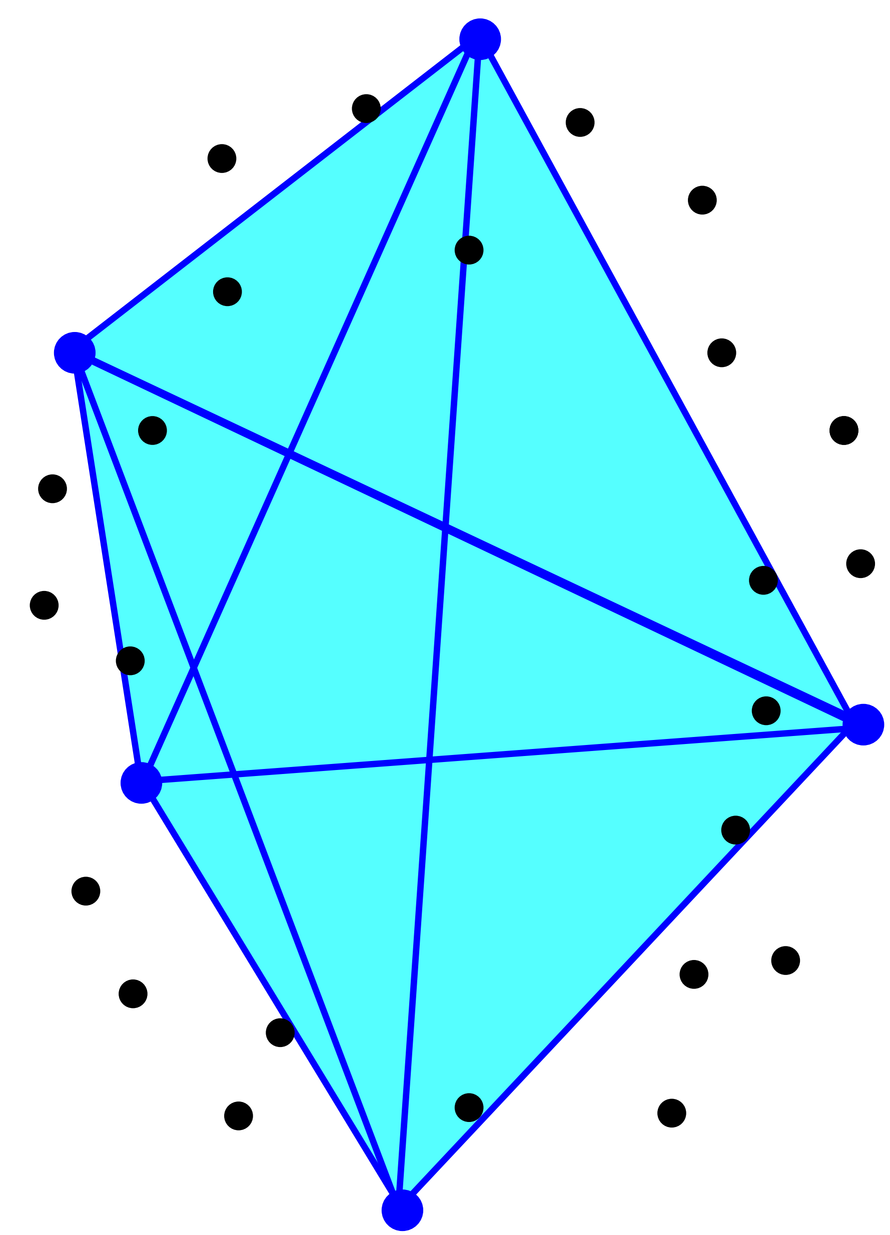
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L=\emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w,L)$ in L
update $\varepsilon = \max_{w\in W} d(w,L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

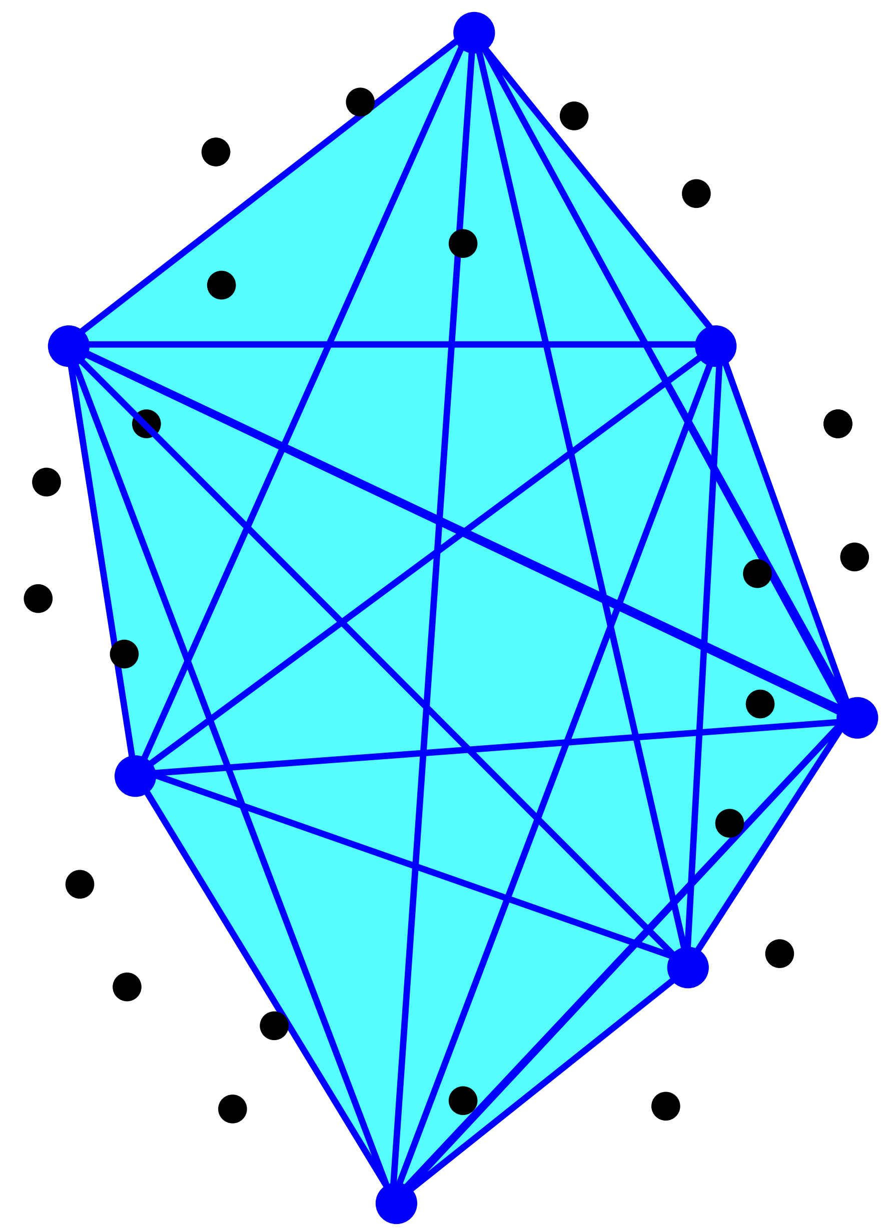
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w'\in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

WHILE $L \subset W$
insert $p = argmax_{w\in W} d(w,L)$ in L
update $\varepsilon = \max_{w\in W} d(w,L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference

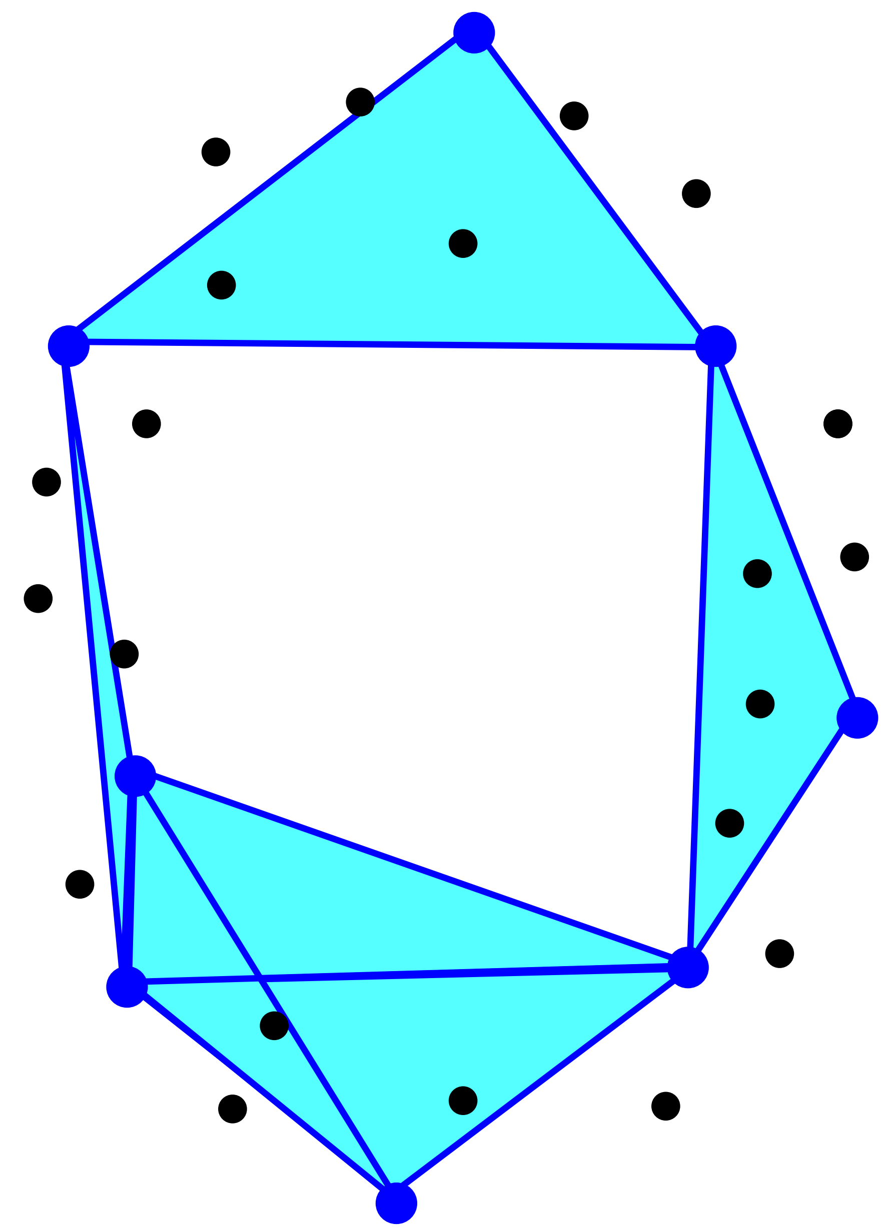
Input: A point cloud W and its pairewise distances $\{d(w,w')\}_{w,w' \in W}$.
$\rightarrow$ Maintain a nested pair $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

Init.: $L = \emptyset$; $\varepsilon = +\infty$

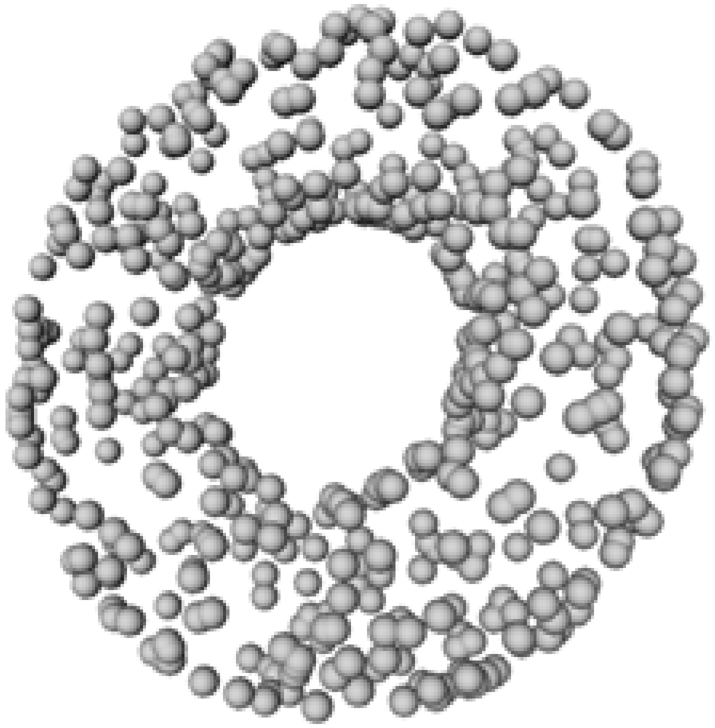
WHILE $L \subset W$
insert $p = argmax_{w \in W} d(w, L)$ in L
update $\varepsilon = \max_{w \in W} d(w, L)$
update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$
Persistence($\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$)
END_WHILE

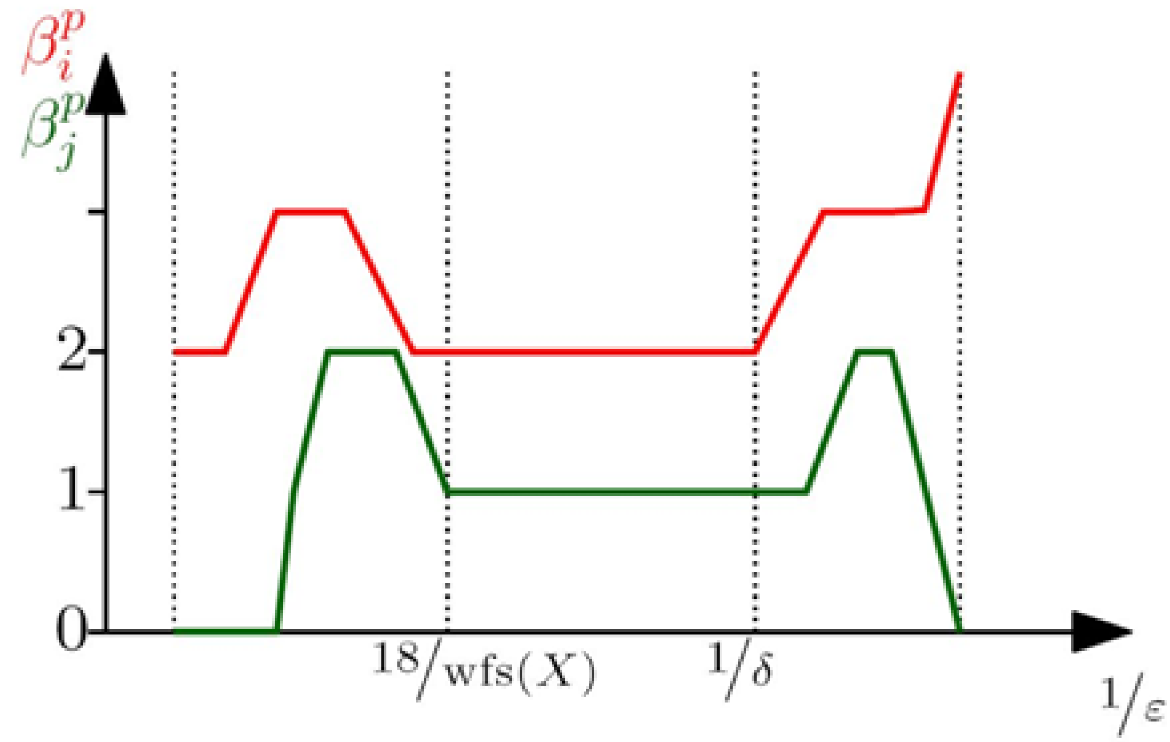
Rank of the map induced at homology level

Output: Sequence of persistent Betti numbers of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$



Multiscale inference



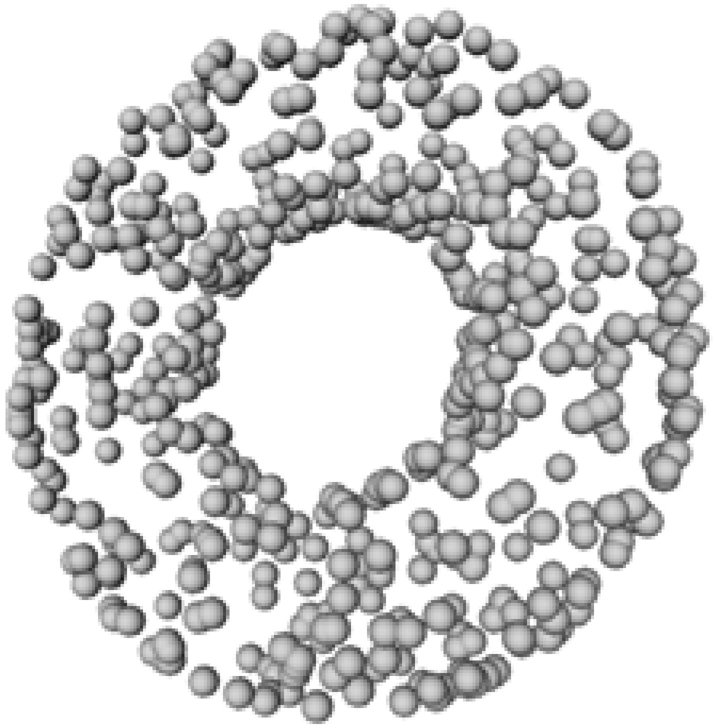


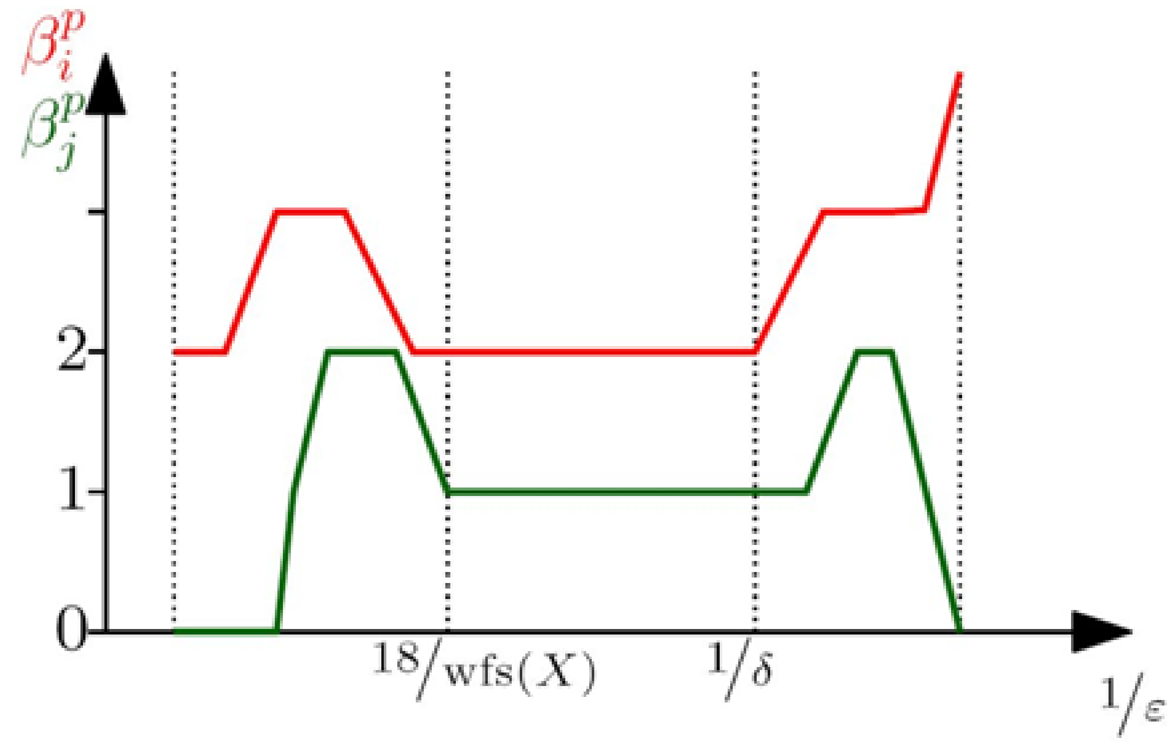
Theorem: [C-Oudot'08]
If $d_H(W, X) < \delta$ for $\delta < \frac{1}{18}\mathrm{wfs}(X)$, then at every iteration of the algorithm such that $\delta < \varepsilon < \frac{1}{18}\mathrm{wfs}(X)$,

$$\beta_k(X^\lambda) = \dim H_k(X^\lambda) = rk(H_k(\mathcal{R}^{4\varepsilon}(L)) \to H_k(\mathcal{R}^{4\varepsilon}(L)))$$

for any $\lambda \in (0, \mathrm{wfs}(X))$ and any $k \in \mathbb{N}$.

Multiscale inference





Complexity of the algorithm:

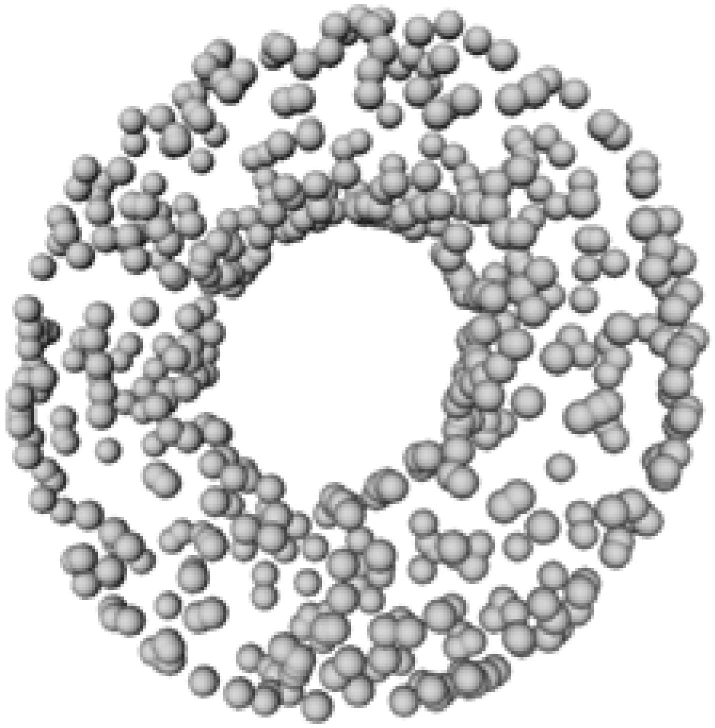
- If $X \subset \mathbb{R}^d$ is non smooth the running time of the algorithm is

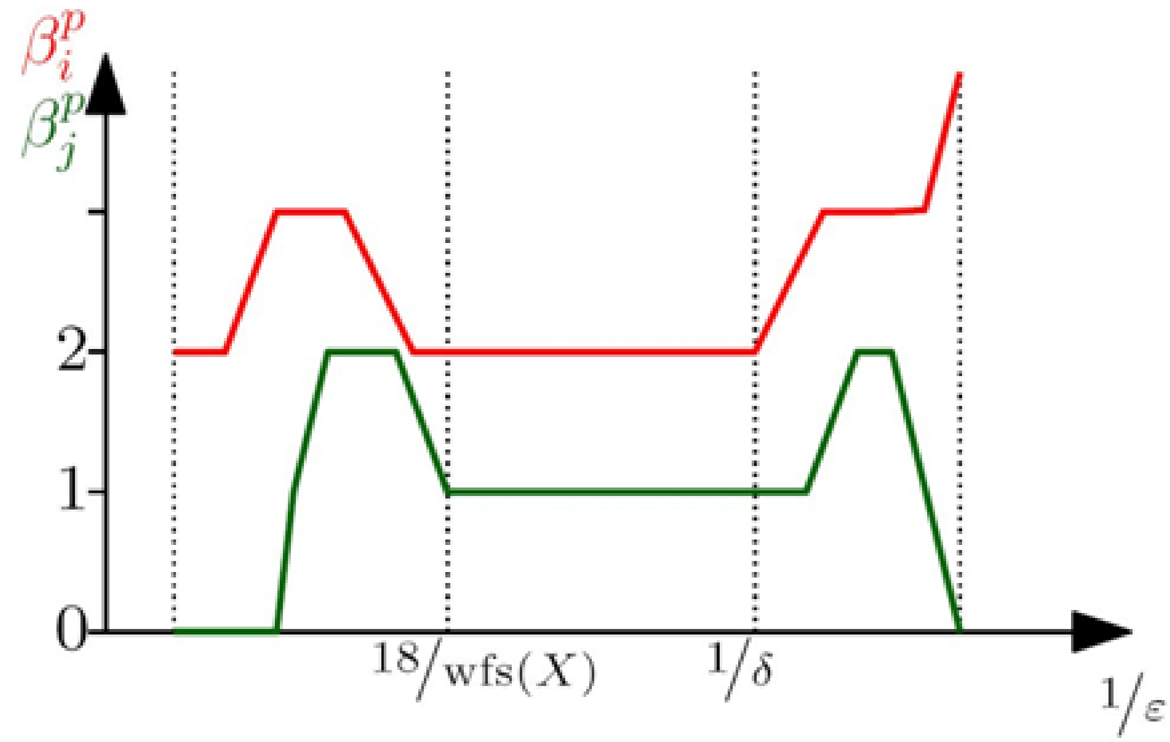
$$O(8^{33^d}|W|^5)$$

- If X is a smooth submanifold of $\mathbb{R}^d$ dimension m the running time is

$$O(8^{35^m}|W|)$$

Multiscale inference





Complexity of the algorithm:

- If $X \subset \mathbb{R}^d$ is non smooth the running time of the algorithm is

$$O(8^{33^d}|W|^5)$$

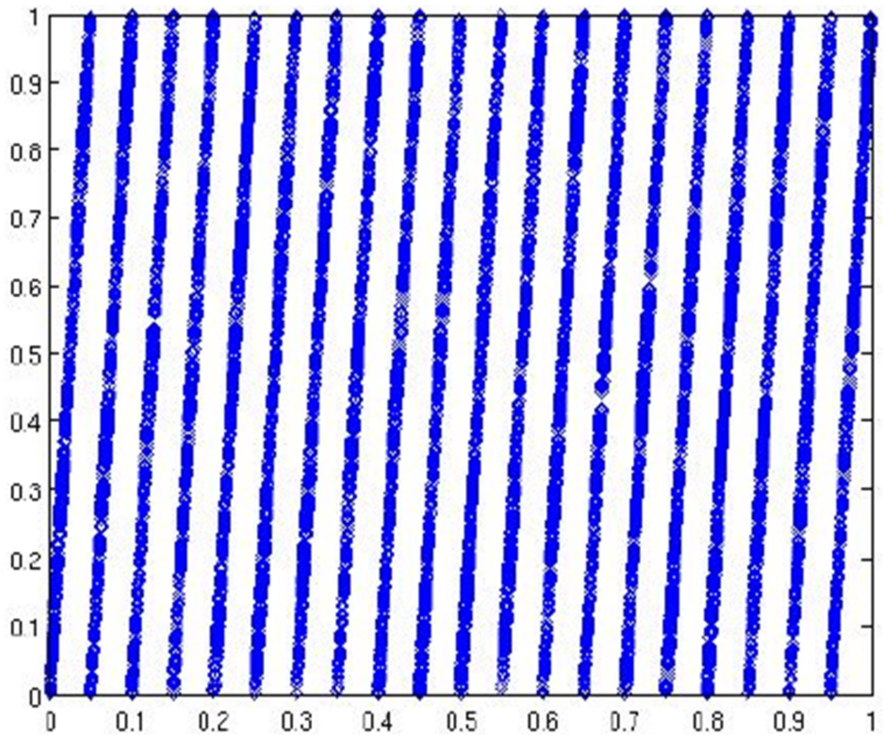
- If X is a smooth submanifold of $\mathbb{R}^d$ dimension m the running time is

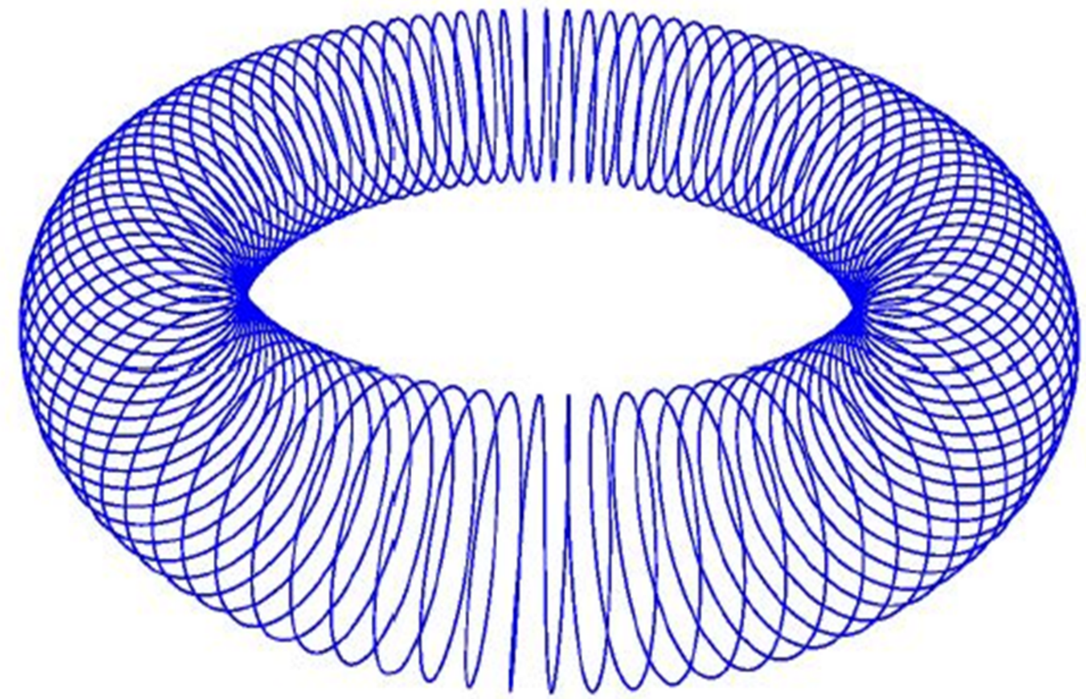
$$O(8^{35^m}|W|)$$

Depend on the intrinsic dimension of X

A synthetic example

$[0,1] \times [0,1]$ $\mathbb{R}^{1000}$

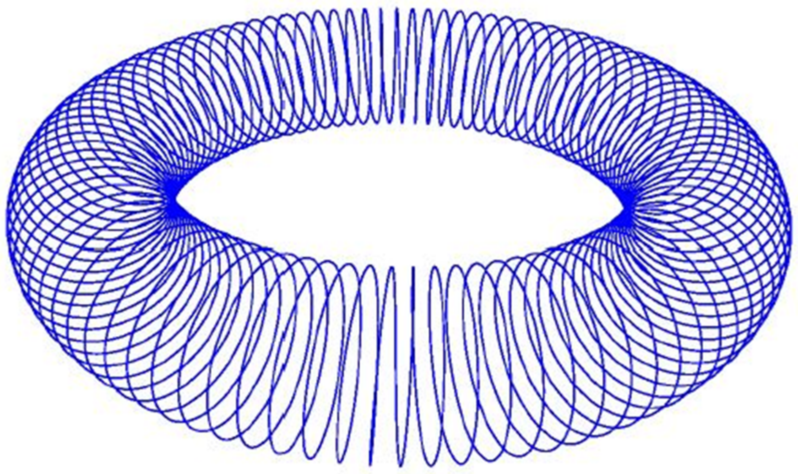


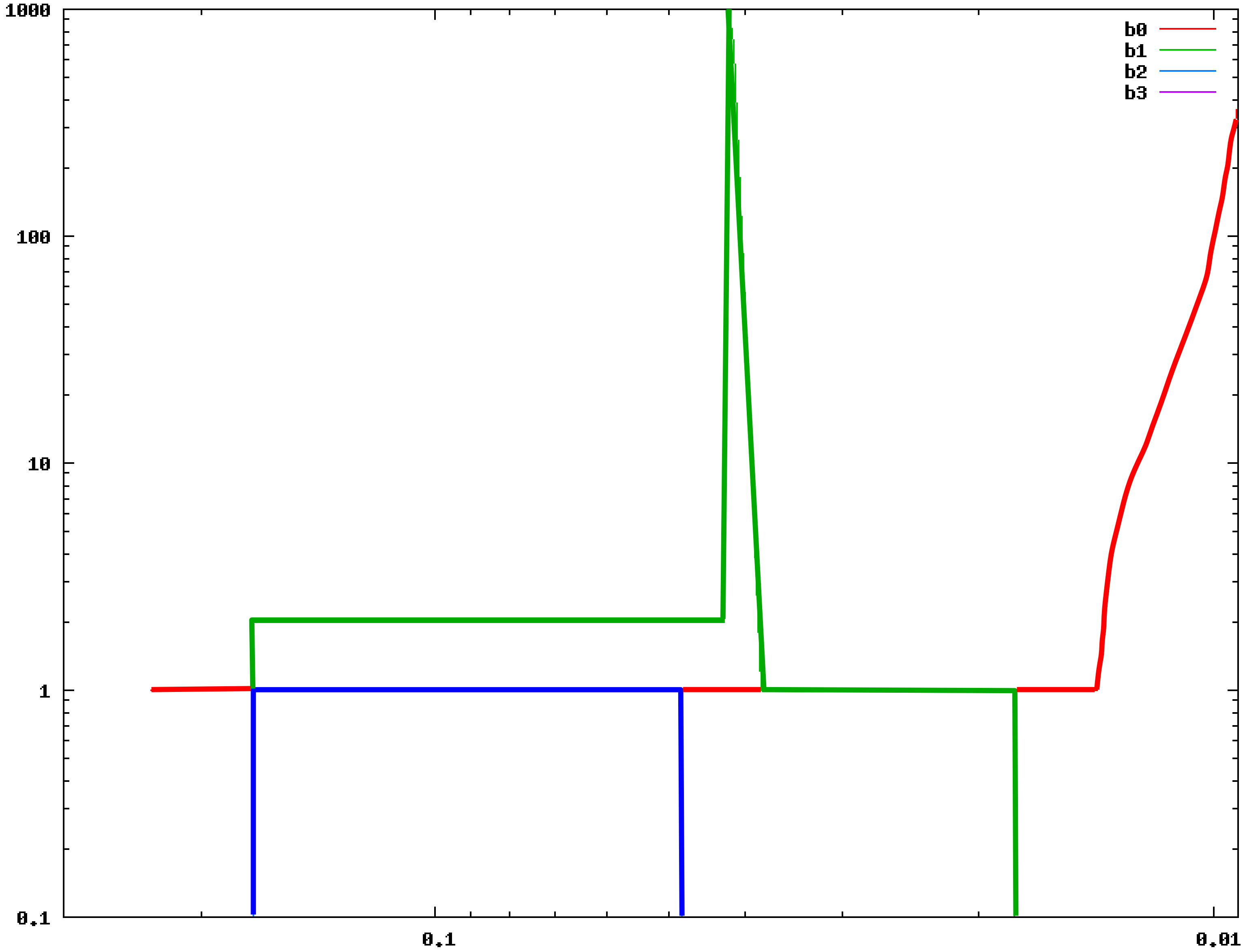


Non-linear embedding of $\mathbb{S}^1 \times \mathbb{S}^1$ in $\mathbb{R}^{1000}$

$50,000$ points sampled uniformly at random from a curve drawn on the 2-torus $\mathbb{S}^1 \times \mathbb{S}^1$.

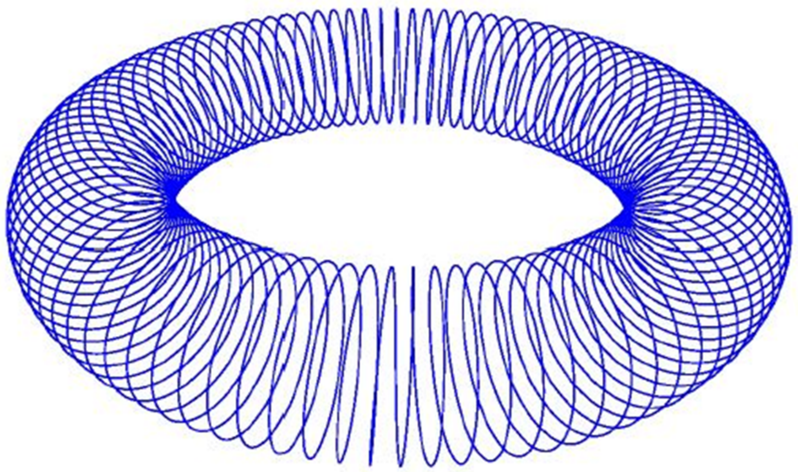
A synthetic example

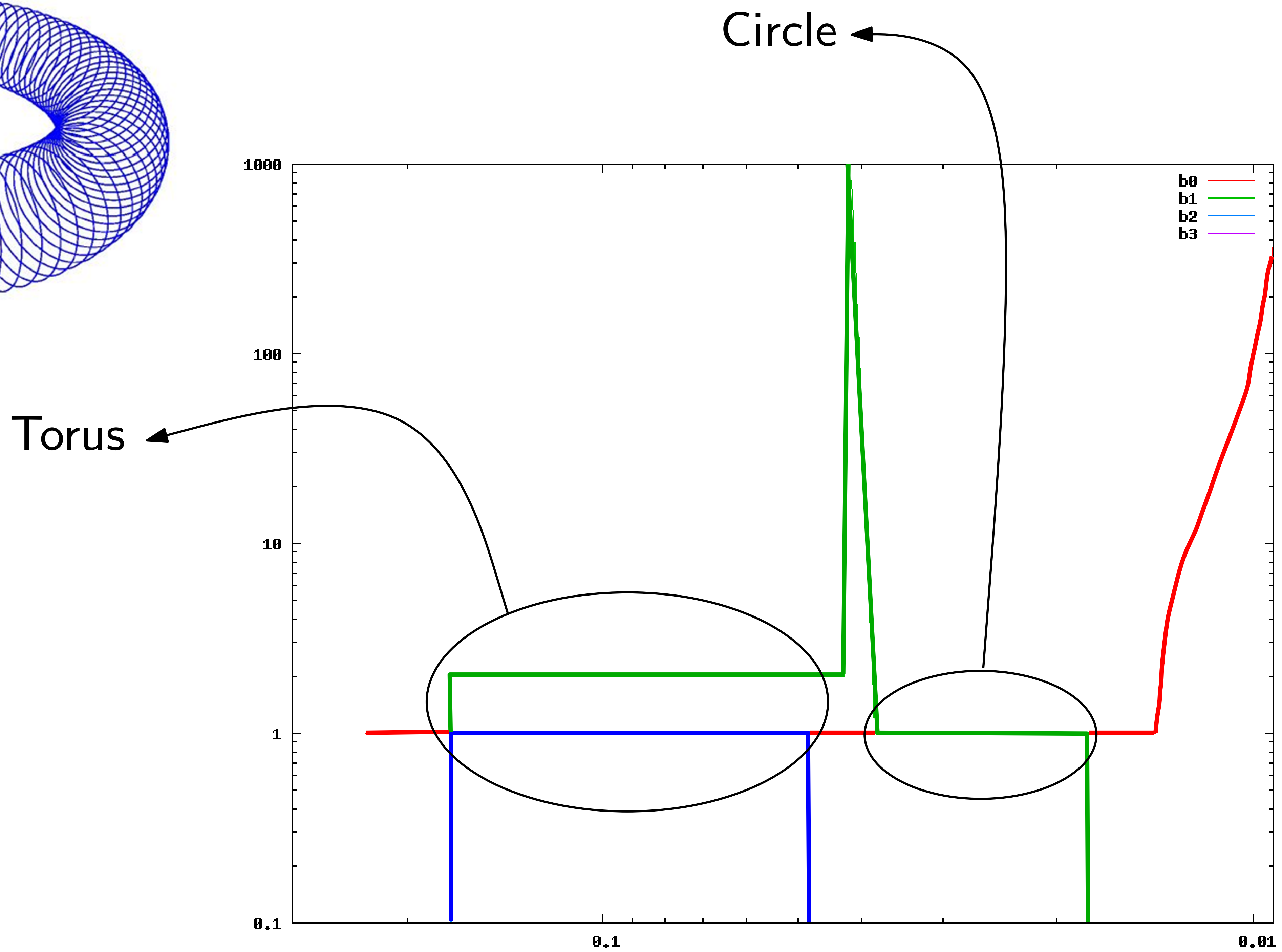




Output: sequence of Betti numbers on a log-log scale

A synthetic example





Output: sequence of Betti numbers on a log-log scale

An algorithm to compute Betti numbers

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.

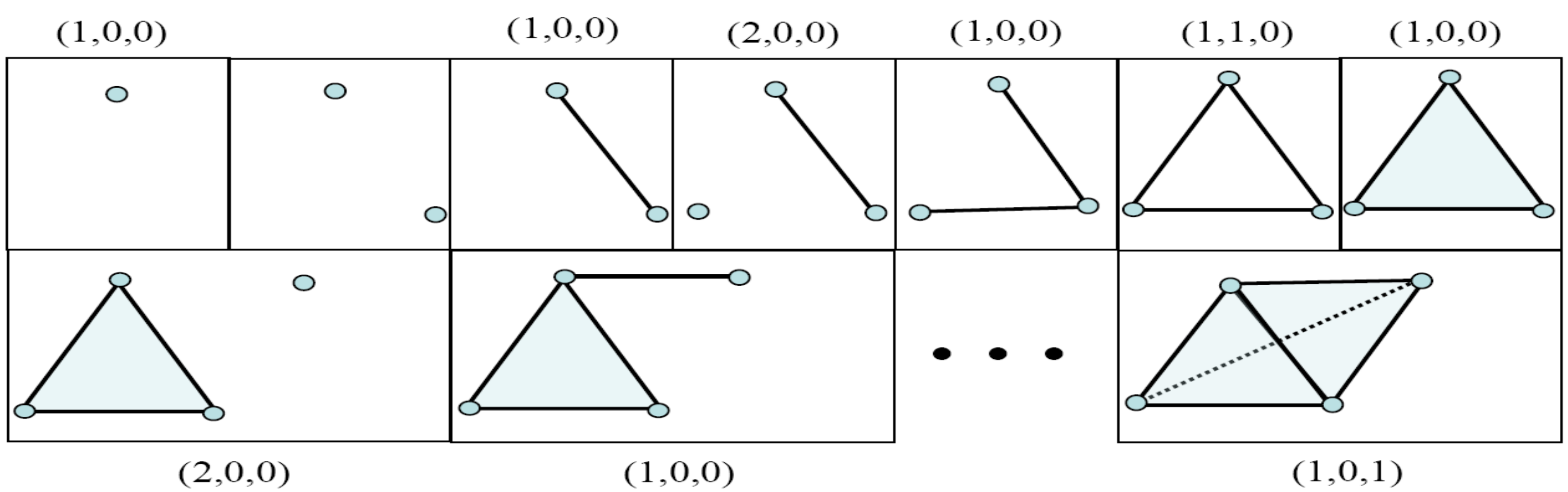
> $\beta_0 = \beta_1 = \cdots = \beta_d = 0$;
> for $i = 1$ to m
> $k = \dim \sigma^i - 1$;
>   if σ^i is contained in a $(k+1)$-cycle in K^i
>   then $\beta_{k+1} = \beta_{k+1} + 1$;
>   else $\beta_k = \beta_k - 1$;
>   end if;
> end for;
> output $(\beta_0, \beta_1, \cdots, \beta_d)$;

An algorithm to compute Betti numbers

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



An algorithm to compute Betti numbers

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

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}

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

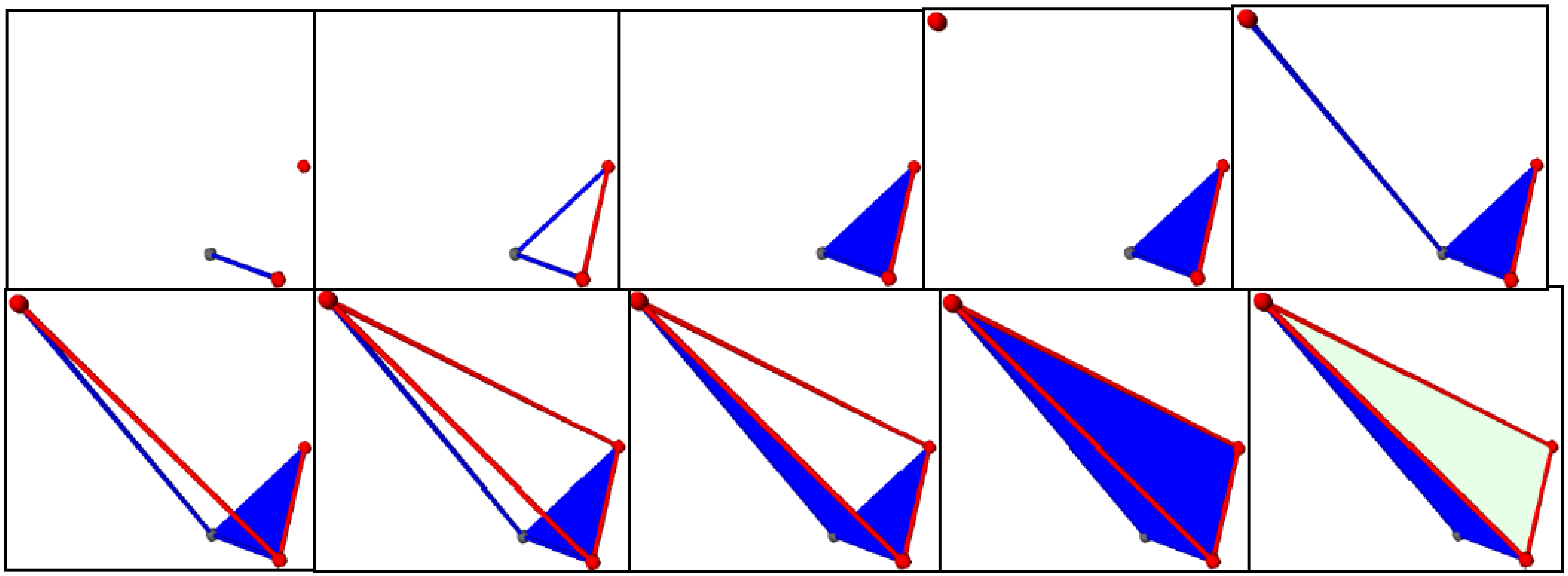
- If σ^i is not contained in a $(k+1)$-cycle c in K^i, then $\partial\sigma^i$ is not a boundary in K^{i-1}

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

- the previous inequalities are equalities.

Positive and negative simplices

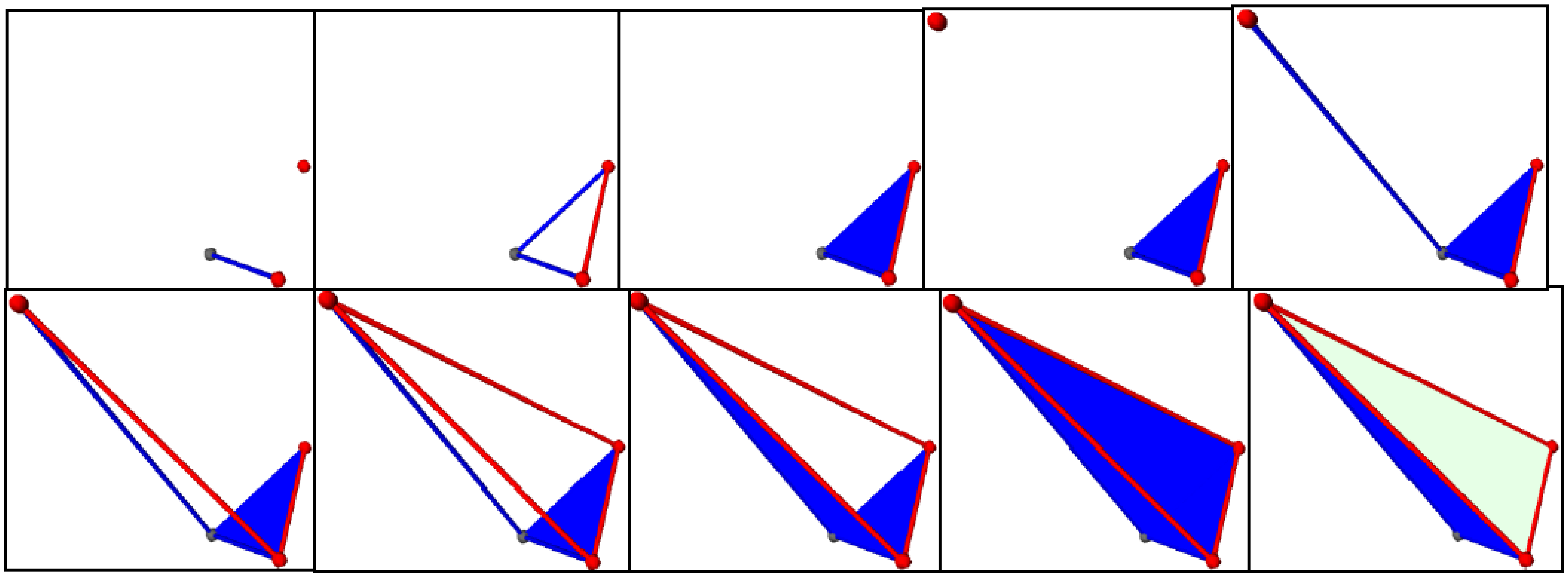
Let $\emptyset = K^0 \subset K^1 \subset \cdots \subset K^m = K$ be a filtration of a simplicial complex K s. t. $K^{i+1} = K^i \cup \sigma^{i+1}$ where σ^{i+1} is a simplex of K.



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.

Positive and negative simplices

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.

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

negative → 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

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

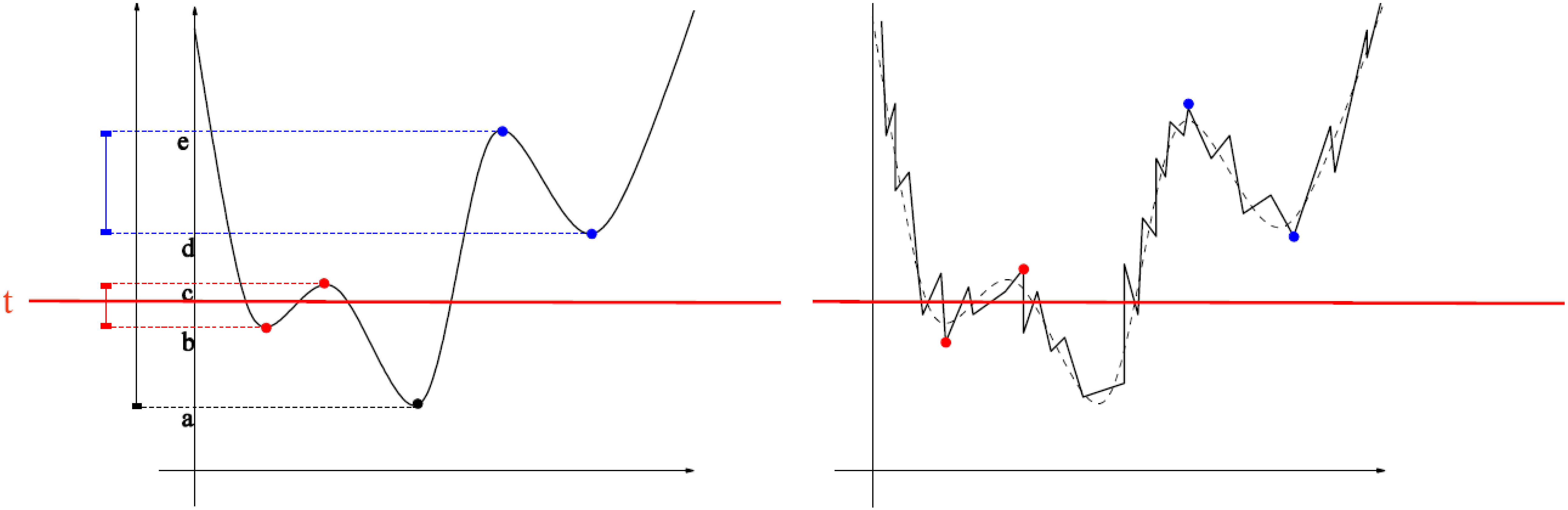
- How to keep track of the evolution of the topology all along the filtration?
- What are the created/destroyed cycles?
- What is the lifetime of a cycle?
- How to compute rank$(H_k(K^i) \to H_k(K^j))$?

This is where topological persistence comes into play!

Topological persistence

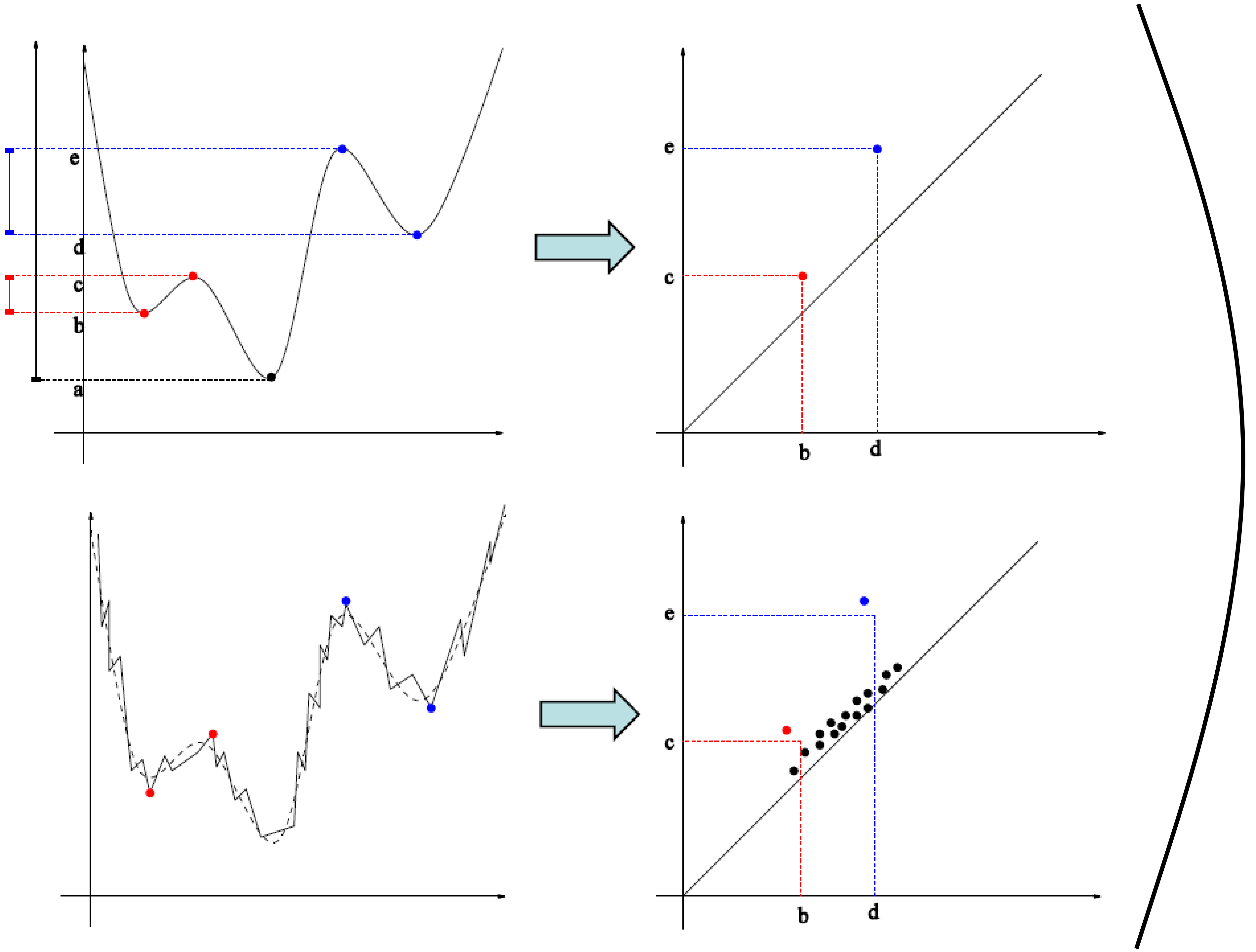
- a tool to study topological properties of data (represented by real valued functions on topological spaces).
- A method that allow to separate information from topological noise.
- References:
  - H. Edelsbrunner, D. Letscher and A. Zomorodian. *Topological persistence and simplification*. Discrete Comput. Geom., 28:511-533, 2002.
  - D. Cohen-Steiner and H. Edelsbrunner and J. Harer, *Stability of Persistence Diagrams*, Proc. 21st ACM Sympos. Comput. Geom. 2005.
  - F. Chazal and D. Cohen-Steiner and L. J. Guibas and M. Glisse and S. Y. Oudot, *Proximity of Persistence Modules and their Diagrams*, Proc. 25th ACM Sympos. Comput. Geom. 2009.

A simple example



- What is the relevant number of connected components of $f^{-1}((-\infty, t])$?
- More generally, study the topology of the sublevel sets $f^{-1}((-\infty, t])$ as t varies.

A simple example: filter out topological noise



Persistence diagrams

Functions defined over higher dimensional spaces

- $f : X \to \mathbb{R}$ continuous where X is a topological space
- Not only connected components but also cycles, voids, etc... $\to$ persistence of homological features / evolution of $H_k(f^{-1}((-\infty, t]))$

Relation between fonctions and filtrations:

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

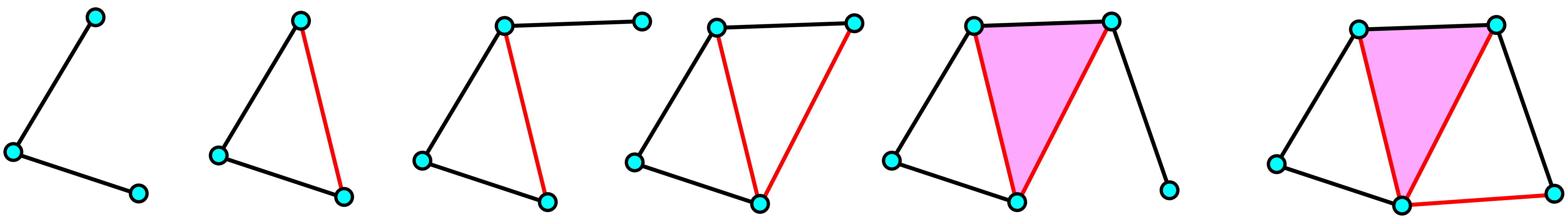
- For $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\cdots,k} f(v_i)$
- The simplices of K are ordered according increasing f values (and dimension in case of equal values on different simplices).

Notations

In the following:

- 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.
- $Z_k^i =$ the k-cylcles of K^i, $B_k^i =$ the k-boundaries of K^i and $H_k^i =$ the k^{th}-homology group of K^i.
- $Z_k^0 \subseteq Z_k^1 \subseteq \cdots \subseteq Z_k^i \subseteq \cdots \subseteq Z_k^m = Z_k(K)$
- $B_k^0 \subseteq B_k^1 \subseteq \cdots \subseteq B_k^i \subseteq \cdots \subseteq B_k^m = B_k(K)$

Cycle associated to a positive simplex

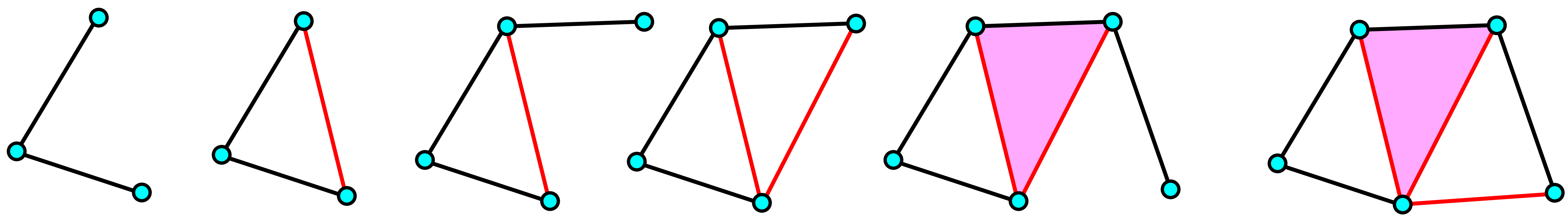


Lemma: If σ^i is a positive k-cycle, then there exists a k-cycle c_σ s.t.:
- c_σ is not a boundary in K^i,
- c_σ contains σ^i but no other positive k-simplex.

The cycle c^σ is unique.

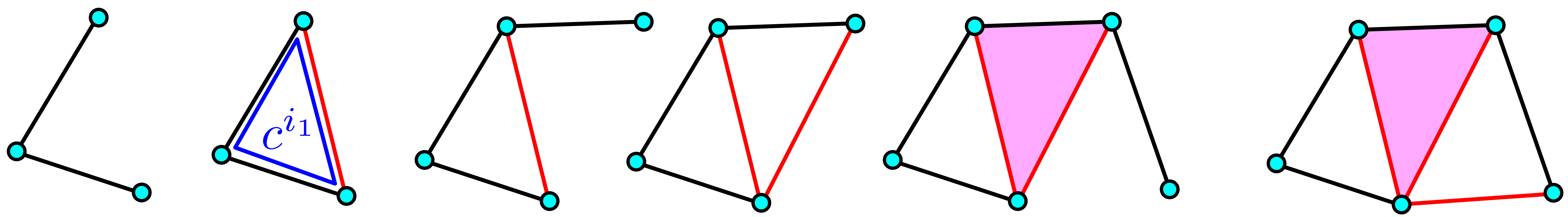
Proof:
By induction on the order of appearence of the simplices in the filtration.

Homology basis



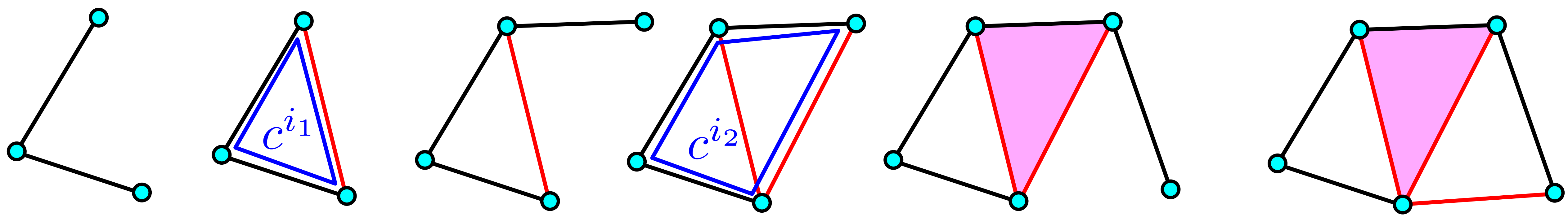
- At the beginning: the basis of H_k^0 is empty.
- If a basis of H_k^{i-1} has been built and σ^i is a positive k-simplex then one adds the homology class of the cycle c^i associated to σ^i to the basis of $H_k^{i-1} \Rightarrow$ basis of H_k^i.
- If a basis of H_k^{j-1} has been built and σ^j is a negative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

Homology basis



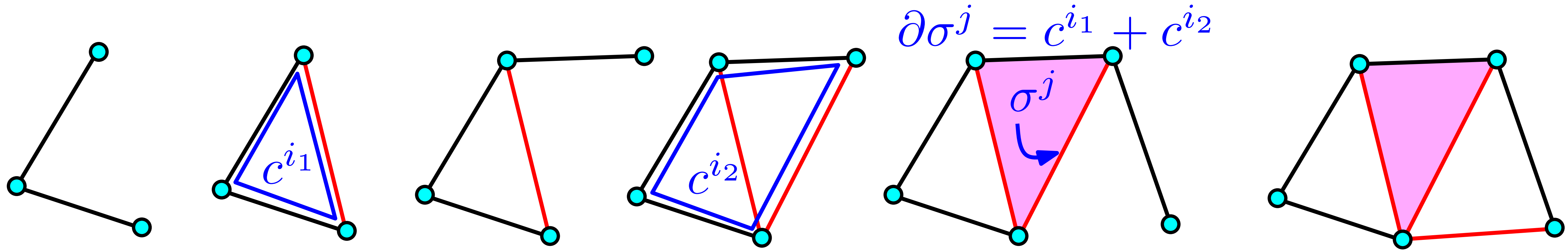
- At the beginning: the basis of H_k^0 is empty.
- If a basis of H_k^{i-1} has been built and σ^i is a positive k-simplex then one adds the homology class of the cycle c^i associated to σ^i to the basis of $H_k^{i-1} \Rightarrow$ basis of H_k^i.
- If a basis of H_k^{j-1} has been built and σ^j is a negative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

Homology basis



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

Homology basis



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

Pairing simplices

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

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

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

The persistence algorithm: first version

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

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

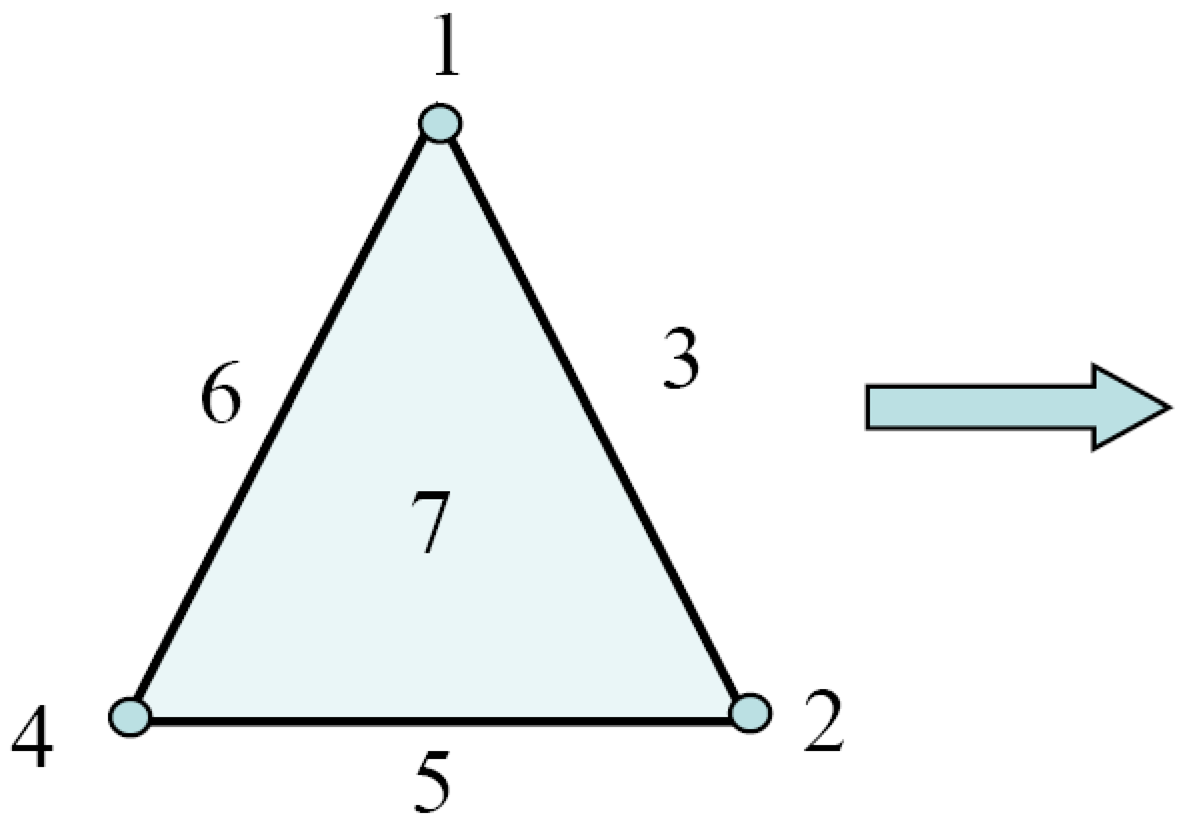
The persistence algorithm: first version

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

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

How to test this condition?

The matrix of the boundary operator



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

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

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

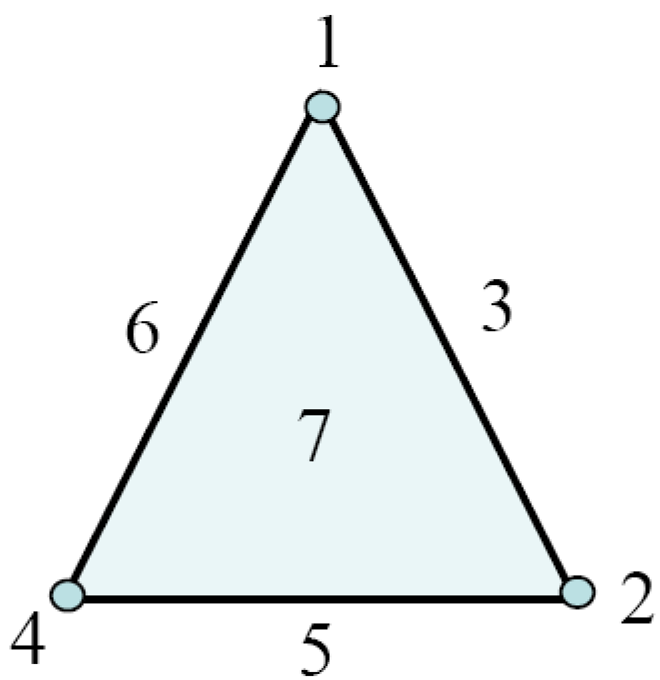
The persistence algorithm: second version

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.

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.

A very simple example



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

$C_5 + C_6$

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

$C_3 + C_6$

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

Pairs: $(2,3)$ $(4,5)$ $(7,7)$

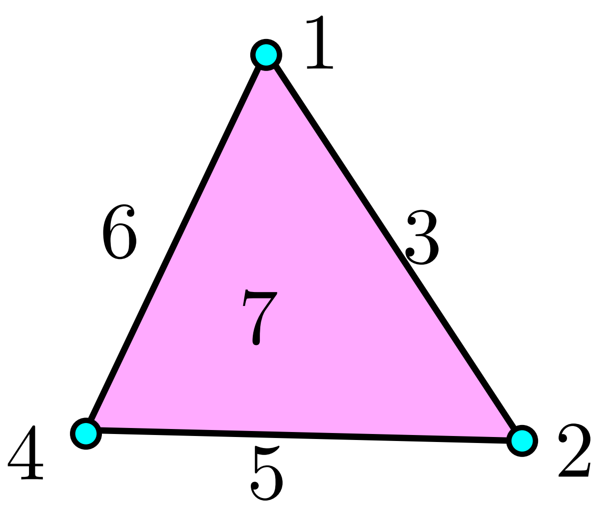
Correctness of the second algorithm

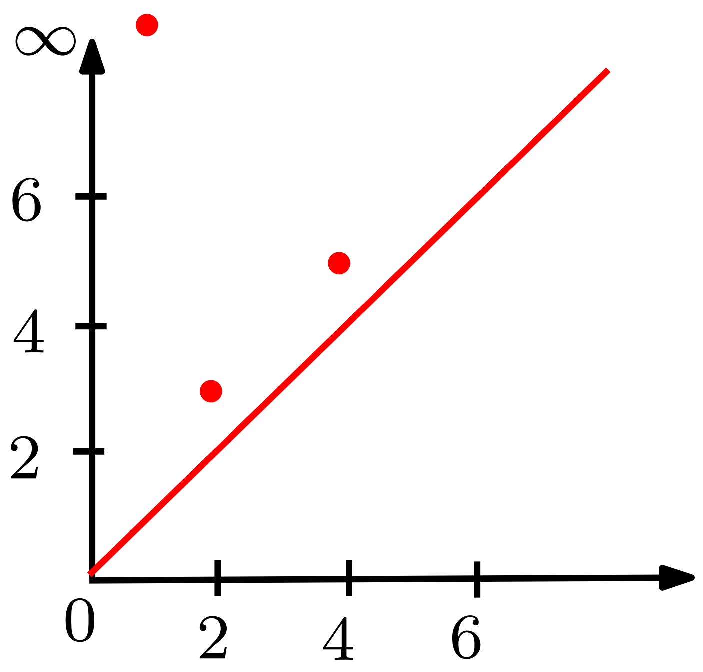
Proposition: the second algorithm outputs the persistence pairs.

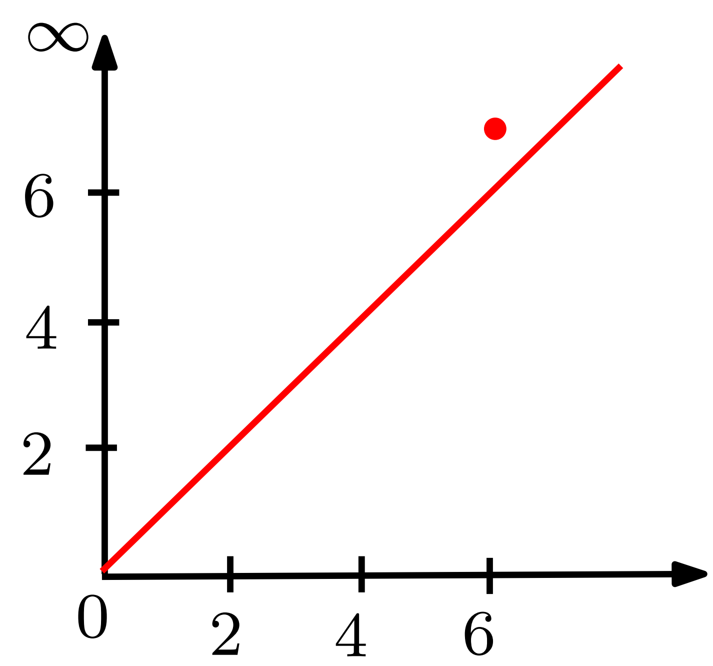
Proof: follows from the four remarks below.

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

Persistence diagrams





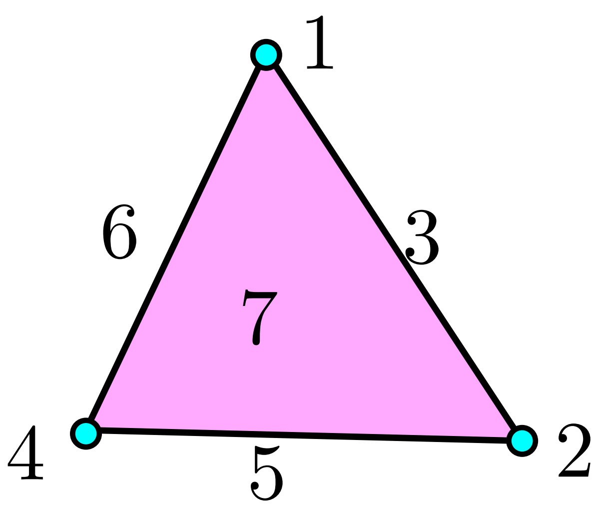


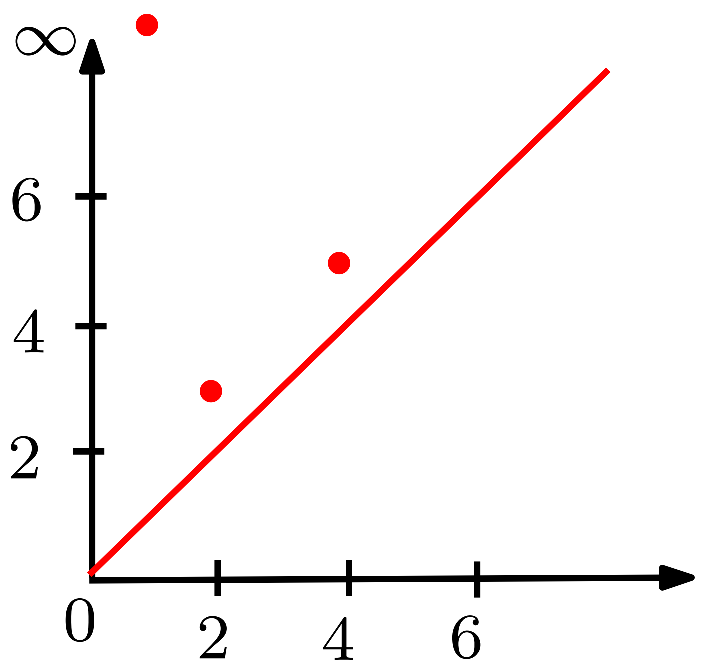
- each pair $(\sigma^{l(j)}, \sigma^j)$ is represented by $(l(j), j)$ or $(f(\sigma^{l(j)}), f(\sigma^j)) \in \mathbb{R}^2$ when considering filtrations induced by functions.

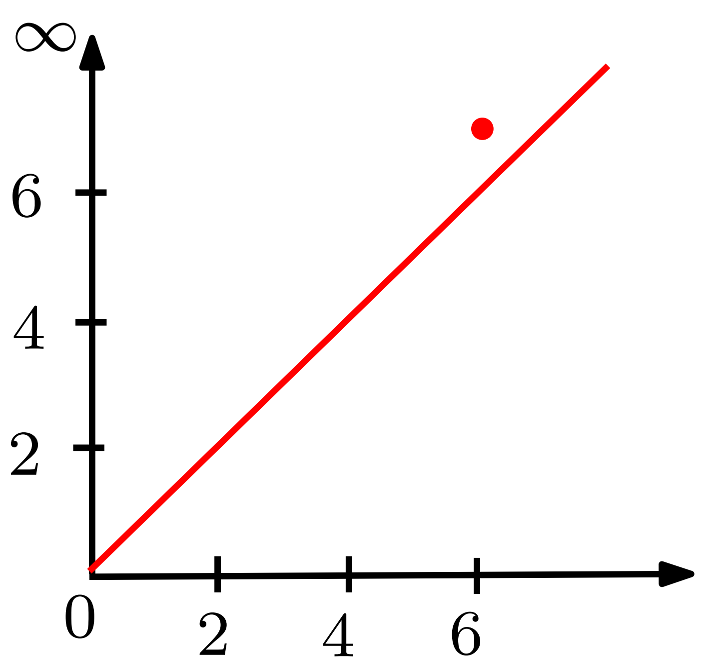
- The diagonal $\{y = x\}$ is added to the persistence diagram.

- Unpaired positive simplex $\sigma^i \to (i, +\infty)$.

Persistence diagrams



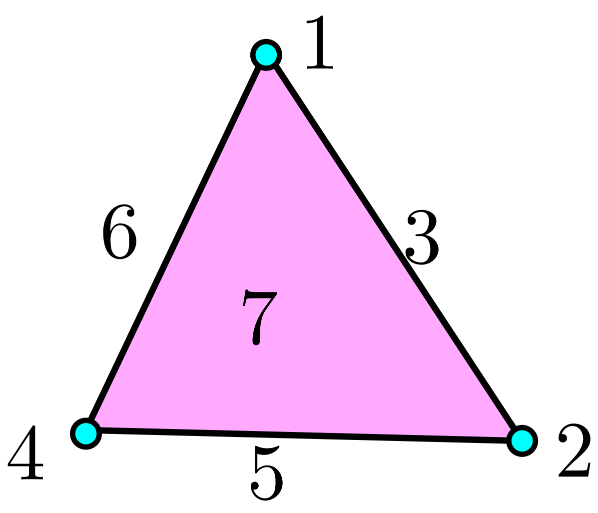


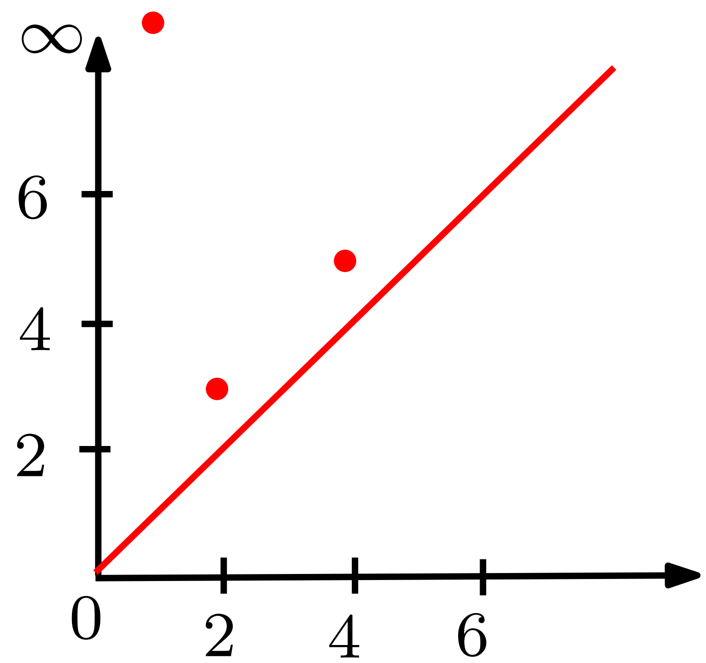


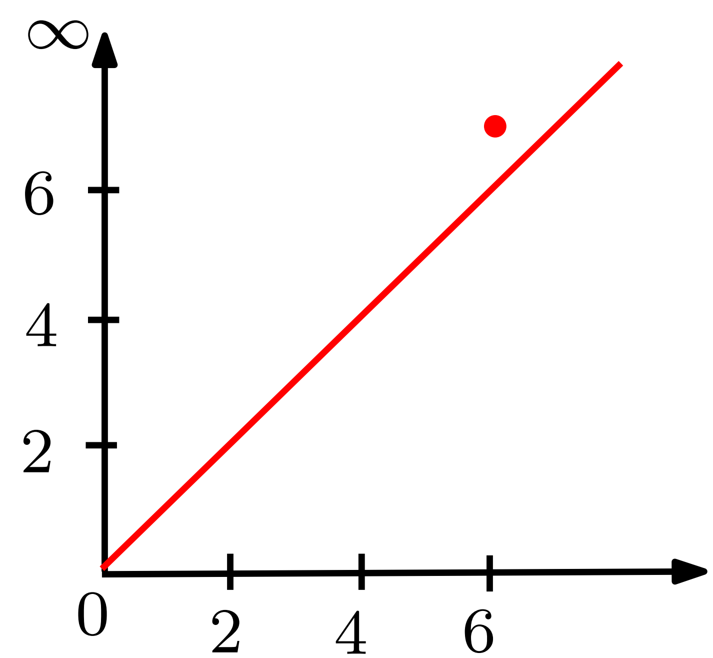
- each pair $(\sigma^{l(j)}, \sigma^j)$ is represented by $(l(j), j)$ or $(f(\sigma^{l(j)}), f(\sigma^j)) \in \mathbb{R}^2$ when considering filtrations induced by functions.
- The diagonal $\{y = x\}$ is added to the persistence diagram.
- Unpaired positive simplex $\sigma^i \to (i, +\infty)$.

Warning: in this case, points may have multiplicity.

Persistence diagrams

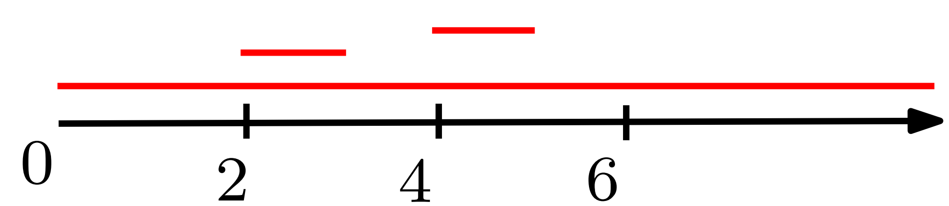




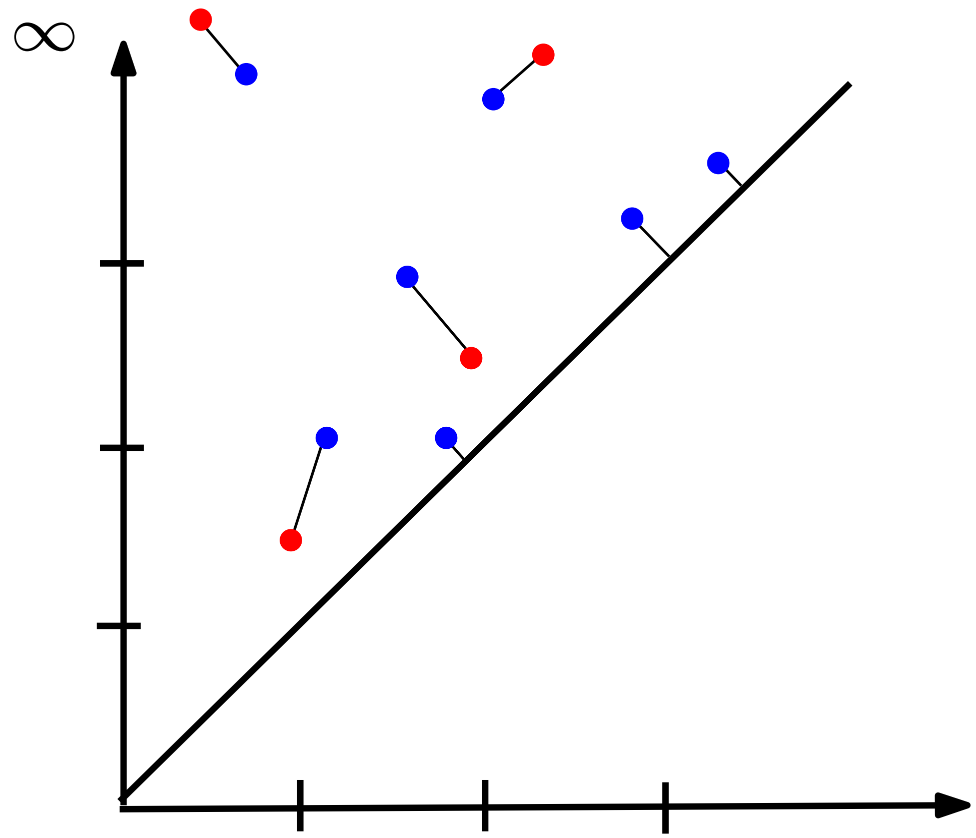


- each pair $(\sigma^{l(j)}, \sigma^j)$ is represented by $(l(j), j)$ or $(f(\sigma^{l(j)}), f(\sigma^j)) \in \mathbb{R}^2$ when considering filtrations induced by functions.
- The diagonal $\{y = x\}$ is added to the persistence diagram.
- Unpaired positive simplex $\sigma^i \to (i, +\infty)$.

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



Distance between persistence diagrams



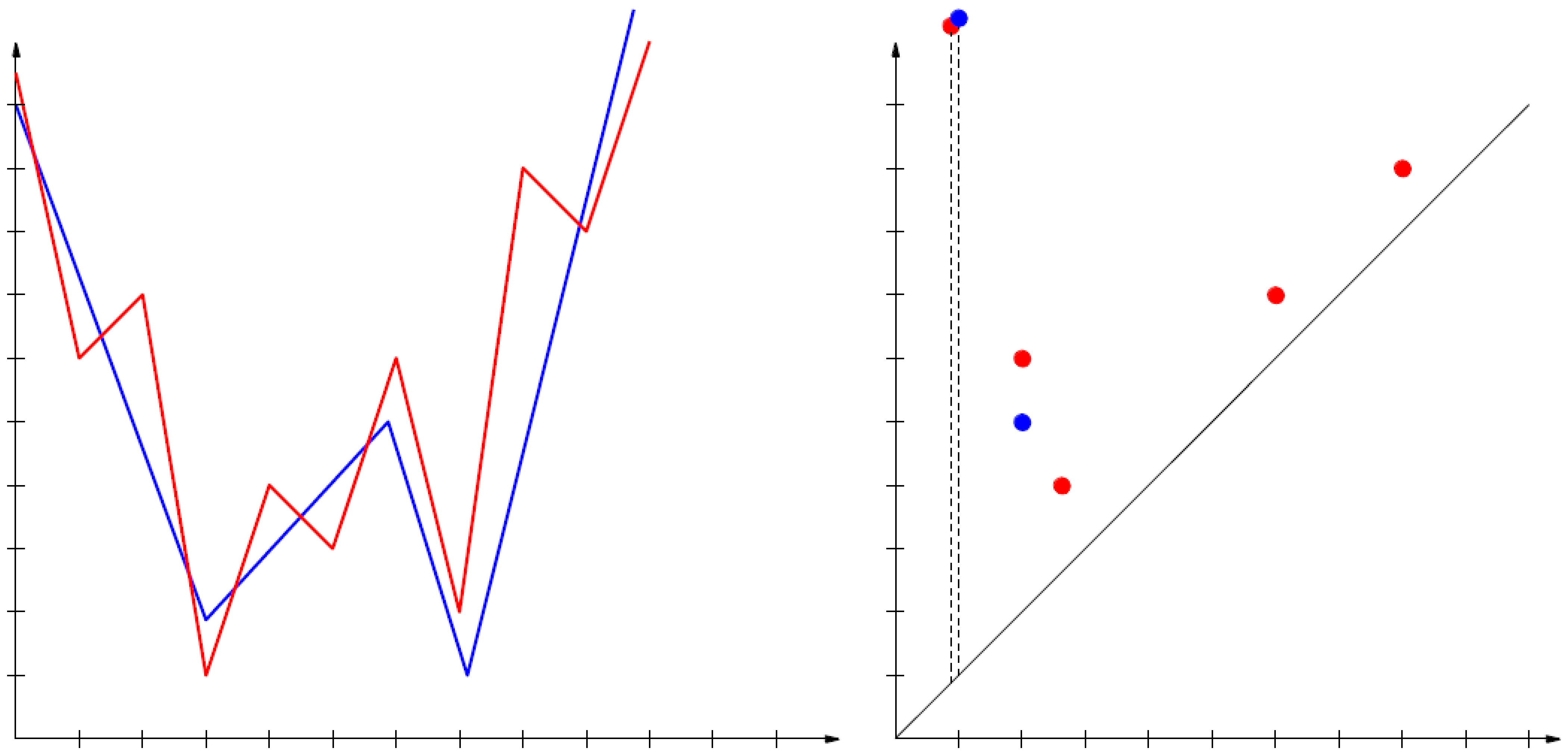
Let K be a simplicial complex and f, g two functions defined on the vertices of K. Let D_f and D_g be the persistence diagrams of f and g.

The bottleneck distance between D_f and D_g is

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

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

Stability of persistence diagrams



Theorem: Let K be a simplicial complex and let $f, g : K \to \mathbb{R}$.

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

where $\|f - g\|_\infty = \sup_{v \in vertices(K)} |f(v) - g(v)|$.

Stability of persistence diagrams

- Let K and K' be two simplicial complexes homeomorphic to a topological space X.
- Let $\phi : K \to X$ and $\phi' : K' \to X$ be homeomorphisms
- Let $f : X \to \mathbb{R}$ be a continuous function and $D_f(K)$ (resp. $D_f(K')$) the persistence diagram of $f \circ \phi$ (resp. $f \circ \phi'$).

Theorem: Let $\varepsilon > 0$ be such that for any simplex $\sigma \in K$ (resp. $\in K'$), $\sup_{x,y\in\sigma} |f \circ \phi(x) - f \circ \phi(y)| < \varepsilon$ (resp. $\sup_{x,y\in\sigma} |f \circ \phi'(x) - f \circ \phi'(y)| < \varepsilon$). Then one has

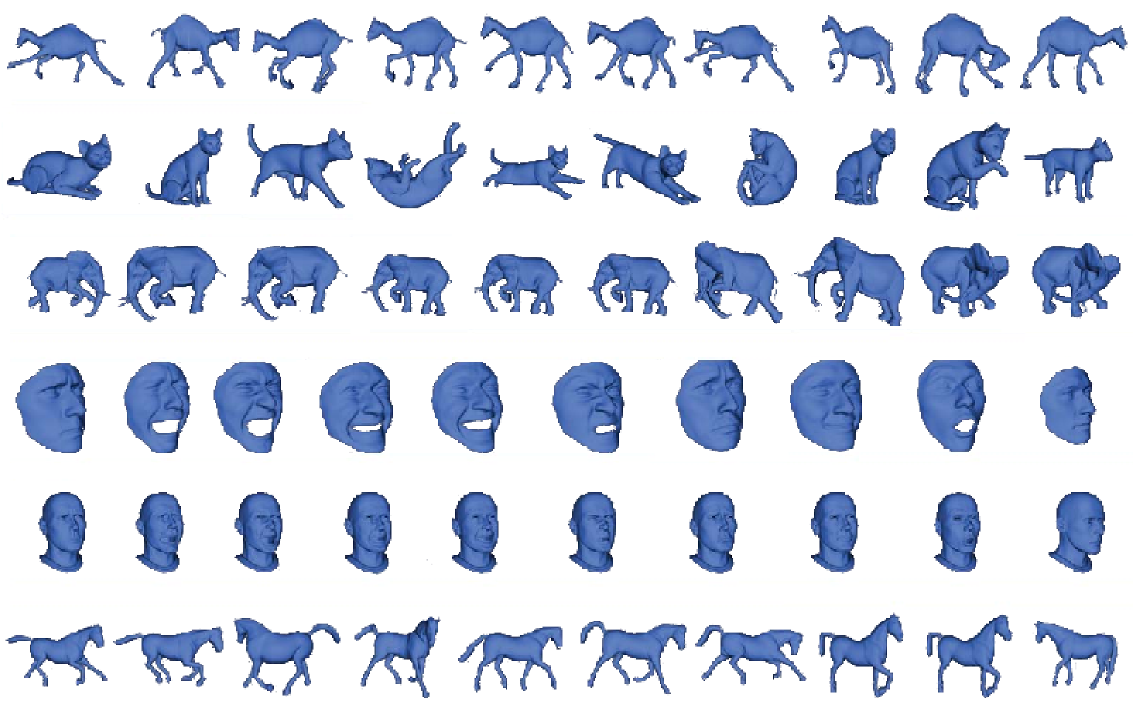
$$d_B(D_f(K), D_f(K')) \leq 2\varepsilon$$

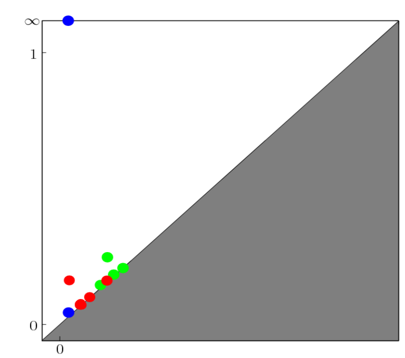
Remark: this is a particular (and weaker) version of a much more general result. See:

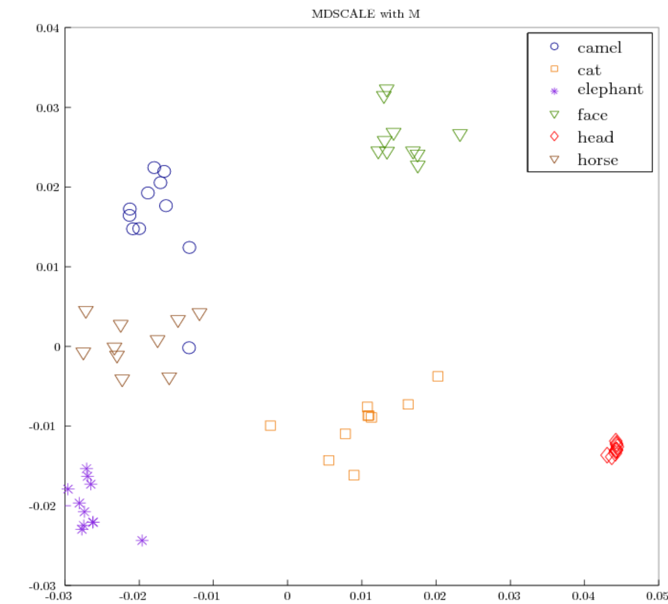
- D. Cohen-Steiner and H. Edelsbrunner and J. Harer, *Stability of Persistence Diagrams*, Proc. 21st ACM Sympos. Comput. Geom. 2005.
- F. Chazal and D. Cohen-Steiner and L. J. Guibas and M. Glisse and S. Y. Oudot, *Proximity of Persistence Modules and their Diagrams*, Proc. 25th ACM Sympos. Comput. Geom. 2009.

Consequences of the stability

- Persistence diagrams are defined and stable for a large class of continuous functions defined over (pre-)compact metric spaces.


3D shapes database



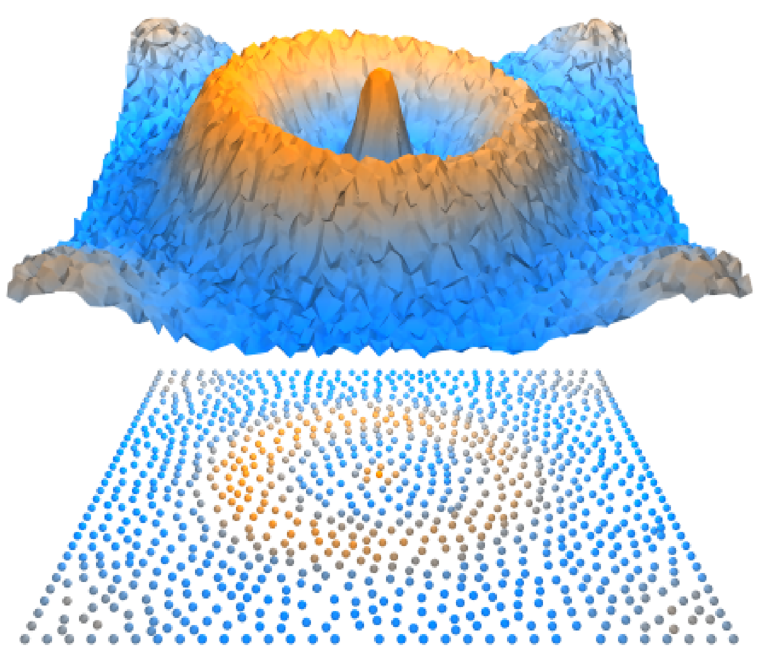


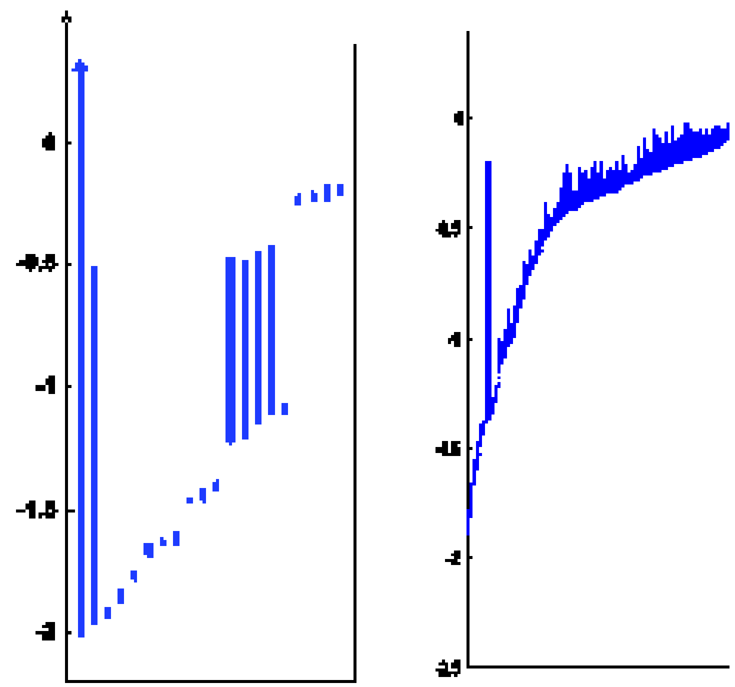
→ definition stable (Gromov-Hausdorff distance) topological signatures for compact metric spaces.
→ Efficient algorithm to compute signatures.
→ applications to shape classification.

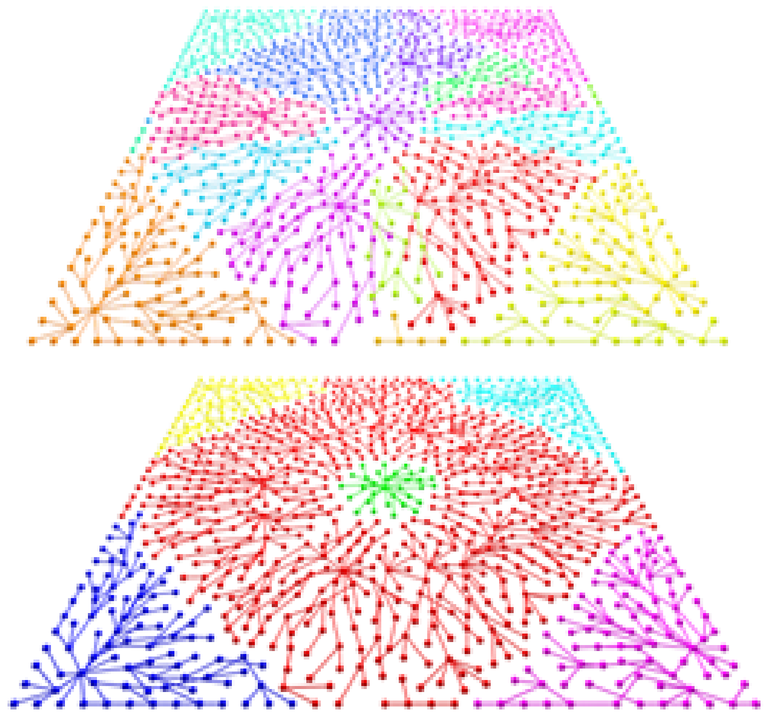
Ref: F. Chazal, D. Cohen-Steiner, L. J. Guibas, F. Mémoli, S. Oudot, Gromov-Hausdorff Stable Signatures for Shapes using Persistence, Computer GraphicsForum (proc. SGP 2009), pp. 1393-1403, 2009.

Consequences of the stability

- Persistence diagrams can be reliably estimated from data (functions known through a point cloud data set approximating a topological space).







Previous approach can be generalized, leading to robust algorithms to compute the topological persistence of functions defined over point clouds sampled around unknown shapes

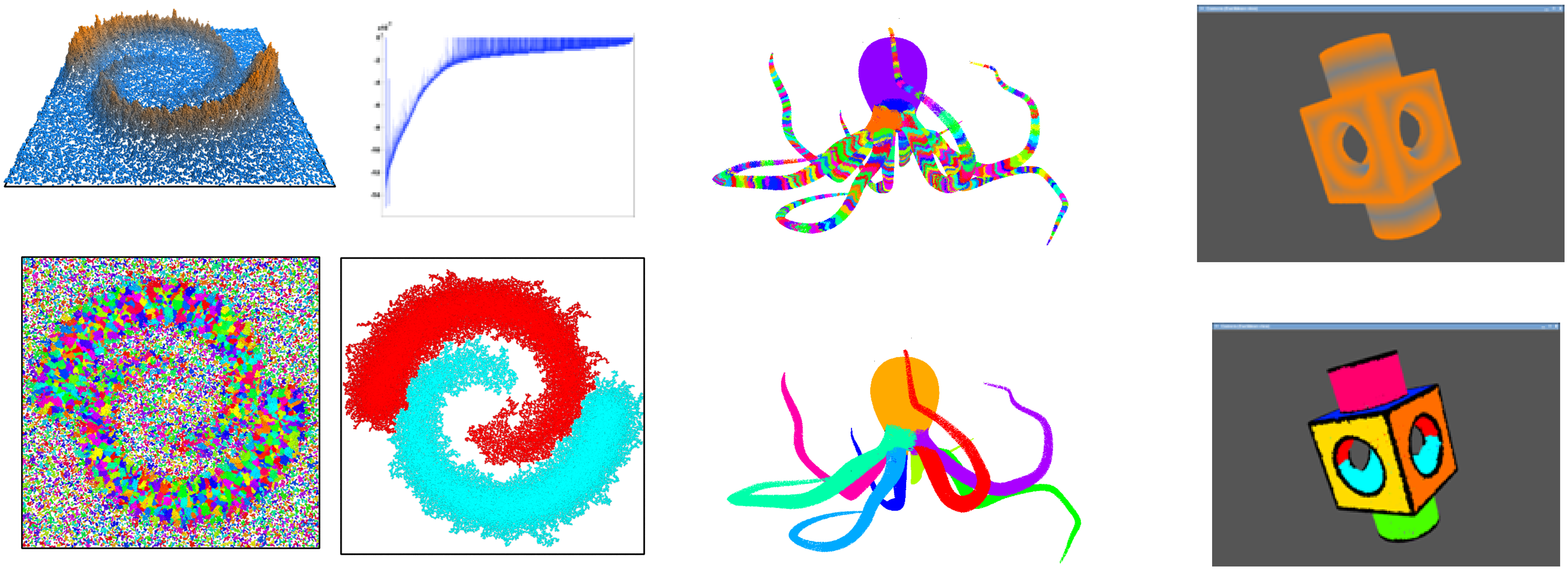
Ref:

- F. Chazal, L. Guibas, S. Oudot, P. Skraba, *Analysis of Scalar Fields over Point Cloud Data*, proc. ACM Symposium on Discrete Algorithms 2009.

- F. Chazal, S. Oudot, *Toward Persistence-Based Reconstruction in Euclidean Spaces*, proc. ACM Symposium on Computational Geometry 2008.

Consequences of the stability

- Persistence diagrams can be reliably estimated from data (functions known through a point cloud data set approximating a topological space).



Applications to clustering, segmentations, sensor networks,...

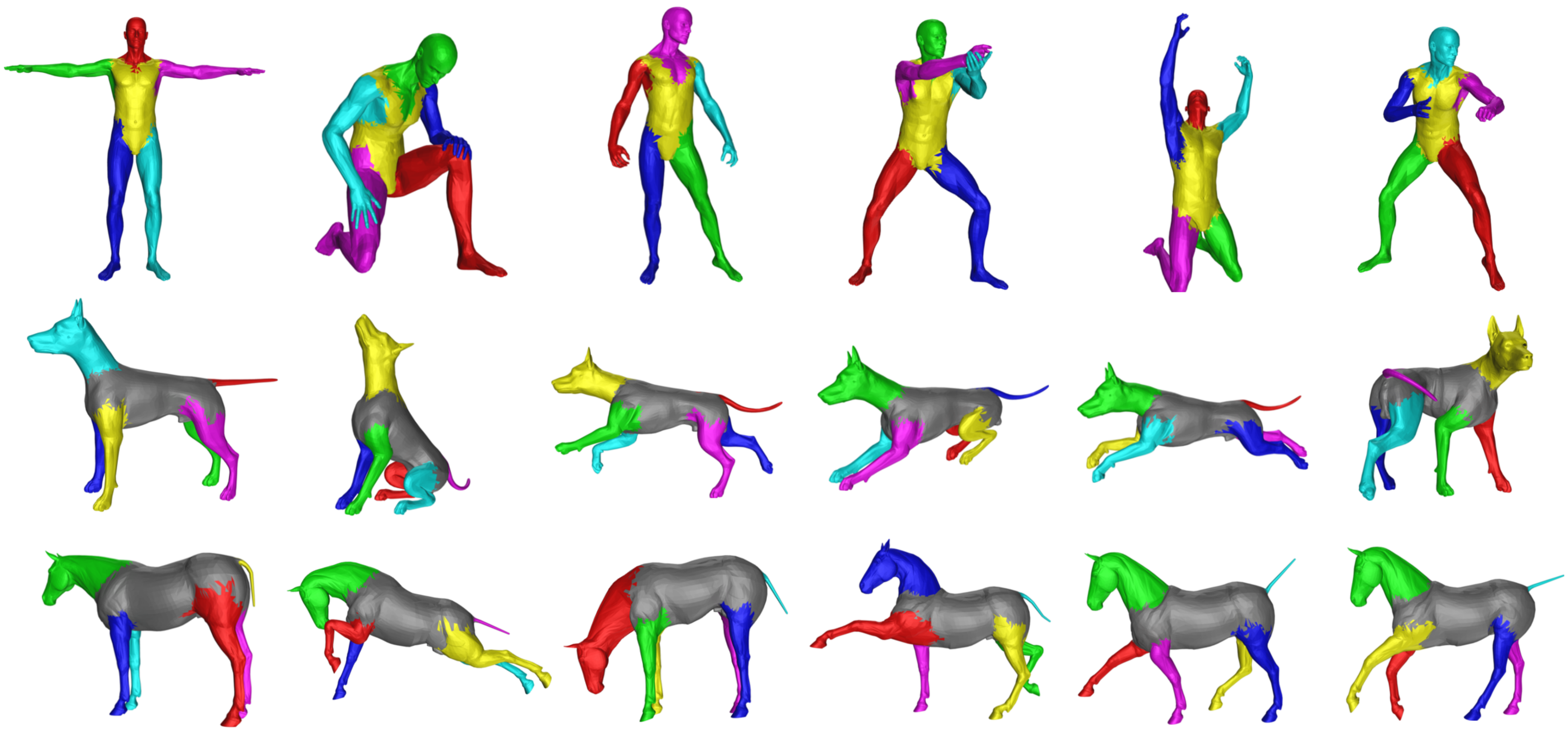
Ref:

- F. Chazal, L. Guibas, S. Oudot, P. Skraba, *Analysis of Scalar Fields over Point Cloud Data*, proc. ACM Symposium on Discrete Algorithms 2009.

- F. Chazal, S. Oudot, *Toward Persistence-Based Reconstruction in Euclidean Spaces*, proc. ACM Symposium on Computational Geometry 2008.

Consequences of the stability

- Persistence diagrams can be reliably estimated from data (functions known through a point cloud data set approximating a topological space).



Applications to non rigid shapes segmentation

Ref:

- P. Skraba, M. Ovsjanikov, F. Chazal, L. Guibas, Persistence-Based Segmentation of Deformable Shapes, Proc. Workshop on Nonrigid Shape Analysis and Deformable Image Alignment (NORDIA), Proc. CVPR 2010