Homology and topological persistence

F. Chazal
Geometrika Group
INRIA Saclay

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

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
v_0, v_1, \ldots, v_k \in \mathbb{R}^d are affinely independent 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 \( \sigma = [v_0, v_1, \ldots, v_k] \) is a simplex of dimension \( d \). A simplex generated by a subset of the vertices \( v_0, v_1, \ldots, v_k \) of \( \sigma \) is a face of \( \sigma \).
A (finite) simplicial complex $C$ is a (finite) union of simplices s.t.

i) for any $\sigma \in C$, all the faces of $\sigma$ 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.
A (finite) simplicial complex $C$ is a (finite) union of simplices s.t.

i) for any $\sigma \in C$, all the faces of $\sigma$ 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.
A (finite) simplicial complex $C$ is a (finite) union of simplices s.t.

i) for any $\sigma \in C$, all the faces of $\sigma$ 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 homogeneous of dimension $n$ if any of its faces is a face of a $n$-dimensional simplex.
A filtration of a (finite) simplicial complex $K$ is a sequence of subcomplexes such that

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

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

Filtrations of simplicial complexes
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 $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\ldots,k} f(v_i)$
- 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 $\sigma = [v_0, \cdots, v_k] \in K$, $f(\sigma) = \max_{i=0,\ldots,k} f(v_i)$
- 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 \( \sigma = [v_0, \ldots, v_k] \in K \), \( f(\sigma) = \max_{i=0,\ldots,k} f(v_i) \)
- 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

\[ \text{Example: filtration associated to a function} \]

\[ \begin{align*}
\bullet ~ f & \text{ a real valued function defined on the vertices of } K \\
\bullet ~ \text{For } \sigma = [v_0, \ldots, v_k] \in K, \ f(\sigma) = \max_{i=0, \ldots, k} f(v_i) \\
\bullet ~ \text{The simplices of } K \text{ are ordered according increasing } f \text{ values (and dimension in case of equal values on different simplices).}
\end{align*} \]

\[ \Rightarrow \text{ The sublevel sets } \text{filtration}. \]

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 $\sigma = [v_0, \ldots, v_k] \in K$, $f(\sigma) = \max_{i=0,\ldots,k} f(v_i)$
- 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 Čech complex

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

- The Čech 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 $U_i$
  - $[U_{i_0}, \ldots, U_{i_k}]$ is a $k$-simplex in $C(\mathcal{U})$ iff $\cap_{j=0}^k U_{i_j} \neq \emptyset$. 
Example: The Čech complex

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

\[ \Rightarrow \text{The combinatorics of the covering (a simplicial complex) carries the topology of the space.} \]
Example: The Čech complex

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

⇒ 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 Čech 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 $p_0, \cdots p_k \in L$,

$$\sigma = [p_0p_1 \cdots p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i, j \in \{0, \cdots k\}, \ d(p_i, p_j) \leq \alpha$$

- 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 $\mathbb{Z}/2$

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:

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

Sum of $k$-chains:

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

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:**

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

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:

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

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 $\sigma$ is the sum of its $(k-1)$-faces. This is a $(k-1)$-chain.

If $\sigma = [v_0, \cdots, v_k]$ 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

$$\partial : \mathcal{C}_k(K) \rightarrow \mathcal{C}_{k-1}(K)$$
$$c \rightarrow \partial c = \sum_{\sigma \in c} \partial \sigma$$
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.

\[
\partial \partial \sigma = \partial \left( \sum_{i=0}^{k} [v_0 \cdots \hat{v_i} \cdots v_k] \right) \\
= \sum_{i=0}^{k} \partial [v_0 \cdots \hat{v_i} \cdots v_k] \\
= \sum_{j<i} [v_0 \cdots \hat{v_j} \cdots \hat{v_i} \cdots v_k] + \sum_{j>i} [v_0 \cdots \hat{v_i} \cdots \hat{v_j} \cdots v_k] \\
= 0
\]
Cycles and boundaries

The chain complex associated to a complex $K$ of dimension $d$

$$
\emptyset \to C_d(K) \xrightarrow{\partial} C_{d-1}(K) \xrightarrow{\partial} \cdots C_{k+1}(K) \xrightarrow{\partial} C_k(K) \xrightarrow{\partial} \cdots C_1(K) \xrightarrow{\partial} C_0(K) \xrightarrow{\partial} \emptyset
$$

$k$-cycles:

$$Z_k(K) := \ker(\partial : C_k \to C_{k-1}) = \{c \in C_k : \partial c = \emptyset\}$$

$k$-boundaries:

$$B_k(K) := \text{im}(\partial : C_{k+1} \to C_k) = \{c \in C_k : \exists c' \in C_{k+1}, c = \partial c'\}$$

$$B_k(K) \subset Z_k(K) \subset C_k(K)$$
Cycles and boundaries

Non homologous 1-cycles
Cycles and boundaries

Non homologous 1-cycles

A 1-boundary
Cycles and boundaries

Non homologous 1-cycles

Two homologous 1-cycles

A 1-boundary
Cycles and boundaries

Not a cycle

Non homologous 1-cycles

Two homologous 1-cycles

A 1-boundary
Homology groups and Betti numbers

\[ B_k(K) \subset Z_k(K) \subset C_k(K) \]

- 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: $\beta_0 = \text{number of connected components of } K$
Elementary examples

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

\[
\begin{align*}
\beta_0 &= 2 \\
\beta_1 &= 0 \\
\beta_2 &= 0
\end{align*}
\]
Elementary examples

Remark: $\beta_0 = \text{number of connected components of } K$

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

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

Remark: $\beta_0 = \text{number of connected components of } K$

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

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

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

\[ \beta_0 = 2 \]
\[ \beta_1 = 2 \]
\[ \beta_2 = 1 \text{ if empty and } \beta_2 = 0 \text{ if filled} \]
\[ \beta_3 = 0 \]
Theorem: If $K$ and $K'$ are two simplicial complexes with homeomorphic supports then their homology groups are isomorphic and their Betti numbers are equal.

\[ \beta_0 = 1, \beta_1 = 2, \beta_2 = 0 \]

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.
Let $\Delta_k$ 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 → **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!
Let $\Delta_k$ 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:

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

- if $f : X \to Y$ is an homeomorphism or an homotopy equivalence then $f_\#$ 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 \text{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 $0 < r < \text{wfs}(X)$ from $L$. 

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 $\varepsilon > 0$.

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

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

$$H_k(X^r) \cong \text{im} \left( i_* : H_k(L^{\alpha + \varepsilon}) \rightarrow H_k(L^{\alpha + 3\varepsilon}) \right)$$
An algorithm for geometric inference

Proof:

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

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

\[ H_k(X^\alpha) \rightarrow H_k(L^{\alpha+\varepsilon}) \rightarrow H_k(X^{\alpha+2\varepsilon}) \rightarrow H_k(L^{\alpha+3\varepsilon}) \rightarrow H_k(X^{\alpha+4\varepsilon}) \rightarrow \cdots \]

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

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

Cannot be directly computed!

$\text{rank} = \dim H_k(X^\alpha)$
Using the Čech complex

The Čech complex $C^\alpha(L)$:

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

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

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

Nerve theorem: For any $\alpha > 0$, $L^\alpha$ 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 $C^\alpha(L)$:

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

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

\[ \cdots \rightarrow H_k(L^{\alpha+\varepsilon}) \rightarrow H_k(L^{\alpha+3\varepsilon}) \rightarrow \cdots \]

\[ \cdots \rightarrow H_k(C^{\alpha+\varepsilon}(L)) \rightarrow H_k(C^{\alpha+3\varepsilon}(L)) \rightarrow \cdots \]

Allow to work with simplicial complexes but... still too difficult to compute
Using the Rips complex

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

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

\[ C^{\frac{\alpha}{2}}(L) \subseteq R^\alpha(L) \subseteq C^\alpha(L) \subseteq R^{2\alpha}(L) \subseteq \cdots \]
Using the Rips complex

\[ \sigma = [p_0p_1 \cdots p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i, j \in \{0, \cdots k\}, \ d(p_i, p_j) \leq \alpha \]

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

\[ \sigma = [p_0p_1 \cdots p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i, j \in \{0, \cdots k\}, \ d(p_i, p_j) \leq \alpha \]

**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 \( \varepsilon < \frac{1}{9} \) \( \text{wfs}(X) \). Then for all \( \alpha \in [2\varepsilon, \frac{1}{4}(\text{wfs}(X) - \varepsilon)] \) and all \( \lambda \in (0, \text{wfs}(X)) \), one has: \( \forall k \in \mathbb{N} \)

\[ \beta_k(X^\lambda) = \dim(H_k(X^\lambda)) = \text{rk}(\mathcal{R}^\alpha(L) \to \mathcal{R}^{4\alpha}(L)) \]
Using the Rips complex

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

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

**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 $\alpha \in [2\varepsilon, \frac{1}{4}(\text{wfs}(X) - \varepsilon)]$ and all $\lambda \in (0, \text{wfs}(X))]$, one has: $\forall k \in \mathbb{N}$

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

Easy to compute using persistence algo.
Using the Rips complex

The Rips complex $\mathcal{R}^\alpha(L)$: for $p_0, \ldots, p_k \in L$,
\[ \sigma = [p_0 p_1 \cdot \cdot p_k] \in \mathcal{R}^\alpha(L) \text{ iff } \forall i, j \in \{0, \ldots, k\}, \ d(p_i, p_j) \leq \alpha \]

**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 $\alpha \in [2\varepsilon, \frac{1}{4}(\wfs(X) - \varepsilon)]$ and all $\lambda \in (0, \wfs(X))]$, one has: $\forall k \in \mathbb{N}$
\[ \beta_k(X^\lambda) = \dim(H_k(X^\lambda)) = \text{rk}(\mathcal{R}^\alpha(L) \rightarrow \mathcal{R}^{4\alpha}(L)) \]

**Pb:** Choice of $\alpha$ when $\wfs(X)$ is unknown?
Multiscale inference

**Input:** A point cloud $W$ and its pairwise distances $\{d(w, w')\}_{w, w' \in W}$.

→ Maintain a nested pair $\mathcal{R}^{4\epsilon}(L) \hookrightarrow \mathcal{R}^{16\epsilon}(L)$ where $L = L(\epsilon)$.

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

**WHILE** $L \subset W$

insert $p = \arg\max_{w \in W} d(w, L)$ in $L$

update $\epsilon = \max_{w \in W} d(w, L)$

update $\mathcal{R}^{4\epsilon}(L)$ and $\mathcal{R}^{16\epsilon}(L)$

Persistence($\mathcal{R}^{4\epsilon}(L) \hookrightarrow \mathcal{R}^{16\epsilon}(L)$)

**END WHILE**

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

**Input:** A point cloud $\mathcal{W}$ and its pairwise distances $\{d(w, w')\}_{w, w' \in \mathcal{W}}$. 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** $L \subset \mathcal{W}$

insert $p = \arg \max_{w \in \mathcal{W}} d(w, L)$ in $L$
update $\varepsilon = \max_{w \in \mathcal{W}} d(w, L)$
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**

**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 pairwise distances $\{d(w, w')\}_{w, w' \in W}$.

→ 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 $p = \arg\max_{w \in W} d(w, L)$ 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**

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 pairwise distances $\{d(w, w')\}_{w, w' \in W}$.

→ Maintain a nested pair $R^{4\varepsilon}(L) \hookrightarrow R^{16\varepsilon}(L)$ where $L = L(\varepsilon)$.

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

**WHILE** $L \subset W$

insert $p = \arg\max_{w \in W} d(w, L)$ in $L$

update $\varepsilon = \max_{w \in W} d(w, L)$

update $R^{4\varepsilon}(L)$ and $R^{16\varepsilon}(L)$

Persistence($R^{4\varepsilon}(L) \hookrightarrow R^{16\varepsilon}(L)$)

**END WHILE**

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

**Input:** A point cloud $W$ and its pairwise distances $\{d(w, w')\}_{w, w' \in W}$.

→ 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 $p = \arg \max_{w \in W} d(w, L)$ 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**

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

Rank of the map induced at homology level
Multiscale inference

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

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

**WHILE** $L \subset W$

insert $p = \arg\max_{w \in W} d(w, L)$ in $L$

update $\epsilon = \max_{w \in W} d(w, L)$

update $\mathcal{R}^{4\epsilon}(L)$ and $\mathcal{R}^{16\epsilon}(L)$

Persistence($\mathcal{R}^{4\epsilon}(L) \hookrightarrow \mathcal{R}^{16\epsilon}(L)$)

**END** \textbf{WHILE}

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

**Input:** A point cloud \( W \) and its pairwise distances \( \{d(w, w')\}_{w, w' \in W} \).

→ 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 \( p = \arg\max_{w \in W} d(w, L) \) 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**

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

Rank of the map induced at homology level
Multiscale inference

Input: A point cloud $W$ and its pairwise distances $\{d(w, w')\}_{w, w' \in W}$.

→ Maintain a nested pair $R^4(\varepsilon)(L) \hookrightarrow R^{16}(\varepsilon)(L)$ where $L = L(\varepsilon)$.

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

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

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

Rank of the map induced at homology level
**Theorem:** [C-Oudot’08]
If \( d_H(W, X) < \delta \) for \( \delta < \frac{1}{18} \text{wfs}(X) \), then at every iteration of the algorithm such that \( \delta < \varepsilon < \frac{1}{18} \text{wfs}(X) \),

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

for any \( \lambda \in (0, \text{wfs}(X)) \) 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

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

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

50,000 points sampled uniformly at random from a curve drawn on the 2-torus $S^1 \times S^1$. 
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 $\sigma^{i+1}$ is a simplex of $K$.

Output: The Betti numbers $\beta_0, \beta_1, \cdots, \beta_d$ of $K$.

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

\[
\text{then } \beta_{k+1} = \beta_{k+1} + 1;
\]

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 $\sigma^{i+1}$ 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 \( \sigma^i \) is contained in a \((k + 1)\)-cycle in \( K^i \), this cycle is not a boundary in \( K^i \).

• If \( \sigma^i \) 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 \( \sigma^i \) 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 $\sigma^{i+1}$ is a simplex of $K$.

**Definition:** A $(k+1)$-simplex $\sigma^i$ 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 $\sigma^{i+1}$ is a simplex of $K$.

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

$$\beta_k(K) = \#(\text{positive simplices}) - \#(\text{negative simplices})$$
Getting more information

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

\[
\beta_k(K) = \#(\text{positive simplices}) - \#(\text{negative simplices})
\]

- 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\((H_k(K^i) \rightarrow H_k(K^j))\)?
Getting more information

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

\[ \beta_k(K) = \#(\text{positive simplices}) - \#(\text{negative simplices}) \]

- 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 \(\text{rank}(H_k(K^i) \rightarrow H_k(K^j))\)?

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:
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
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 functions 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 $\sigma^{i+1}$ is a simplex of $K$.

- $Z^i_k = \text{the } k\text{-cycles of } K^i$, $B^i_k = \text{the } k\text{-boundaries of } K^i$ and $H^i_k = \text{the } k^{th}\text{-homology group of } K^i$.

- $Z^0_k \subset Z^1_k \subset \cdots \subset Z^i_k \subset \cdots \subset Z^m_k = Z_K(K)$

- $B^0_k \subset B^1_k \subset \cdots \subset B^i_k \subset \cdots \subset B^m_k = B_K(K)$
**Lemma:** If $\sigma^i$ is a positive $k$-cycle, then there exists a $k$-cycle $c_\sigma$ s.t.:
- $c_\sigma$ is not a boundary in $K^i$,
- $c_\sigma$ contains $\sigma^i$ but no other positive $k$-simplex.

The cycle $c_\sigma$ is unique.

**Proof:**
By induction on the order of appearance 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 $\sigma^i$ is a positive $k$-simplex then one adds the homology class of the cycle $c^i$ associated to $\sigma^i$ 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 $\sigma^j$ is a negative $(k+1)$-simplex:
  
  - let $c^{i_1}, \ldots, c^{i_p}$ be the cycles associated to the positive simplices $\sigma^{i_1}, \ldots, \sigma^{i_p}$ 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 $\sigma^i$ is a positive $k$-simplex then one adds the homology class of the cycle $c^i$ associated to $\sigma^i$ 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 $\sigma^j$ is a negative $(k+1)$-simplex:
  - let $c^{i_1}, \ldots, c^{i_p}$ be the cycles associated to the positive simplices $\sigma^{i_1}, \ldots, \sigma^{i_p}$ 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 $\sigma^i$ is a positive $k$-simplex then one adds the homology class of the cycle $c^i$ associated to $\sigma^i$ 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 $\sigma^j$ is a negative $(k+1)$-simplex:
  - let $c_1^i, \cdots , c_p^i$ be the cycles associated to the positive simplices $\sigma_1^i, \cdots , \sigma_p^i$ 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 $\sigma^i$ is a positive $k$-simplex then one adds the homology class of the cycle $c^i$ associated to $\sigma^i$ 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 $\sigma^j$ 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 $H_{k-1}^j$ has been built and $\sigma^j$ is a negative $(k+1)$-simplex:
  - let $c^1, \ldots, c^p$ be the cycles associated to the positive simplices $\sigma^i_1, \ldots, \sigma^i_p$ that form a basis of $H_{k-1}^j$
  - $d = \partial \sigma^j = \sum_{k=1}^{p} \varepsilon_k c^i_k + b$
  - $l(j) = \max\{i_k : \varepsilon_k = 1\}$
  - Remove the homology class of $c^{l(j)}$ from the basis of $H_{k-1}^j$ ⇒ basis of $H_k^j$.

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

**Remark:** filtrations of $K$ can be indexed by increasing sequences $\alpha_i$ 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 $\sigma^{i+1}$ 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 $L_0, L_1, \ldots, L_{d-1}$;
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 $\sigma^{i+1}$ 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 matrix of the boundary operator

- \( M = (m_{ij})_{i,j=1,...,m} \) 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 $\sigma^{i+1}$ 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 $(l(j), j);$
A very simple example

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

$$\partial \left( \sigma^j + \sum_{i<j} \varepsilon_i \sigma^i \right)$$

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 $\sigma^j$ 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 \(\sigma^i \rightarrow (i, +\infty)\).
- 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 \(\sigma^i \rightarrow (i, +\infty)\).

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 \(\sigma^i \rightarrow (i, +\infty)\).

**Barcodes:** an alternative (equivalent) representation where each pair \((i, j)\) is represented by the interval \([i, j]\)
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 $\Gamma$ 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 $\|f - g\|_{\infty} = \sup_{v \in \text{vertices}(K)} |f(v) - g(v)|$. 

\[ \text{Stability of persistence diagrams} \]
Stability of persistence diagrams

- Let $K$ and $K'$ be two simplicial complexes homeomorphic to a topological space $X$.

- Let $\phi : K \rightarrow X$ and $\phi' : K' \rightarrow X$ be homeomorphisms.

- Let $f : X \rightarrow \mathbb{R}$ 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 \quad \text{(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:

Consequences of the stability

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

→ definition stable (Gromov-Hausdorff distance) topological signatures for compact metric spaces.

→ Efficient algorithm to compute signatures.

→ applications to shape classification.

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:

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:
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: