Stable vectorization of Multiparameter Persistent Homology using Signed Barcodes as Measures

joint work with L.Scoccola, M. Carrière, S. Oudot, and M.Botnan

Young Topologists Meeting

EPFL

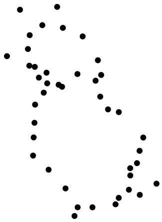
Inria

UNIVERSITÉ CÔTE D'AZUR | INTERDISCIPLINARY INSTITUTE FOR ARTIFICIAL INTELLIGENCE 3IA CÔTE D'AZUR

FRANCE 2030

Topological Data Analysis introduction

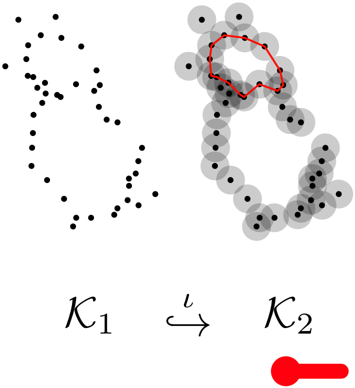
Persistent homology, for Čech's filtration



$\mathcal{K}_1$

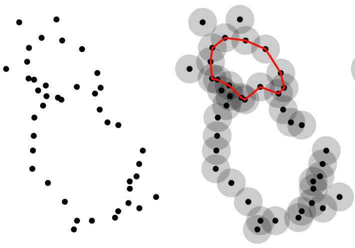
Topological Data Analysis introduction

Persistent homology, for Čech's filtration



Topological Data Analysis introduction

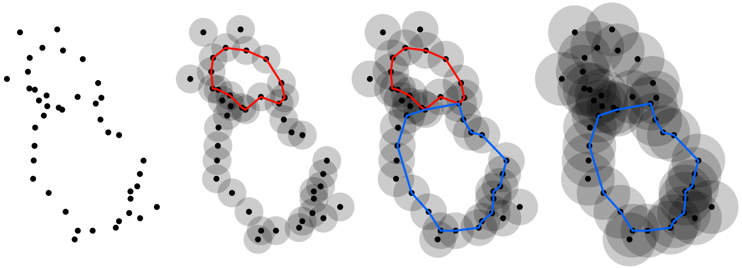
Persistent homology, for Čech's filtration



$$\mathcal{K}_1 \overset{\iota}{\hookrightarrow} \mathcal{K}_2 \overset{\iota}{\hookrightarrow} \mathcal{K}_3$$

Topological Data Analysis introduction

Persistent homology, for Čech's filtration



$$\mathcal{K}_1 \overset{\iota}{\hookrightarrow} \mathcal{K}_2 \overset{\iota}{\hookrightarrow} \mathcal{K}_3 \overset{\iota}{\hookrightarrow} \mathcal{K}_4$$

Topological Data Analysis introduction

Formally

Topological Filtration

$$X_t = \{x \in \mathbb{R}^n : d(x,P) \leq t\} = \bigcup_{p \in P} B(x,t)$$

$$X = \left\{ (P = X_0) \overset{\iota}{\hookrightarrow} \cdots \overset{\iota}{\hookrightarrow} X_{r_1} \overset{\iota}{\hookrightarrow} \cdots \overset{\iota}{\hookrightarrow} X_{r_2} \overset{\iota}{\hookrightarrow} \cdots \right\}$$

Topological Data Analysis introduction

Formally

Topological Filtration

$$X_t = \{x \in \mathbb{R}^n : d(x,P) \leq t\} = \bigcup_{p \in P} B(x,t)$$

$$X = \left\{ (P = X_0) \overset{\iota}{\hookrightarrow} \cdots \overset{\iota}{\hookrightarrow} X_{r_1} \overset{\iota}{\hookrightarrow} \cdots \overset{\iota}{\hookrightarrow} X_{r_2} \overset{\iota}{\hookrightarrow} \cdots \right\}$$

Persistent Homology or Persistent Module

$$PH_\bullet(X) = \left\{ H_\bullet(X_{r_0}) \overset{\iota_\star}{\to} \cdots \overset{\iota_\star}{\to} H_\bullet(X_{r_1}) \overset{\iota_\star}{\to} \cdots \overset{\iota_\star}{\to} H_\bullet(X_{r_2}) \overset{\iota_\star}{\to} \cdots \right\}$$

Topological Data Analysis introduction

Formally

Topological Filtration

$$X_t = \{x \in \mathbb{R}^n : d(x,P) \leq t\} = \bigcup_{p \in P} B(x,t)$$

$$X = \left\{(P = X_0) \overset{\iota}{\hookrightarrow} \cdots \overset{\iota}{\hookrightarrow} X_{r_1} \overset{\iota}{\hookrightarrow} \cdots \overset{\iota}{\hookrightarrow} X_{r_2} \overset{\iota}{\hookrightarrow} \cdots \right\}$$

Persistent Homology or Persistent Module

$$PH_\bullet(X) = \left\{H_\bullet(X_{r_0}) \overset{\iota_\star}{\to} \cdots \overset{\iota_\star}{\to} H_\bullet(X_{r_1}) \overset{\iota_\star}{\to} \cdots \overset{\iota_\star}{\to} H_\bullet(X_{r_2}) \overset{\iota_\star}{\to} \cdots \right\}$$

In general, a Persistent Module is a family of *vector spaces* $\mathbb{V} = (V_t)_{t \in \mathbb{R}}$, with *linear maps* $v_{s \to t} : V_s \to V_t$,

Topological Data Analysis introduction

Formally

Topological Filtration

$$X_t = \{x \in \mathbb{R}^n : d(x, P) \leq t\} = \bigcup_{p \in P} B(x, t)$$

$$X = \left\{(P = X_0) \overset{\iota}{\hookrightarrow} \cdots \overset{\iota}{\hookrightarrow} X_{r_1} \overset{\iota}{\hookrightarrow} \cdots \overset{\iota}{\hookrightarrow} X_{r_2} \overset{\iota}{\hookrightarrow} \cdots\right\}$$

Persistent Homology or Persistent Module

$$PH_\bullet(X) = \left\{H_\bullet(X_{r_0}) \overset{\iota_\star}{\to} \cdots \overset{\iota_\star}{\to} H_\bullet(X_{r_1}) \overset{\iota_\star}{\to} \cdots \overset{\iota_\star}{\to} H_\bullet(X_{r_2}) \overset{\iota_\star}{\to} \cdots\right\}$$

In general, a Persistent Module is a family of *vector spaces* $\mathbb{V} = (V_t)_{t \in \mathbb{R}}$, with *linear maps* $v_{s \to t} : V_s \to V_t$, s.t.

$$v_{s \to r} \circ v_{t \to s} = v_{t \to r} \quad \text{and} \quad v_{t \to t} \equiv \mathrm{id}_{V_t}.$$

Pillars of Topological Data Analysis

Theorem (Krull,Remak,Schmidt,Azumaya,Gabriel)

Let M be a pointwise finite dimensional persistence module. Then, M is interval-decomposable, i.e.,

$$M \simeq \bigoplus_i k_{[b_i,d_i)}.$$

Pillars of Topological Data Analysis

Theorem (Krull,Remak,Schmidt,Azumaya,Gabriel)

Let M be a pointwise finite dimensional persistence module.
Then, M is interval-decomposable, i.e.,

$$M \simeq \bigoplus_i k_{[b_i,d_i)}.$$

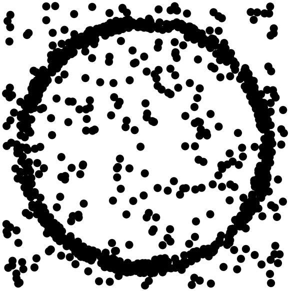
Theorem (Cohen-Steiner, Edelsbrunner, Harer)

If X is a triangulable space, with continuous tame functions $f, g\colon X \to \mathbb{R}$,

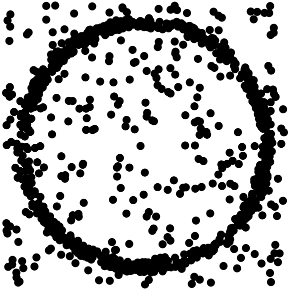
$$d_I(H_\star(X_f), H_\star(X_g)) \leq \|f - g\|_\infty$$

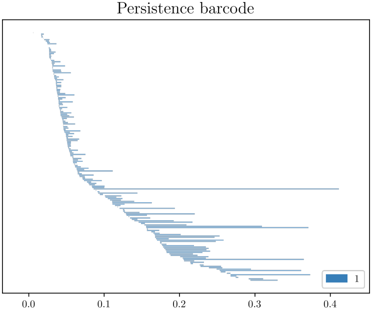
where $X_h := (\{x \in X \mid h(x) \leq t\})_{t\in\mathbb{R}}$.

Easy multi-persistence motivation : noise

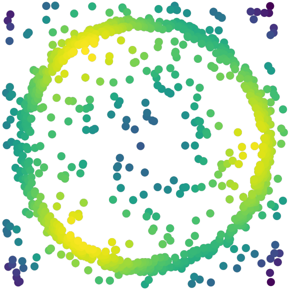


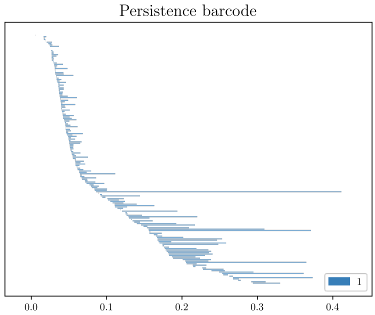
Easy multi-persistence motivation : noise



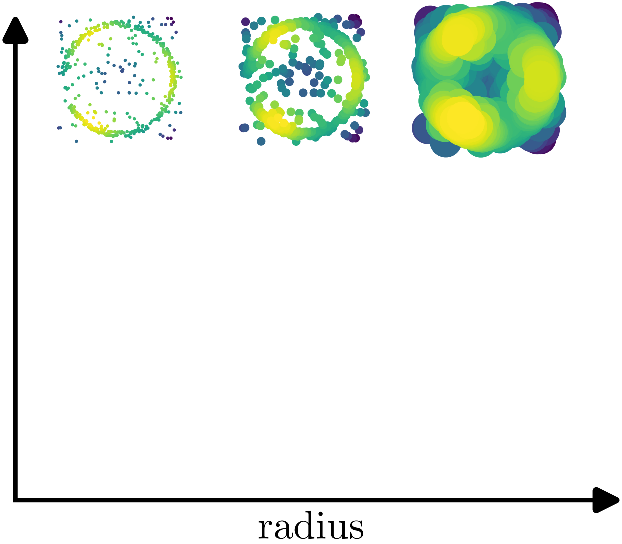


Easy multi-persistence motivation : noise

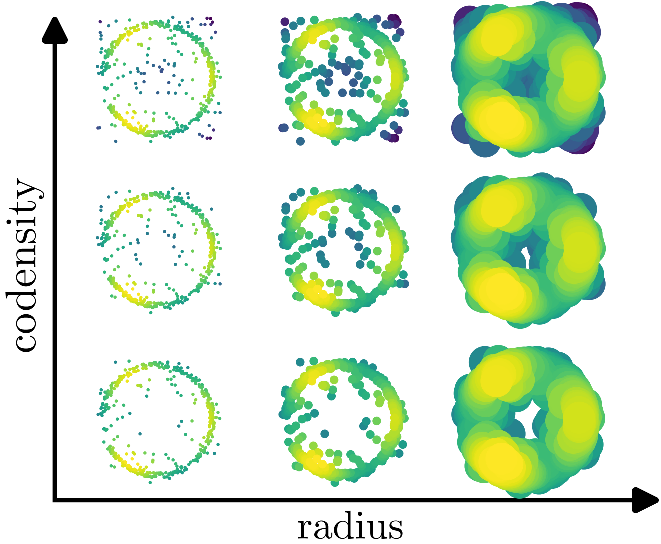




2-persistence with noise 101



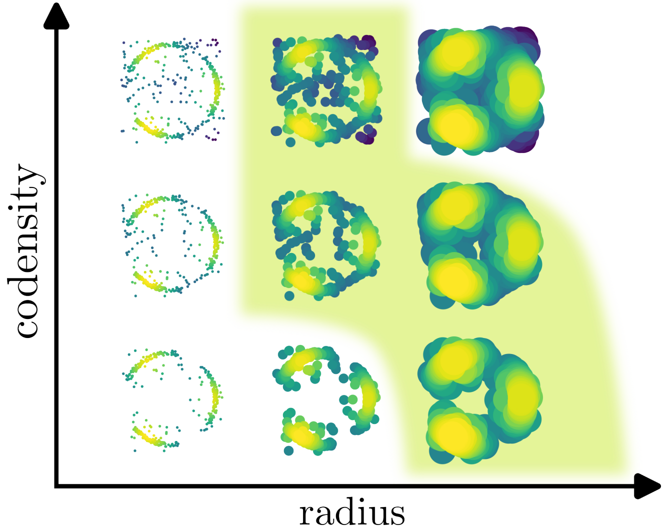
2-persistence with noise 101



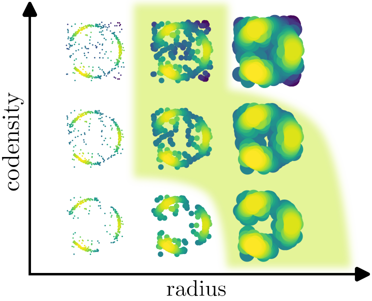
2-persistence with noise 101

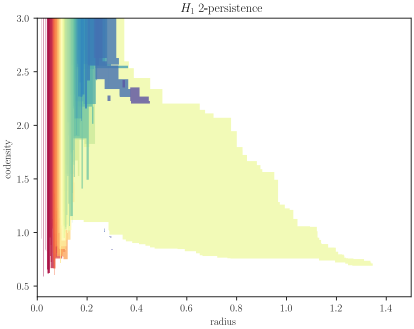
$$
\begin{array}{ccccccccc}
\mathcal{K}_1^k & \xhookrightarrow{\iota} & \mathcal{K}_2^k & \xhookrightarrow{\iota} & \mathcal{K}_3^k & \xhookrightarrow{\iota} & \cdots & \xhookrightarrow{\iota} & \mathcal{K}_n^k \\
\iota\big\uparrow & & \iota\big\uparrow & & \iota\big\uparrow & & \iota\big\uparrow & & \iota\big\uparrow \\
\cdots & \xhookrightarrow{\iota} & \cdots & \xhookrightarrow{\iota} & \cdots & \xhookrightarrow{\iota} & \cdots & \xhookrightarrow{\iota} & \cdots \\
\iota\big\uparrow & & \iota\big\uparrow & & \iota\big\uparrow & & \iota\big\uparrow & & \iota\big\uparrow \\
\mathcal{K}_1^2 & \xhookrightarrow{\iota} & \mathcal{K}_2^2 & \xhookrightarrow{\iota} & \mathcal{K}_3^2 & \xhookrightarrow{\iota} & \cdots & \xhookrightarrow{\iota} & \mathcal{K}_n^2 \\
\iota\big\uparrow & & \iota\big\uparrow & & \iota\big\uparrow & & \iota\big\uparrow & & \iota\big\uparrow \\
\mathcal{K}_1 & \xhookrightarrow{\iota} & \mathcal{K}_2 & \xhookrightarrow{\iota} & \mathcal{K}_3 & \xhookrightarrow{\iota} & \cdots & \xhookrightarrow{\iota} & \mathcal{K}_n
\end{array}
$$

2-persistence with noise 101

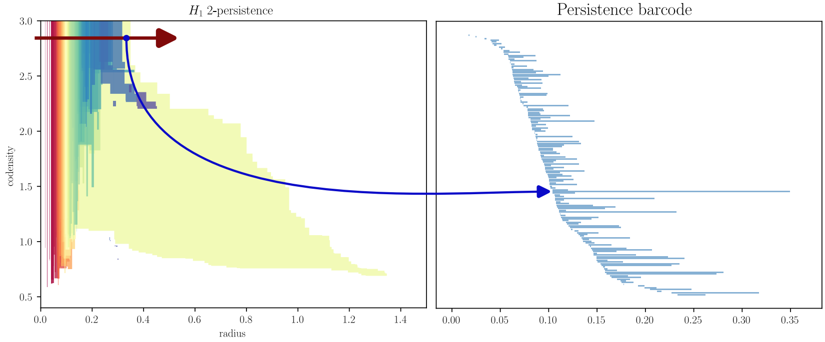


2-persistence with noise 101



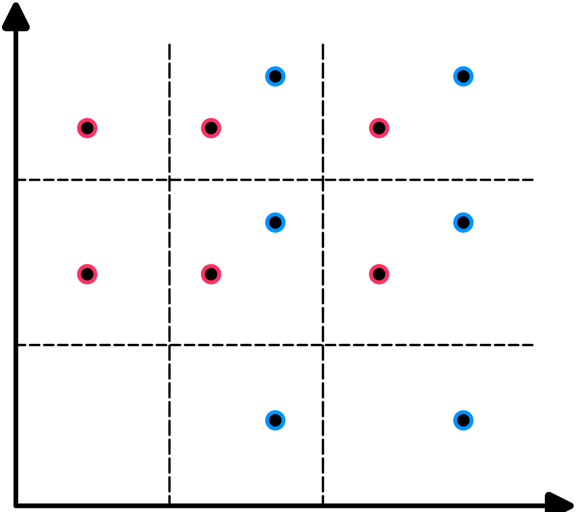


2-persistence with noise 101



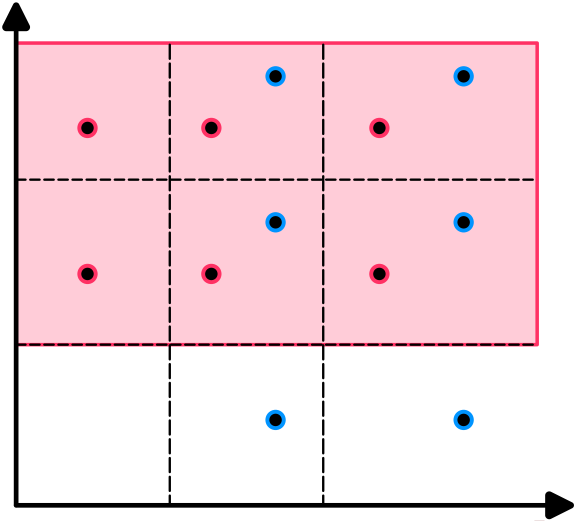
Multiparameter curse

Connected components



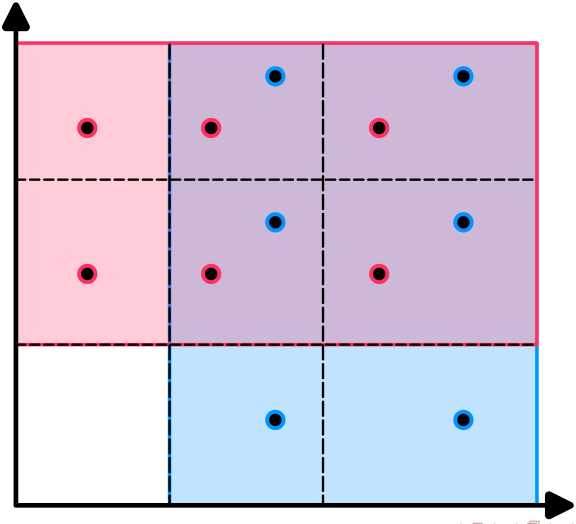
Multiparameter curse

Connected components



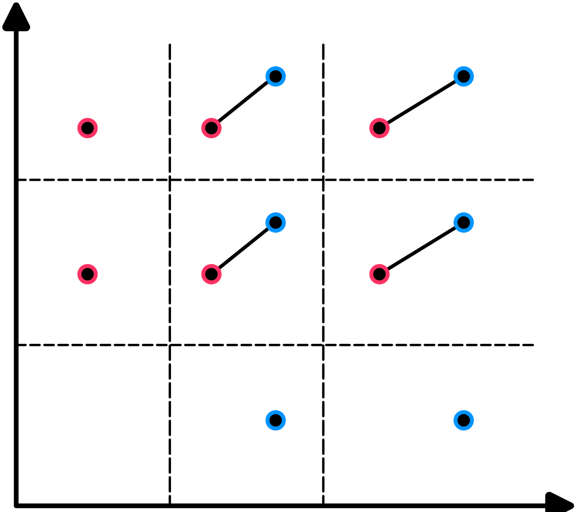
Multiparameter curse

Connected components



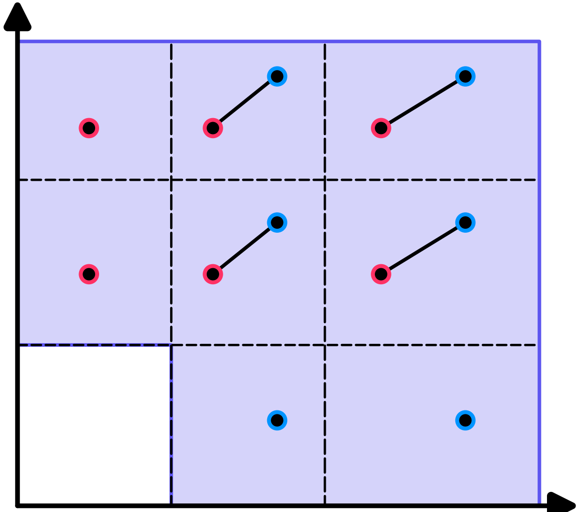
Multiparameter curse

Connected components



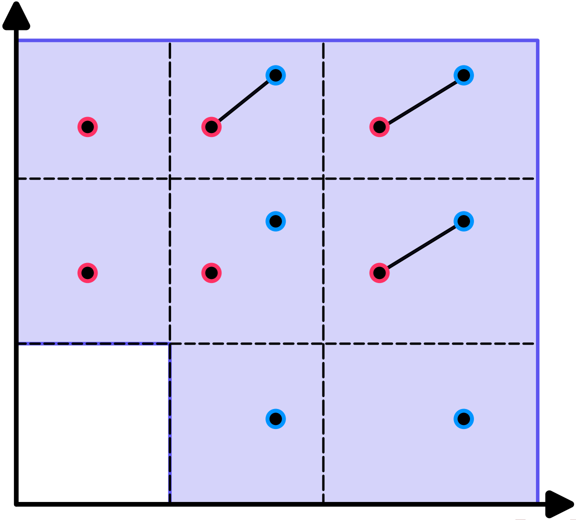
Multiparameter curse

Connected components



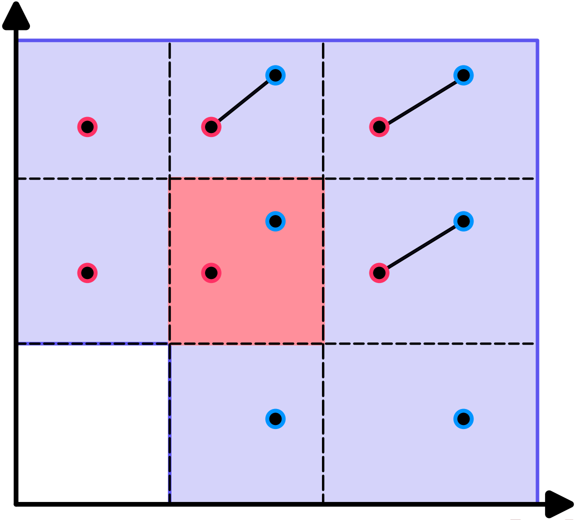
Multiparameter curse

Connected components

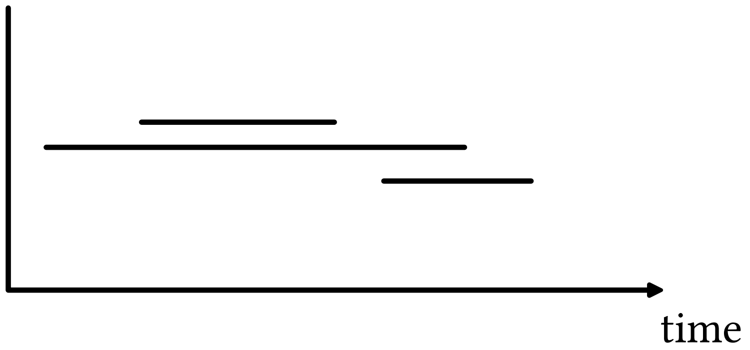


Multiparameter curse

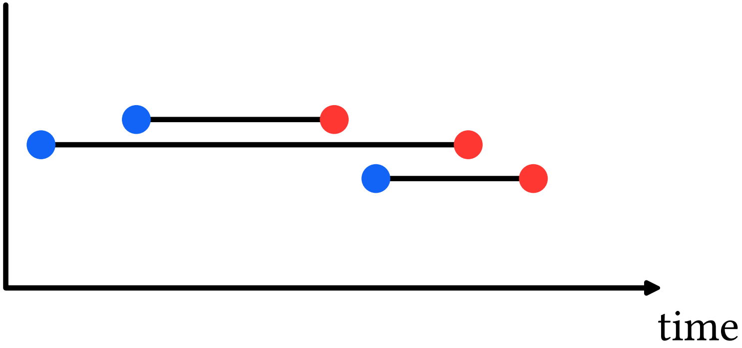
Connected components



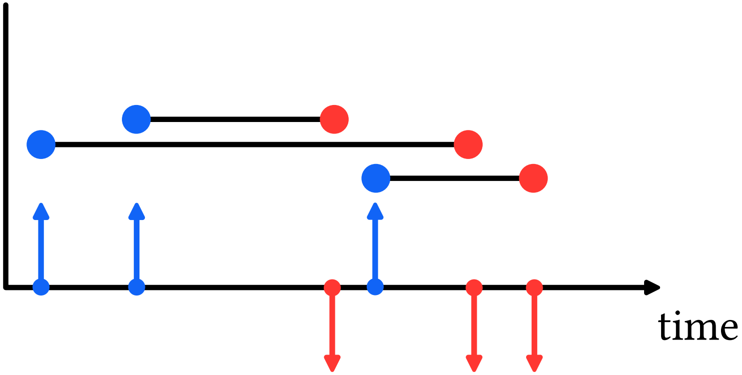
Let's forget about this matching : 1-parameter pictures



Let’s forget about this matching : 1-parameter pictures

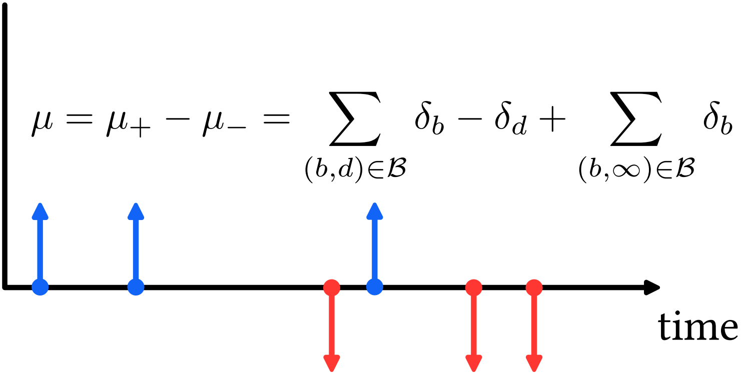


Let’s forget about this matching : 1-parameter pictures

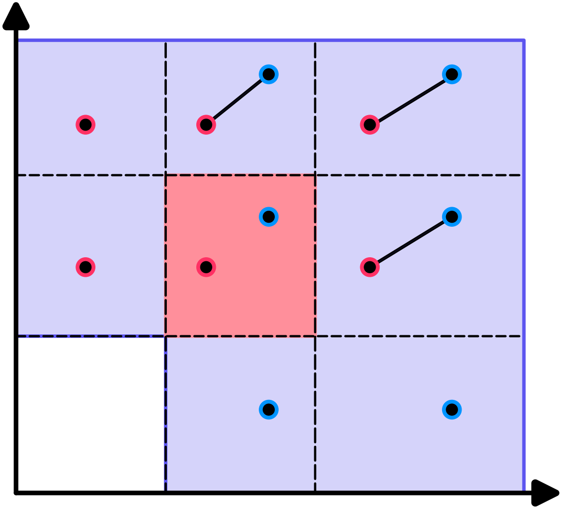


Let's forget about this matching : 1-parameter pictures

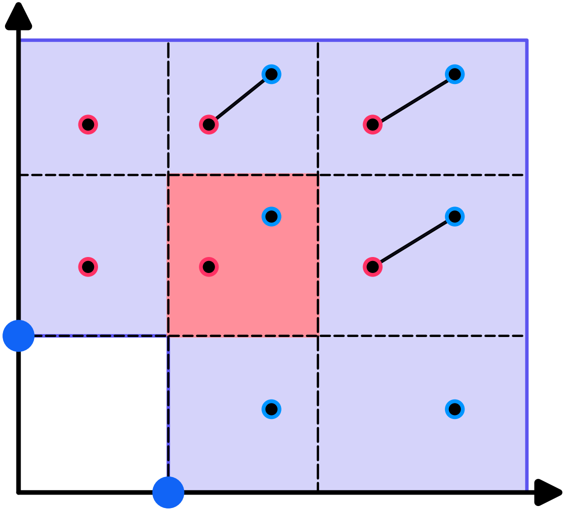
$$\mu = \mu_+ - \mu_- = \sum_{(b,d)\in\mathcal{B}} \delta_b - \delta_d + \sum_{(b,\infty)\in\mathcal{B}} \delta_b$$



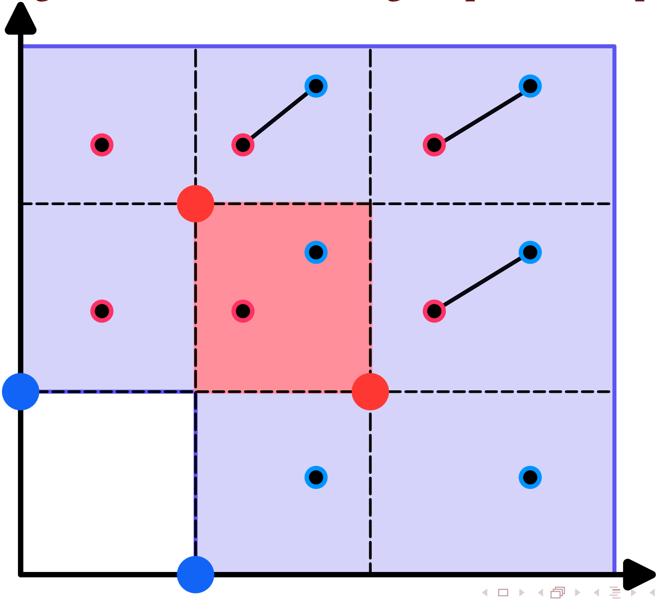
Let’s forget about this matching : 2-parameter pictures



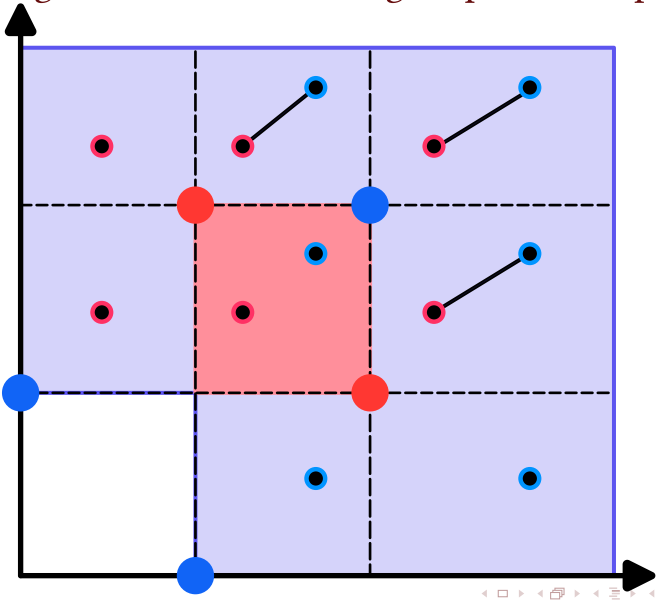
Let's forget about this matching : 2-parameter pictures



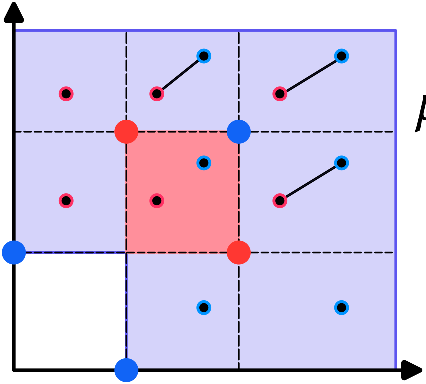
Let's forget about this matching : 2-parameter pictures



Let's forget about this matching : 2-parameter pictures



Let's forget about this matching : 2-parameter pictures



$$\mu = \delta_{(1,0)} + \delta_{(0,1)} - \delta_{(2,1)} - \delta_{(1,2)} + \delta_{(2,2)}$$

Re-extension to Multiparameter Persistence

Definition (Hilbert decomposition signed measure)

For all *finitely presentable* multiparameter persistence module M, there *exists* a *unique discrete radon* measure $\boldsymbol{\mu}_M$ such that

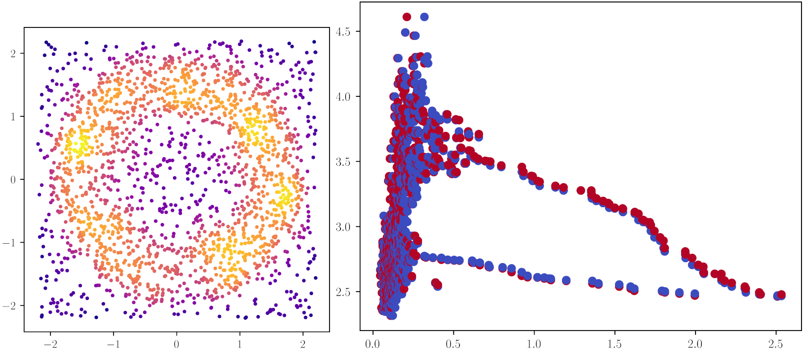
$$\forall x \in \mathbb{R}^n, \quad \dim(M_x) = \mu_M\left(\{y \in \mathbb{R}^n : y \leq x\}\right)$$

Re-extension to Multiparameter Persistence

Definition (Hilbert decomposition signed measure)

For all *finitely presentable* multiparameter persistence module M, there *exists* a *unique discrete radon* measure μ_M such that

$$\forall x \in \mathbb{R}^n, \quad \dim(M_x) = \mu_M\left(\{y \in \mathbb{R}^n : y \leq x\}\right)$$

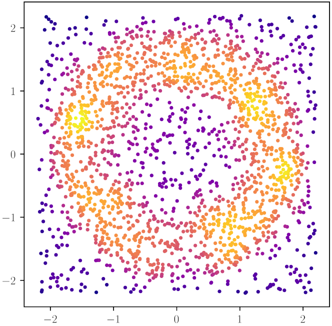


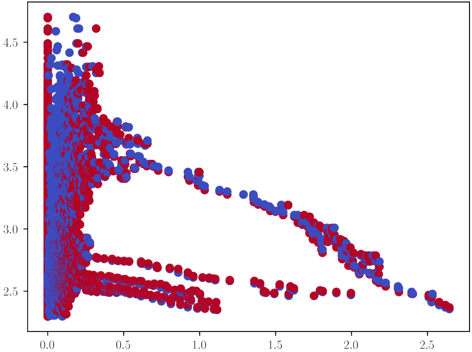
Re-extension to Multiparameter Persistence

Definition (Euler decomposition signed measure)

For a *finite* multi-filtered simplicial complex (S, f),

$$\mu_{\chi(f)} := \sum_{i \in \mathbb{N}} (-1)^i \mu_{H_i(f)}.$$



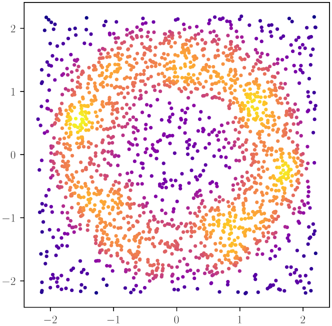


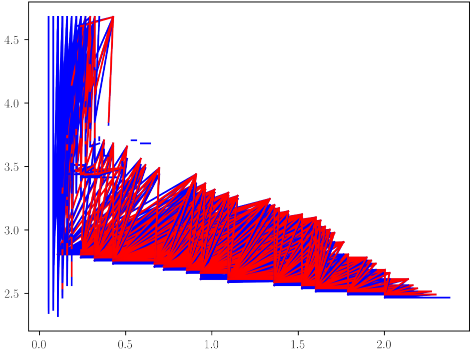
Re-extension to Multiparameter Persistence

Definition (Rank decomposition signed measure)

For all *finitely presentable* multiparameter persistence module M, there *exists* a *unique discrete radon* measure $\boldsymbol{\mu}_M$ such that

$$\forall x, y \in \mathbb{R}^n, \quad \dim(M(x \leq y)) = \mu_M\left(\{z \in \mathbb{R}^n : z \leq x\} \times \{z \in \mathbb{R}^n : z \leq y\}\right)$$





Distances

Let $\boldsymbol{\mu} = \mu_+ - \mu_-$ be a zero mass discrete measure.

Distances

Let $\boldsymbol{\mu} = \mu_+ - \mu_-$ be a zero mass discrete measure.
Define, if $\mu_+ = \sum_i \delta_{x_i}$ and $\mu_- = \sum_i \delta_{y_i}$

$$\|\boldsymbol{\mu}\| :=$$

Distances

Let $\boldsymbol{\mu} = \mu_+ - \mu_-$ be a zero mass discrete measure.
Define, if $\mu_+ = \sum_i \delta_{x_i}$ and $\mu_- = \sum_i \delta_{y_i}$

$$\|\boldsymbol{\mu}\| := d_{W^1}(\mu_+, \mu_-) = \min_{\sigma \in \mathfrak{S}_{|\mu_+|}} \left\{ \sum_{1 \leq i \leq |\mu_+|} \left\| x_i - y_{\sigma(i)} \right\|_1 \right\}$$

Distances

Let $\boldsymbol{\mu} = \mu_+ - \mu_-$ be a zero mass discrete measure.
Define, if $\mu_+ = \sum_i \delta_{x_i}$ and $\mu_- = \sum_i \delta_{y_i}$

$$\|\boldsymbol{\mu}\| := d_{W^1}(\mu_+, \mu_-) = \min_{\sigma \in \mathfrak{S}_{|\mu_+|}} \left\{ \sum_{1 \leq i \leq |\mu_+|} \left\| x_i - y_{\sigma(i)} \right\|_1 \right\}$$

If $\boldsymbol{\nu}$ is another such measure, then

$$\boldsymbol{\mu} - \boldsymbol{\nu} = (\mu_+ + \nu_-) - (\mu_- + \nu_+).$$

Distances

Let $\boldsymbol{\mu} = \mu_+ - \mu_-$ be a zero mass discrete measure.
Define, if $\mu_+ = \sum_i \delta_{x_i}$ and $\mu_- = \sum_i \delta_{y_i}$

$$\|\boldsymbol{\mu}\| := d_{W^1}(\mu_+, \mu_-) = \min_{\sigma \in \mathfrak{S}_{|\mu_+|}} \left\{ \sum_{1 \leq i \leq |\mu_+|} \left\| x_i - y_{\sigma(i)} \right\|_1 \right\}$$

If $\boldsymbol{\nu}$ is another such measure, then

$$\boldsymbol{\mu} - \boldsymbol{\nu} = (\mu_+ + \nu_-) - (\mu_- + \nu_+).$$

In particular, we can consider *Wasserstein* norms on signed measures

$$\|\boldsymbol{\mu} - \boldsymbol{\nu}\|_1^K := d_{W^1}(\mu_+ + \nu_-, \mu_- + \nu_+)$$

Stability

Theorem

Let $\mathbf{n} \in \mathbb{N}$, S be a finite simplicial complex and (S, f), (S, g) two $\mathbf{n}$-filtrations.

1. *For $\mathbf{n} \in \{1, 2\}$ and $i \in \mathbb{N}$,*

$$\left\| \mu_{H_i(f)} - \mu_{H_i(g)} \right\|_1^K \leq \mathbf{n} \cdot \|f - g\|_1$$

2. *For all $\mathbf{n} \in \mathbb{N}$,*

$$\left\| \mu_{\chi(f)} - \mu_{\chi(g)} \right\|_1^K \leq \|f - g\|_1$$

Stability

Theorem

Let $\boldsymbol{n} \in \mathbb{N}$, S be a finite simplicial complex and (S, f), (S, g) two $\boldsymbol{n}$-filtrations.

1. *For $\boldsymbol{n} \in \{1, 2\}$ and $i \in \mathbb{N}$,*

$$\left\| \mu_{H_i(f)} - \mu_{H_i(g)} \right\|_1^K \leq \boldsymbol{n} \cdot \|f - g\|_1$$

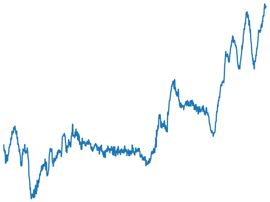
2. *For all $\boldsymbol{n} \in \mathbb{N}$,*

$$\left\| \mu_{\chi(f)} - \mu_{\chi(g)} \right\|_1^K \leq \|f - g\|_1$$

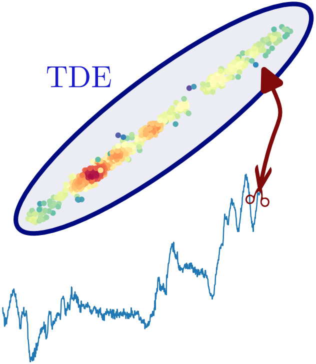
where

$$\|f - g\|_1 = \sum_{\sigma \in S} \|f(\sigma) - g(\sigma)\|_1 .$$

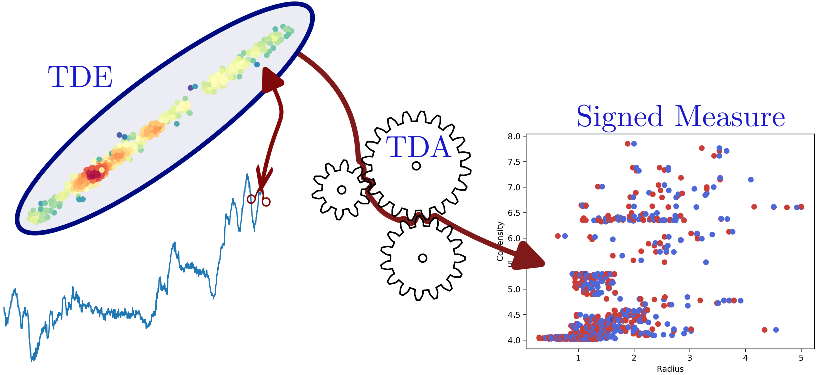
Convolutions



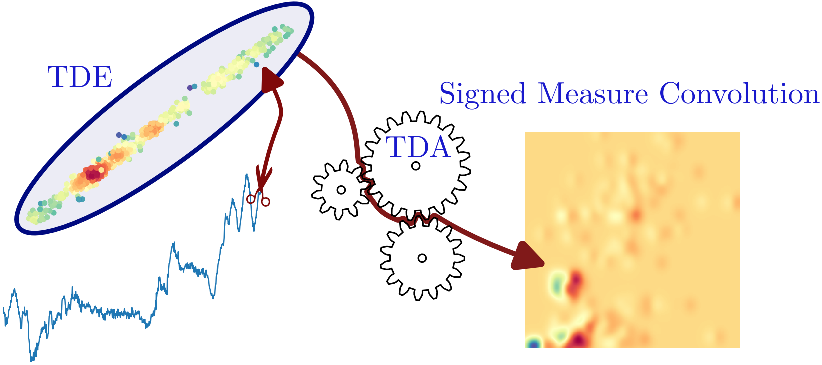
Convolutions



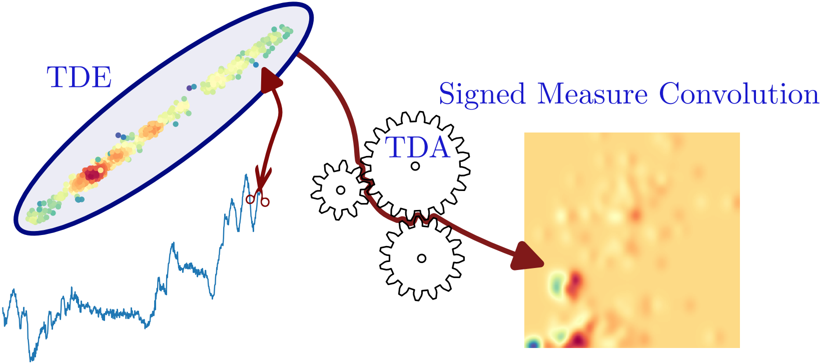
Convolutions



Convolutions



Convolutions

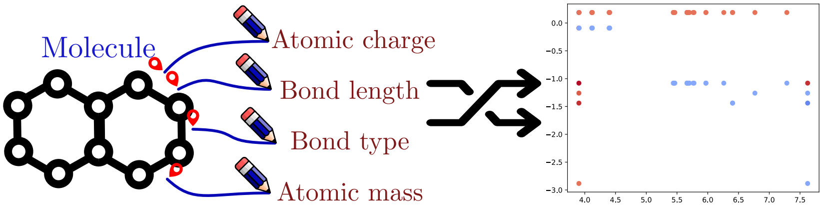


Proposition

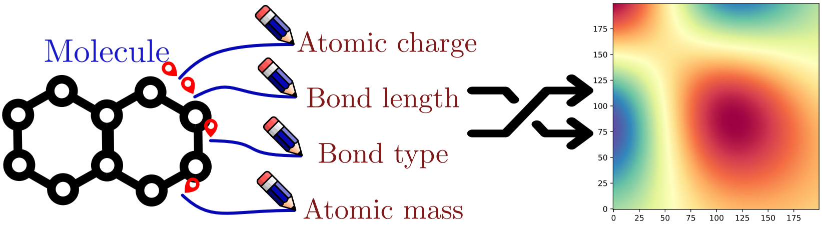
Let $K:\mathbb{R}^n\to\mathbb{R}$ be a kernel satisfying $\|K_x-K_y\|_2\lesssim\|x-y\|_2$. Then if $\mu,\nu\in\mathcal{M}$ have the same total mass,

$$\|K*\mu-K*\nu\|_2\lesssim\|\mu-\nu\|_2^K$$

Not limited to 2 parameters :)



Not limited to 2 parameters :)



Kernels with Wasserstein distances on slices

Kernels with Wasserstein distances on slices

Definition (Sliced Wasserstein on signed measures)

For any measure α on S^{n-1}, and $\mu, \nu \in \mathcal{M}_0$, define

$$\mathrm{SW}^{\alpha}(\mu,\nu) := \int \left\| \pi^{\theta}_{*}\mu - \pi^{\theta}_{*}\nu \right\|_1^K \mathrm{d}\alpha(\theta), \text{ and, } k^{\alpha}_{\mathrm{SW}} = \exp(-\mathrm{SW}^{\alpha}(\mu,\nu)).$$

where $\pi^{\theta} : \mathbb{R}^n \to \mathbb{R}$ is the orthogonal projection on the line of slope θ .

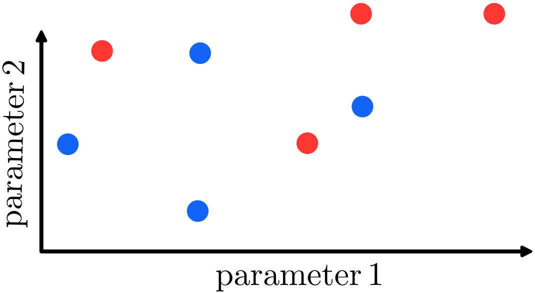
Kernels with Wasserstein distances on slices

Definition (Sliced Wasserstein on signed measures)

For any measure α on S^{n-1}, and $\mu, \nu \in \mathcal{M}_0$, define

$$\mathrm{SW}^\alpha(\mu,\nu) := \int \left\|\pi^\theta_* \mu - \pi^\theta_* \nu\right\|_1^K \mathrm{d}\alpha(\theta), \text{ and, } k^\alpha_{\mathrm{SW}} = \exp(-\mathrm{SW}^\alpha(\mu,\nu)).$$

where $\pi^\theta : \mathbb{R}^n \to \mathbb{R}$ is the orthogonal projection on the line of slope θ .



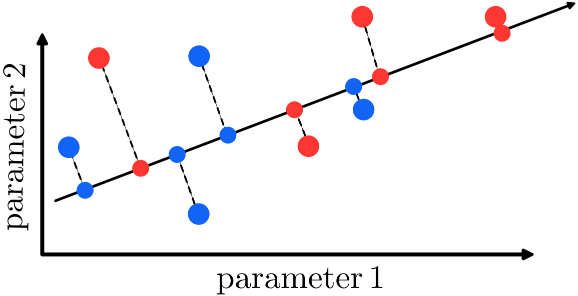
Kernels with Wasserstein distances on slices

Definition (Sliced Wasserstein on signed measures)

For any measure α on S^{n-1}, and $\mu, \nu \in \mathcal{M}_0$, define

$$\mathrm{SW}^\alpha(\mu,\nu) := \int \left\| \pi_*^\theta \mu - \pi_*^\theta \nu \right\|_1^K \mathrm{d}\alpha(\theta), \text{ and, } k_{\mathrm{SW}}^\alpha = \exp(-\mathrm{SW}^\alpha(\mu,\nu)).$$

where $\pi^\theta : \mathbb{R}^n \to \mathbb{R}$ is the orthogonal projection on the line of slope θ .



Kernels with Wasserstein distances on slices

Definition (Sliced Wasserstein on signed measures)

For any measure α on S^{n-1}, and $\mu, \nu \in \mathcal{M}_0$, define

$$\mathrm{SW}^{\alpha}(\mu,\nu) := \int \left\|\pi^{\theta}_{*}\mu - \pi^{\theta}_{*}\nu\right\|_1^K \mathrm{d}\alpha(\theta), \text{ and, } k^{\alpha}_{\mathrm{SW}} = \exp(-\mathrm{SW}^{\alpha}(\mu,\nu)).$$

where $\pi^{\theta} : \mathbb{R}^n \to \mathbb{R}$ is the orthogonal projection on the line of slope θ .

Proposition

For any $n \in \mathbb{N}$, there exists a Hilbert space $\mathcal{H}$ and a map $\Phi^{\alpha}_{\mathrm{SW}} : \mathcal{M}_0(\mathbb{R}^n) \to \mathcal{H}$ such that, for any $\mu, \nu \in \mathcal{M}_0$,

$$\|\Phi^{\alpha}_{\mathrm{SW}}(\mu) - \Phi^{\alpha}_{\mathrm{SW}}(\nu)\|_{\mathcal{H}} \leq 2\alpha(S^{n-1}) \|\mu - \nu\|_2^K$$

and

$$k^{\alpha}_{\mathrm{SW}}(\mu,\nu) = \langle \Phi^{\alpha}_{\mathrm{SW}}(\mu), \Phi^{\alpha}_{\mathrm{SW}}(\nu)\rangle_{\mathcal{H}}$$

Outline

- By *forgetting topological matchings* we define signed measures based on various topological invariants, e.g., *dimension vector, Euler characteristic, rank invariant*,

Outline

- By *forgetting topological matchings* we define signed measures based on various topological invariants, e.g., *dimension vector, Euler characteristic, rank invariant*,
- Nice *stability* properties,

Outline

- By *forgetting topological matchings* we define signed measures based on various topological invariants, e.g., *dimension vector, Euler characteristic, rank invariant*,
- Nice *stability* properties,
- *Diagram*-like structure $\Longrightarrow$ ***easy*** vectorizations,

Outline

- By *forgetting topological matchings* we define signed measures based on various topological invariants, e.g., *dimension vector, Euler characteristic, rank invariant*,
- Nice *stability* properties,
- *Diagram*-like structure $\Longrightarrow$ ***easy*** vectorizations,
- Available at [arXiv:2306.03801]

Outline

- By *forgetting topological matchings* we define signed measures based on various topological invariants, e.g., *dimension vector, Euler characteristic, rank invariant*,
- Nice *stability* properties,
- *Diagram*-like structure $\Longrightarrow$ ***easy*** vectorizations,
- Available at [arXiv:2306.03801]

Future work

- *Differentiation* of this construction,

Outline

- By *forgetting topological matchings* we define signed measures based on various topological invariants, e.g., *dimension vector, Euler characteristic, rank invariant*,
- Nice *stability* properties,
- *Diagram*-like structure $\Longrightarrow$ ***easy*** vectorizations,
- Available at [arXiv:2306.03801]

Future work

- *Differentiation* of this construction,
- *More* machine learning technics on theses measures,

Outline

- By *forgetting topological matchings* we define signed measures based on various topological invariants, e.g., *dimension vector, Euler characteristic, rank invariant*,
- Nice *stability* properties,
- *Diagram*-like structure $\Longrightarrow$ ***easy*** vectorizations,
- Available at [arXiv:2306.03801]

Future work

- *Differentiation* of this construction,
- *More* machine learning technics on theses measures,
- Work on specific data, e.g., *molecular data*,

Outline

- By *forgetting topological matchings* we define signed measures based on various topological invariants, e.g., *dimension vector, Euler characteristic, rank invariant*,
- Nice *stability* properties,
- *Diagram*-like structure $\Longrightarrow$ ***easy*** vectorizations,
- Available at [arXiv:2306.03801]

Future work

- *Differentiation* of this construction,
- *More* machine learning technics on theses measures,
- Work on specific data, e.g., *molecular data*,
- More stability, and filtration dependent guarantees.

Outline

- By *forgetting topological matchings* we define signed measures based on various topological invariants, e.g., *dimension vector, Euler characteristic, rank invariant*,
- Nice *stability* properties,
- *Diagram*-like structure $\Longrightarrow$ ***easy*** vectorizations,
- Available at [arXiv:2306.03801]

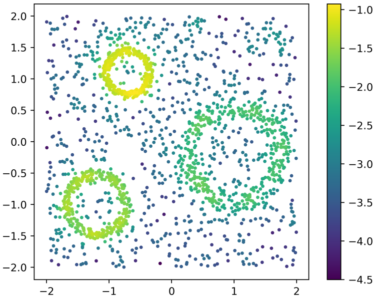
Future work

- *Differentiation* of this construction,
- *More* machine learning technics on theses measures,
- Work on specific data, e.g., *molecular data*,
- More stability, and filtration dependent guarantees.

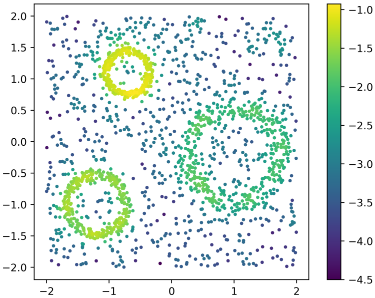
Code

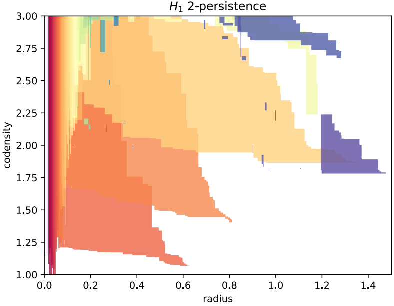
- *Python* library, available as a गुढी GUDHI Geometry Understanding in Higher Dimensions extension at : `https://github.com/DavidLapous/multipers-signed-measure`

Another motivation for multipers

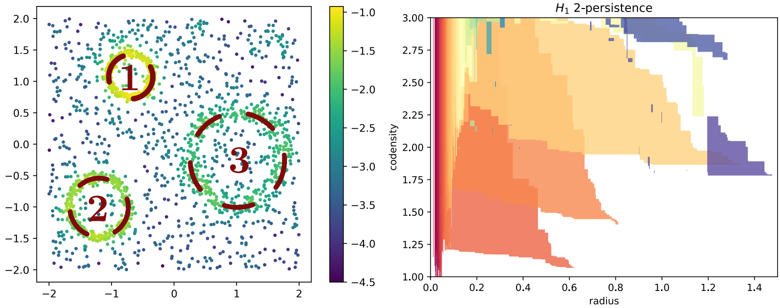


Another motivation for multipers

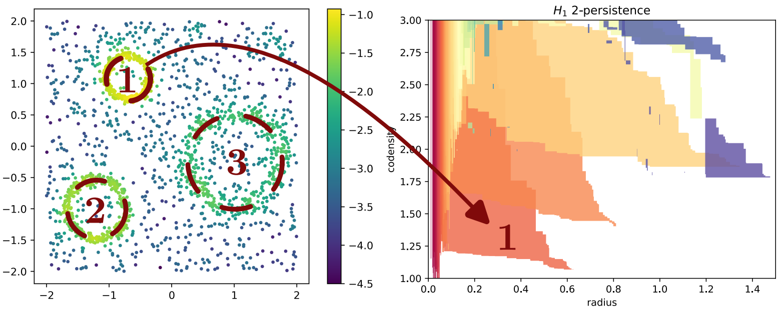




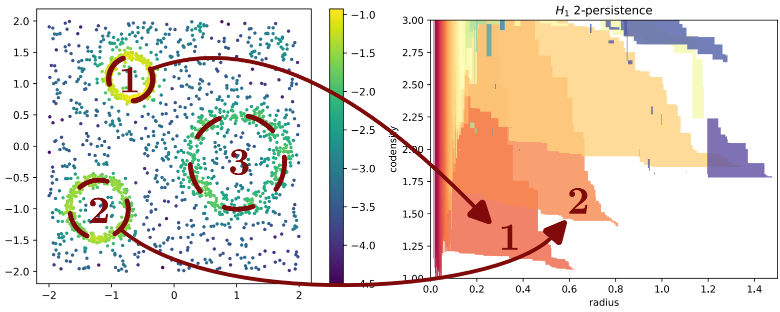
Another motivation for multipers



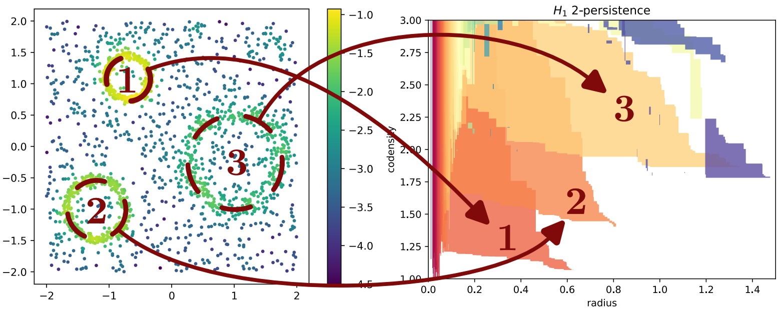
Another motivation for multipers



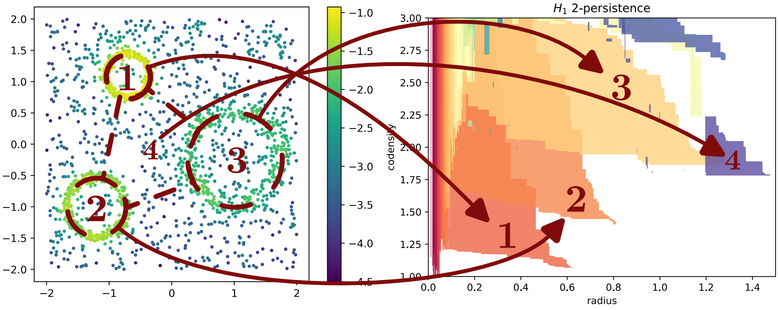
Another motivation for multipers



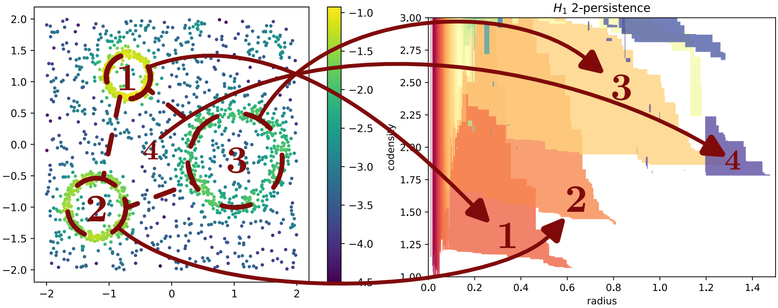
Another motivation for multipers



Another motivation for multipers

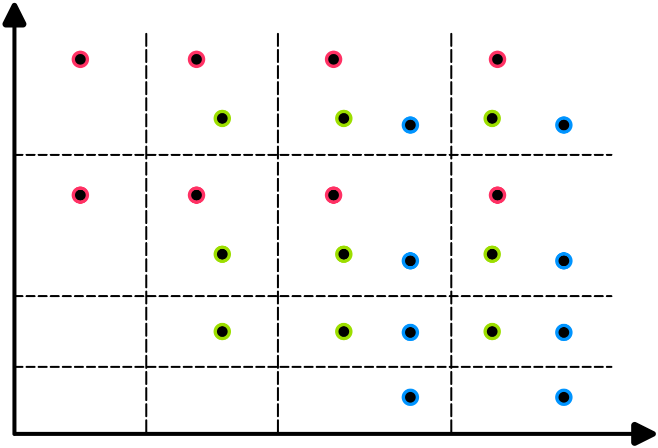


Another motivation for multipers

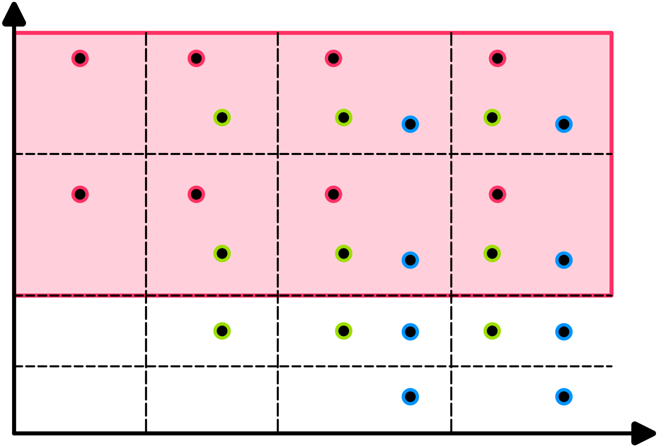


At every scale, *and at **every** concentrations !*

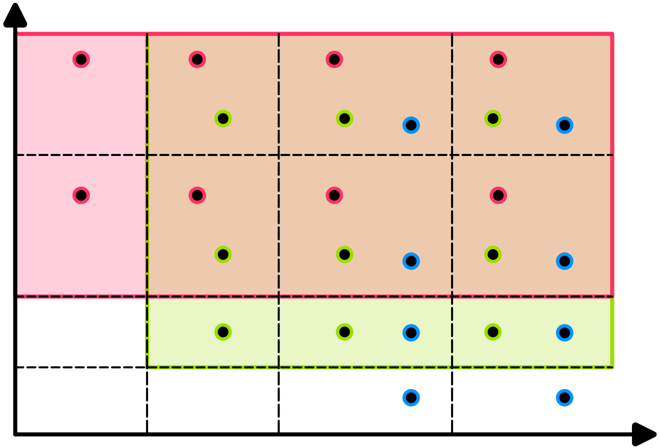
Another indecomposable



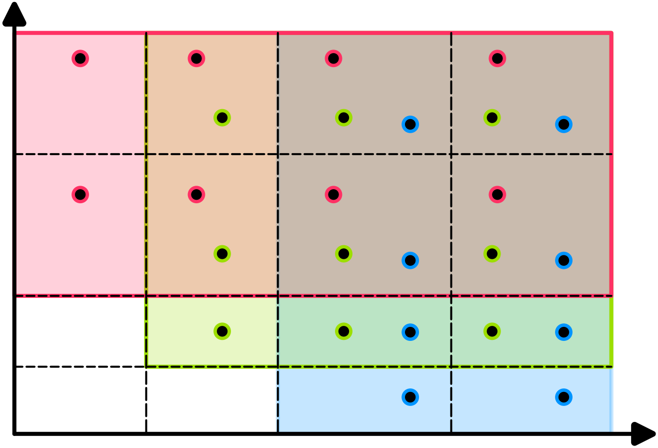
Another indecomposable



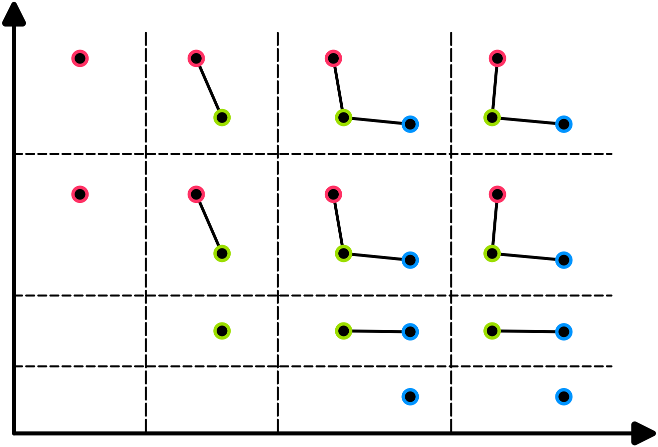
Another indecomposable



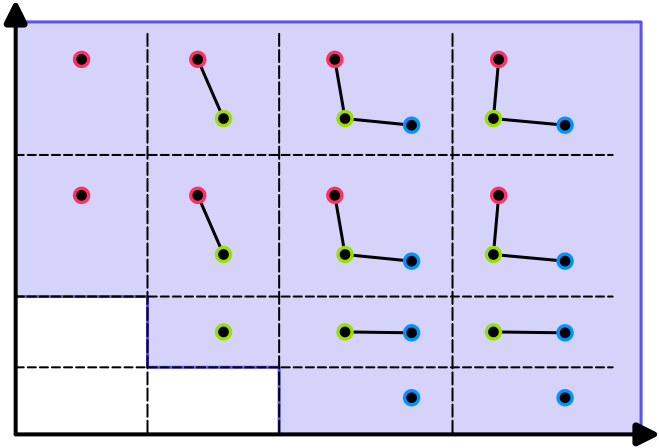
Another indecomposable



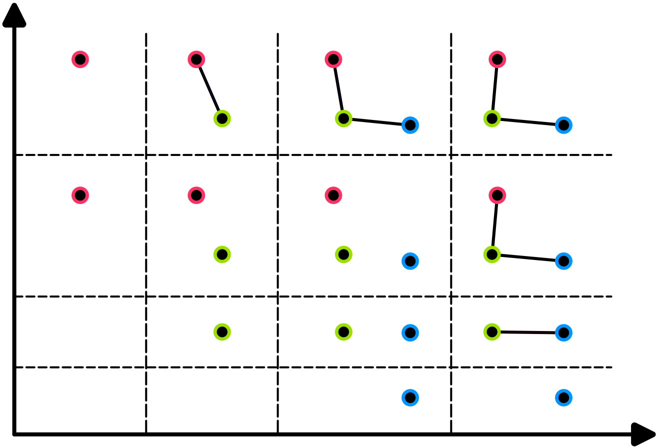
Another indecomposable



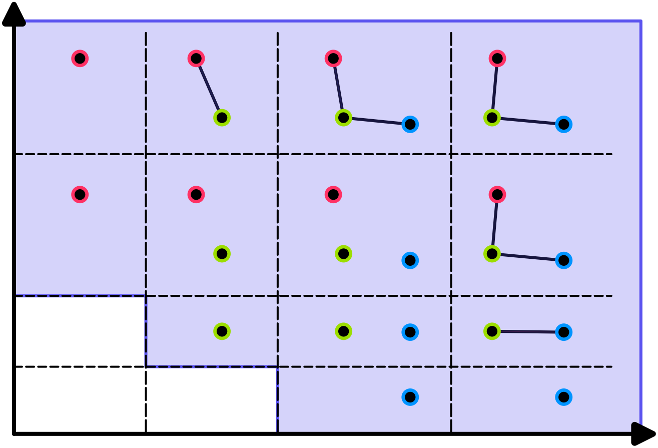
Another indecomposable



Another indecomposable



Another indecomposable



Another indecomposable