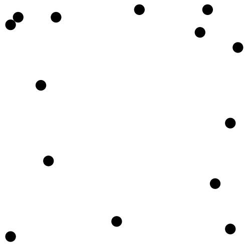
Efficient and Robust Persistent Homology for Measures

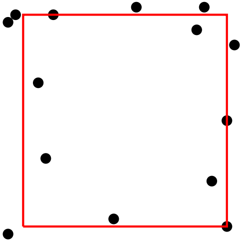
Mickaël Buchet

joint work with F. Chazal, S. Oudot and D. Sheehy

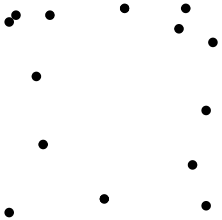
Topological inference



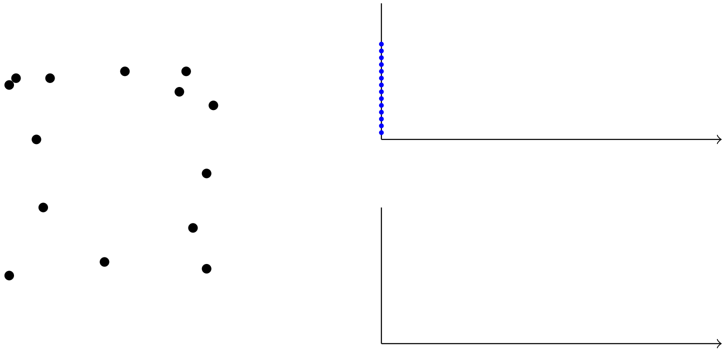
Topological inference



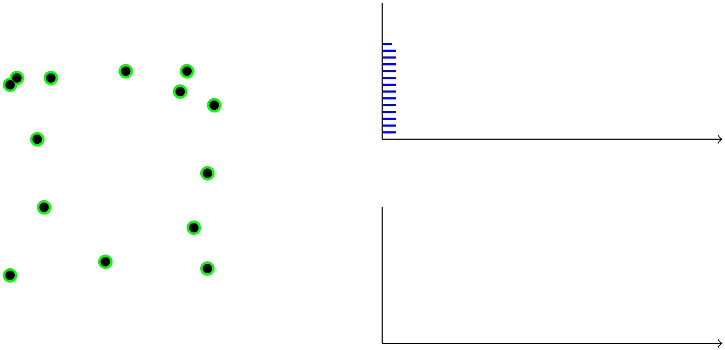
Persistence



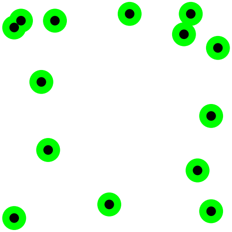
Persistence

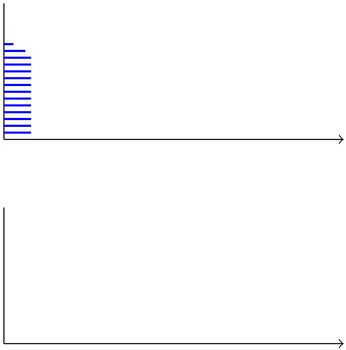


Persistence

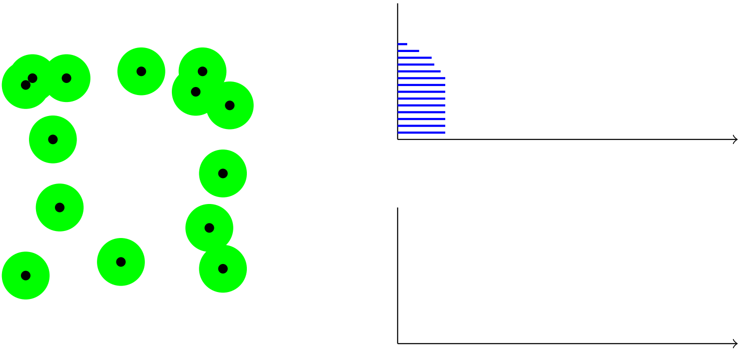


Persistence

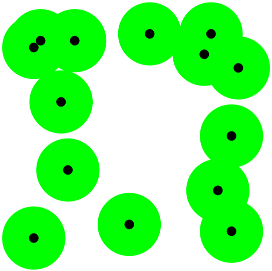


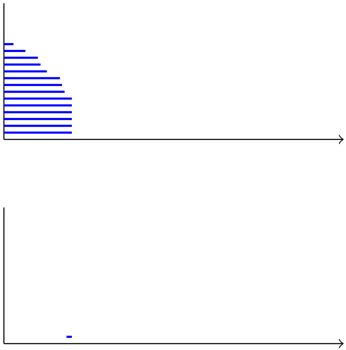


Persistence

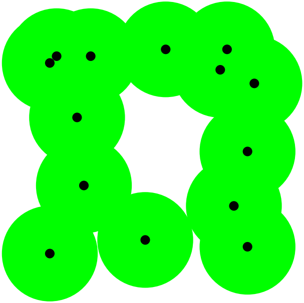


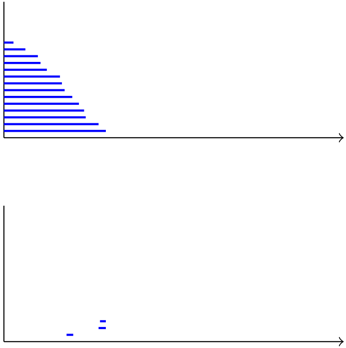
Persistence



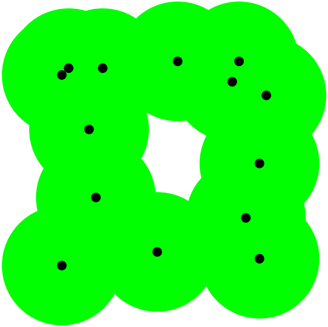


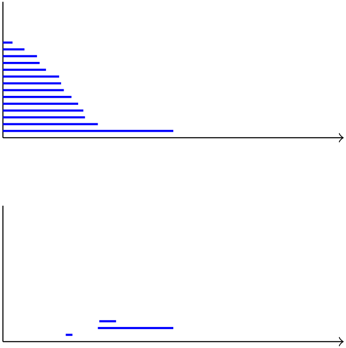
Persistence



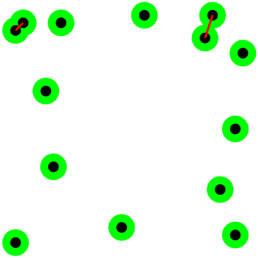


Persistence

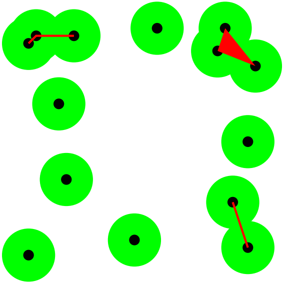




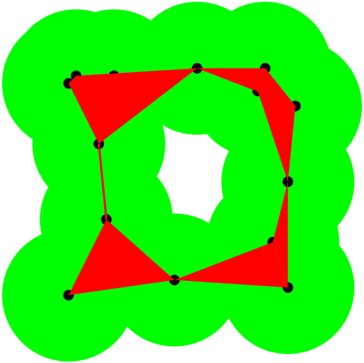
Practical application



Practical application



Practical application



Complexity

- Computing the persistence diagram of a filtered simplicial complex is equivalent to reducing a $N \times N$ matrix, where N is the total number of simplexes.

Complexity

- Computing the persistence diagram of a filtered simplicial complex is equivalent to reducing a $N \times N$ matrix, where N is the total number of simplexes.
- In practice, the matrix is sparse the computation time is linear in N.

Complexity

- Computing the persistence diagram of a filtered simplicial complex is equivalent to reducing a $N \times N$ matrix, where N is the total number of simplexes.
- In practice, the matrix is sparse the computation time is linear in N.
- A d-dimensional Rips complex has $\Theta(n^{d+1})$ simplexes.

Complexity

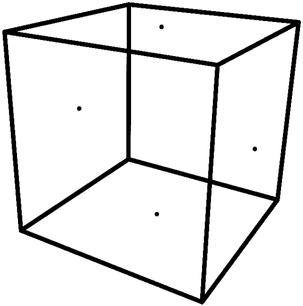
- Computing the persistence diagram of a filtered simplicial complex is equivalent to reducing a $N \times N$ matrix, where N is the total number of simplexes.
- In practice, the matrix is sparse the computation time is linear in N.
- A d-dimensional Rips complex has $\Theta(n^{d+1})$ simplexes.
- The computation of the persistence diagram of n points has complexity $O(n^{d+1})$.

Complexity

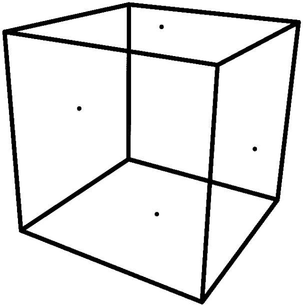
- Computing the persistence diagram of a filtered simplicial complex is equivalent to reducing a $N \times N$ matrix, where N is the total number of simplexes.
- In practice, the matrix is sparse the computation time is linear in N.
- A d-dimensional Rips complex has $\Theta(n^{d+1})$ simplexes.
- The computation of the persistence diagram of n points has complexity $O(n^{d+1})$.

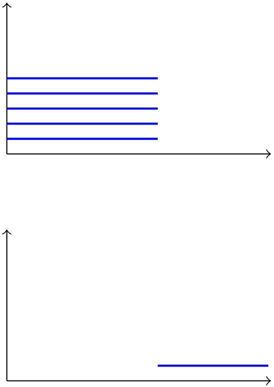
- Unusable in high dimensions.

Outliers

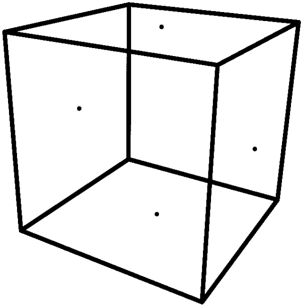


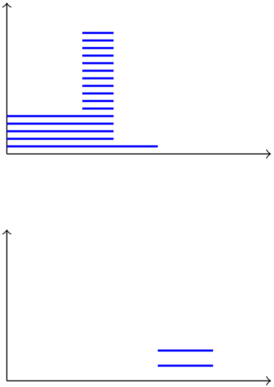
Outliers





Outliers





Computation of persistence diagrams

Distance to the point cloud and Vietoris-Rips filtration:

- Complexity: $O(n^{d+1})$
- Robustness to outliers: none

Distance to a measure

Definition (Chazal, Cohen-Steiner, Mérigot, 2011)

Let μ be a measure and $m \in]0,1[$, then

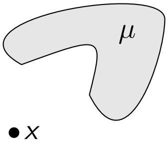
$$d_{\mu,m}(x) = \frac{1}{\sqrt{m}} \inf_{\nu \in \mathrm{Sub}_m(\mu)} W_2(m\delta_x, \nu)$$

Distance to a measure

Definition (Chazal, Cohen-Steiner, Mérigot, 2011)

Let μ be a measure and $m \in]0,1[$, then

$$d_{\mu,m}(x) = \frac{1}{\sqrt{m}} \inf_{\nu \in \mathrm{Sub}_m(\mu)} W_2(m\delta_x, \nu)$$

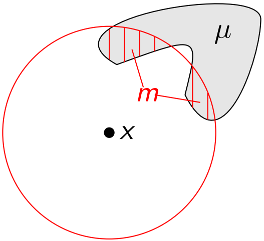


Distance to a measure

Definition (Chazal, Cohen-Steiner, Mérigot, 2011)

Let μ be a measure and $m \in]0, 1[$, then

$$d_{\mu,m}(x) = \frac{1}{\sqrt{m}} \inf_{\nu \in \mathrm{Sub}_m(\mu)} W_2(m\delta_x, \nu)$$



Case of an empirical measure

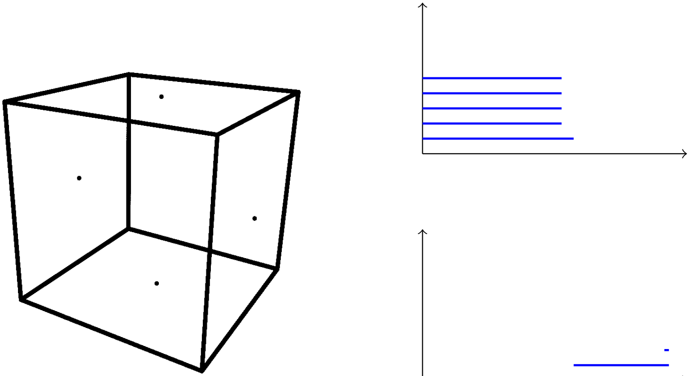
Theorem

Let μ be an empirical P and $k = m|P|$ is an integer then:

$$d_{\mu,m}(x) = \sqrt{\frac{1}{k}\sum_{i=1}^{k} d_{\mathbb{X}}(x, p_i(x))^2}$$

where $p_i(x)$ is the i^{th}-neighbour of x in P.

Results



Barycentric decomposition

Theorem

In Euclidean spaces, the distance to an empirical measure is a power distance.

Barycentric decomposition

Theorem

In Euclidean spaces, the distance to an empirical measure is a power distance.

$$d_{\mu,m}(x)^2 = \sum_{i=1}^{k} ||x - p_i(x)||^2$$

Barycentric decomposition

Theorem

In Euclidean spaces, the distance to an empirical measure is a power distance.

$$d_{\mu,m}(x)^2 = \sum_{i=1}^{k} ||x - p_i(x)||^2$$

$$= ||x - bar(x)||^2 + \sum_{i=1}^{k} ||p_i(x) - bar(x)||^2$$

where $bar(x) = \sum_{i=1}^{k} p_i(x)$

Barycentric decomposition

Theorem

In Euclidean spaces, the distance to an empirical measure is a power distance.

$$d_{\mu,m}(x)^2 = \sum_{i=1}^{k} ||x - p_i(x)||^2$$

$$= ||x - bar(x)||^2 + \sum_{i=1}^{k} ||p_i(x) - bar(x)||^2$$

$$= ||x - bar(x)||^2 + w_{bar(x)}^2$$

where $bar(x) = \sum_{i=1}^{k} p_i(x)$

Barycentric decomposition

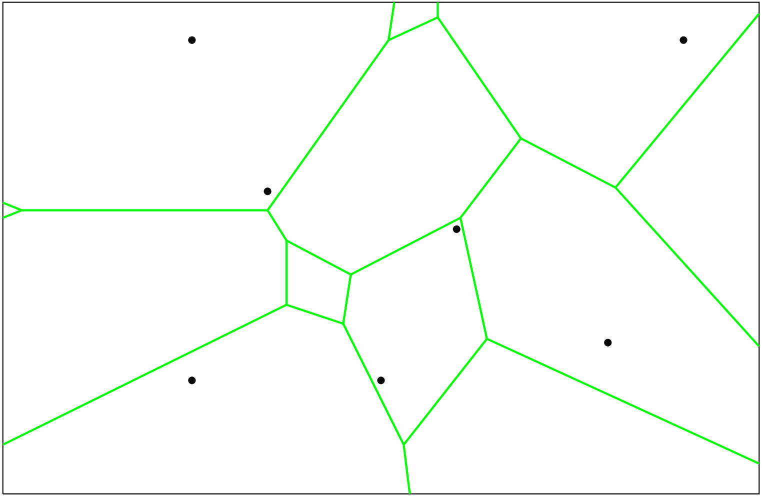
Theorem

In Euclidean spaces, the distance to an empirical measure is a power distance.

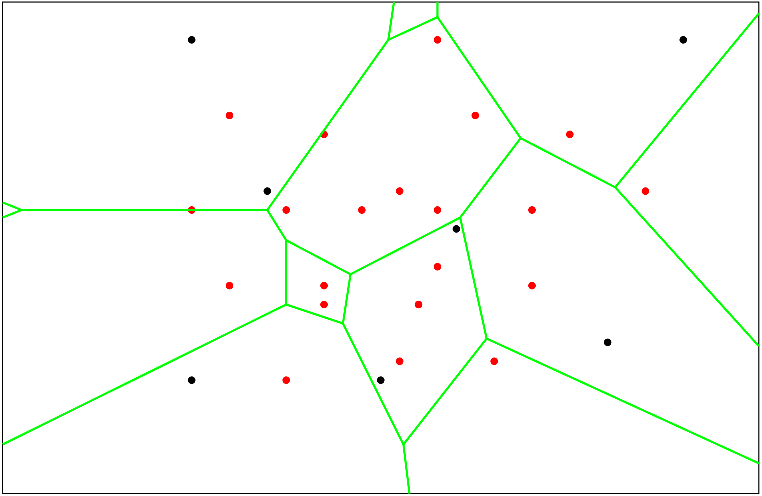
$$
\begin{aligned}
d_{\mu,m}(x)^2 &= \sum_{i=1}^{k} ||x - p_i(x)||^2 \\
&= ||x - bar(x)||^2 + \sum_{i=1}^{k} ||p_i(x) - bar(x)||^2 \\
&= ||x - bar(x)||^2 + w_{bar(x)}^2 \\
&= \min_{b \in B} \left(||x - b||^2 + w_b^2 \right)
\end{aligned}
$$

where $bar(x) = \sum_{i=1}^{k} p_i(x)$ and B is the set of all barycentres of k points.

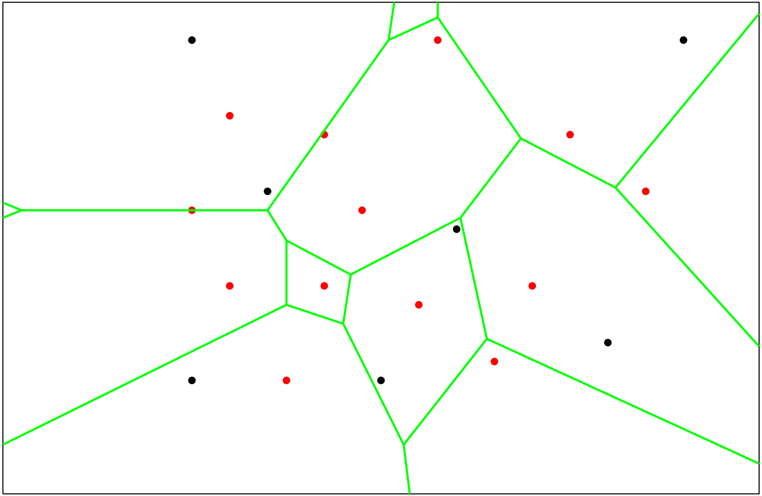
k^{th}-order Voronoi diagram



k^{th}-order Voronoi diagram



k^{th}-order Voronoi diagram



Size of k^{th}-order Voronoi diagram

Sub-level sets of $d_{\mu,m}$ are unions of balls.

$$d_{\mu,m}^{-1}(]-\infty,\alpha]) = \bigcup_{b \in B} \bar{B}(b, \sqrt{\alpha^2 - w_b^2})$$

Theorem (Clarkson, Shor, 1989)

The number of non-empty cells can be as large as:

$$O\left(n^{\lfloor \frac{d+1}{2} \rfloor} k^{\lceil \frac{d+1}{2} \rceil}\right).$$

Computation of persistence diagrams

Distance to the empirical measure and weighted Rips filtration:

- Complexity: $O(n^{\left\lfloor \frac{3(d+1)}{2} \right\rfloor} k^{\left\lceil \frac{d+1}{2} \right\rceil})$
- Robustness to outliers: yes

Witnessed k-distance

Approximation by sampling barycentres.

Definition (Guibas, Mérigot, Morozov, 2011)

$$d^{W}_{\mu,m}(x) = \min_{p \in P} \sqrt{||x - bar(p)||^2 + w^2_{bar(p)}}$$

Witnessed *k*-distance

Approximation by sampling barycentres.

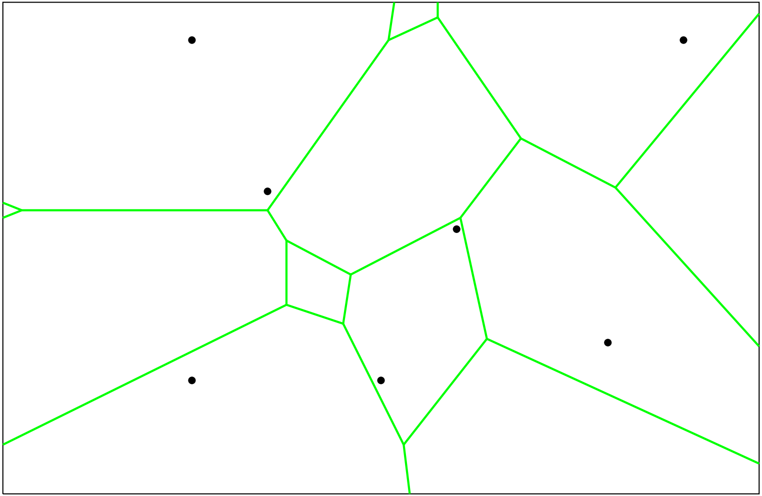
Definition (Guibas, Mérigot, Morozov, 2011)

$$d^W_{\mu,m}(x) = \min_{p \in P} \sqrt{||x - bar(p)||^2 + w^2_{bar(p)}}$$

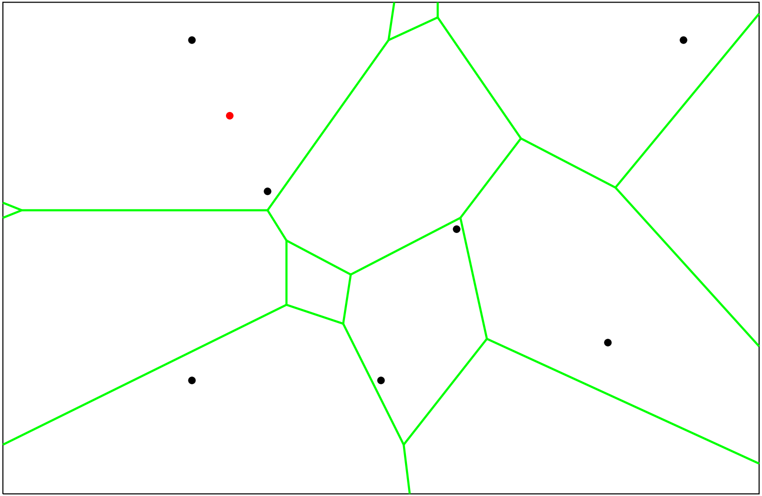
Theorem

$$d_{\mu,m} \leq d^W_{\mu,m} \leq \sqrt{6} d_{\mu,m}$$

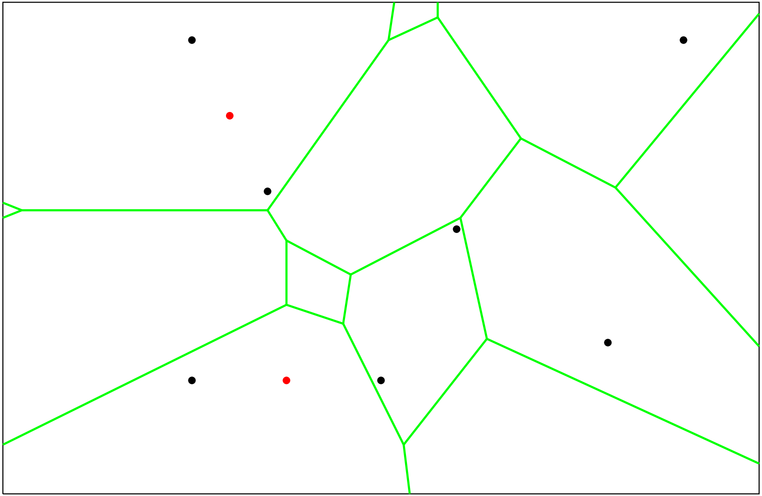
Building the witnessed k-distance



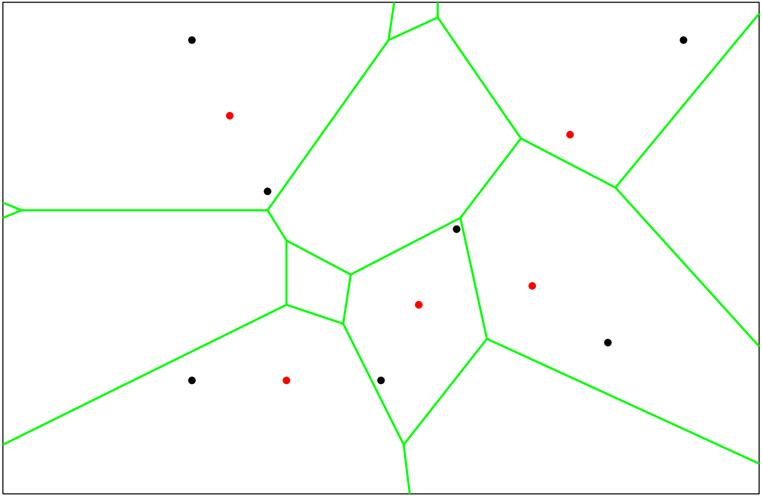
Building the witnessed *k*-distance



Building the witnessed *k*-distance



Building the witnessed k-distance



Approximation supported by the points

Using a power distance supported by input points.

Definition

$$d^{P}_{\mu,m}(x) = \min_{p \in P} \sqrt{d_{\mathbb{X}}(x,p)^2 + d_{\mu,m}(p)^2}$$

Approximation supported by the points

Using a power distance supported by input points.

Definition

$$d^P_{\mu,m}(x) = \min_{p \in P} \sqrt{d_{\mathbb{X}}(x,p)^2 + d_{\mu,m}(p)^2}$$

Theorem

In Euclidean space:

$$\frac{1}{\sqrt{2}} d_{\mu,m} \leq d^P_{\mu,m} \leq \sqrt{3} d_{\mu,m}$$

Approximation supported by the points

Using a power distance supported by input points.

Definition

$$d^P_{\mu,m}(x) = \min_{p \in P} \sqrt{d_{\mathbb{X}}(x,p)^2 + d_{\mu,m}(p)^2}$$

Theorem

In Euclidean space:

$$\frac{1}{\sqrt{2}} d_{\mu,m} \leq d^P_{\mu,m} \leq \sqrt{3} d_{\mu,m}$$

In any metric space:

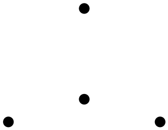
$$\frac{1}{\sqrt{2}} d_{\mu,m} \leq d^P_{\mu,m} \leq \sqrt{5} d_{\mu,m}$$

Computation of persistence diagrams

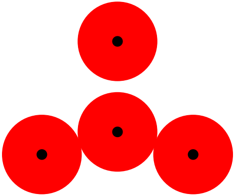
Approximation to the distance to the empirical measure and weighted Rips filtration:

- Complexity: $O(n^{d+1})$
- Robustness to outliers: yes

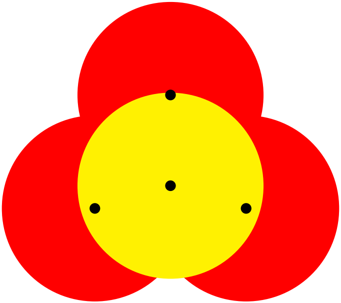
Sparse Rips



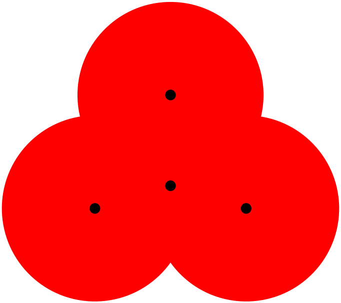
Sparse Rips



Sparse Rips

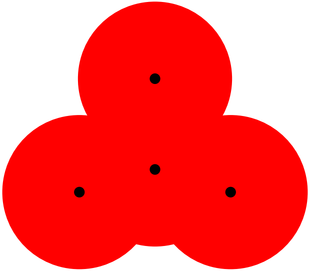


Sparse Rips



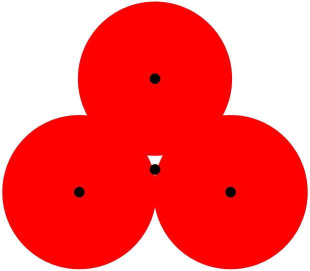
Topological noise

Naively removing points can create topological noise.



Topological noise

Naively removing points can create topological noise.



Construction (I)

- Let $(p_1, \ldots, p_n)$ be a futhest point sampling of P.

Construction (I)

- Let $(p_1, \ldots, p_n)$ be a futhest point sampling of P.
  - p_1 is arbitrary and $\lambda_1 = \infty$.

Construction (I)

- Let $(p_1, \ldots, p_n)$ be a futhest point sampling of P.
  - p_1 is arbitrary and $\lambda_1 = \infty$.
  - $p_i = \mathrm{argmax}_{p \in P \setminus P_{i-1}} d_{\mathbb{X}}(p, P_{i-1},)$ and $P_{i-1} = (p_1, \ldots, p_{i-1})$.

Construction (I)

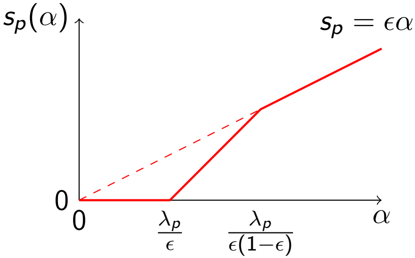
- Let $(p_1, \ldots, p_n)$ be a futhest point sampling of P.
    - p_1 is arbitrary and $\lambda_1 = \infty$.
    - $p_i = \mathrm{argmax}_{p \in P \setminus P_{i-1}} d_{\mathbb{X}}(p, P_{i-1},)$ and $P_{i-1} = (p_1, \ldots, p_{i-1})$.
    - Fix $\lambda_i = d_{\mathbb{X}}(p_i, P_{i-1})$.

Construction (I)

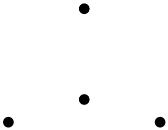
- Let $(p_1, \ldots, p_n)$ be a futhest point sampling of P.
  - p_1 is arbitrary and $\lambda_1 = \infty$.
  - $p_i = \mathrm{argmax}_{p \in P \setminus P_{i-1}} d_{\mathbb{X}}(p, P_{i-1},)$ and $P_{i-1} = (p_1, \ldots, p_{i-1})$.
  - Fix $\lambda_i = d_{\mathbb{X}}(p_i, P_{i-1})$.
- Fix $\bar{N}_\gamma = \{p \in P | \lambda_p \geq \gamma\}$.

Construction (I)

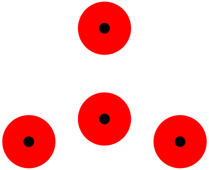
- Let $(p_1, \ldots, p_n)$ be a futhest point sampling of P.
  - p_1 is arbitrary and $\lambda_1 = \infty$.
  - $p_i = \mathrm{argmax}_{p \in P \setminus P_{i-1}} d_{\mathbb{X}}(p, P_{i-1},)$ and $P_{i-1} = (p_1, \ldots, p_{i-1})$.
  - Fix $\lambda_i = d_{\mathbb{X}}(p_i, P_{i-1})$.
- Fix $\bar{N}_\gamma = \{p \in P | \lambda_p \geq \gamma\}$.
- Perturbated metric : $f_\alpha(p,q) = d_{\mathbb{X}}(p,q) + s_p(\alpha) + s_q(\alpha)$.



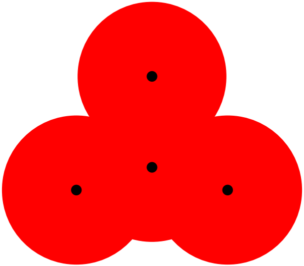
Effects of perturbated metric



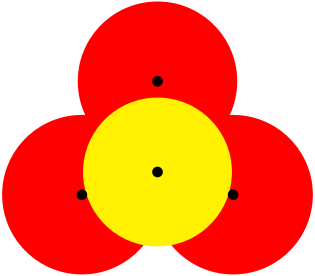
Effects of perturbated metric



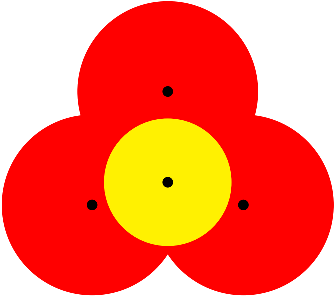
Effects of perturbated metric



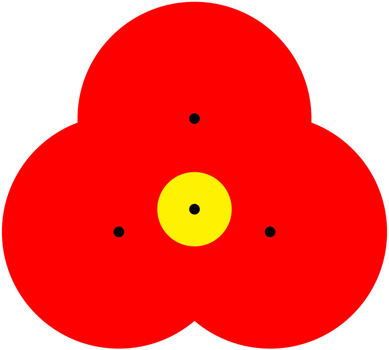
Effects of perturbated metric



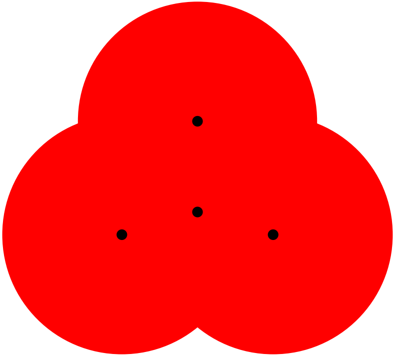
Effects of perturbated metric



Effects of perturbated metric



Effects of perturbated metric



Construction (II)

Definition

The sparse Rips complex is given by:

$$Q_\alpha = \{\sigma \subset \bar{N}_{\epsilon(1-\epsilon)\alpha} | \forall p, q \in \sigma,\ f_\alpha(p, q) < 2\alpha\}$$

Construction (II)

Definition

The sparse Rips complex is given by:

$$Q_\alpha = \{\sigma \subset \bar{N}_{\epsilon(1-\epsilon)\alpha} | \forall p, q \in \sigma,\ f_\alpha(p,q) < 2\alpha\}$$

Definition

The sparse Rips filtration is given by:

$$S_\beta = \bigcup_{\alpha \leq \beta} Q_\alpha.$$

Properties of sparse Rips

Theorem

$\{S_\alpha\}$ contains $O(C^l n)$ simplexes where l is the intrinsic dimension of the underlying object.

Theorem

$\{S_\alpha\}$ is $\frac{1}{1-\epsilon}$-interleaved with the Rips filtration $\{R_\alpha\}$.

Using the weighted metric

We used a weighted Rips filtration to compute the persistence diagram:

$$R_\alpha = \left\{\sigma \subset P | \forall p, q \in P,\ d_{\mathbb{X}}(p, q) \leq \sqrt{\alpha^2 - w_p^2} + \sqrt{\alpha^2 - w_q^2}\right\}$$

Using the weighted metric

We used a weighted Rips filtration to compute the persistence diagram:

$$R_\alpha = \left\{\sigma \subset P | \forall p, q \in P,\ d_{\mathbb{X}}(p,q) \leq \sqrt{\alpha^2 - w_p^2} + \sqrt{\alpha^2 - w_q^2}\right\}$$

This structure induces a metric $\tilde{d}$.

Using the weighted metric

We used a weighted Rips filtration to compute the persistence diagram:

$$R_\alpha = \left\{\sigma \subset P | \forall p, q \in P,\ d_{\mathbb{X}}(p, q) \leq \sqrt{\alpha^2 - w_p^2} + \sqrt{\alpha^2 - w_q^2}\right\}$$

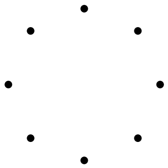
This structure induces a metric $\tilde{d}$.
Replacing $d_{\mathbb{X}}$ by $\tilde{d}$ causes loss of properties on the size of the sparse filtration.

Using the weighted metric

We used a weighted Rips filtration to compute the persistence diagram:

$$R_\alpha = \left\{ \sigma \subset P | \forall p, q \in P,\ d_{\mathbb{X}}(p,q) \leq \sqrt{\alpha^2 - w_p^2} + \sqrt{\alpha^2 - w_q^2} \right\}$$

This structure induces a metric $\tilde{d}$.
Replacing $d_{\mathbb{X}}$ by $\tilde{d}$ causes loss of properties on the size of the sparse filtration.



Adapting to weighted Rips

Definition

The sparse weighted RIps is defined by:

$$T_\alpha = R_\alpha \bigcap S_\alpha.$$

Adapting to weighted Rips

Definition

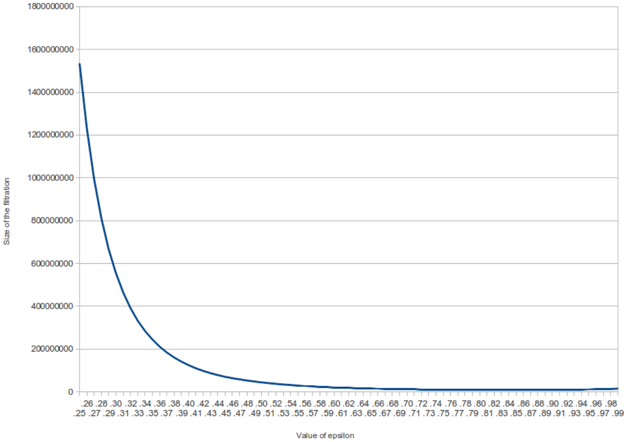
The sparse weighted RIps is defined by:

$$T_\alpha = R_\alpha \bigcap S_\alpha.$$

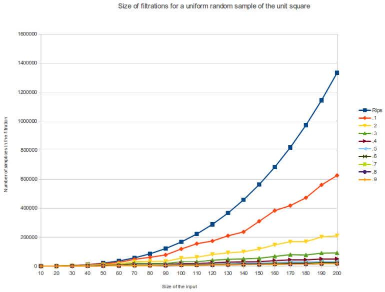
Theorem

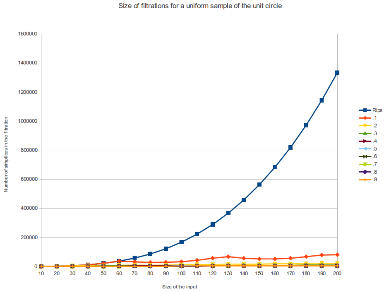
R_α and T_α are κ-interleaved where $\kappa = 1 + \frac{\sqrt{1+t^2\epsilon}}{1-\epsilon}$, id est :

SIze of the filtration depending on ϵ

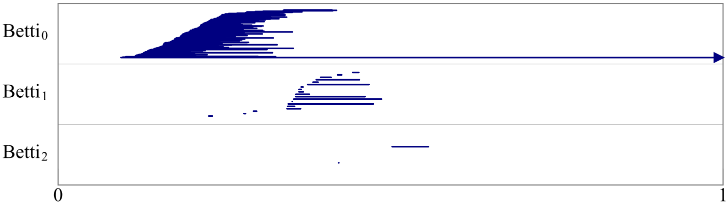


Intrinsic dimension influence





Cube skeleton with Gaussian noise



Take home

- The distance to a measure can handle outliers.

Take home

- The distance to a measure can handle outliers.
- The sparse wieghted Rips filtration and the approximation to the distance to a measure make computing its peristence diagram tractable.

Efficient and Robust Persistent Homology for Measures

Mickaël Buchet

joint work with F. Chazal, S. Oudot and D. Sheehy