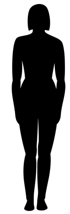
Topological inference from measures and vector bundles

Raphaël Tinarrage

Under the supervision of Marc Glisse and Frédéric Chazal

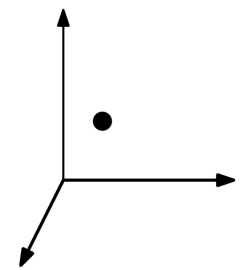
Introduction to topological inference

We aim at studying datasets represented as point clouds.

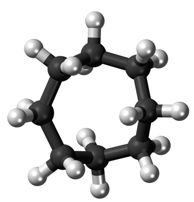


Breast cancer

[Nicolau, Levine and Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. 2011]

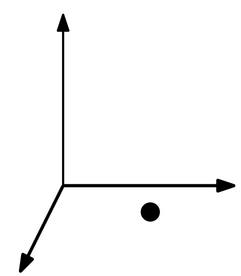


$\mathbb{R}^{262}$



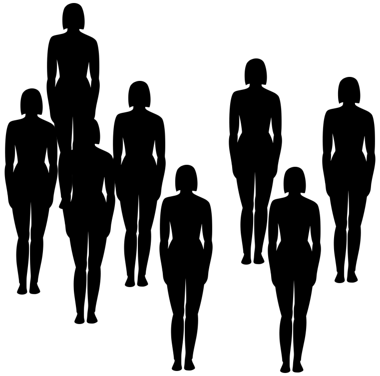
Conformation of cyclo-octane

[Martin, Thompson, Coutsias and Watson. Topology of cyclo-octane energy landscape. 2010]

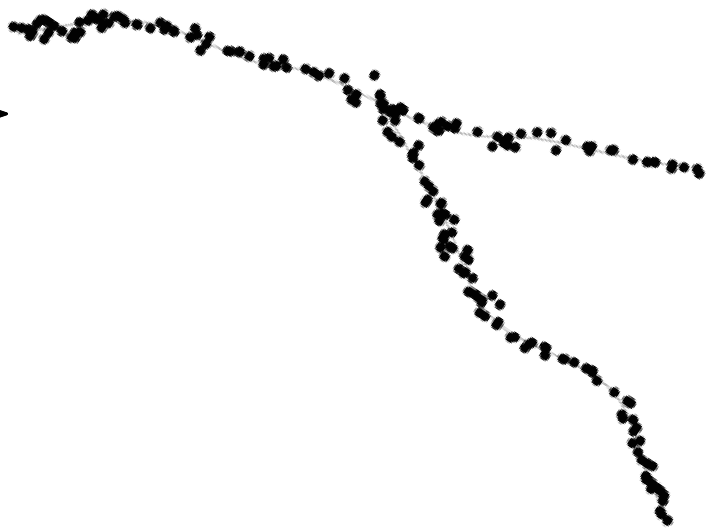


$\mathbb{R}^{72}$

We aim at studying datasets represented as point clouds.

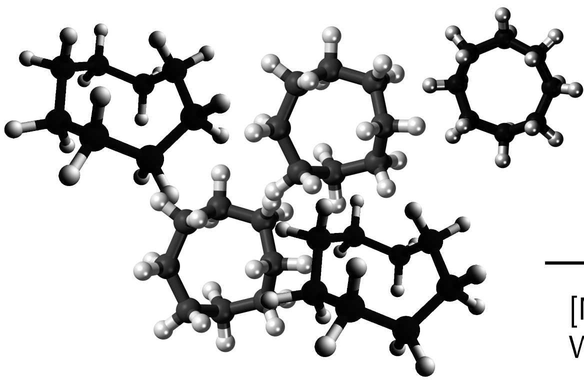


Breast cancer

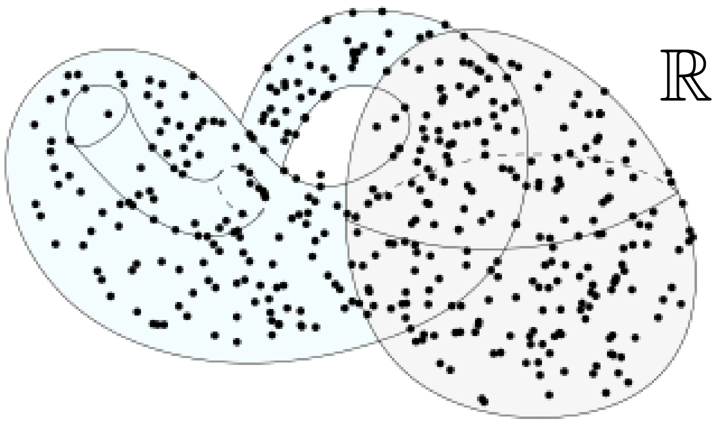


$\mathbb{R}^{262}$

[Nicolau, Levine and Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. 2011]



Conformation of cyclo-octane



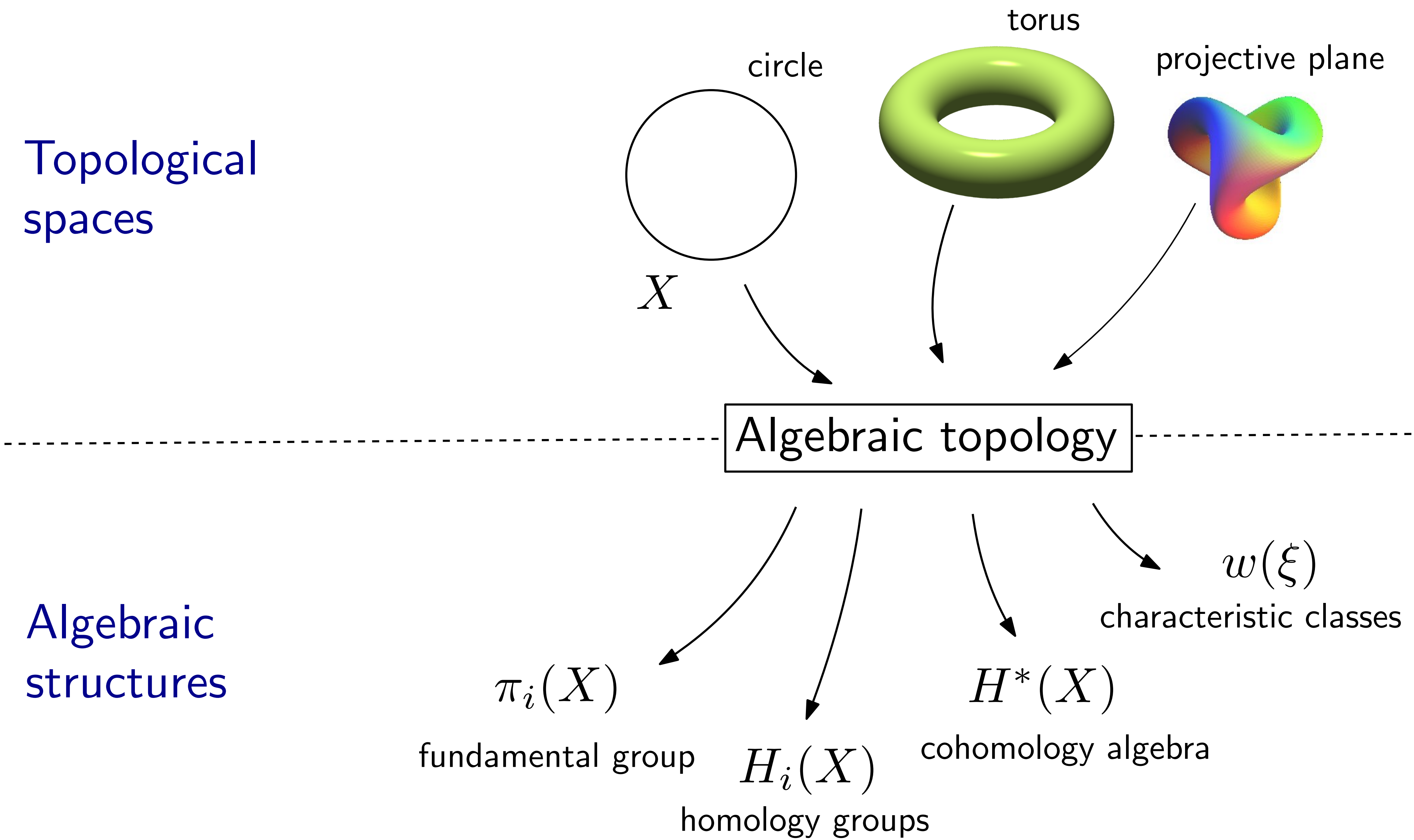
$\mathbb{R}^{72}$

[Martin, Thompson, Coutsias and Watson. Topology of cyclo-octane energy landscape. 2010]

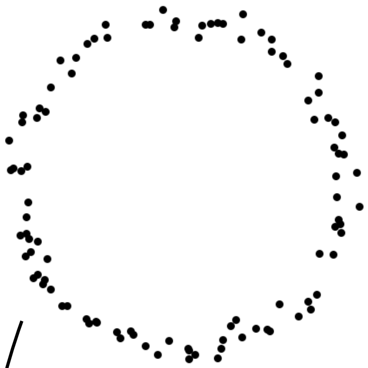
Principle of topological inference: The data is sampled near a shape whose topology is worth understanding.

Algebraic invariants

Algebraic topology allows to transform topological problems into algebraic ones.



Algebraic topology allows to transform topological problems into algebraic ones.
But it does not work anymore if the input is a sample.



Topological spaces

Algebraic topology

Algebraic structures

not interesting

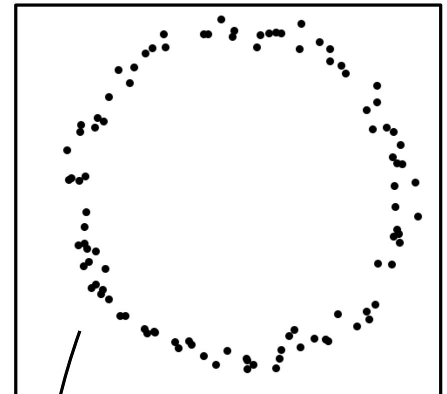
not interesting

not interesting

not interesting

Algebraic invariants

Algebraic topology allows to transform topological problems into algebraic ones.
But it does not work anymore if the input is a sample.
This is where persistent homology comes in.



Topological spaces

Persistent homology

Algebraic topology

Algebraic structures

persistence module

not interesting

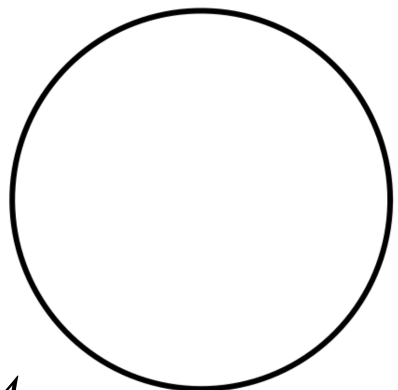
not interesting

not interesting

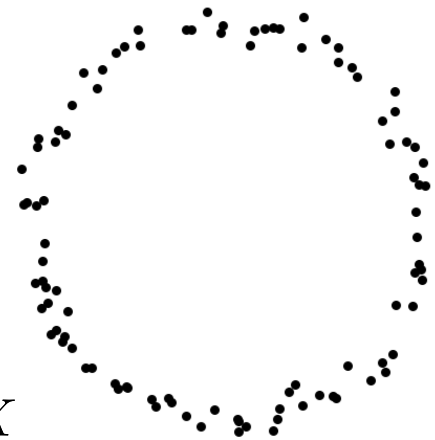
not interesting

Thickenings

Let $\mathcal{M} \subset \mathbb{R}^n$ be a submanifold, and $X \subset \mathbb{R}^n$ a point cloud.
How to recover the homotopy type of $\mathcal{M}$ from X?



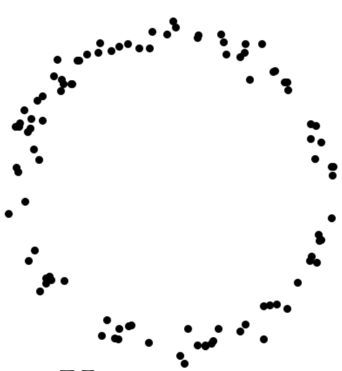
$\mathcal{M}$



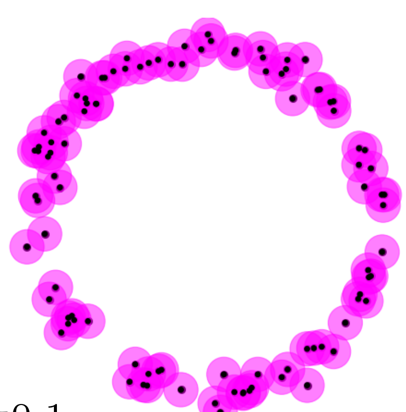
X

For all $t \geq 0$, define the t-thickening of X:

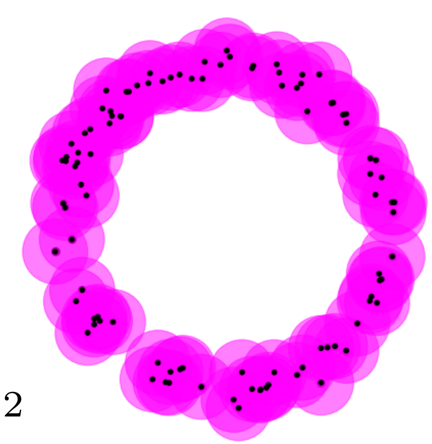
$$X^t = \{y \in \mathbb{R}^n, \exists x \in X, \|x - y\| \leq t\}$$



$X^0 = X$



$X^{0,1}$



$X^{0,2}$

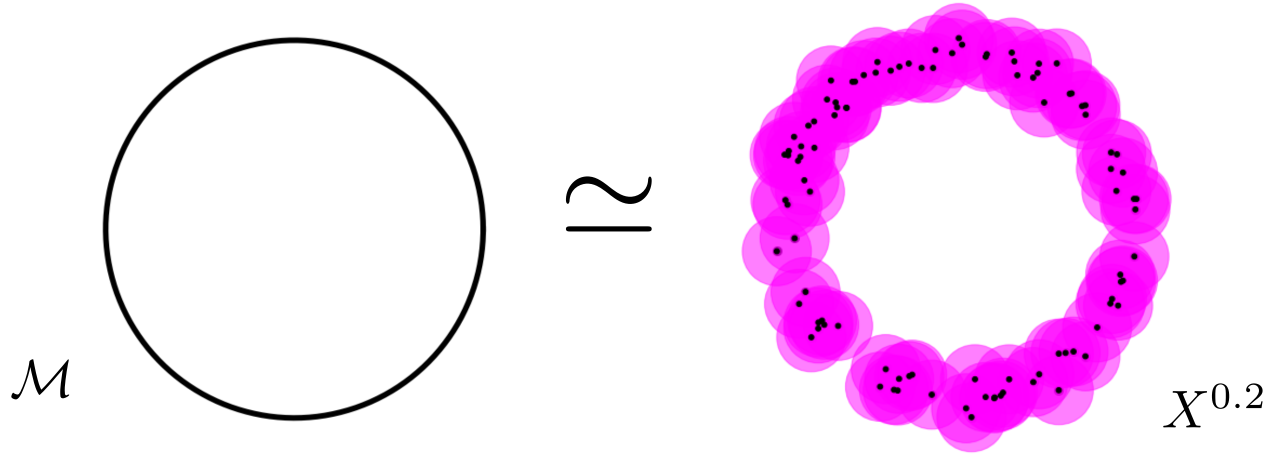
Thickenings

Theorem (Chazal, Cohen-Steiner, Lieutier, 2009)

Let $\mathcal{M}, X$ be subsets of $\mathbb{R}^n$.
Suppose that $\mathrm{reach}\,(\mathcal{M}) > 0$ and $\mathrm{d_H}\,(X, \mathcal{M}) \leq \frac{1}{17}\mathrm{reach}\,(\mathcal{M})$. Let

$$t \in [4\mathrm{d_H}\,(X, \mathcal{M})\,, \mathrm{reach}\,(\mathcal{M}) - 3\mathrm{d_H}\,(X, \mathcal{M}))\,.$$

Then X^t and $\mathcal{M}$ are homotopy equivalent.



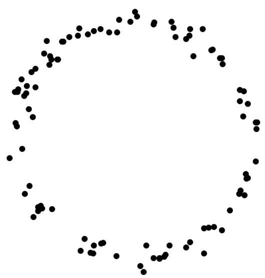
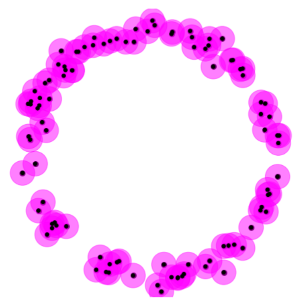
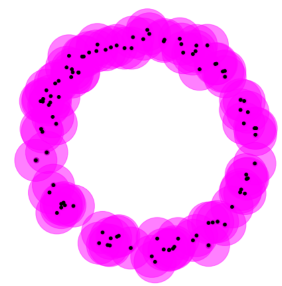
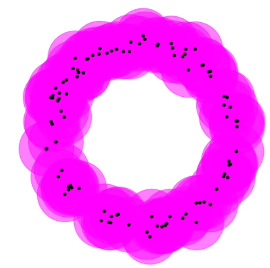
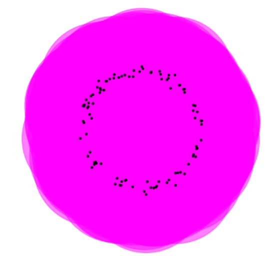
Definition

The *Čech filtration* of X is the collection:

$$V[X] = \left(X^t\right)_{t \geq 0}\,.$$

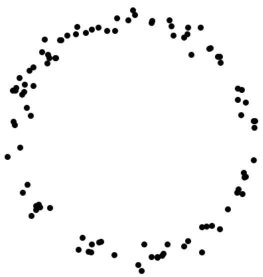
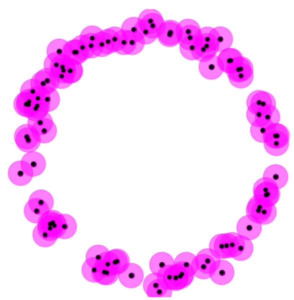
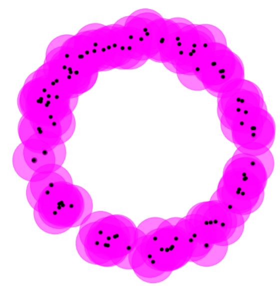
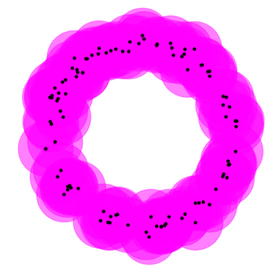
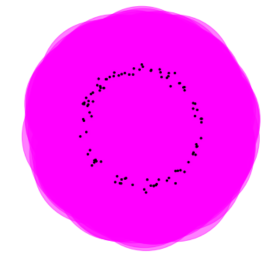
Construction of persistence modules

We compute the singular homology of the thickenings of X over $\mathbb{Z}/2\mathbb{Z}$.

X^t	 $X^0 = X$	 $X^{0,1}$	 $X^{0,2}$	 $X^{0,3}$	 X^1
$H_0(X^t)$	$(\mathbb{Z}/2\mathbb{Z})^{100}$	$(\mathbb{Z}/2\mathbb{Z})^5$	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$
$H_1(X^t)$	0	0	$\mathbb{Z}/2\mathbb{Z}$	$\mathbb{Z}/2\mathbb{Z}$	0

Construction of persistence modules

We compute the singular homology of the thickenings of X over $\mathbb{Z}/2\mathbb{Z}$.

inclusions i_s^t		$\xrightarrow{i_0^{0.1}}$		$\xrightarrow{i_{0.1}^{0.2}}$		$\xrightarrow{i_{0.2}^{0.3}}$		$\xrightarrow{i_{0.3}^{1}}$	
X^t	 $X^0 = X$		 $X^{0,1}$		 $X^{0,2}$		 $X^{0,3}$		 X^1
$H_0(X^t)$	$(\mathbb{Z}/2\mathbb{Z})^{100}$		$(\mathbb{Z}/2\mathbb{Z})^{5}$		$\mathbb{Z}/2\mathbb{Z}$		$\mathbb{Z}/2\mathbb{Z}$		$\mathbb{Z}/2\mathbb{Z}$
$H_1(X^t)$	0		0		$\mathbb{Z}/2\mathbb{Z}$		$\mathbb{Z}/2\mathbb{Z}$		0
		$\xrightarrow{(i_0^{0.1})_*}$		$\xrightarrow{(i_{0.1}^{0.2})_*}$		$\xrightarrow{(i_{0.2}^{0.3})_*}$		$\xrightarrow{(i_{0.3}^{1})_*}$	

The data of $(H_i(X^t))_{t\geq 0}$ and $((i_s^t)_*)_{s\leq t}$, is called a *persistence module*.

Persistence modules

Definition

A *persistence module* $\mathbb{V}$ over $\mathbb{R}^+$ is a family of $\mathbb{Z}/2\mathbb{Z}$-vector spaces $(V^t)_{t\geq 0}$, and a family of a family of linear maps $(v_s^t \colon V^s \to V^t)_{0\leq s\leq t}$ such that:

- for every $t \geq 0$, $v_t^t \colon V^t \to V^t$ is the identity map,
- for every $r, s, t \geq 0$ such that $r \leq s \leq t$, we have $v_s^t \circ v_r^s = v_r^t$.

$$\begin{tikzcd} V^r \arrow[r, "v_r^s"] \arrow[rr, bend right, "v_r^t"'] & V^s \arrow[r, "v_s^t"] & V^t \end{tikzcd}$$

Main construction of persistence modules:

A *filtration* of $\mathbb{R}^n$ is a collection of subsets $(X_t)_{t\geq 0}$ such that $X_s \subset X_t$ when $s \leq t$.

$$\dashrightarrow X_{t_1} \xhookrightarrow{i_{t_1}^{t_2}} X_{t_2} \xhookrightarrow{i_{t_2}^{t_3}} X_{t_3} \xhookrightarrow{i_{t_3}^{t_4}} X_{t_4} \dashrightarrow$$

Applying the i^{th} singular homology functor give rise to a persistence module.

$$\dashrightarrow H_i(X_{t_1}) \xrightarrow{(i_{t_1}^{t_2})_*} H_i(X_{t_2}) \xrightarrow{(i_{t_2}^{t_3})_*} H_i(X_{t_3}) \xrightarrow{(i_{t_3}^{t_4})_*} H_i(X_{t_4}) \dashrightarrow$$

Persistence modules

Theorem (Crawley-Boevey, 2015)

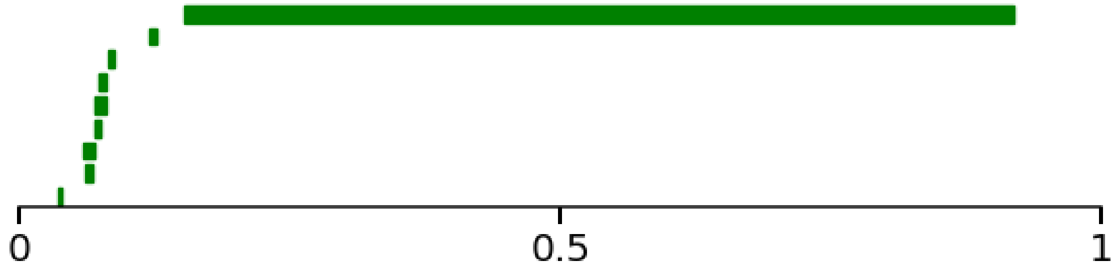
A (pointwise finite-dimensional) persistence module is isomorphic to a unique sum of interval modules.

This multi-set of intervals is called the *persistence barcode*. It is a complete invariant of pointwise finite dimensional persistence modules.

Persistence module: $\mathbb{V}$

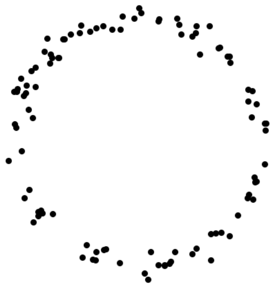
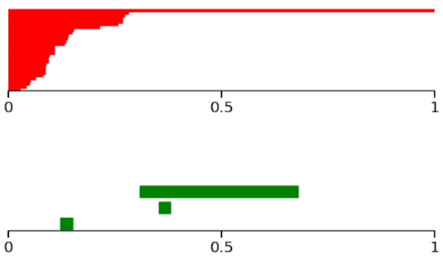
Barcode: $\{$ [0.171, 0.897), [0.035, 0.049), [0.037, 0.046), [0.072, 0.078), [0.077, 0.083), [0.046, 0.050), [0.050, 0.054), [0.036, 0.040), [0.089, 0.092) $\}$

Graphical representation:

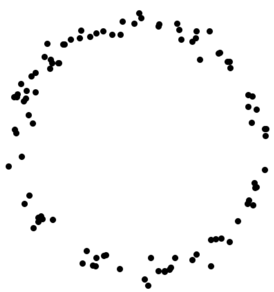
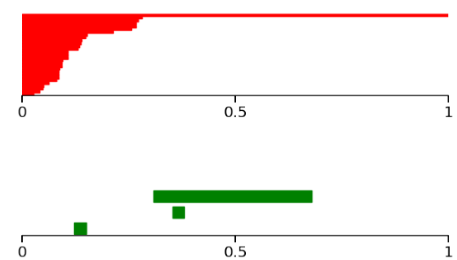


Persistence barcodes

Stability of persistence modules

Subset	Filtration	Persistence module	Barcode
	$(X_t)_{t\geq 0}$	$(V^t)_{t\geq 0}$	
Hausdorff distance $\mathrm{d_H}$	Interleaving distance for filtrations $\mathrm{d_i}$	Interleaving distance for persistence modules $\mathrm{d_i}$	Bottleneck distance $\mathrm{d_b}$

Stability of persistence modules

Subset	Filtration	Persistence module	Barcode
	$(X_t)_{t\geq 0}$	$(V^t)_{t\geq 0}$	
Hausdorff distance $\mathrm{d_H}$	Interleaving distance for filtrations $\mathrm{d_i}$	Interleaving distance for persistence modules $\mathrm{d_i}$	Bottleneck distance $\mathrm{d_b}$

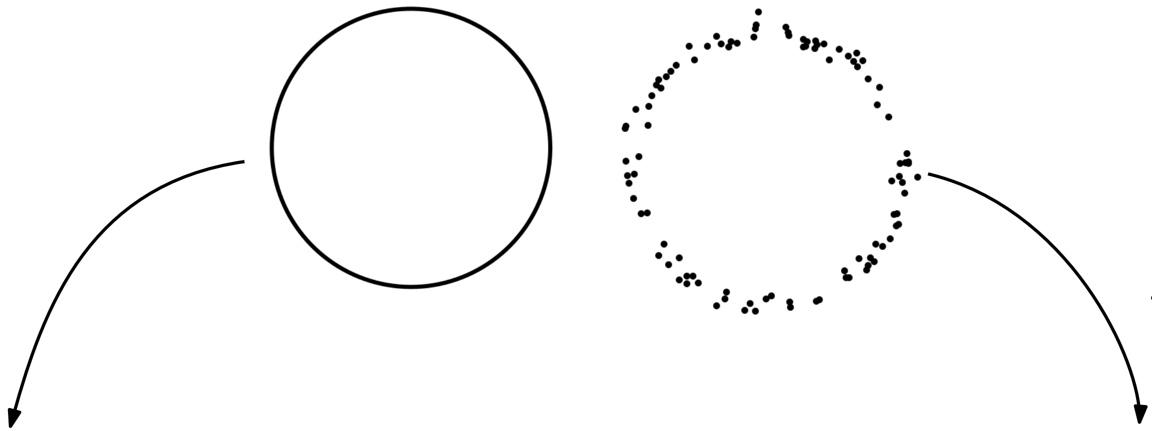
Isometry theorem (Chazal, Cohen-Steiner, Glisse, Guibas, Oudot, 2009 - Lesnik 2011)

Between pointwise finite-dimensional persistence modules, the bottleneck distance and the interleaving distance are equal.

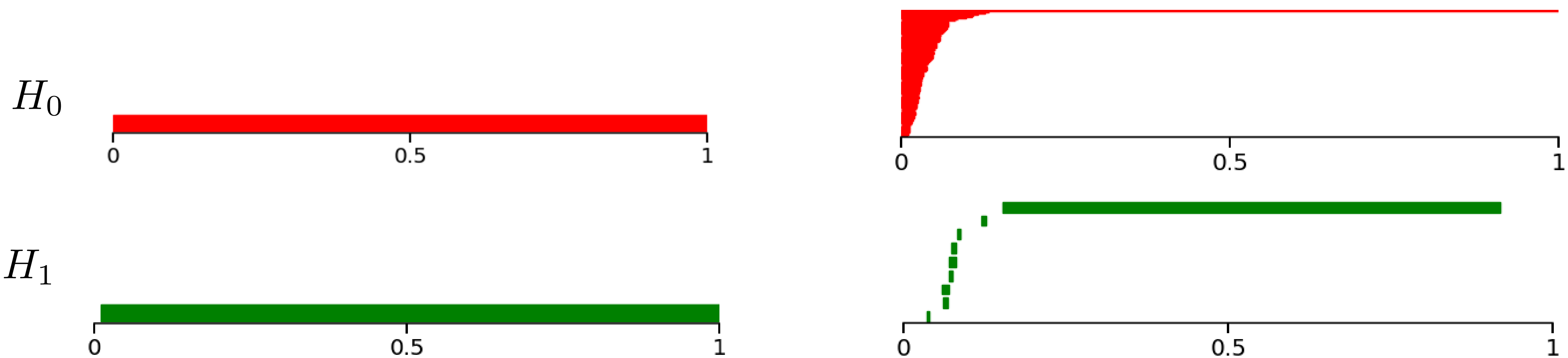
Stability theorem (Edelsbrunner, Harer, Cohen-Steiner, 2005)

Let $X, Y \subset \mathbb{R}^n$ be two compact subsets, and denote $\mathbb{V}, \mathbb{W}$ the persistence modules of i^{th} homology of their Čech filtrations. Then $\mathrm{d_b}(\mathbb{V}, \mathbb{W}) \leq \mathrm{d_H}(X, Y)$.

Stability of persistence modules



Barcode of the persistence module associated to the Čech filtration

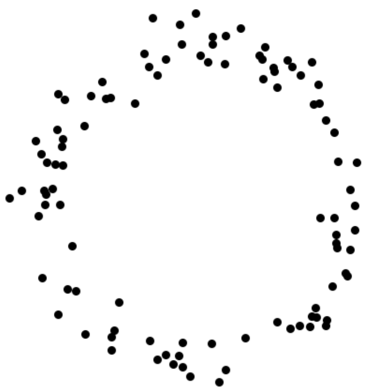


Stability theorem (Edelsbrunner, Harer, Cohen-Steiner, 2005)

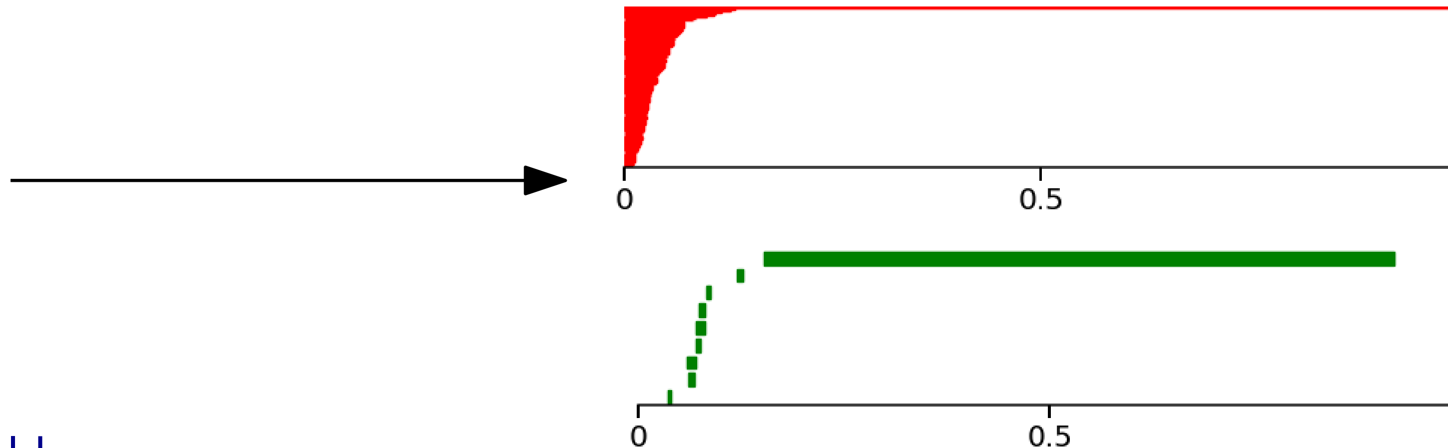
Let $X, Y \subset \mathbb{R}^n$ be two compact subsets, and denote $\mathbb{V}, \mathbb{W}$ the persistence modules of i^{th} homology of their Čech filtrations. Then $\mathrm{d_b}(\mathbb{V}, \mathbb{W}) \leq \mathrm{d_H}(X, Y)$.

Contributions

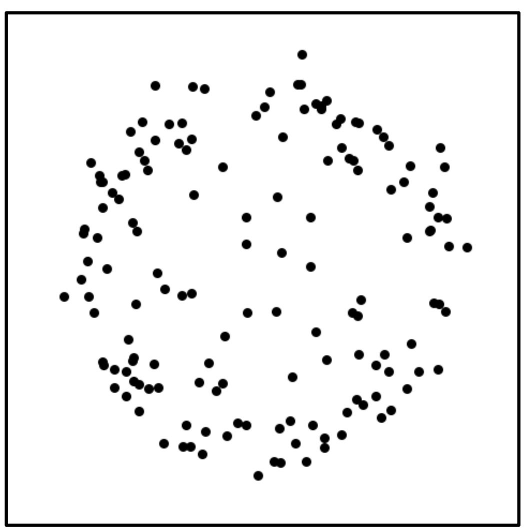
Usual framework of persistent homology:



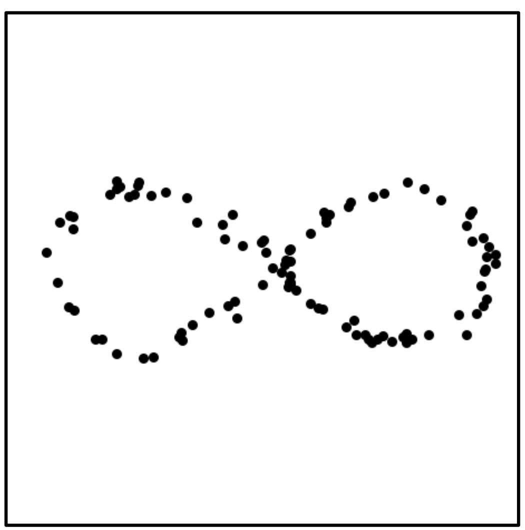
Sample of a submanifold



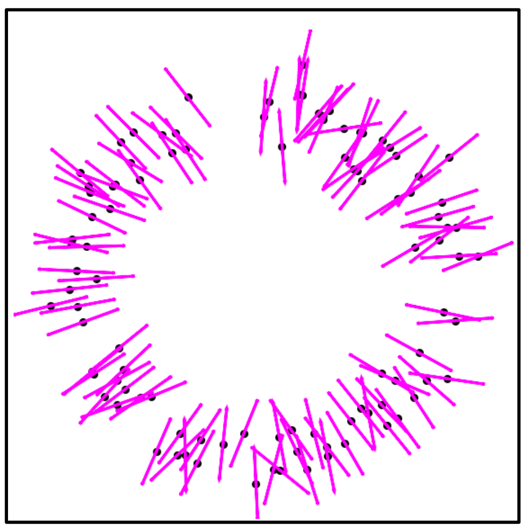
Problems studied in this thesis:



Sample of a submanifold with anomalous points



Sample of an immersed manifold

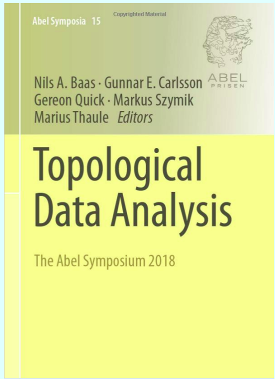


Sample of a vector bundle

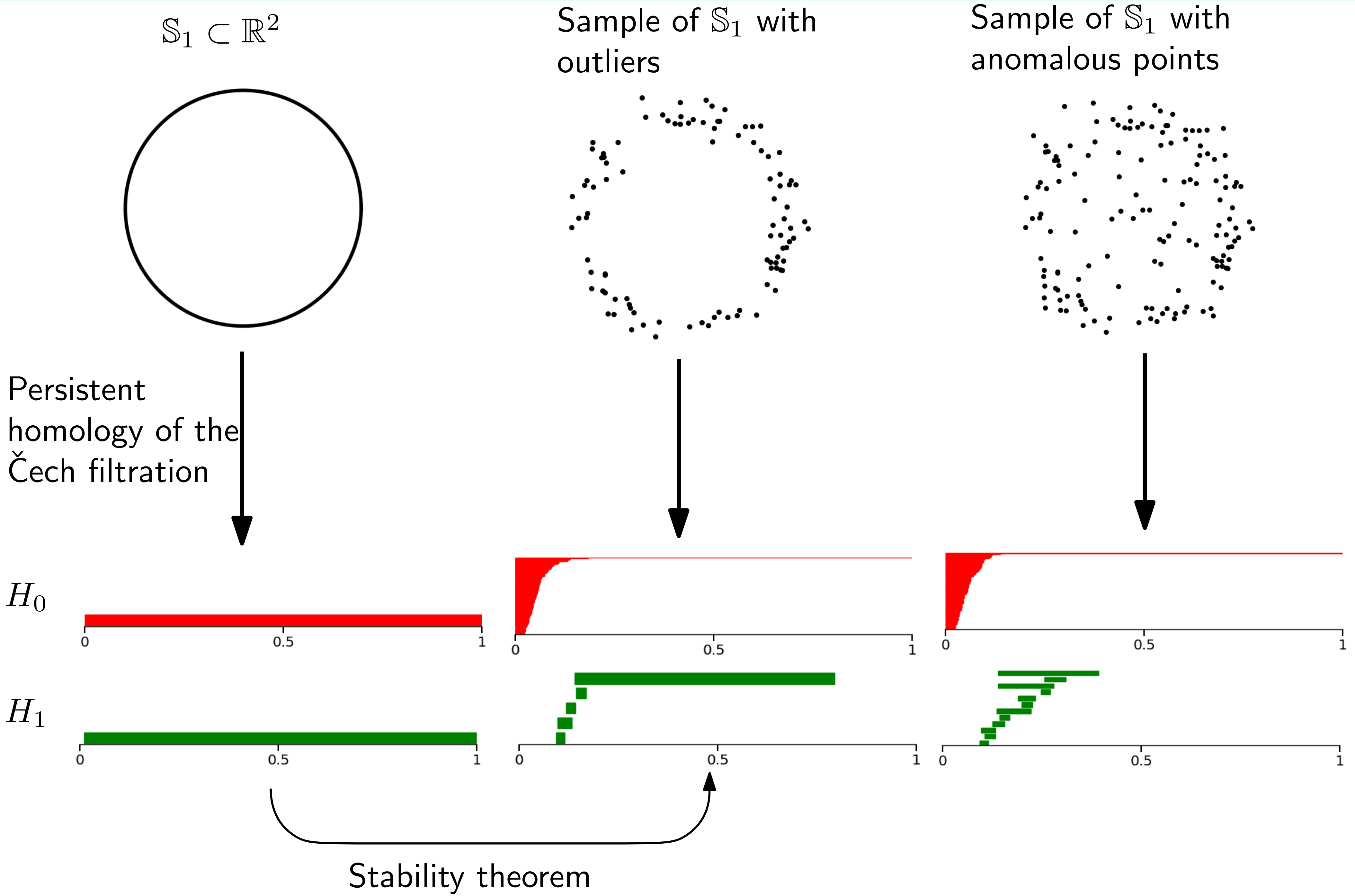
DTM-based filtrations

Joint work with Hirokazu Anai, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hiroya Inakoshi and Yuhei Umeda
Experimented in the setting of an industrial research project

Published in the proceedings of Symposium of Computational Geometry (June 2019) and in the proceedings of Abel Symposium (2018)



Statement of the problem



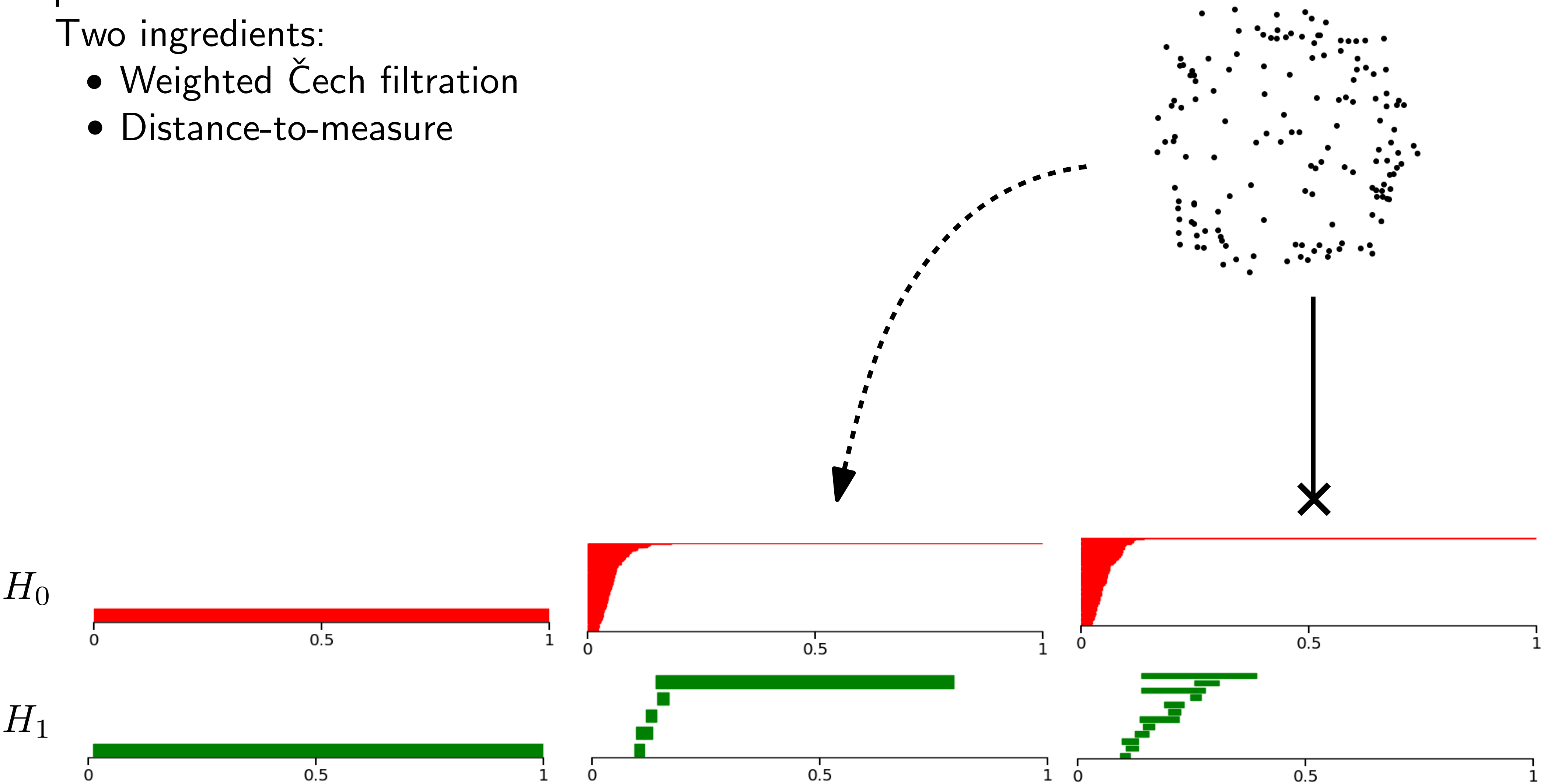
Statement of the problem

Goal: build a filtration that is robust to anomalous points.

Two ingredients:

- Weighted Čech filtration
- Distance-to-measure

Sample of $\mathbb{S}_1$ with anomalous points



Weighted Čech filtration

Input point cloud: $X \subset \mathbb{R}^n$

Reminder: The Čech filtration of X is the collection $V[X] = (X^t)_{t \geq 0}$, where

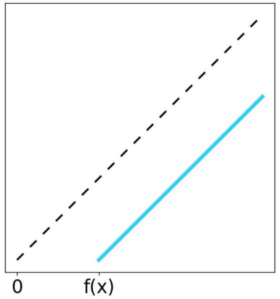
$$X^t = \bigcup_{x \in X} \overline{\mathcal{B}}\,(x, t)\,.$$

Let $f \colon X \to \mathbb{R}^+$ be any map.

Definition

The *weighted Čech filtration* of X with parameter f is the collection $V[X, f] = (V^t[X, f])_{t \geq 0}$, where

$$V^t[X, f] = \bigcup_{x \in X} \overline{\mathcal{B}}\,(x, t - f(x))\,.$$



$t \mapsto t - f(x)$

Weighted Čech filtration

Input point cloud: $X \subset \mathbb{R}^n$

Reminder: The Čech filtration of X is the collection $V[X] = (X^t)_{t \geq 0}$, where

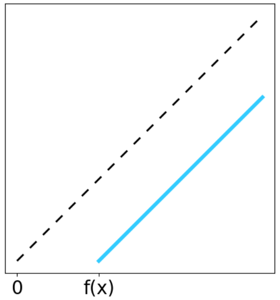
$$X^t = \bigcup_{x \in X} \overline{\mathcal{B}}(x, t).$$

Let $f \colon X \to \mathbb{R}^+$ be any map.

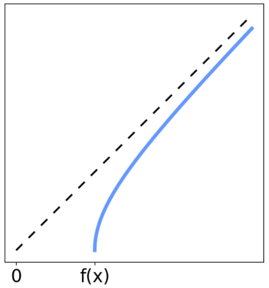
Definition

The *weighted Čech filtration* of X with parameter f is the collection $V[X, f] = (V^t[X, f])_{t \geq 0}$, where

$$V^t[X, f] = \bigcup_{x \in X} \overline{\mathcal{B}}\left(x, \sqrt{t^2 - f(x)^2}\right).$$



$t \mapsto t - f(x)$



$t \mapsto \sqrt{t^2 - f(x)^2}$

[Buchet et al., SODA 2015]

Weighted Čech filtration

Input point cloud: $X \subset \mathbb{R}^n$

Reminder: The Čech filtration of X is the collection $V[X] = (X^t)_{t \geq 0}$, where

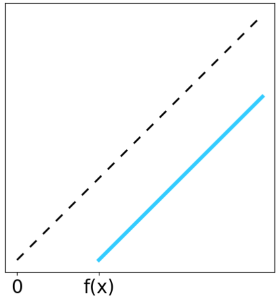
$$X^t = \bigcup_{x \in X} \overline{\mathcal{B}}(x, t).$$

Let $f\colon X \to \mathbb{R}^+$ be any map, and $p \in [1, +\infty)$.

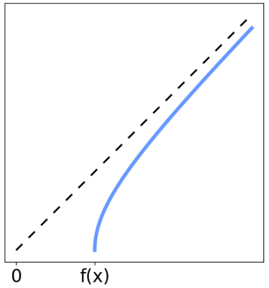
Definition

The *weighted Čech filtration* of X with parameters p and f is the collection $V[X, f, p] = (V^t[X, f, p])_{t \geq 0}$, where

$$V^t[X, f, p] = \bigcup_{x \in X} \overline{\mathcal{B}}\left(x, (t^p - f(x)^p)^{\frac{1}{p}}\right).$$

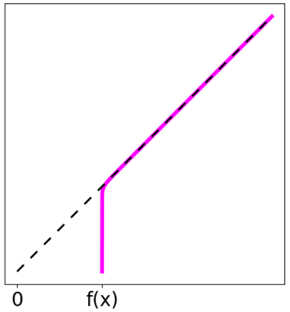


$t \mapsto t - f(x)$



$t \mapsto \sqrt{t^2 - f(x)^2}$

[Buchet et al., SODA 2015]



$t \mapsto \left(t^{30} - f(x)^{30}\right)^{\frac{1}{30}}$

Distance-to-measure (DTM)

Introduced in [Chazal, Cohen-Steiner, Mérigot. Geometric inference for probability measures, 2011].

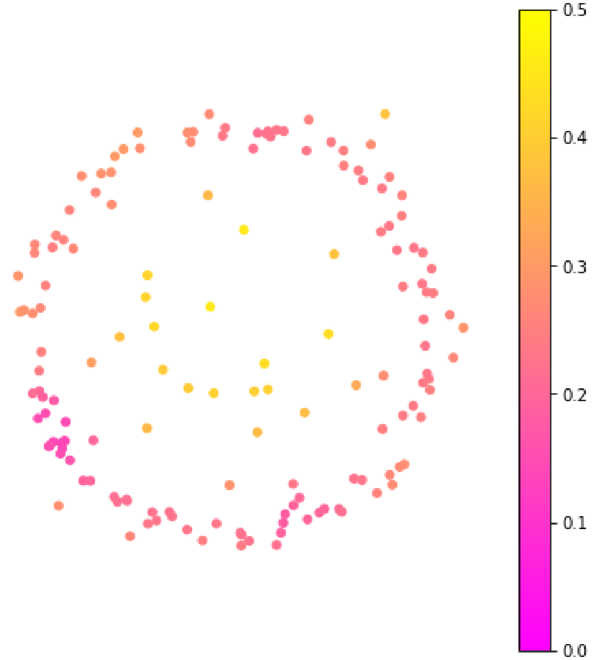
Let μ be a probability measure. For $x \in \mathbb{R}^n$ and $t \in [0,1)$, define

$$\delta_{\mu,t}(x) = \inf\{r \geq 0, \mu(\overline{\mathcal{B}}(x,r)) > t\}.$$

Definition

Let $m \in [0,1[$. The *DTM* μ with parameter m is the function:

$$\begin{array}{rccl} \mathrm{d}_{\mu,m}: & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \sqrt{\frac{1}{m}\int_0^m \delta_{\mu,t}^2(x)dt} \end{array}$$



$\mathrm{d}_{\mu,m}$ with $m = 0.1$

μ empirical measure on X

Distance-to-measure (DTM)

Introduced in [Chazal, Cohen-Steiner, Mérigot. Geometric inference for probability measures, 2011].

Let μ be a probability measure. For $x \in \mathbb{R}^n$ and $t \in [0,1)$, define

$$\delta_{\mu,t}(x) = \inf\{r \geq 0, \mu(\overline{\mathcal{B}}(x,r)) > t\}.$$

Definition

Let $m \in [0,1[$. The *DTM* μ with parameter m is the function:

$$\begin{array}{cccc} \mathrm{d}_{\mu,m}: & \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \sqrt{\frac{1}{m}\int_0^m \delta^2_{\mu,t}(x)dt} \end{array}$$

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

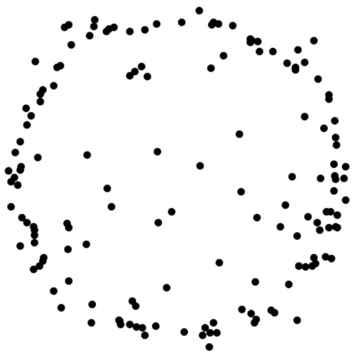
For every probability measures μ, ν and $m \in (0,1)$, we have

$$\|\mathrm{d}_{\mu,m} - \mathrm{d}_{\nu,m}\|_\infty \leq m^{-\frac{1}{2}} \mathrm{W}_2(\mu,\nu),$$

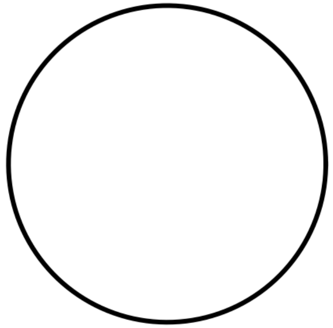
where W_2 denotes the Wasserstein distance.

Distance-to-measure (DTM)

We shall now adopt a measure point of view. It requires to see our subsets as probability measures.



X finite → $\mu = \frac{1}{|X|} \sum_{x \in X} \delta_x$
empirical measure



$\mathcal{M}$ submanifold → $\nu = \frac{1}{\mathcal{H}^d(\mathcal{M})} \mathcal{H}^d_{|\mathcal{M}}$
Hausdorff measure

X and $\mathcal{M}$ are not close in Hausdorff distance...

But μ and ν are close in Wasserstein distance!

DTM-filtrations

Definition

Let μ be a probability measure, $m \in [0,1)$ and $p \geq 1$.
The *DTM-filtration* with parameters μ, m and p is the weighted Čech filtration $V[X, f, p]$ with parameters:

- $X = \mathrm{supp}(\mu)$
- $f = \mathrm{d}_{\mu,m}$

It is denoted $W[\mu, m, p]$.

Explicitely, $W[\mu, m, p] = (W^t[\mu, m, p])_{t \geq 0}$ with:

$$W^t[\mu, m, p] = \bigcup_{x \in \mathrm{supp}(\mu)} \overline{\mathcal{B}}\left(x, (t^p - d_{\mu,m}(x)^p)^{\frac{1}{p}}\right)$$

DTM-filtrations

DTM-filtrations

Case $p = 1$: $W^t[\mu, m, p] = \bigcup \overline{\mathcal{B}}\left(x, t - \mathrm{d}_{\mu,m}(x)\right)$

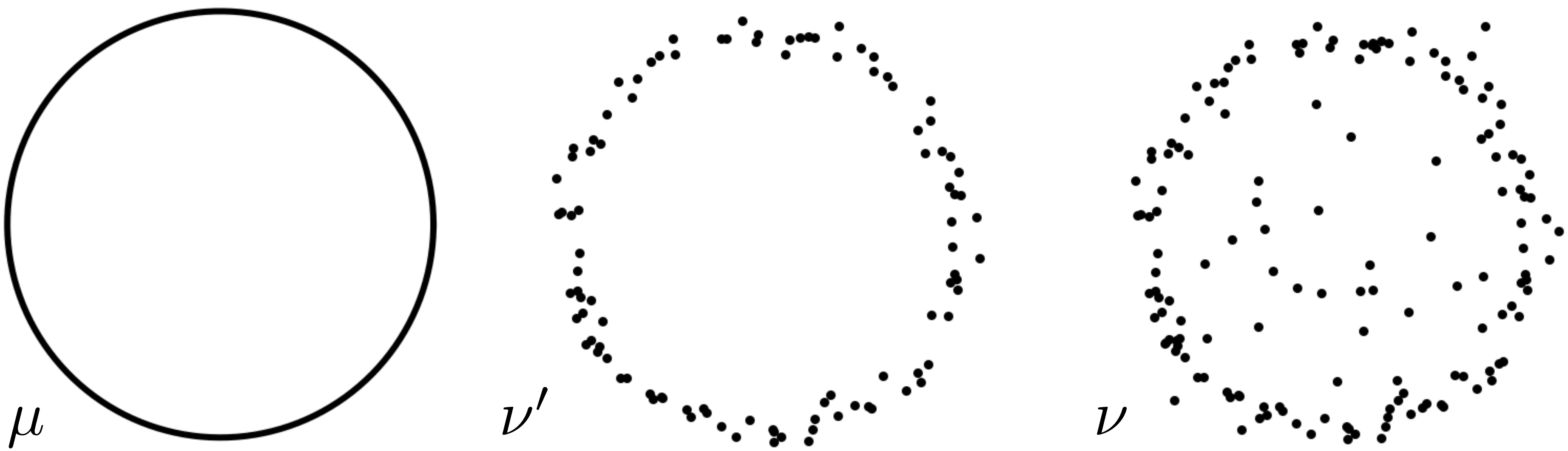
Define $c(\mu, m, p) = \sup_{x \in \mathrm{supp}(\mu)} \mathrm{d}_{\mu,m}(x)$.

The quantity c is small if the measure μ is close to the Hausdorff measure restricted to a submanifold.

Theorem (Anai, Chazal, Glisse, Ike, Inakoshi, T. and Umeda, 2020)

Let μ, ν be probability measures. Let ν' be any probability measure with compact support included in $\mathrm{supp}(\nu)$.
The interleaving distance between the (set) filtrations $W[\mu, m, p]$ and $W[\nu, m, p]$ is bounded by:

$$m^{-\frac{1}{2}} \mathrm{W}_2\left(\mu, \nu'\right) + m^{-\frac{1}{2}} \mathrm{W}_2\left(\nu', \nu\right) + c(\mu, m, p) + c(\nu', m, p)$$



DTM-filtrations

Case $p > 1$: $W^t[\mu, m, p] = \bigcup \overline{\mathcal{B}}\left(x, (t^p - d_{\mu,m}(x)^p)^{\frac{1}{p}}\right)$

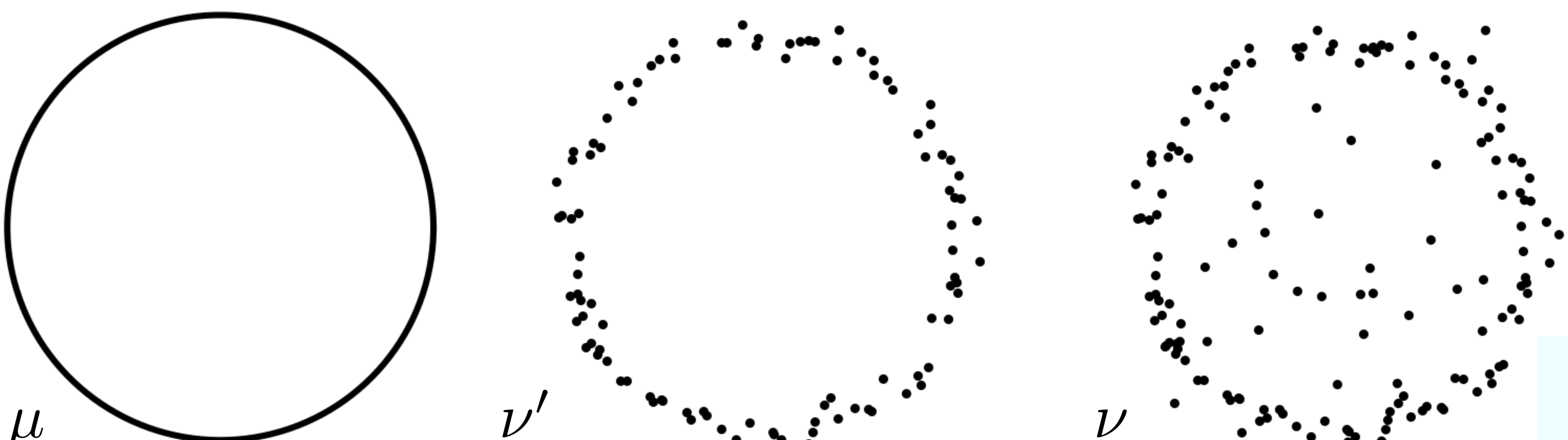
Define $c(\mu, m, p) = \sup_{x \in \mathrm{supp}(\mu)} \mathrm{d}_{\mu,m}(x)$. $+2(1-\frac{1}{p})\mathrm{diam}(\mathrm{supp}(\mu))$

The quantity c is small if the measure μ is close to the Hausdorff measure restricted to a submanifold.

Theorem (Anai, Chazal, Glisse, Ike, Inakoshi, T. and Umeda, 2020)

Let μ, ν be probability measures. Let ν' be any probability measure with compact support included in $\mathrm{supp}(\nu)$.
The interleaving distance between the ~~(set) filtrations~~ persistence modules $W[\mu, m, p]$ and $W[\nu, m, p]$ is bounded by:

$$m^{-\frac{1}{2}} \mathrm{W}_2(\mu, \nu') + m^{-\frac{1}{2}} \mathrm{W}_2(\nu', \nu) + c(\mu, m, p) + c(\nu', m, p)$$



Recovering the homology of immersed manifolds

Presented at the Young Researcher Forum of Symposium of Computational Geometry (June 2020)

CG Week

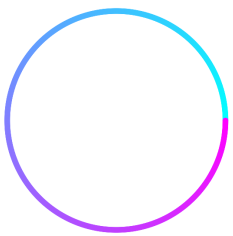
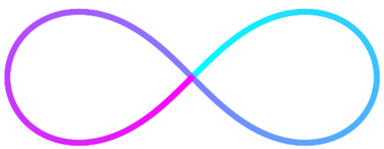
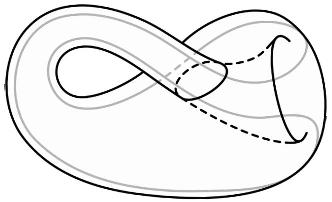
Statement of the problem

We are observing an immersed manifold $\mathcal{M} \subset \mathbb{R}^n$.

Abstract manifold	Immersion	Immersed manifold
$\mathcal{M}_0$	$\xrightarrow{\quad u \quad}$	$\mathcal{M} = u(\mathcal{M}_0) \quad \subset \mathbb{R}^n$

Statement of the problem

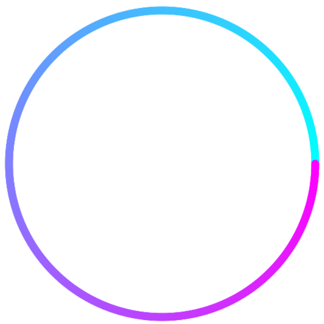
We are observing an immersed manifold $\mathcal{M} \subset \mathbb{R}^n$.

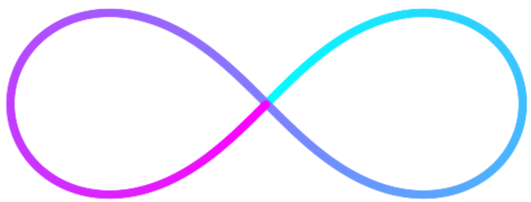
Abstract manifold	Immersion	Immersed manifold
$\mathcal{M}_0$	$\xrightarrow{u}$	$\mathcal{M} = u(\mathcal{M}_0) \quad \subset \mathbb{R}^n$
	$\longrightarrow$	
Klein bottle	$\longrightarrow$	
	$\longrightarrow$	

Statement of the problem

We are observing an immersed manifold $\mathcal{M} \subset \mathbb{R}^n$.

Abstract manifold	Immersion	Immersed manifold
$\mathcal{M}_0$	$\xrightarrow{\quad u \quad}$	$\mathcal{M} = u(\mathcal{M}_0) \qquad \subset \mathbb{R}^n$





$\mathcal{M}$

X

Homology groups

$$H_0 = \mathbb{Z}/2\mathbb{Z}$$

$$H_1 = \mathbb{Z}/2\mathbb{Z}$$

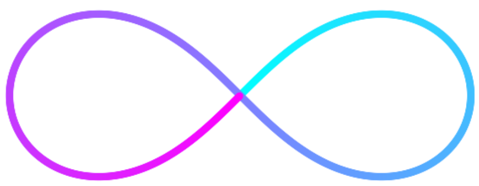
Problem:

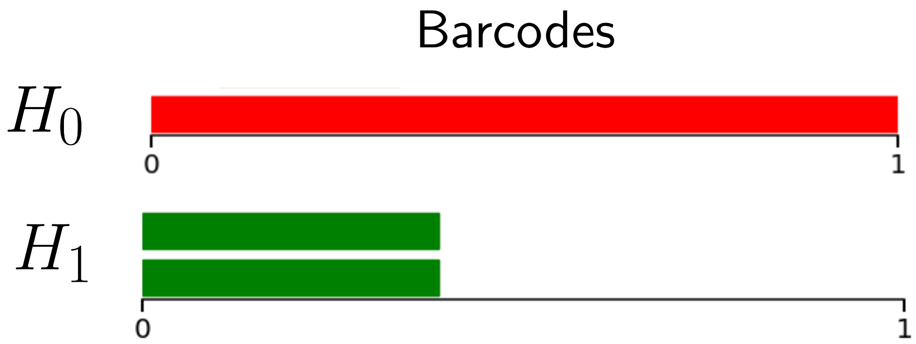
Given a point cloud $X \subset \mathbb{R}^n$ close to $\mathcal{M}$, compute the homology groups of $\mathcal{M}_0$.

Our method

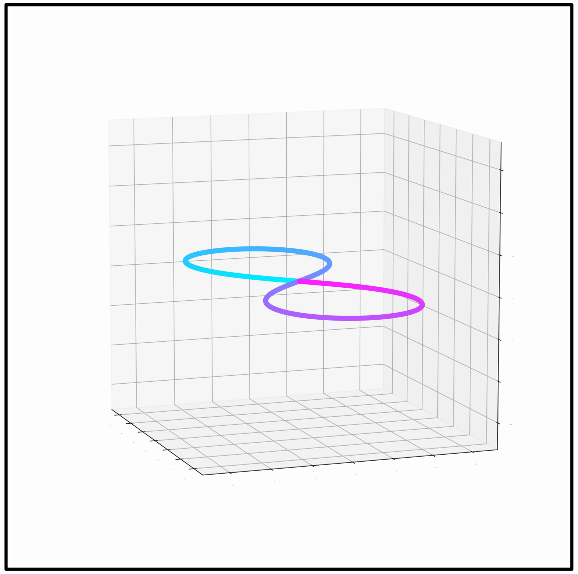
We will use persistent homology.

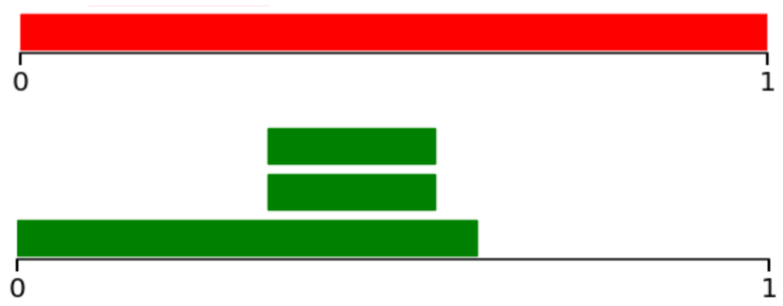
Unfortunately, the persistent homology of the Čech filtration of $\mathcal{M}$ does not reveal the homology of $\mathcal{M}_0$.





We will lift $\mathcal{M}$ in a higher dimensional space, where the Čech filtration reveals a circle.





Our method

How to lift $\mathcal{M}$?

$$
\begin{array}{ccc}
\mathcal{M}_0 & \xrightarrow{\quad u \quad} & \mathcal{M} \subset \mathbb{R}^n \\
x_0 & \longmapsto & u(x_0)
\end{array}
$$

$$
\begin{array}{ccc}
\mathcal{M}_0 & \xrightarrow{\quad \check{u} \quad} & \check{\mathcal{M}} \subset \mathbb{R}^n \times \mathbb{R}^m \\
x_0 & \longmapsto & (u(x_0), f(x_0))
\end{array}
$$

Choose f such that $\check{u}$ is an embedding.

Our method

How to lift $\mathcal{M}$?

$$\begin{array}{ccc} \mathcal{M}_0 & \xrightarrow{\qquad u \qquad} & \mathcal{M} \subset \mathbb{R}^n \\ x_0 & \longmapsto & u(x_0) \end{array}$$

$$\begin{array}{ccc} \mathcal{M}_0 & \xrightarrow{\qquad \check{u} \qquad} & \check{\mathcal{M}} \subset \mathbb{R}^n \times \mathbb{R}^m \\ x_0 & \longmapsto & (u(x_0), f(x_0)) \end{array}$$

Choose f such that $\check{u}$ is an embedding.

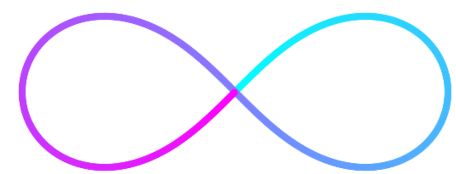
Our choice is $\boxed{f : x_0 \longmapsto T_{x_0}\mathcal{M}_0}$ (tangent space of $\mathcal{M}_0$ at x_0)

- $\check{u}$ is an embedding under a reasonable assumption
- we are actually estimating the tangent bundle of $\mathcal{M}_0$

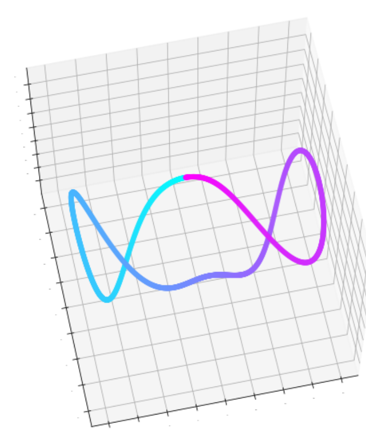
Our method

Notations:

- $u\colon \mathcal{M}_0 \to \mathcal{M} \subset \mathbb{R}^n$ is an immersion
- For $x_0 \in \mathcal{M}_0$, $x = u(x_0)$
- For $x_0 \in \mathcal{M}_0$, $T_x\mathcal{M}$ denotes the tangent space of $\mathcal{M}_0$ seen in $\mathbb{R}^n$
- $\mathrm{M}(\mathbb{R}^n)$ denotes the space of $n \times n$ matrices
- $p_{T_x\mathcal{M}} \in \mathrm{M}(\mathbb{R}^n)$ denotes the orthogonal projection matrix on $T_x\mathcal{M}$
- *Lift space*: $\mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n)$
- *Lifted manifold*: $\check{\mathcal{M}} = \{(x, p_{T_x\mathcal{M}}), x_0 \in \mathcal{M}_0\} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n)$
- *Lifting map*: $\check{u}\colon \mathcal{M}_0 \to \check{\mathcal{M}}$



$$\begin{tikzcd} & \mathcal{M} \subset \mathbb{R}^n \\ \mathcal{M}_0 \arrow[ur, "u"] \arrow[dr, "\check{u}"'] & \\ & \check{\mathcal{M}} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n) \end{tikzcd}$$



$$\check{u}\colon x_0 \longmapsto (x, p_{T_x\mathcal{M}})$$

Recipe in practice

We aim at estimating the set

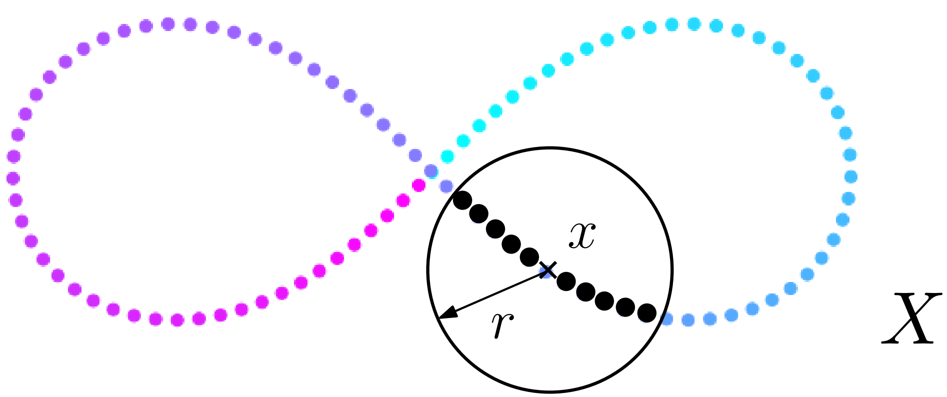
$$\check{\mathcal{M}} = \{(x, p_{T_x\mathcal{M}}), x_0 \in \mathcal{M}_0\} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n).$$

- We observe a point cloud $X \subset \mathbb{R}^n$ close to $\mathcal{M}$.
- Let $r > 0$ be a parameter.
  For every $x \in X$, compute a *local covariance matrix*

$$\Sigma_X(x,r) = \frac{1}{|X \cap \overline{\mathcal{B}}(x,r)|} \sum_{y \in X \cap \overline{\mathcal{B}}(x,r)} (x-y)^{\otimes 2} \qquad \in \mathrm{M}(\mathbb{R}^n)$$

- Consider the set

$$\check{X} = \{(x, \Sigma_X(x,r)), x \in X\} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n).$$



Recipe in practice

We aim at estimating the set

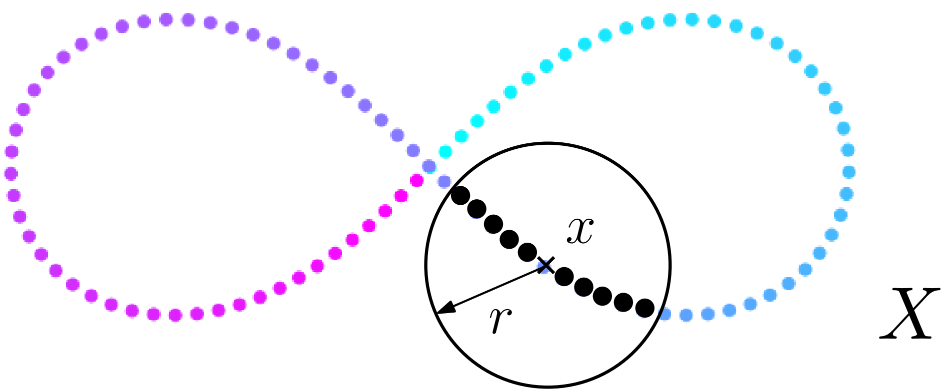
$$\check{\mathcal{M}} = \{(x, p_{T_x\mathcal{M}}), x_0 \in \mathcal{M}_0\} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n).$$

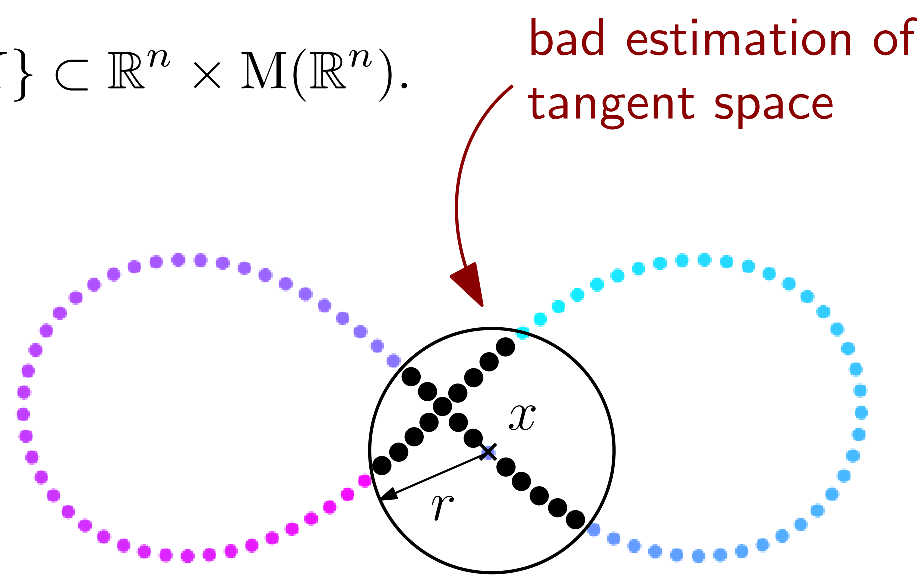
- We observe a point cloud $X \subset \mathbb{R}^n$ close to $\mathcal{M}$.
- Let $r > 0$ be a parameter.
  For every $x \in X$, compute a *local covariance matrix*

$$\Sigma_X(x,r) = \frac{1}{|X \cap \overline{\mathcal{B}}(x,r)|} \sum_{y \in X \cap \overline{\mathcal{B}}(x,r)} (x-y)^{\otimes 2} \qquad \in \mathrm{M}(\mathbb{R}^n)$$

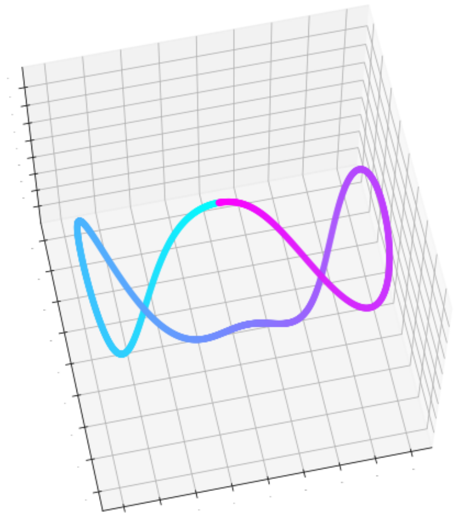
- Consider the set

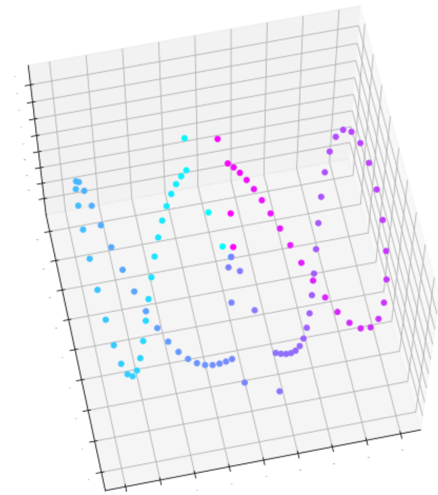
$$\check{X} = \{(x, \Sigma_X(x,r)), x \in X\} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n).$$





Recipe in practice





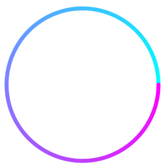
$$\check{\mathcal{M}} = \{(x, p_{T_x\mathcal{M}}), x_0 \in \mathcal{M}_0\}$$

$$\check{X} = \{(x, \Sigma_X(x, r)), x \in \mathcal{M}\}$$

$\check{\mathcal{M}}$ and $\check{X}$ are not close in Hausdorff distance...

But they are in Wasserstein distance!

A measure-theoretic setting

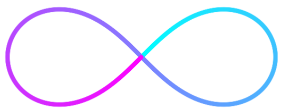


$\mathcal{M}_0$ $\xrightarrow{u}$ $\mathcal{M} \subset \mathbb{R}^n$

μ_0 measure on $\mathcal{M}_0$

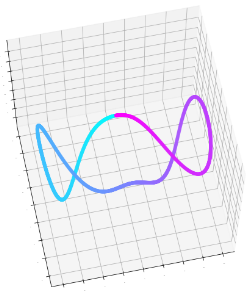
μ push-forward measure on $\mathcal{M}$

ν measure on $\mathbb{R}^n$



$\mathcal{M}_0$ $\xrightarrow{\check{u}}$ $\check{\mathcal{M}} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^m)$

$\check{\mu}_0$ push-forward measure on $\check{\mathcal{M}}$



where $\check{u}\colon x_0 \longmapsto \left(x, \frac{1}{d+2} p_{T_x\mathcal{M}}\right)$

$\check{\mu}_0$ can be defined as follows: for every test function $\phi\colon \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n) \to \mathbb{R}$,

$$\int \phi(x, A) \cdot \mathrm{d}\check{\mu}_0(x, A) = \int \phi\left(x, \frac{1}{d+2} p_{T_x\mathcal{M}}\right) \cdot \mathrm{d}\mu_0(x_0).$$

A measure-theoretic setting

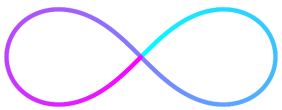
μ_0 measure on $\mathcal{M}_0$

$$\begin{tikzcd} \mathcal{M}_0 \arrow[r, "u"] \arrow[dr, "\check{u}"] & \mathcal{M} \subset \mathbb{R}^n \\ & \check{\mathcal{M}} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^m) \end{tikzcd}$$

$\mathcal{M} \subset \mathbb{R}^n$

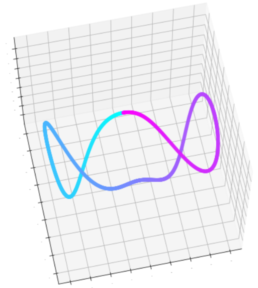
μ push-forward measure on $\mathcal{M}$

ν measure on $\mathbb{R}^n$



$\check{\mathcal{M}} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^m)$

$\check{\mu}_0$ push-forward measure on $\check{\mathcal{M}}$



where $\check{u}\colon x_0 \longmapsto \left(x, \frac{1}{d+2} p_{T_x\mathcal{M}}\right)$

$\check{\mu}_0$ can be defined as follows: for every test function $\phi\colon \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n) \to \mathbb{R}$,

$$\int \phi(x, A) \cdot \mathrm{d}\check{\mu}_0(x, A) = \int \phi\left(x, \frac{1}{d+2} p_{T_x\mathcal{M}}\right) \cdot \mathrm{d}\mu_0(x_0).$$

Now, we are observing a measure ν close to μ

Define $\check{\nu}$ as follows: for every test function $\phi\colon \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n) \to \mathbb{R}$,

$$\int \phi(x, A) \cdot \mathrm{d}\check{\nu}(x, A) = \int \phi\left(x, \frac{1}{r^2} \Sigma_\nu(x, r)\right) \cdot \mathrm{d}\nu(x),$$

where $\Sigma_\nu(x, r)$ is the local covariance matrix.

A measure-theoretic setting

Theorem

Let ν be any probability measure on $\mathbb{R}^n$. Suppose that $\mathrm{W}_1(\mu, \nu)$ and r are small enough. Under technical assumptions on $\mathcal{M}_0$ and μ_0, we have

$$\mathrm{W}_p(\check{\nu}, \check{\mu}_0) \leq \text{constant} \cdot r^{\frac{1}{p}}$$

where W_p denote the p-Wasserstein distance.

$\check{\mu}_0$ can be defined as follows: for every test function $\phi \colon \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n) \to \mathbb{R}$,

$$\int \phi(x, A) \cdot \mathrm{d}\check{\mu}_0(x, A) = \int \phi\left(x, \frac{1}{d+2} p_{T_x \mathcal{M}}\right) \cdot \mathrm{d}\mu_0(x_0).$$

Now, we are observing a measure ν close to μ

Define $\check{\nu}$ as follows: for every test function $\phi \colon \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n) \to \mathbb{R}$,

$$\int \phi(x, A) \cdot \mathrm{d}\check{\nu}(x, A) = \int \phi\left(x, \frac{1}{r^2} \Sigma_\nu(x, r)\right) \cdot \mathrm{d}\nu(x),$$

where $\Sigma_\nu(x, r)$ is the local covariance matrix.

A measure-theoretic setting

Theorem

Let ν be any probability measure on $\mathbb{R}^n$. Suppose that $\mathrm{W}_1(\mu, \nu)$ and r are small enough. Under technical assumptions on $\mathcal{M}_0$ and μ_0, we have

$$\mathrm{W}_p(\check{\nu}, \check{\mu}_0) \leq \text{constant} \cdot r^{\frac{1}{p}}$$

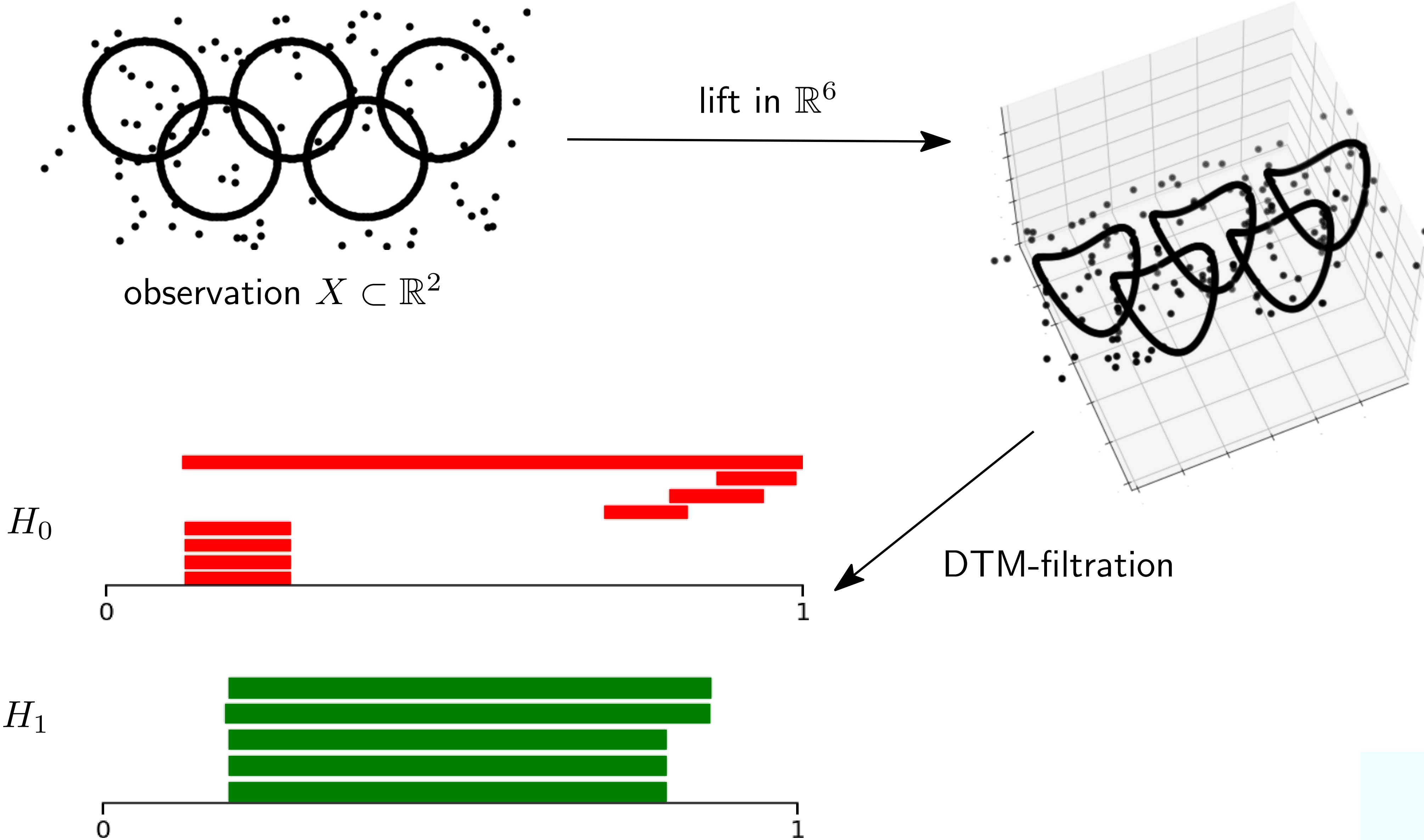
where W_p denote the p-Wasserstein distance.

Corollary

Let ν be any probability measure on $\mathbb{R}^n$. Suppose that $\mathrm{W}_2(\mu, \nu)$ and r are small enough.
Under technical assumptions on $\mathcal{M}_0$ and μ_0, there exists $\epsilon > 0$ such that, for every $t \in \left[4\epsilon, \mathrm{reach}\left(\check{\mathcal{M}}\right) - 3\epsilon\right]$, the sublevel set of the DTM $\mathrm{d}_{\check{\nu},m}^{-1}([0,t])$ is homotopy equivalent to $\mathcal{M}_0$.

DTM-Filtration on the lifted measure

Let $\check{\nu}$ be the lifted measure, on $\mathbb{R}^n \times \mathrm{M}(\mathbb{R}^n)$. We apply the DTM-filtration to it.

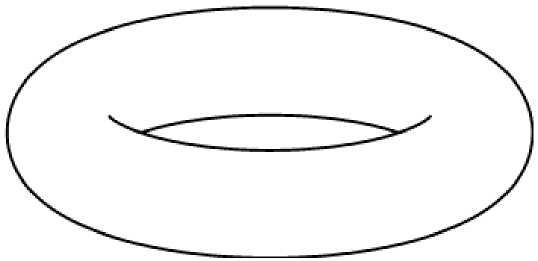


Persistent Stiefel-Whitney classes

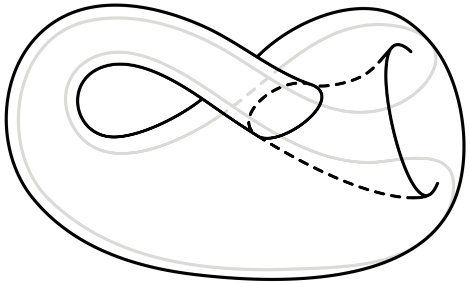
Statement of the problem

Persistent homology allows to estimate the *homology* of a space.

Over $\mathbb{Z}/2\mathbb{Z}$, homology may not be fine enough to distinguish between non-homeomorphic spaces.



Torus



Klein bottle

The Stiefel-Whitney classes are a refinement of cohomology, that allows to differenciate such spaces.

They are defined for any topological space X endowed with a vector bundle ξ. For all $i \in \mathbb{N}$, the i^{th} Stiefel-Whitney class is denoted

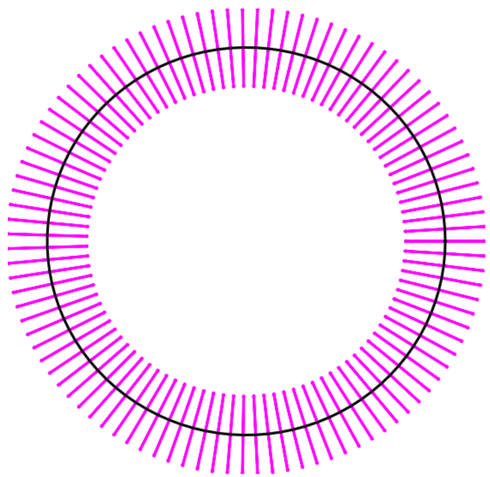
$$w_i(\xi) \in H^i(X).$$

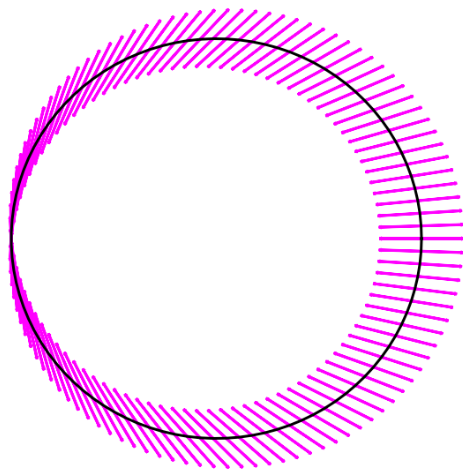
How to estimate the Stiefel-Whitney classes from a point cloud?

A practical definition of vector bundles

(First) Definition: A vector bundle over X is a surjection $\pi : E \to X$ whose fibers are vector spaces and which satisfies a local triviality condition.

Normal bundle of the circle





Möbius strip

- If V is a vector space, the Grassmannian of d-planes of V is denoted $\mathcal{G}_d(V)$. It is the set of d-dimensional subspaces of V.
- Let $\mathbb{R}^\infty$ denote the space of sequences of $\mathbb{R}$ that are 0 from some point.

Thanks to the correspondance between vector bundles and classifying maps, we have an alternative definition.

(Second) Definition

A vector bundle over X is a continuous map $\pi\colon X \to \mathcal{G}_d(\mathbb{R}^\infty)$ or $\pi : X \to \mathcal{G}_d(\mathbb{R}^m)$.

A practical definition of vector bundles

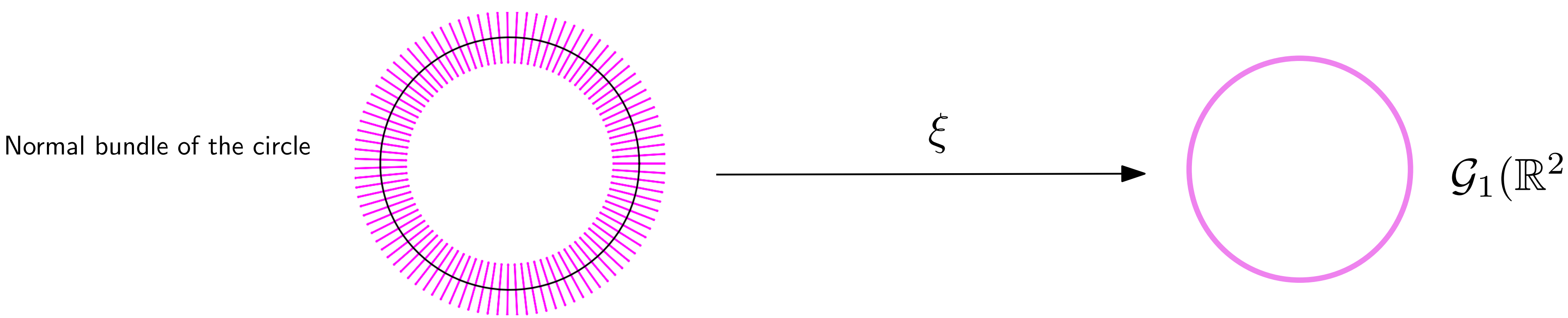
Recall that the Grassmannian has $\mathbb{Z}/2\mathbb{Z}$-cohomology

$$H^*(\mathcal{G}_d(\mathbb{R}^\infty)) = \mathbb{Z}/2\mathbb{Z}[w_1, ..., w_d]$$

where w_i has degree i.

The Stiefel-Whitney classes of the vector bundle $\xi\colon X \to \mathcal{G}_d(\mathbb{R}^\infty)$ can be defined as

$$w_i(\xi) = \xi^*(\omega_i).$$



$$\begin{array}{ccc} H^*(\mathbb{S}_1) & \xleftarrow{\xi^*} & H^*(\mathcal{G}_d(\mathbb{R}^\infty)) \\ w_1(\xi) = 0 & \longleftarrow\!\shortmid & w_1 \end{array}$$

Filtrations of vector bundles

- Suppose that $X \subset \mathbb{R}^n$. For any vector bundle $\xi \colon X \to \mathcal{G}_d(\mathbb{R}^m)$, consider the lifted space

$$\check{X} = \left\{ (x, \xi(x)), x \in X \right\}.$$

It is a subset of $\mathbb{R}^n \times \mathcal{G}_d(\mathbb{R}^m)$.

- We embed $\mathcal{G}_d(\mathbb{R}^m)$ in $\mathrm{M}(\mathbb{R}^m)$ via

$$T \mapsto \text{projection matrix onto } T.$$

Then $\check{X}$ can be seen as a subset of $\mathbb{R}^n \times \mathrm{M}(\mathbb{R}^m)$.

- Let $V[\check{X}] = \left(\check{X}^t\right)_{t \geq 0}$ be the Čech filtration of $\check{X}$ in the space $\mathbb{R}^n \times \mathrm{M}(\mathbb{R}^m)$ endowed with the norm $\|(x, A)\| = \sqrt{\|x\|^2 + \|A\|_{\mathrm{F}}^2}$.

The thickening $\check{X}^t$ is endowed with a natural vector bundle structure $\xi^t \colon \check{X}^t \to \mathcal{G}_d(\mathbb{R}^m)$ defined as

$$(x, A) \in \check{X}^t \to \mathrm{proj}\left(A, \mathcal{G}_d(\mathbb{R}^m)\right).$$

Filtrations of vector bundles

Definition

The i^{th} *persistent Stiefel-Whitney class* of $\check{X}$ is the collection

$$w_i(\check{X}) = \left(w_i^t(\check{X})\right)_t$$

where $w_i^t(\check{X}) = (\xi^t)^*(\omega_i)$ is the i^{th} Stiefel-Whitney class of the vector bundle $\xi^t \colon \check{X}^t \to \mathcal{G}_d(\mathbb{R}^m)$.

Issue: the map ξ^t is not well-defined for every value of $t \geq 0$: A must not be in the medial axis of $\mathcal{G}_d(\mathbb{R}^m)$ in $\mathrm{M}(\mathbb{R}^m)$.

The thickening $\check{X}^t$ is endowed with a natural vector bundle structure $\xi^t \colon \check{X}^t \to \mathcal{G}_d(\mathbb{R}^m)$ defined as

$$(x, A) \in \check{X}^t \to \mathrm{proj}\left(A, \mathcal{G}_d(\mathbb{R}^m)\right).$$

Filtrations of vector bundles

Definition
The i^{th} *persistent Stiefel-Whitney class* of $\check{X}$ is the collection

$$w_i(\check{X}) = \left(w_i^t(\check{X})\right)_t$$

where $w_i^t(\check{X}) = (\xi^t)^*(\omega_i)$ is the i^{th} Stiefel-Whitney class of the vector bundle $\xi^t \colon \check{X}^t \to \mathcal{G}_d(\mathbb{R}^m)$.

Issue: the map ξ^t is not well-defined for every value of $t \geq 0$: A must not be in the medial axis of $\mathcal{G}_d(\mathbb{R}^m)$ in $\mathrm{M}(\mathbb{R}^m)$.

Lemma
For any $A \in \mathrm{M}(\mathbb{R}^m)$, let A^s denote the matrix $A^s = \frac{1}{2}(A + {}^tA)$, and let $\lambda_1(A^s), ..., \lambda_n(A^s)$ be the eigenvalues of A^s in decreasing order. The distance from A to $\mathrm{med}\left(\mathcal{G}_d(\mathbb{R}^m)\right)$ is $\frac{\sqrt{2}}{2}\left|\lambda_d(A^s) - \lambda_{d+1}(A^s)\right|$.

The thickening $\check{X}^t$ is endowed with a natural vector bundle structure $\xi^t \colon \check{X}^t \to \mathcal{G}_d(\mathbb{R}^m)$ defined as

$$(x, A) \in \check{X}^t \to \mathrm{proj}\left(A, \mathcal{G}_d(\mathbb{R}^m)\right).$$

Lifebar of a persistent class

Let $X \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^m)$ and $w_i(X)$.

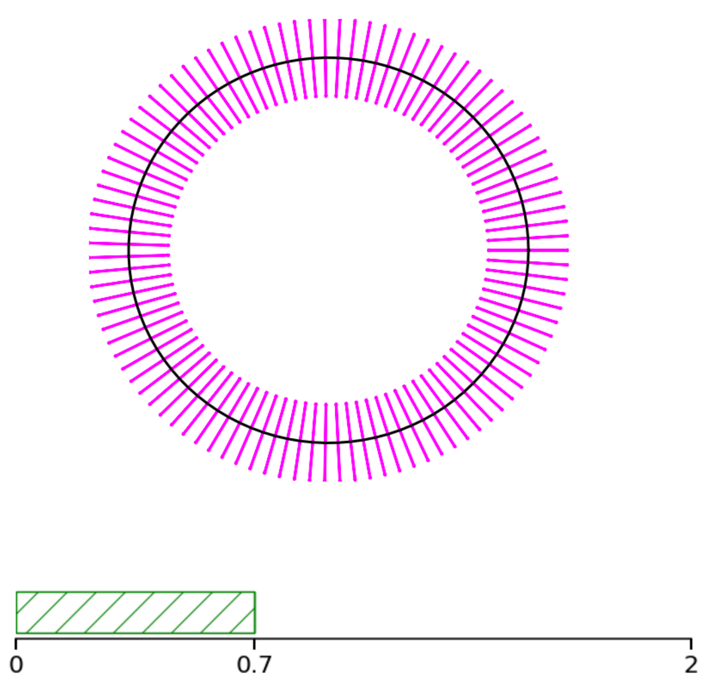
Definition

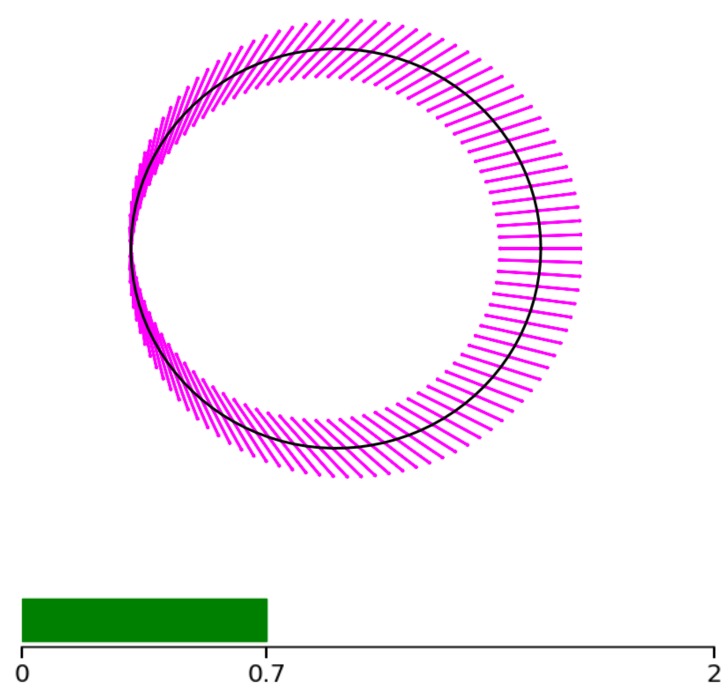
The *lifebar* of the persistent Stiefel-Whitney class $w_i(X)$ is the set

$$\left\{t < t_{\mathsf{max}}, w_i^t(X) \neq 0\right\}.$$

the lifebar is an interval!

Example: lifebars of first persistent Stiefel-Whitney classes

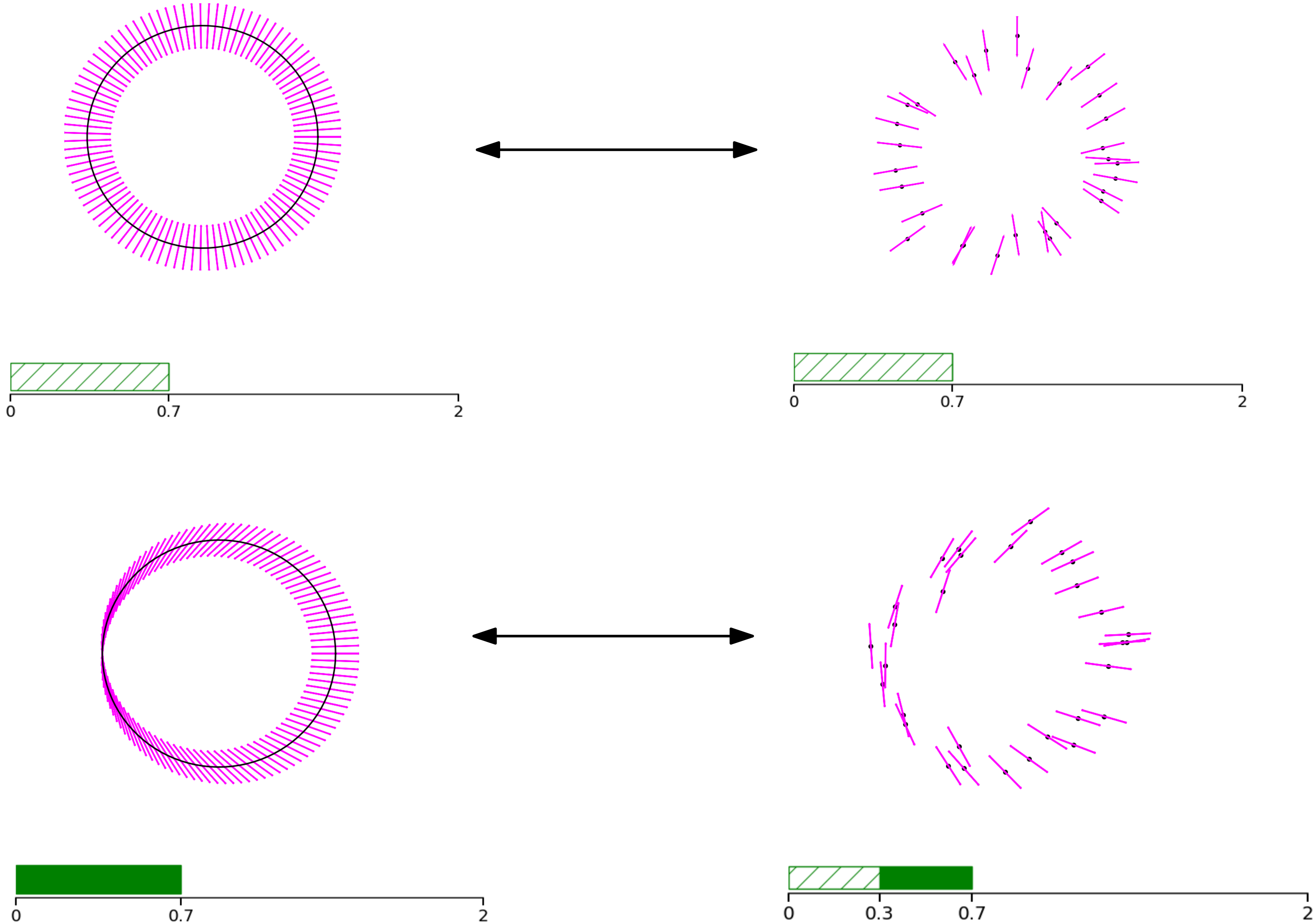




Stability

Theorem

If two subsets $X, Y \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^m)$ satisfies $\mathrm{d_H}\,(X, Y) \leq \epsilon$, then for all $i \geq 0$, the lifebars of their i^{th} Stiefel-Whitney classes are ϵ-close.



Consistency

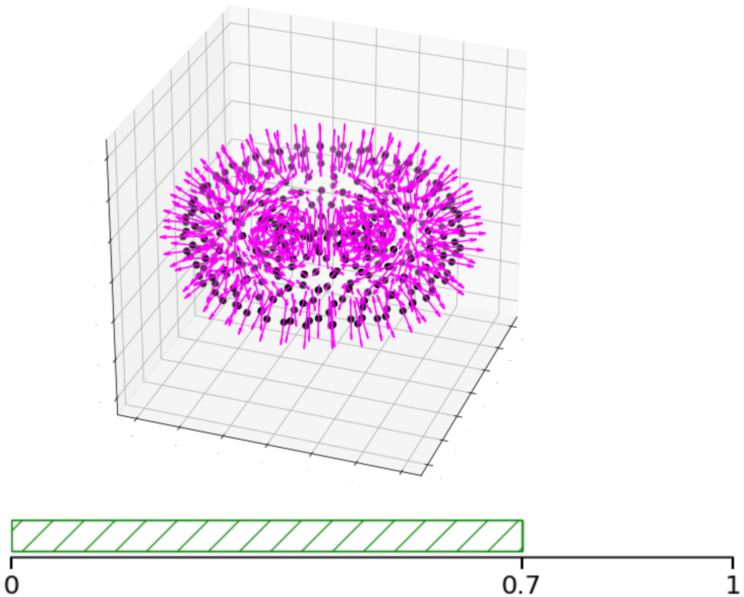
If $u\colon \mathcal{M}_0 \to \mathcal{M} \subset \mathbb{R}^n$ is an immersion and $\xi\colon \mathcal{M}_0 \to \mathcal{G}_d(\mathbb{R}^m)$ a vector bundle, consider the set

$$\check{\mathcal{M}} = \{(u(x_0), \xi(x_0)), x_0 \in \mathcal{M}_0\} \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^m).$$

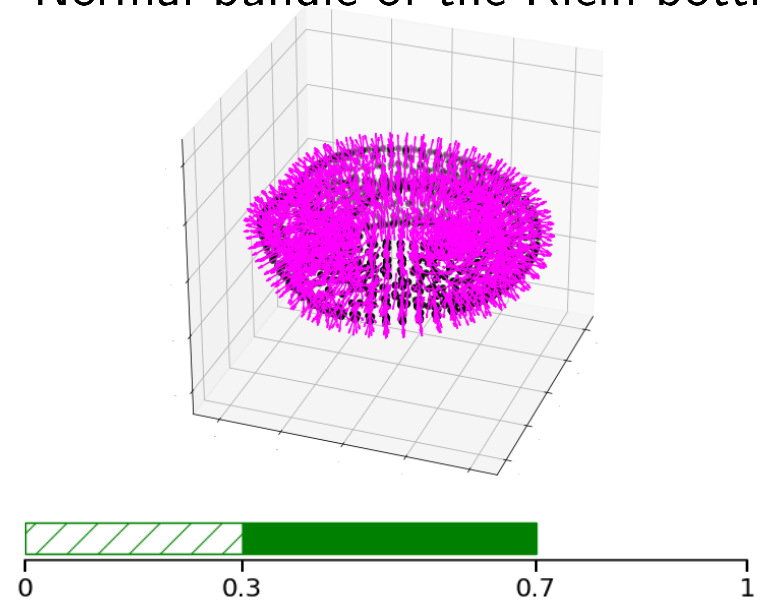
Theorem

Let $X \subset \mathbb{R}^n \times \mathrm{M}(\mathbb{R}^m)$ be any subset such that $\mathrm{d_H}\left(X, \check{\mathcal{M}}\right) \leq \epsilon$. Then for every $t \in [4\epsilon, \mathrm{reach}\left(\check{\mathcal{M}}\right) - 3\epsilon)$, the composition of inclusions $\mathcal{M}_0 \hookrightarrow \check{\mathcal{M}} \hookrightarrow X^t$ induces an isomorphism $H^*(\mathcal{M}_0) \leftarrow H^*(X^t)$ which sends the i^{th} persistent Stiefel-Whitney class $w_i^t(X)$ of the Čech bundle filtration of X to the i^{th} Stiefel-Whitney class of $(\mathcal{M}_0, p)$.

Normal bundle of the torus

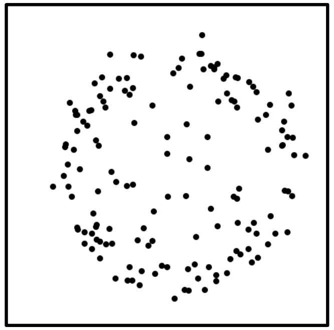


Normal bundle of the Klein bottle

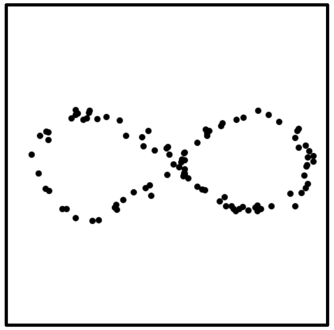


Conclusion

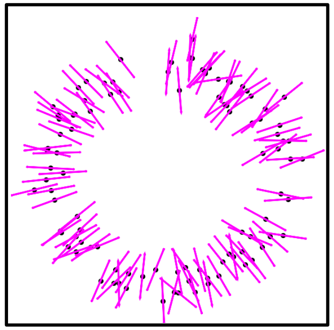
- Persistent homology for point clouds with anomalous points



- Persistent homology for point clouds lying on immersed manifolds
  (normal reach, tangent space estimation via local stability of measures)



- Persistent homology for vector bundles
  (weak star condition, triangulation of the projective spaces)

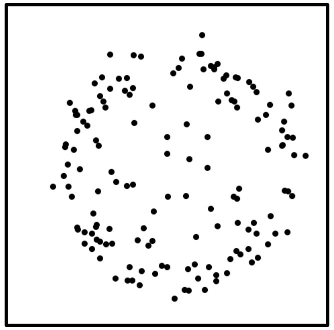


Perspectives:

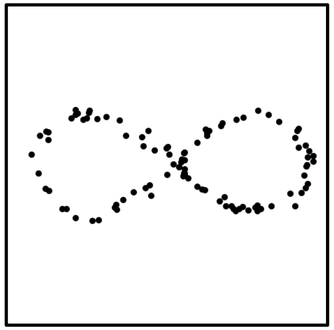
- Study of stratified spaces
- Study of more general fiber bundles, triangulation of Grassmann manifolds

Conclusion

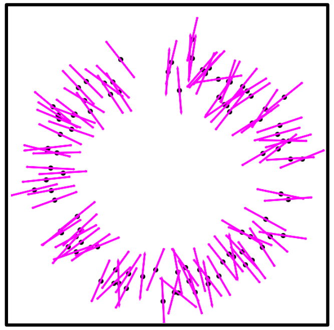
- Persistent homology for point clouds with anomalous points



- Persistent homology for point clouds lying on immersed manifolds
  (normal reach, tangent space estimation via local stability of measures)



- Persistent homology for vector bundles
  (weak star condition, triangulation of the projective spaces)



Perspectives:

- Study of stratified spaces
- Study of more general fiber bundles, triangulation of Grassmann manifolds

Thank you