

# Inférence topologique avec l'homologie persistante

Raphaël Tinarrage

Séminaire MACS, UMPA Lyon

I - Homologie singulière

II - Homologie persistante

III - Exemples

IV - Homologie persistante pour mesures

Soit  $\mathbb{Z}/2\mathbb{Z}$  le corps à deux éléments.

A tout espace topologique  $X$  est associé une suite d'espaces vectoriels sur  $\mathbb{Z}/2\mathbb{Z}$ , notés

$$H_0(X), H_1(X), H_2(X), \dots$$

# Homologie singulière

3/21 (2/7)

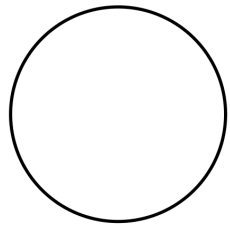
Soit  $\mathbb{Z}/2\mathbb{Z}$  le corps à deux éléments.

A tout espace topologique  $X$  est associé une suite d'espaces vectoriels sur  $\mathbb{Z}/2\mathbb{Z}$ , notés

$$H_0(X), H_1(X), H_2(X), \dots$$

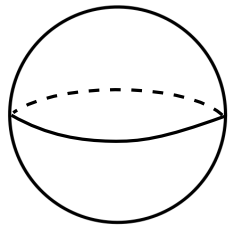
Exemples :

- Si  $X$  est le cercle  $\mathbb{S}_1 \subset \mathbb{R}^2$ , alors



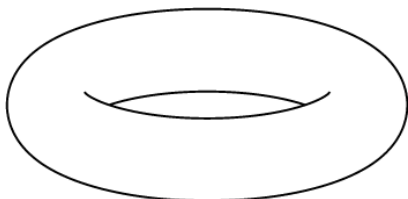
$$H_0(\mathbb{S}_1) = \mathbb{Z}/2\mathbb{Z}, \quad H_1(\mathbb{S}_1) = \mathbb{Z}/2\mathbb{Z}, \quad H_2(\mathbb{S}_1) = 0$$

- Si  $X$  est la sphère  $\mathbb{S}_n \subset \mathbb{R}^{n+1}$ , alors



$$H_0(\mathbb{S}_n) = \mathbb{Z}/2\mathbb{Z}, \quad H_1(\mathbb{S}_n) = 0, \quad \dots, H_{n-1}(\mathbb{S}_n) = 0, \quad H_n(\mathbb{S}_n) = \mathbb{Z}/2\mathbb{Z}$$

- Si  $X$  est le tore  $\mathbb{T} \subset \mathbb{R}^3$ , alors



$$H_0(\mathbb{T}) = \mathbb{Z}/2\mathbb{Z}, \quad H_1(\mathbb{T}) = (\mathbb{Z}/2\mathbb{Z})^2, \quad H_2(\mathbb{T}) = \mathbb{Z}/2\mathbb{Z}$$

Soit  $\mathbb{Z}/2\mathbb{Z}$  le corps à deux éléments.

A tout espace topologique  $X$  est associé une suite d'espaces vectoriels sur  $\mathbb{Z}/2\mathbb{Z}$ , notés

$$H_0(X), H_1(X), H_2(X), \dots$$

## Construction des groupes d'homologie :

Soit  $\Delta^n$  le  $n$ -simplexe (l'espace topologique défini comme l'enveloppe convexe des vecteurs de la base canonique de  $\mathbb{R}^{n+1}$ ). Un  $n$ -simplexe singulier est une application continue  $\sigma: \Delta^n \rightarrow X$ . Pour tout  $i \in [0, n]$ , sa  $i^{\text{ème}}$  face est le  $(n-1)$ -simplexe singulier défini par  $\delta_i \sigma: (t_0, \dots, t_{n-1}) \mapsto \sigma(t_0, \dots, t_i, 0, t_{i+1}, \dots, t_{n-1})$ .

Soit  $C_n(X)$  le groupe libre engendré par les  $n$ -simplexes et avec coefficients dans  $\mathbb{Z}/2\mathbb{Z}$ . On définit l'opérateur de bord  $\partial_n: C_n(X) \rightarrow C_{n-1}(X)$  comme  $\partial_n(\sigma) = \sum_{i=0}^n \delta_i \sigma$ . On a la relation  $\delta_n \circ \delta_{n+1} = 0$ .

Pour tout  $n \geq 0$ , on définit les  $n$ -cycles  $Z_n(X) = \ker(\partial_n)$  et les  $n$ -bords  $B_n(X) = \text{im}(\partial_{n+1})$ . La relation  $Z_n(X) \subseteq B_n(X)$  permet de définir le  $n^{\text{ème}}$  groupe d'homologie  $H_n(X) = Z_n(X)/B_n(X)$ .

Soit  $\mathbb{Z}/2\mathbb{Z}$  le corps à deux éléments.

A tout espace topologique  $X$  est associé une suite d'espaces vectoriels sur  $\mathbb{Z}/2\mathbb{Z}$ , notés

$$H_0(X), H_1(X), H_2(X), \dots$$

Interprétation des groupes d'homologie :

- $H_0(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_0}$ , où  $k_0$  est le nombre de composantes connexes de  $X$
- $H_1(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_1}$ , où  $k_1$  est le nombre de "trous dans  $X$ "
- $H_2(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_2}$ , où  $k_2$  est le nombre de "trous de dimension 2 dans  $X$ "

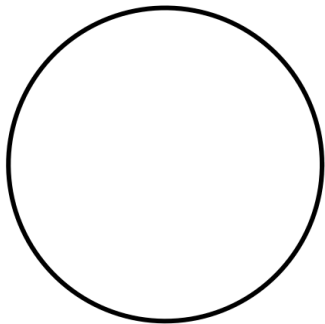
Soit  $\mathbb{Z}/2\mathbb{Z}$  le corps à deux éléments.

A tout espace topologique  $X$  est associé une suite d'espaces vectoriels sur  $\mathbb{Z}/2\mathbb{Z}$ , notés

$$H_0(X), H_1(X), H_2(X), \dots$$

Interprétation des groupes d'homologie :

- $H_0(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_0}$ , où  $k_0$  est le nombre de composantes connexes de  $X$
- $H_1(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_1}$ , où  $k_1$  est le nombre de "trous dans  $X$ "
- $H_2(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_2}$ , où  $k_2$  est le nombre de "trous de dimension 2 dans  $X$ "



$$H_0(\mathbb{S}_1) = \mathbb{Z}/2\mathbb{Z}, \quad H_1(\mathbb{S}_1) = \mathbb{Z}/2\mathbb{Z}, \quad H_2(\mathbb{S}_1) = 0, \quad \dots$$

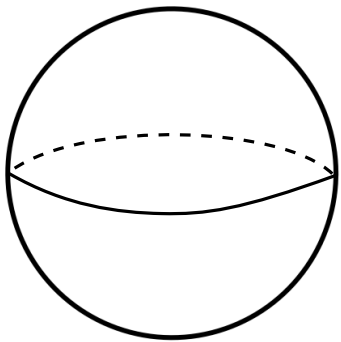
Soit  $\mathbb{Z}/2\mathbb{Z}$  le corps à deux éléments.

A tout espace topologique  $X$  est associé une suite d'espaces vectoriels sur  $\mathbb{Z}/2\mathbb{Z}$ , notés

$$H_0(X), H_1(X), H_2(X), \dots$$

Interprétation des groupes d'homologie :

- $H_0(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_0}$ , où  $k_0$  est le nombre de composantes connexes de  $X$
- $H_1(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_1}$ , où  $k_1$  est le nombre de "trous dans  $X$ "
- $H_2(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_2}$ , où  $k_2$  est le nombre de "trous de dimension 2 dans  $X$ "



$$H_0(\mathbb{S}_2) = \mathbb{Z}/2\mathbb{Z}, \quad H_1(\mathbb{S}_2) = 0, \quad H_2(\mathbb{S}_2) = \mathbb{Z}/2\mathbb{Z}$$

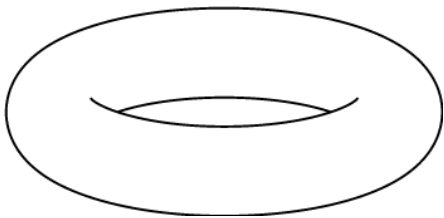
Soit  $\mathbb{Z}/2\mathbb{Z}$  le corps à deux éléments.

A tout espace topologique  $X$  est associé une suite d'espaces vectoriels sur  $\mathbb{Z}/2\mathbb{Z}$ , notés

$$H_0(X), H_1(X), H_2(X), \dots$$

Interprétation des groupes d'homologie :

- $H_0(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_0}$ , où  $k_0$  est le nombre de composantes connexes de  $X$
- $H_1(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_1}$ , où  $k_1$  est le nombre de "trous dans  $X$ "
- $H_2(X) \simeq (\mathbb{Z}/2\mathbb{Z})^{k_2}$ , où  $k_2$  est le nombre de "trous de dimension 2 dans  $X$ "



$$H_0(\mathbb{T}) = \mathbb{Z}/2\mathbb{Z}, \quad H_1(\mathbb{T}) = (\mathbb{Z}/2\mathbb{Z})^2, \quad H_2(\mathbb{T}) = \mathbb{Z}/2\mathbb{Z}$$

# Fonctorialité de l'homologie

4/21 (1/2)

Le  $n^{\text{ème}}$  groupe d'homologie est un foncteur  $\text{Top} \longrightarrow \text{Vect}$ , où  $\text{Top}$  est la catégorie des espaces topologiques, et  $\text{Vect}$  celle des  $\mathbb{Z}/2\mathbb{Z}$ -espaces vectoriels.

Autrement dit, on peut aussi transformer les applications entre espaces topologiques.

$$X \xrightarrow{f} Y$$

$$H_n(X) \xrightarrow{H_n(f)} H_n(Y)$$

Cette opération préserve les diagrammes commutatifs :

$$\begin{array}{ccccc} & & g \circ f & & \\ & \nearrow & & \searrow & \\ X & \xrightarrow{f} & Y & \xrightarrow{g} & Z, \end{array}$$

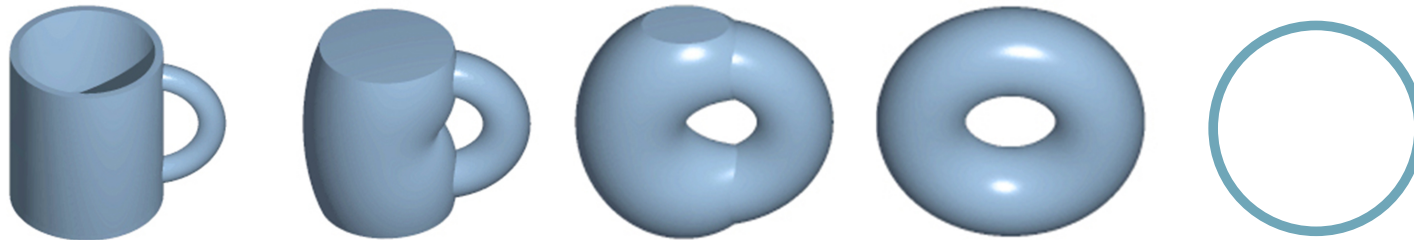
$$\begin{array}{ccccc} & & H_n(g \circ f) & & \\ & \nearrow & & \searrow & \\ H_n(X) & \xrightarrow{H_n(f)} & H_n(Y) & \xrightarrow{H_n(g)} & H_n(Z). \end{array}$$

# Fonctorialité de l'homologie

4/21 (2/2)

Le  $n^{\text{ème}}$  groupe d'homologie est un foncteur  $\text{Top} \longrightarrow \text{Vect}$ , où  $\text{Top}$  est la catégorie des espaces topologiques, et  $\text{Vect}$  celle des  $\mathbb{Z}/2\mathbb{Z}$ -espaces vectoriels.

Conséquence : deux espaces topologique du même type d'homotopie ont les mêmes groupes d'homologie.



$$H_0 = \mathbb{Z}/2\mathbb{Z}, \quad H_1 = \mathbb{Z}/2\mathbb{Z}, \quad H_2 = 0$$

## Invariance du domaine

Montrons que  $\mathbb{R}^n$  et  $\mathbb{R}^m$ , avec  $n \neq m$ , ne sont pas homéomorphes.

Soit  $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$  un homéomorphisme.

Soit  $x \in \mathbb{R}^n$  quelconque, et considérons l'application restreinte

$$h: \mathbb{R}^n \setminus \{x\} \longrightarrow \mathbb{R}^m \setminus \{h(x)\}$$

On a des isomorphismes en homologie :

$$H_i(h): H_i(\mathbb{R}^n \setminus \{x\}) \simeq H_i(\mathbb{R}^m \setminus \{h(x)\})$$

Mais  $\mathbb{R}^n \setminus \{x\} \simeq \mathbb{S}_{n-1}$ , donc on a

$$H_i(\mathbb{S}_{n-1}) \simeq H_i(\mathbb{S}_{m-1})$$

Contradiction si  $n \neq m$ .



## Théorème de Borsuk-Ulam

Soit  $f: \mathbb{S}_n \rightarrow \mathbb{R}^n$  continue. Montrons qu'il existe  $x$  tel que  $f(x) = f(-x)$ .

Supposons que ce ne soit pas le cas. Soit la fonction  $g: \mathbb{S}_n \rightarrow \mathbb{S}_{n-1}$  définie par

$$g(x) = \frac{f(x) - f(-x)}{\|f(x) - f(-x)\|}$$

Soit un grand cercle  $\mathbb{S}_{n-1} \subset \mathbb{S}_n$  et la restriction  $g': \mathbb{S}_{n-1} \rightarrow \mathbb{S}_{n-1}$ . Cette fonction est impaire, et on peut en déduire que l'application induite

$$H_i(g'): H_i(\mathbb{S}_{n-1}) \rightarrow H_i(\mathbb{S}_{n-1})$$

est l'identité.

On obtient un diagramme absurde pour  $i = n - 1$  :

$$\begin{array}{ccccc} & & g' & & \\ & \curvearrowright & & \curvearrowright & \text{id} \\ \mathbb{S}_{n-1} & \xrightarrow{i} & \mathbb{S}_n & \xrightarrow{g} & \mathbb{S}_{n-1} . \\ & & & & \\ H_i(\mathbb{S}_{n-1}) & \xrightarrow{H_i(i)} & H_i(\mathbb{S}_n) & \xrightarrow{H_i(g)} & H_i(\mathbb{S}_{n-1}) . \end{array}$$

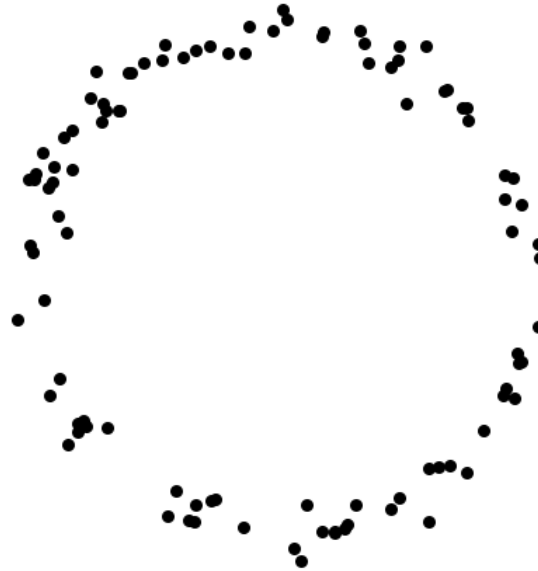
I - Homologie singulière

II - Homologie persistante

III - Exemples

IV - Homologie persistante pour mesures

Après une expérience scientifique, on obtient un nuage de points.



Ici,  $X \subset \mathbb{R}^2$  semble être proche du cercle  $S_1$ .

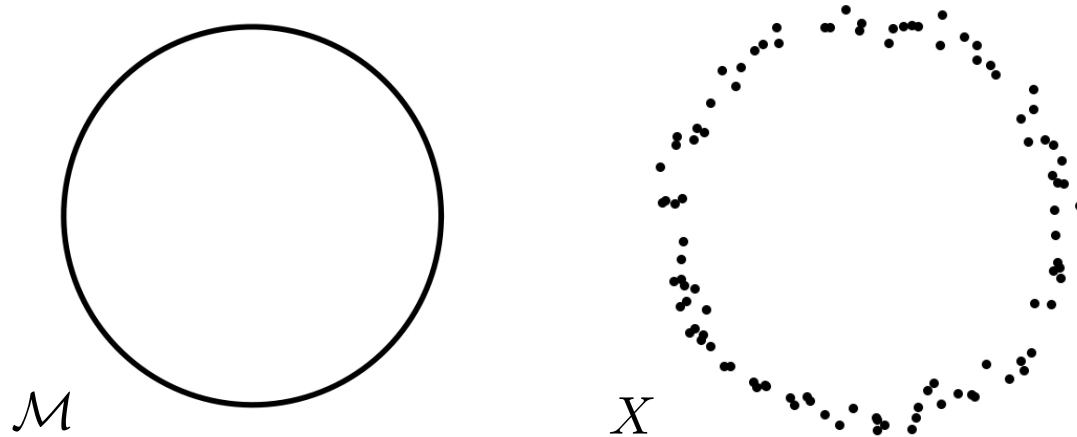
Mais son homologie est décevante :

$$H_0(X) = (\mathbb{Z}/2\mathbb{Z})^{100}, \quad H_1(X) = 0, \quad H_2(X) = 0, \quad \dots$$

# Épaissements

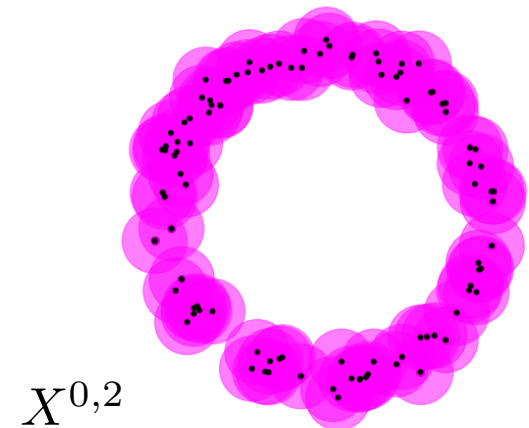
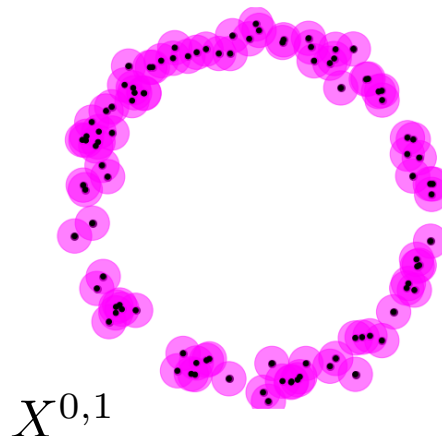
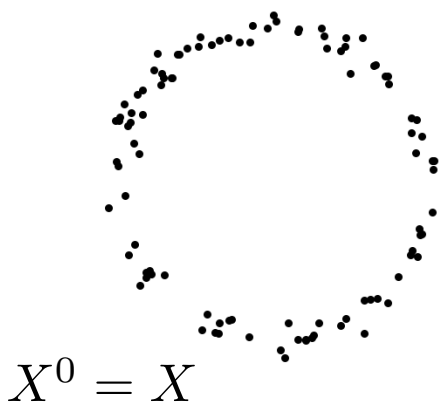
8/21 (1/3)

On va essayer de reconstruire le cercle à partir de  $X$ .



Pour tout  $t \geq 0$ , on définit le  $t$ -épaissement de  $X$ :

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



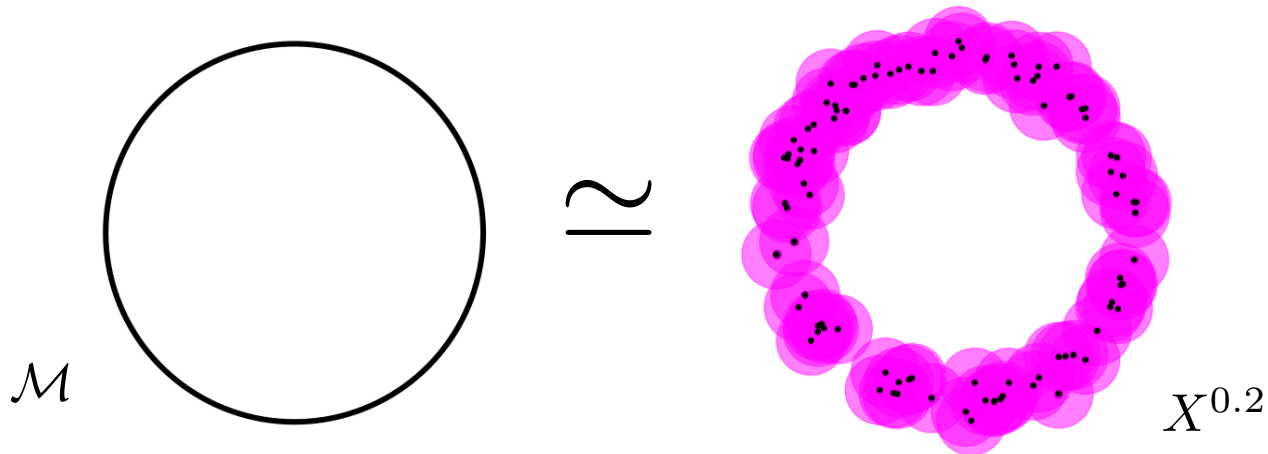
Théorème (Chazal, Cohen-Steiner, Lieutier, 2009)

Soient  $\mathcal{M}, X$  deux sous-ensembles de  $\mathbb{R}^n$ .

On suppose que  $\text{portée}(\mathcal{M}) > 0$  and  $d_H(X, \mathcal{M}) \leq \frac{1}{17} \text{portée}(\mathcal{M})$ . Soit

$$t \in [4d_H(X, \mathcal{M}), \text{portée}(\mathcal{M}) - 3d_H(X, \mathcal{M})].$$

Alors  $X^t$  et  $\mathcal{M}$  ont le même type d'homotopie.



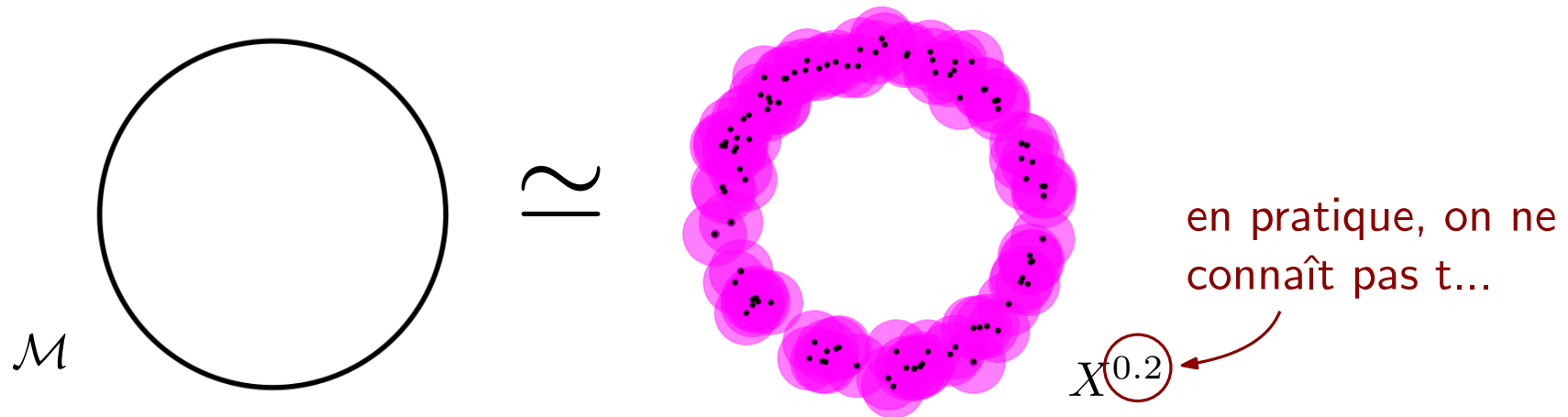
Théorème (Chazal, Cohen-Steiner, Lieutier, 2009)

Soient  $\mathcal{M}, X$  deux sous-ensembles de  $\mathbb{R}^n$ .

On suppose que  $\text{portée}(\mathcal{M}) > 0$  and  $d_H(X, \mathcal{M}) \leq \frac{1}{17} \text{portée}(\mathcal{M})$ . Soit

$$t \in [4d_H(X, \mathcal{M}), \text{portée}(\mathcal{M}) - 3d_H(X, \mathcal{M})].$$

Alors  $X^t$  et  $\mathcal{M}$  ont le même type d'homotopie.



## Definition

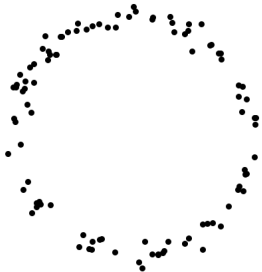
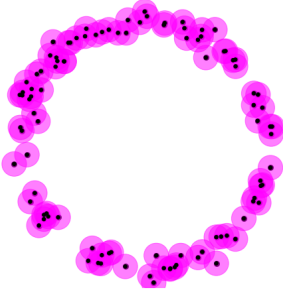
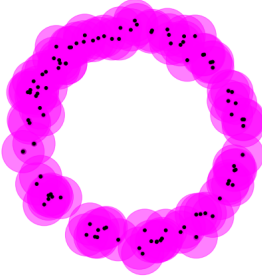
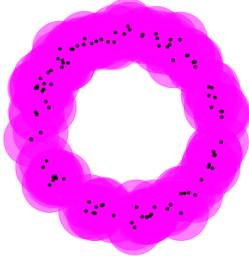
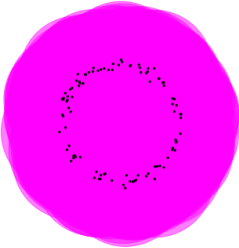
La *filtration de Čech* de  $X$  est la collection :

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

# Homologie de la filtration de Čech

9/21 (1/2)

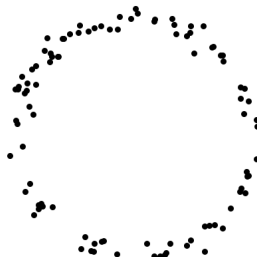
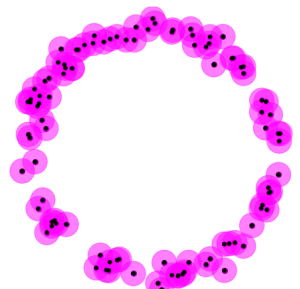
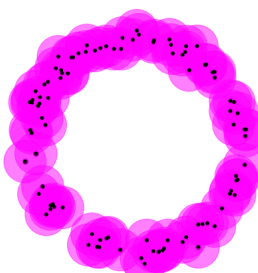
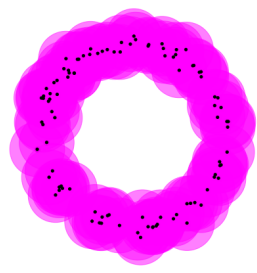
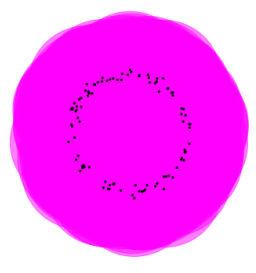
On calcule l'homologie de chacun des épaissements.

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

# Homologie de la filtration de Čech

9/21 (2/2)

On calcule l'homologie de chacun des épaisissements.

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

La donnée de  $(H_i(X^t))_{t \geq 0}$  et  $((i_s^t)_*)_{s \leq t}$  s'appelle un *module de persistance*.

## Définition

Un *module de persistance*  $\mathbb{V}$  sur  $\mathbb{R}^+$  est une famille de  $\mathbb{Z}/2\mathbb{Z}$ -espaces vectoriels  $(V^t)_{t \geq 0}$ , et une famille d'applications linéaires  $(v_s^t: V^s \rightarrow V^t)_{0 \leq s \leq t}$  telles que:

- pour tout  $t \geq 0$ ,  $v_t^t: V^t \rightarrow V^t$  est l'application identité,
- pour tout  $r, s, t \geq 0$  tels que  $r \leq s \leq t$ , on a  $v_s^t \circ v_r^s = v_r^t$ .

$$\begin{array}{ccccc} V^r & \xrightarrow{v_r^s} & V^s & \xrightarrow{v_s^t} & V^t \\ & \searrow & & \nearrow & \\ & & v_r^t & & \end{array}$$

## Construction générale des modules de persistance :

Une *filtration* de  $\mathbb{R}^n$  est une collection de sous-ensembles  $(X_t)_{t \geq 0}$  tels que  $X_s \subset X_t$  quand  $s \leq t$ .

$$\cdots \rightarrow X_{t_1} \xhookrightarrow{i_{t_1}^{t_2}} X_{t_2} \xhookrightarrow{i_{t_2}^{t_3}} X_{t_3} \xhookrightarrow{i_{t_3}^{t_4}} X_{t_4} \hookrightarrow \cdots$$

En appliquant le  $i^{\text{ème}}$  foncteur d'homologie, on obtient un module de persistance.

$$\cdots \rightarrow H_i(X_{t_1}) \xrightarrow{H_i(i_{t_1}^{t_2})} H_i(X_{t_2}) \xrightarrow{H_i(i_{t_2}^{t_3})} H_i(X_{t_3}) \xrightarrow{H_i(i_{t_3}^{t_4})} H_i(X_{t_4}) \hookrightarrow \cdots$$

# Modules de persistance

10/21 (2/3)

## Théorème (Crawley-Boevey, 2015)

Un module de persistance (suffisamment régulier) est isomorphe à une somme de modules-intervalles.

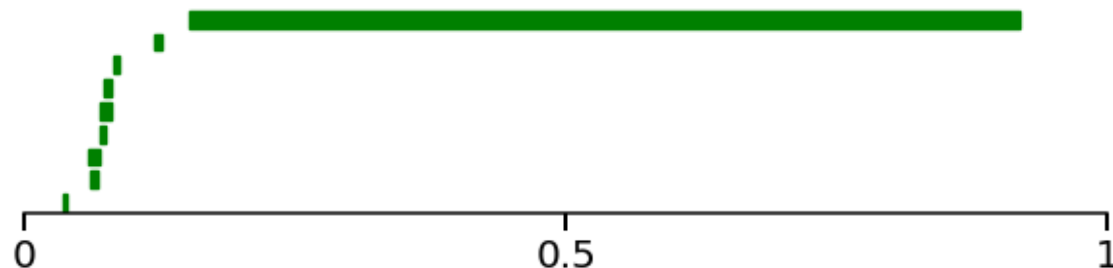
Ce multi-ensemble d'intervalles s'appelle le *code-barre*. C'est un invariant complet des modules de persistance.

Module de persistance :  $\mathbb{V}$

Code-barre :

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

Représentation graphique :



# Modules de persistance

10/21 (3/3)

Codes-barres du module de persistance associé à la filtration de Čech :

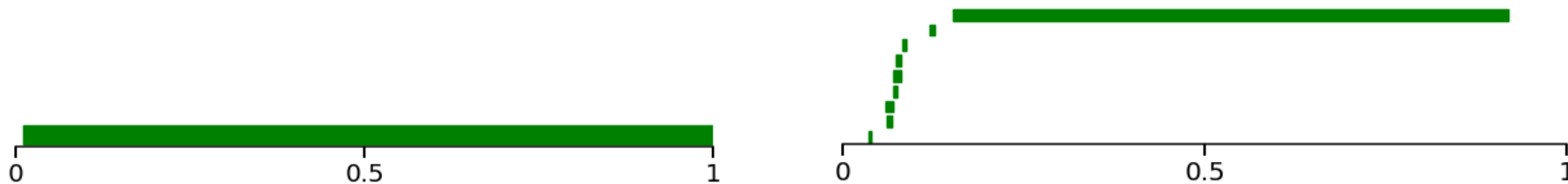
$H_0$  en rouge et  $H_1$  en vert.



# Stabilité des modules de persistance

11/21 (1/7)

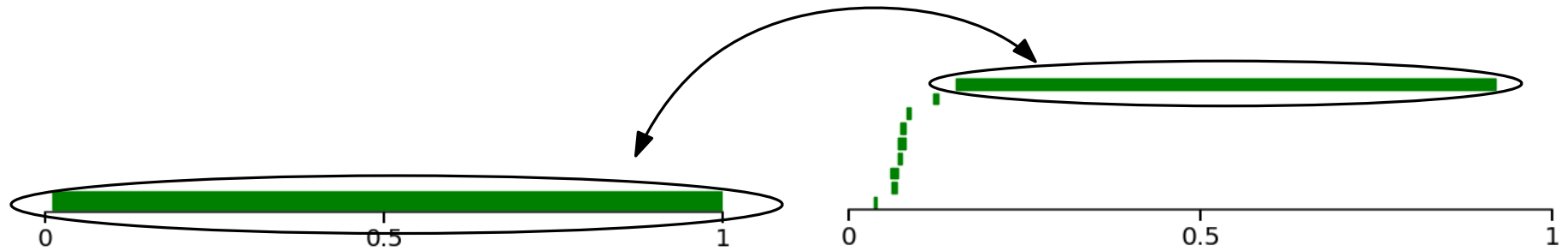
On peut définir une distance sur les codes-barres : la *distance bottleneck*.



# Stabilité des modules de persistance

11/21 (2/7)

On peut définir une distance sur les codes-barres : la *distance bottleneck*.

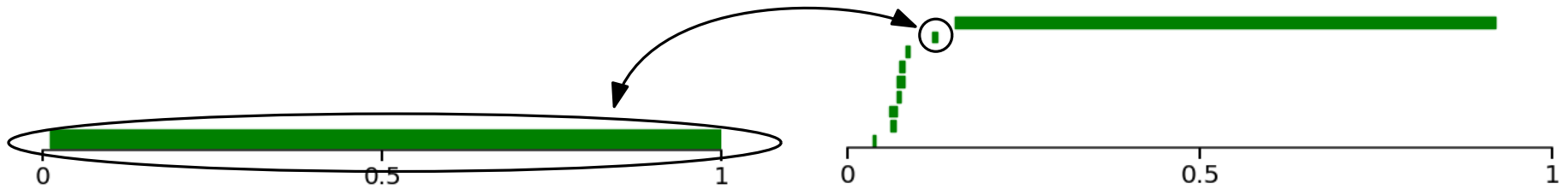


Soit un *appariement partiel* entre deux codes-barres (certaines barres sont appariées, d'autres non).

# Stabilité des modules de persistance

11/21 (3/7)

On peut définir une distance sur les codes-barres : la *distance bottleneck*.

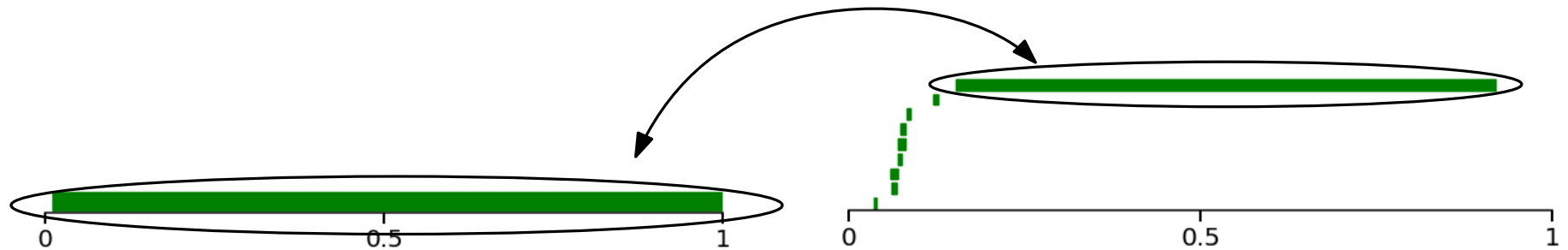


Soit un *appariement partiel* entre deux codes-barres (certaines barres sont appariées, d'autres non).

# Stabilité des modules de persistance

11/21 (4/7)

On peut définir une distance sur les codes-barres : la *distance bottleneck*.



Soit un *appariement partiel* entre deux codes-barres (certaines barres sont appariées, d'autres non).

Son *coût* est le maximum entre :

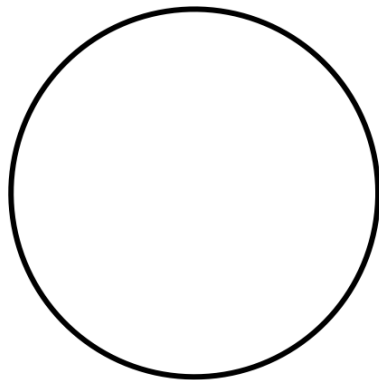
- la différence entre deux barres appariées
- la longueur des barres non-appariées

La distance bottleneck est le plus petit coût parmi les appariements partiels.

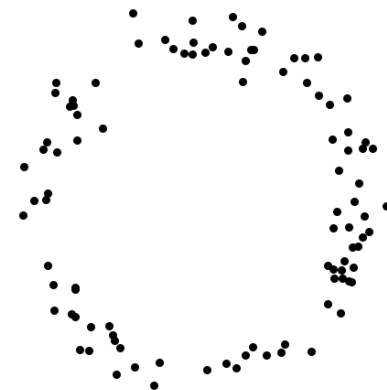
Les petites barres ont peu d'importance pour la distance bottleneck.

# Stabilité des modules de persistance

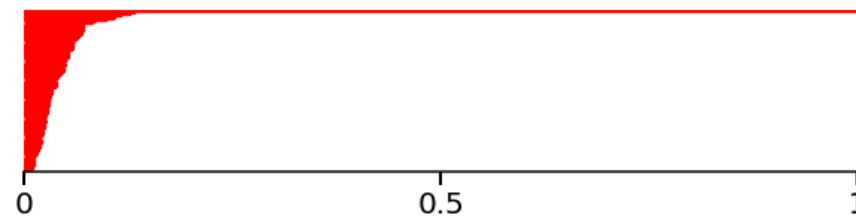
11/21 (5/7)



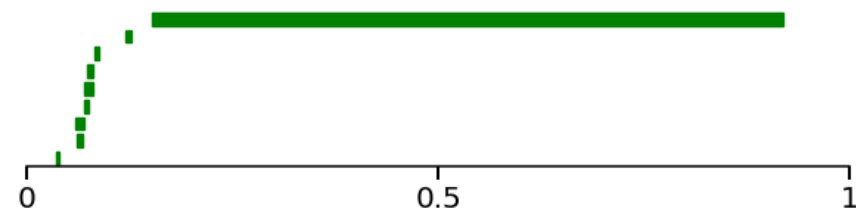
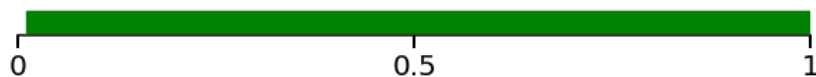
Codes-barres des  
modules de persistance  
associés à la filtration de  
Čech



$H_0$

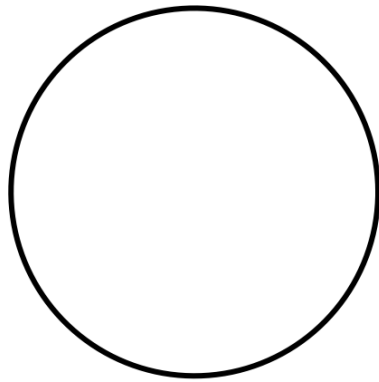


$H_1$

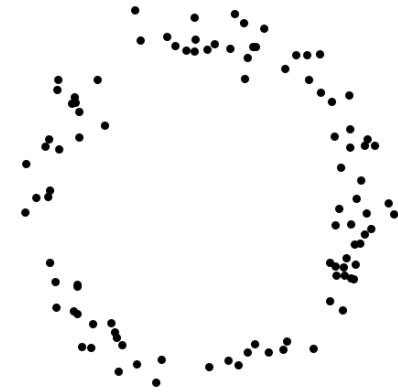


# Stabilité des modules de persistance

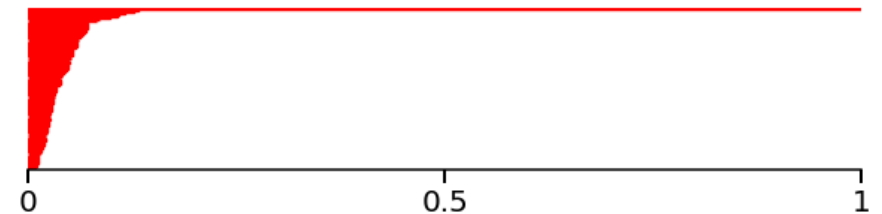
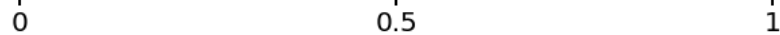
11/21 (6/7)



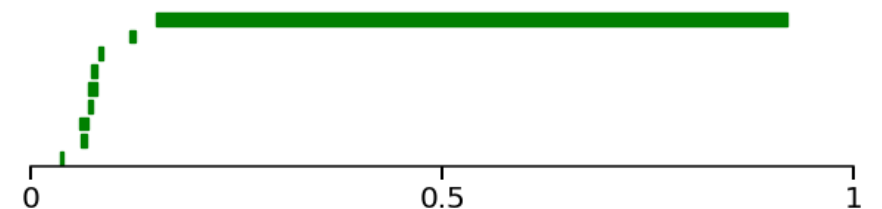
Codes-barres des  
modules de persistance  
associés à la filtration de  
Čech



$H_0$



$H_1$

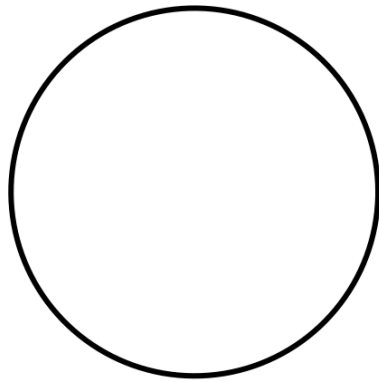


**Théorème de stabilité (Edelsbrunner, Harer, Cohen-Steiner, 2005)**

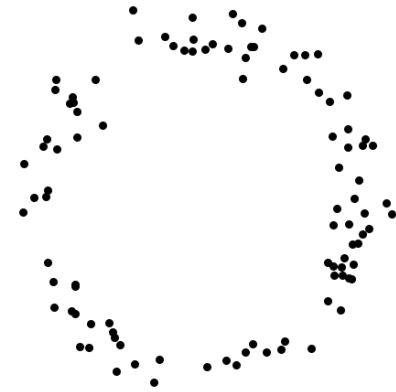
Soit  $X, Y \subset \mathbb{R}^n$  deux sous-ensembles compacts. Alors la distance bottleneck entre les codes-barres de leur filtration de Čech est majorée par la distance de Hausdorff  $d_H(X, Y)$ .

# Stabilité des modules de persistance

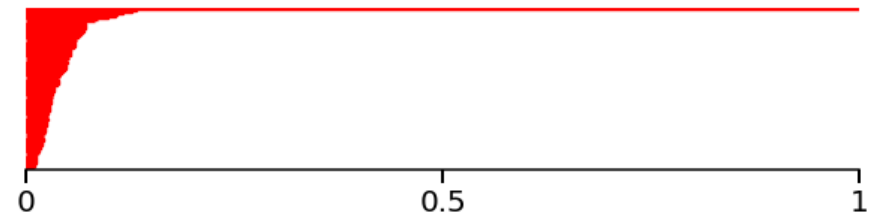
11/21 (7/7)



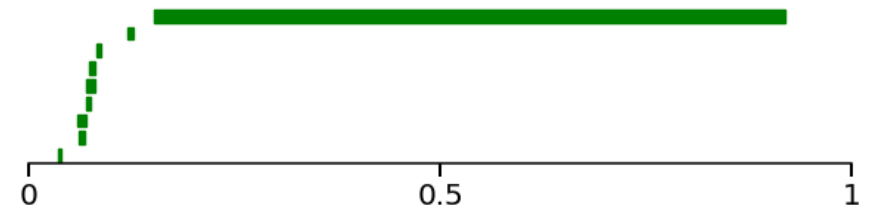
Codes-barres des  
modules de persistance  
associés à la filtration de  
Čech



$H_0$



$H_1$



On déduit l'homologie  $H_0 = \mathbb{Z}/2\mathbb{Z}$  et  $H_1 = \mathbb{Z}/2\mathbb{Z}$ .

I - Homologie singulière

II - Homologie persistante

III - Exemples

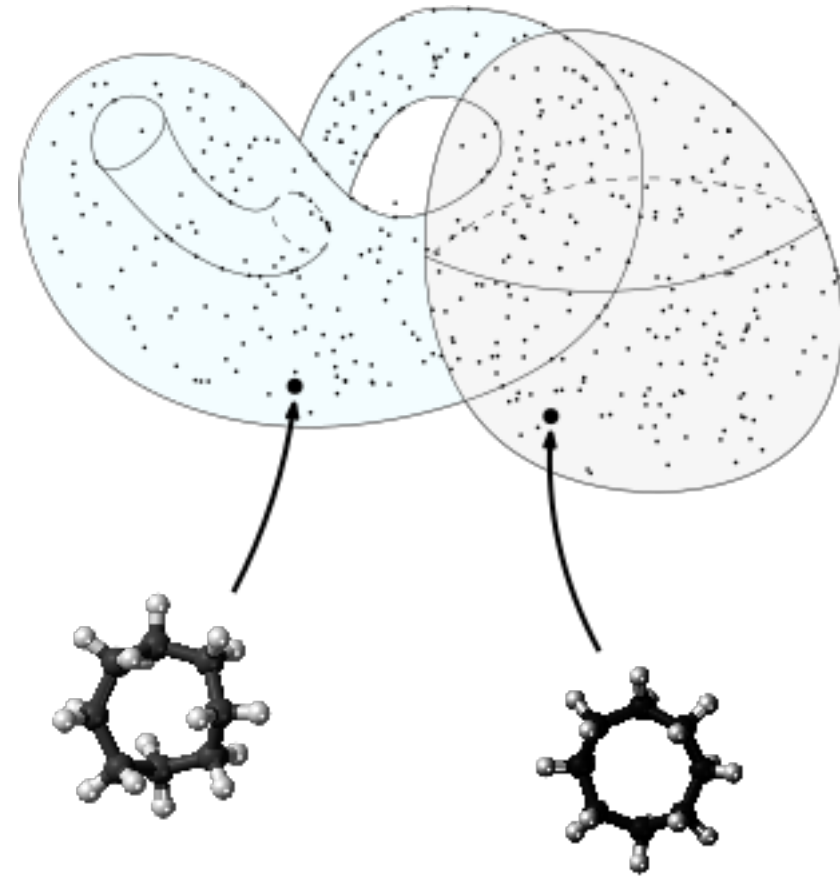
IV - Homologie persistante pour mesures

# Conformations du cyclo-octane

13/21 (1/3)

Depuis [S. Martin, A. Thompson, E. A. Coutsiyas, and J-P. Watson, Topology of cyclo-octane energy landscape, 2010]

Une molécule de cyclooctane  $C_8H_{16}$  peut être transformée en un point dans  $\mathbb{R}^{72}$  ( $3 \times (8 + 16) = 72$ ).



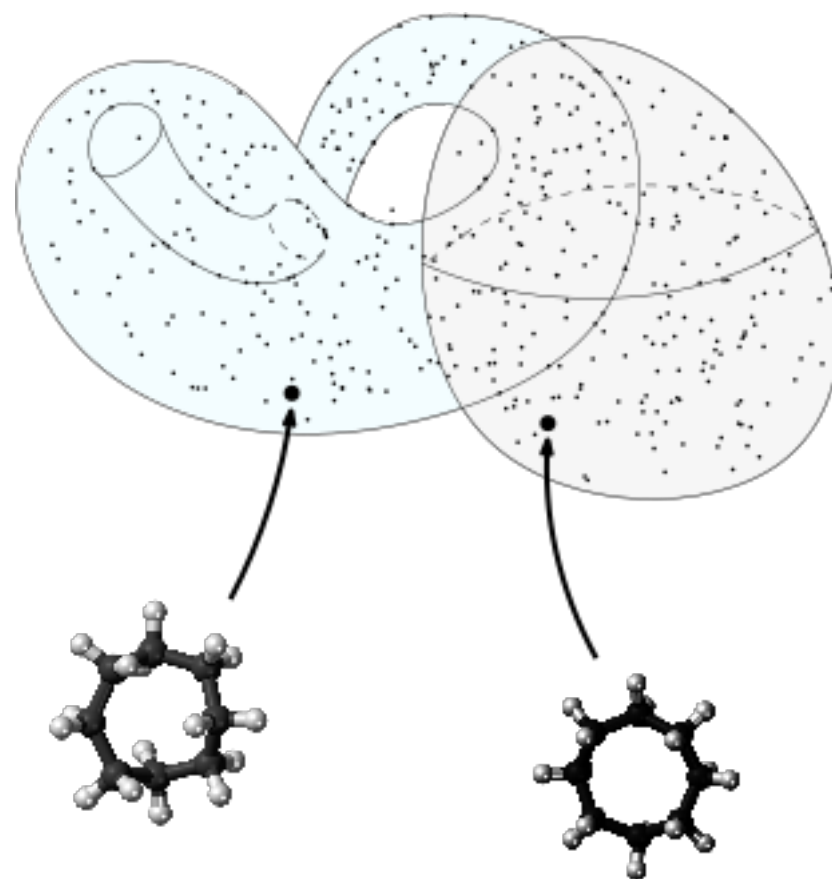
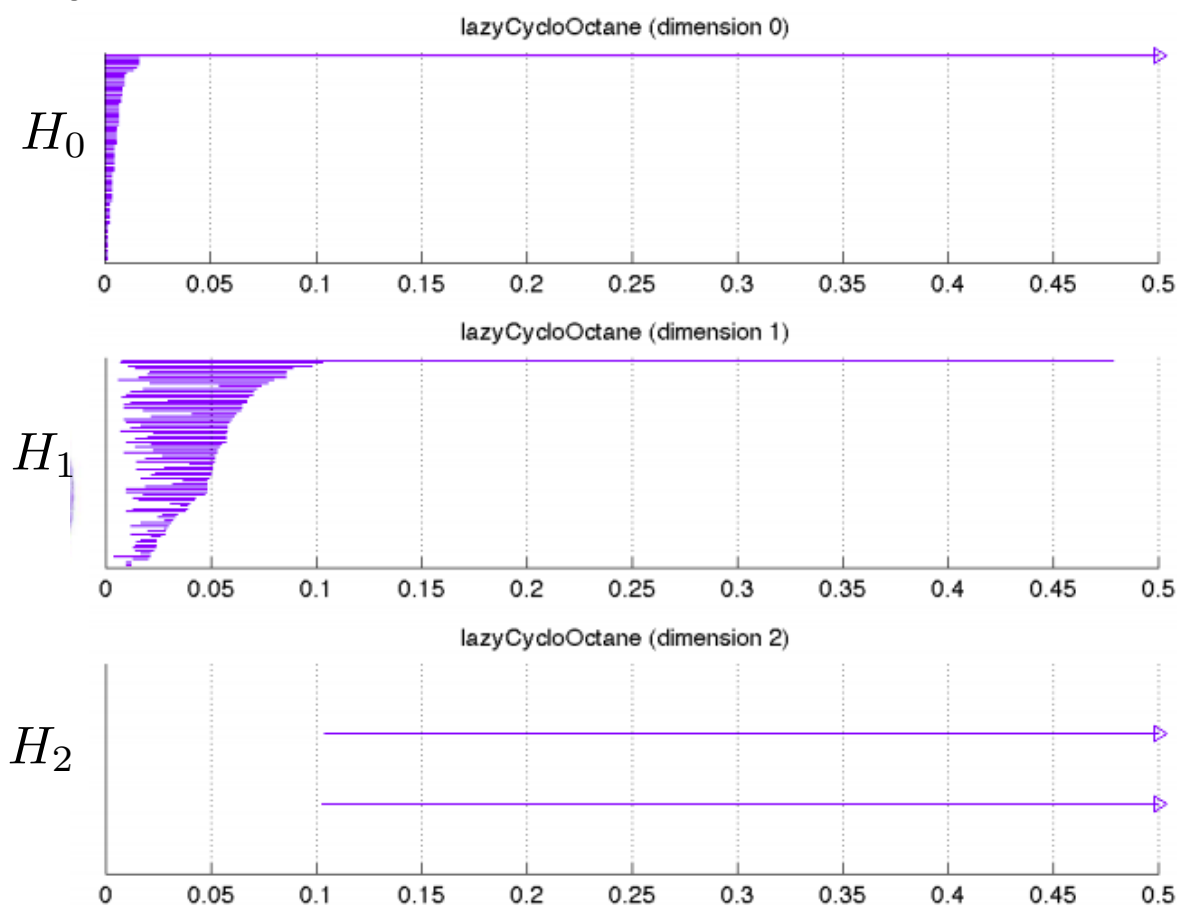
# Conformations du cyclo-octane

13/21 (2/3)

Depuis [S. Martin, A. Thompson, E. A. Coutsiyas, and J-P. Watson, Topology of cyclo-octane energy landscape, 2010]

Une molécule de cyclooctane  $C_8H_{16}$  peut être transformée en un point dans  $\mathbb{R}^{72}$  ( $3 \times (8 + 16) = 72$ ).

On obtient les codes-barres :



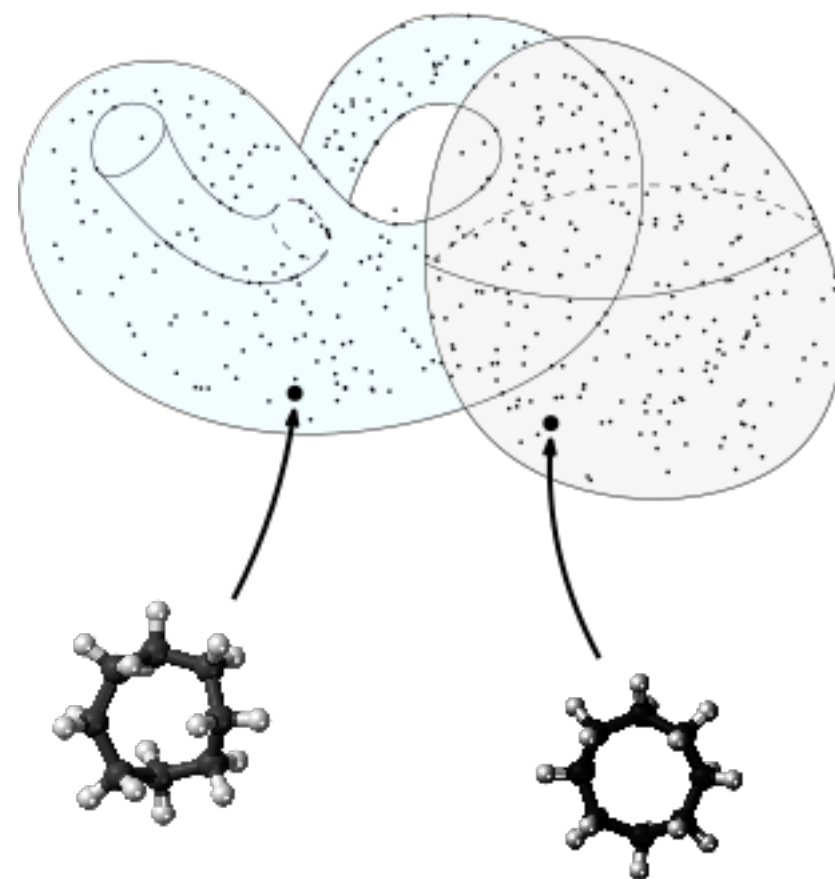
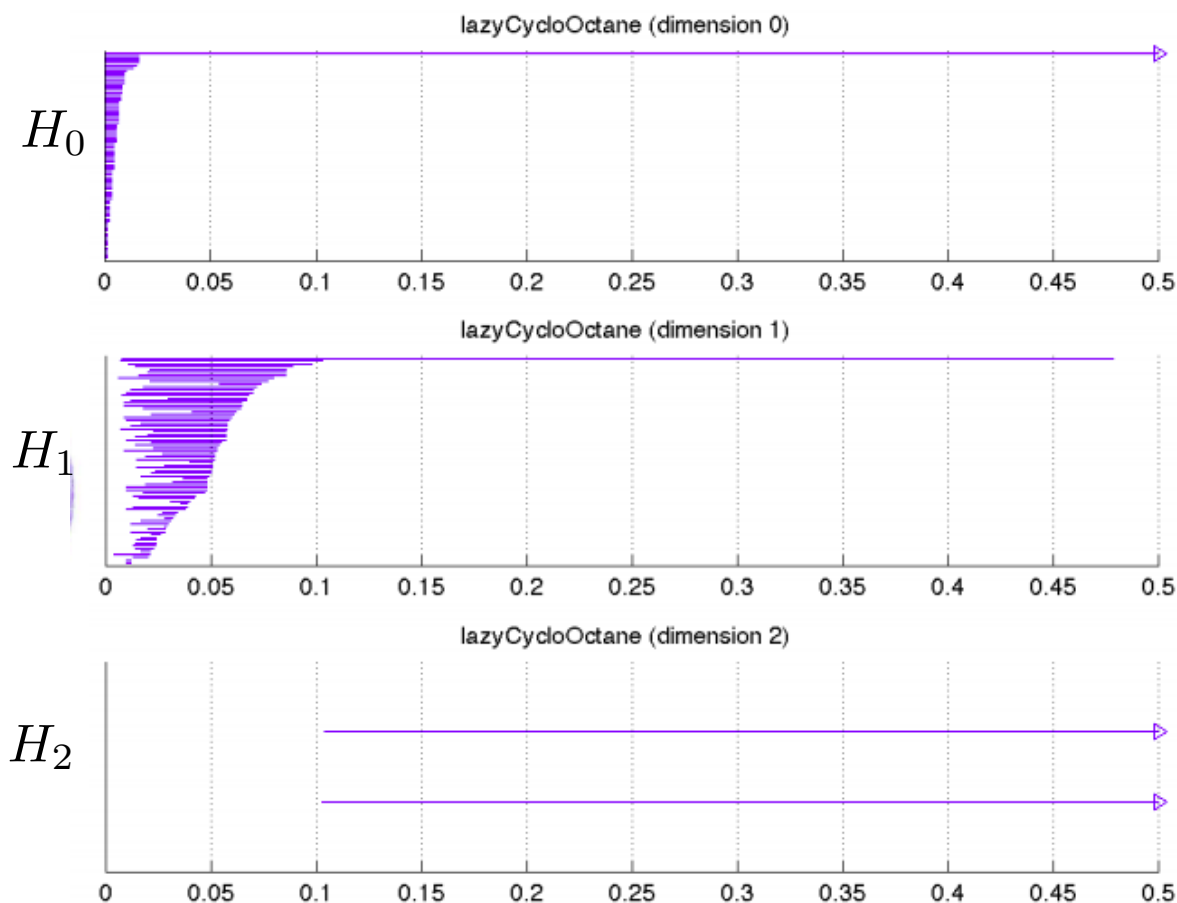
# Conformations du cyclo-octane

13/21 (3/3)

Depuis [S. Martin, A. Thompson, E. A. Coutsiyas, and J-P. Watson, Topology of cyclo-octane energy landscape, 2010]

Une molécule de cyclooctane  $C_8H_{16}$  peut être transformée en un point dans  $\mathbb{R}^{72}$  ( $3 \times (8 + 16) = 72$ ).

On obtient les codes-barres :



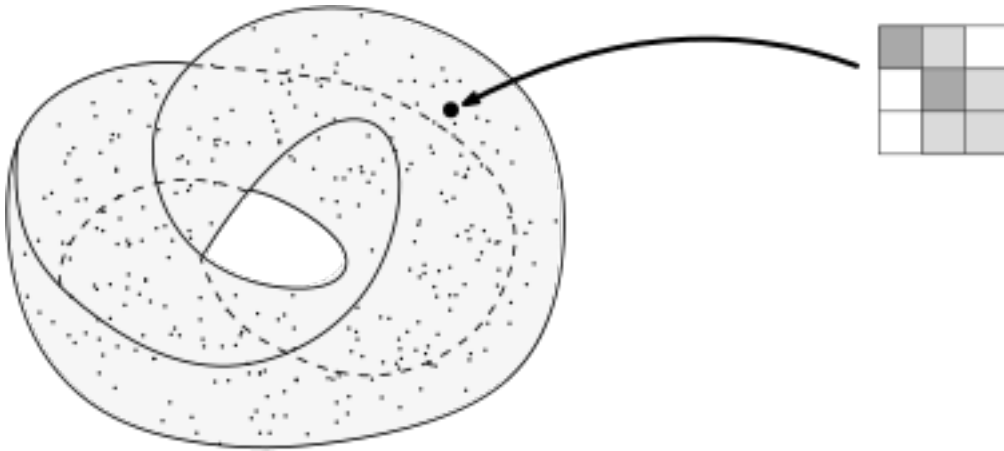
On déduit :  $H_0 = \mathbb{Z}/2\mathbb{Z}$ ,  $H_1 = \mathbb{Z}/2\mathbb{Z}$ ,  $H_2 = (\mathbb{Z}/2\mathbb{Z})^2$

# Images naturelles

14/21 (1/3)

Depuis [On the Local Behavior of Spaces of Natural Images, Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian, 2008.]

À partir d'une large collection d'images, les auteurs extraient des sous-images de taille  $3 \times 3$  pixels. L'ensemble de ces patches peut ainsi être vu comme un nuage de points dans  $\mathbb{R}^9$ .

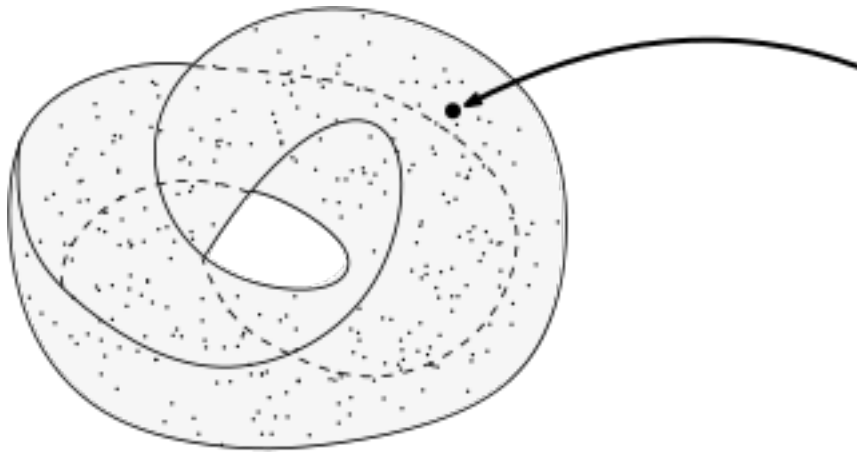


# Images naturelles

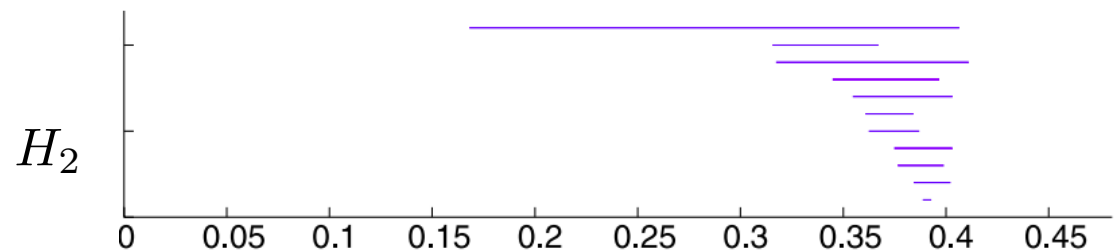
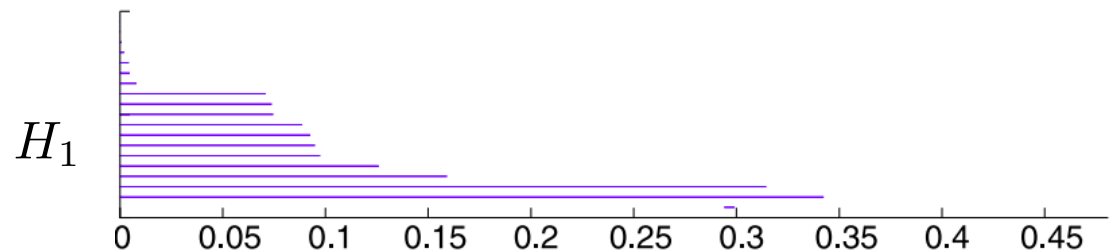
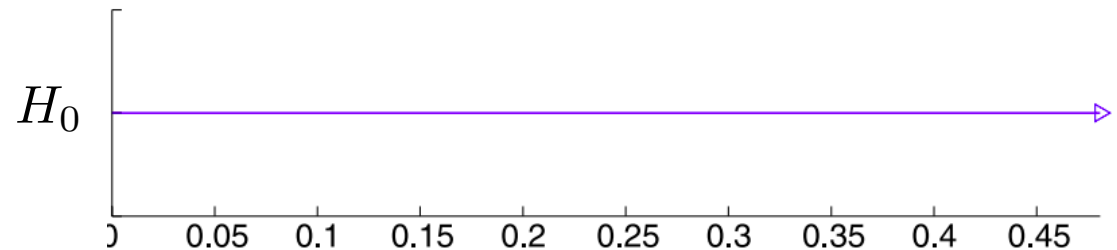
14/21 (2/3)

Depuis [On the Local Behavior of Spaces of Natural Images, Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian, 2008.]

À partir d'une large collection d'images, les auteurs extraient des sous-images de taille  $3 \times 3$  pixels. L'ensemble de ces patches peut ainsi être vu comme un nuage de points dans  $\mathbb{R}^9$ .



On a les codes-barres :

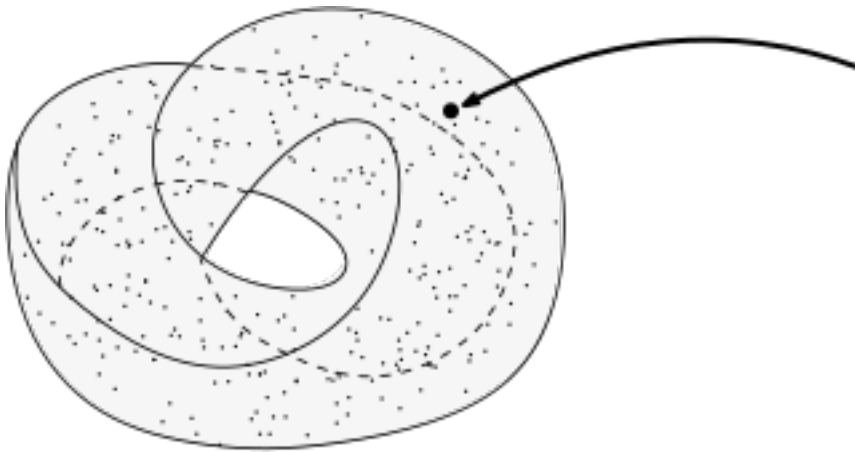


# Images naturelles

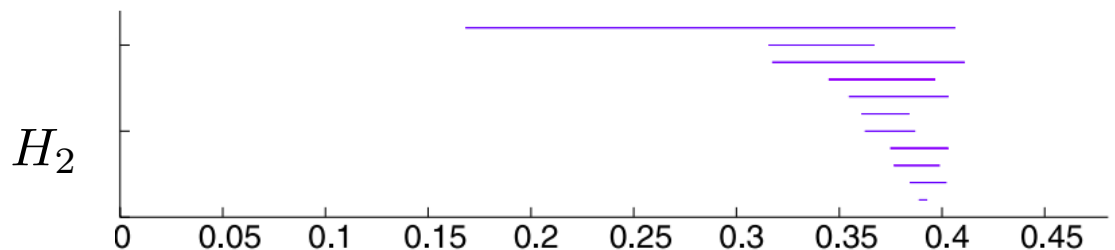
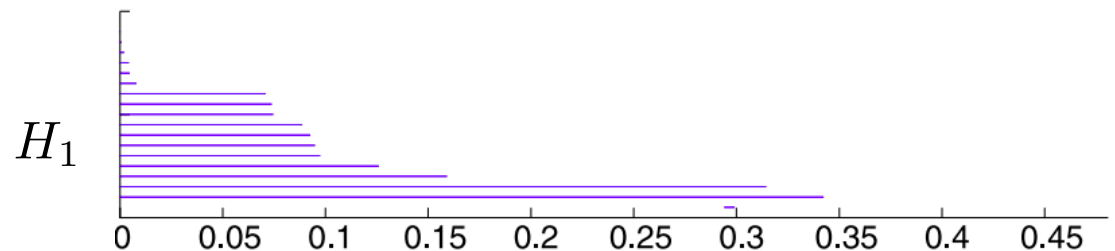
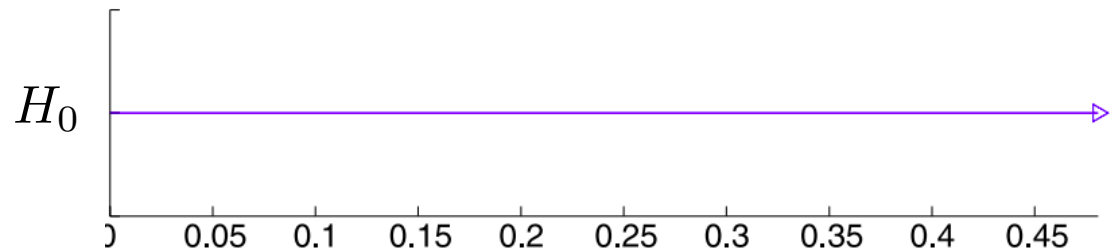
14/21 (3/3)

Depuis [On the Local Behavior of Spaces of Natural Images, Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian, 2008.]

À partir d'une large collection d'images, les auteurs extraient des sous-images de taille  $3 \times 3$  pixels. L'ensemble de ces patches peut ainsi être vu comme un nuage de points dans  $\mathbb{R}^9$ .



On a les codes-barres :



On déduit :

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

$$H_1 = (\mathbb{Z}/2\mathbb{Z})^2,$$

$$H_2 = \mathbb{Z}/2\mathbb{Z}.$$

I - Homologie singulière

II - Homologie persistante

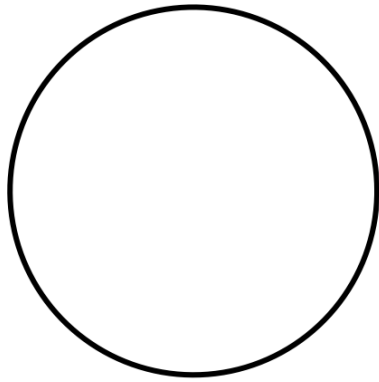
III - Exemples

IV - Homologie persistante pour mesures

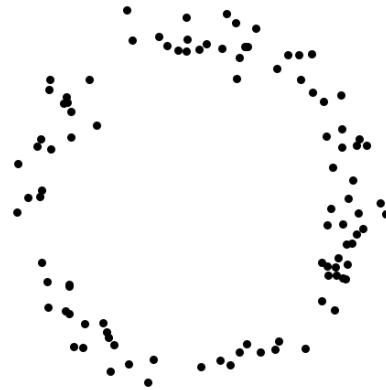
# Le problème des points aberrants

16/21 (1/2)

$$\mathbb{S}_1 \subset \mathbb{R}^2$$



Echantillon de  $\mathbb{S}_1$  avec  
erreurs de mesure

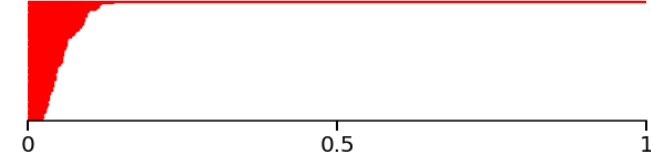
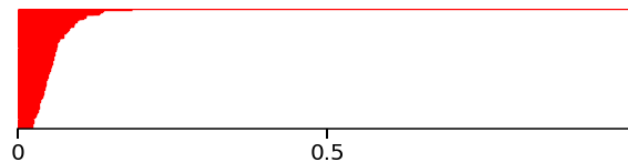
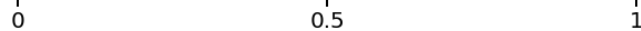


Echantillon de  $\mathbb{S}_1$  avec  
points aberrants

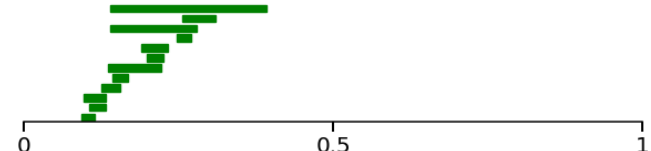
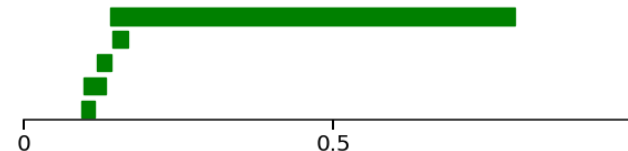
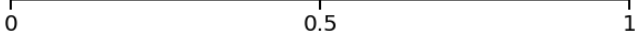


Homologie  
persistante de la  
filtration de Čech

$H_0$



$H_1$



Théorème de stabilité

# Le problème des points aberrants

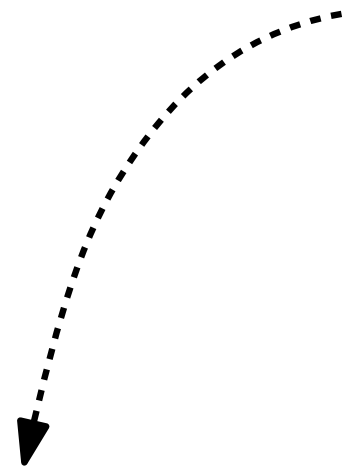
16/21 (2/2)

**Objectif:** construire une filtration robuste aux points aberrants.

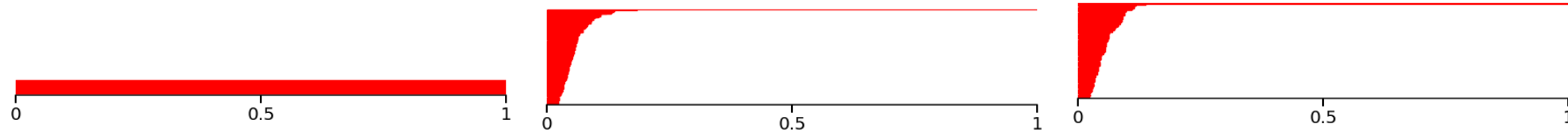
Deux ingrédients:

- Filtration de Čech pondérée
- Distance-à-la-mesure

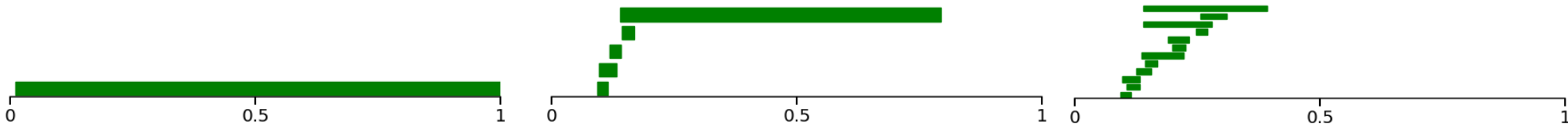
Echantillon de  $\mathbb{S}_1$  avec points aberrants



$H_0$



$H_1$



# Filtration de Čech pondérée

17/21 (1/2)

Données :  $X \subset \mathbb{R}^n$

**Rappel :** La filtration de Čech de  $X$  est la collection  $V[X] = (X^t)_{t \geq 0}$ , où

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

Soit  $f: X \rightarrow \mathbb{R}^+$  une application quelconque.

## Définition

La *filtration de Čech pondérée* de  $X$  avec paramètre  $f$  est la collection  $V[X, f] = (V^t[X, f])_{t \geq 0}$ , où

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

# Filtration de Čech pondérée

17/21 (2/2)

Données :  $X \subset \mathbb{R}^n$

**Rappel :** La filtration de Čech de  $X$  est la collection  $V[X] = (X^t)_{t \geq 0}$ , où

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

Soit  $f: X \rightarrow \mathbb{R}^+$  une application quelconque.

## Définition

La *filtration de Čech pondérée* de  $X$  avec paramètre  $f$  est la collection  $V[X, f] = (V^t[X, f])_{t \geq 0}$ , où

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

Plus  $f(x)$  est grand, plus on retarde l'apparition de  $x$  dans la filtration.  
On voudrait choisir  $f$  qui prenne de grandes valeurs sur les points aberrants.

# Distance-à-la-mesure (DTM)

18/21 (1/3)

Introduite dans [Chazal, Cohen-Steiner, Mérigot. Geometric inference for probability measures, 2011].

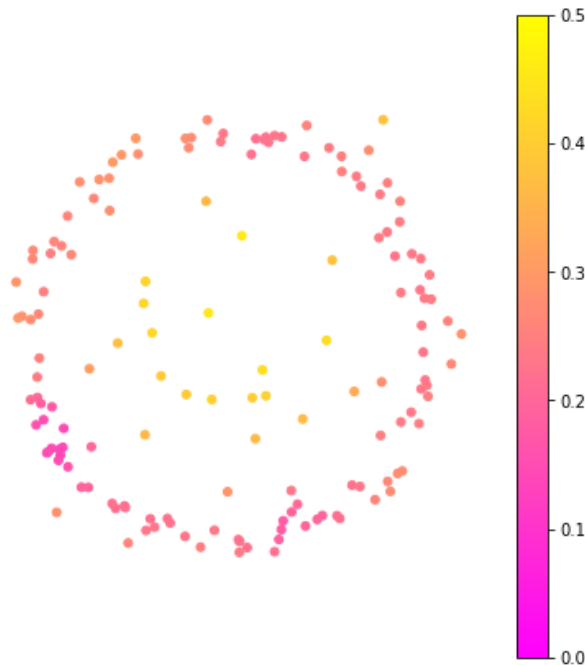
Soit  $\mu$  une mesure de probabilité. Pour  $x \in \mathbb{R}^n$  et  $t \in [0, 1)$ , on définit

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

## Définition

Soit  $m \in [0, 1[$ . La *DTM* de  $\mu$  avec paramètre  $m$  est l'application :

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



$d_{\mu,m}$  avec  $m = 0,1$   
 $\mu$  mesure empirique sur  $X$

# Distance-à-la-mesure (DTM)

18/21 (2/3)

Introduite dans [Chazal, Cohen-Steiner, Mérigot. Geometric inference for probability measures, 2011].

Soit  $\mu$  une mesure de probabilité. Pour  $x \in \mathbb{R}^n$  et  $t \in [0, 1)$ , on définit

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

## Définition

Soit  $m \in [0, 1[$ . La *DTM* de  $\mu$  avec paramètre  $m$  est l'application :

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

## Théorème (Chazal, Cohen-Steiner, Mérigot, 2011)

Pour toute mesures de probabilité  $\mu, \nu$  and  $m \in (0, 1)$ , on a

$$\|d_{\mu,m} - d_{\nu,m}\|_{\infty} \leq m^{-\frac{1}{2}} W_2(\mu, \nu),$$

où  $W_2$  représente la distance de Wasserstein.

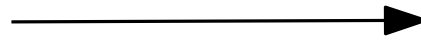
# Distance-à-la-mesure (DTM)

18/21 (3/3)

On va maintenant comparer les nuages de points via leur distance de Wasserstein.

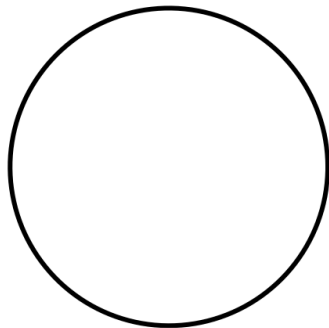


$X$  fini

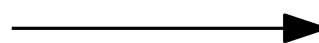


$$\mu = \frac{1}{|X|} \sum_{x \in X} \delta_x$$

mesure empirique



$\mathcal{M}$  sous-variété



$$\nu = \frac{1}{\mathcal{H}^d(\mathcal{M})} \mathcal{H}^d|_{\mathcal{M}}$$

mesure de Hausdorff

$X$  et  $\mathcal{M}$  ne sont pas proches en distance de Hausdorff...

Mais  $\mu$  et  $\nu$  sont proches en distance de Wasserstein !

## Définition

Soit  $\mu$  une mesure de probabilité et  $m \in [0, 1)$ .

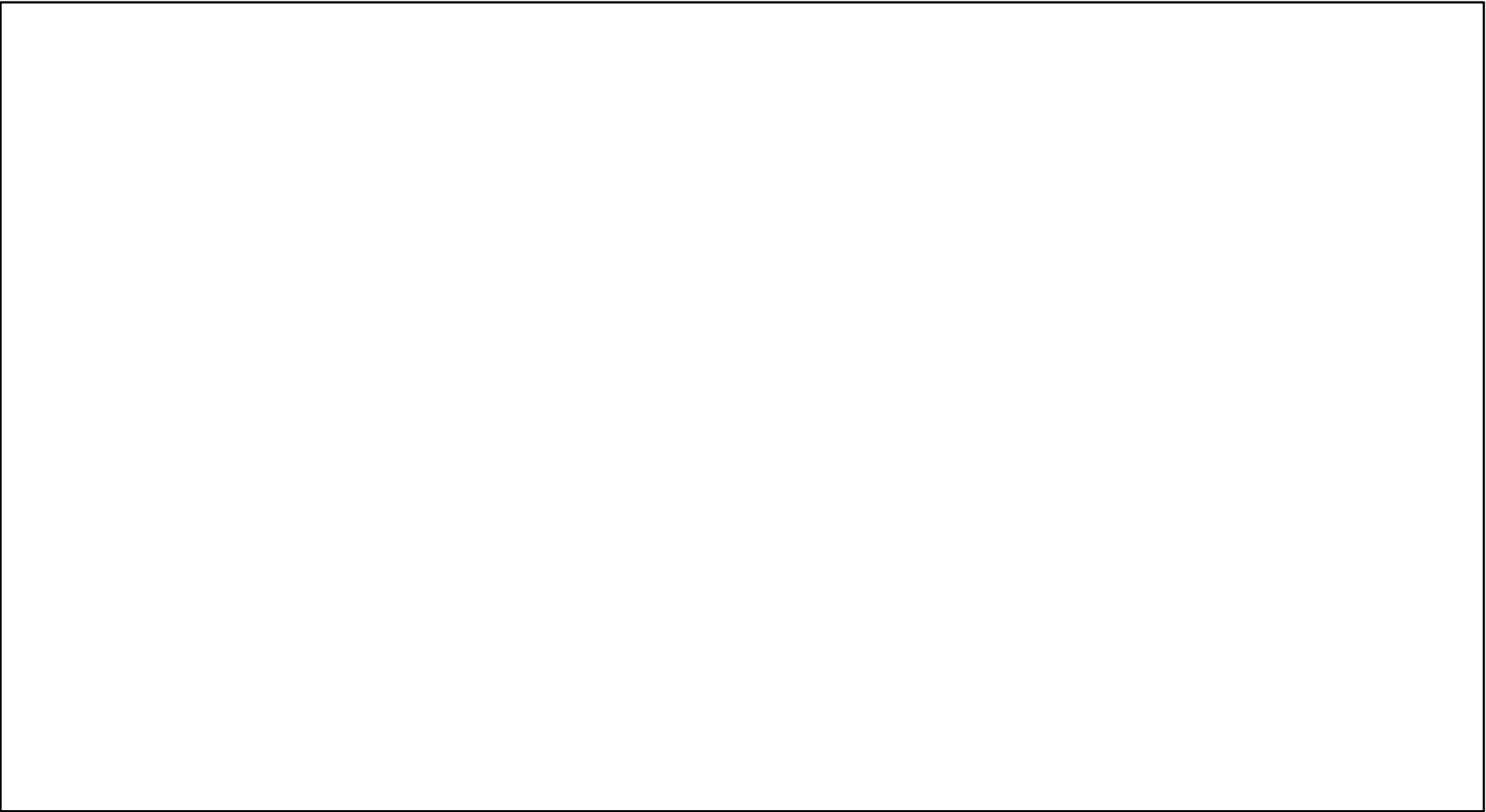
La *filtration-DTM* de paramètres  $\mu$  et  $m$  est la filtration de Čech pondérée  $V[X, f]$  avec les paramètres :

- $X = \text{supp}(\mu)$
- $f = d_{\mu, m}$

Elle est notée  $W[\mu, m]$ .

Explicitement,  $W[\mu, m] = (W^t[\mu, m])_{t \geq 0}$  avec :

$$W^t[\mu, m] = \bigcup_{x \in \text{supp}(\mu)} \overline{\mathcal{B}}(x, t - d_{\mu, m}(x))$$



Définissons  $c(\mu, m) = \sup_{x \in \text{supp}(\mu)} d_{\mu, m}(x)$ .

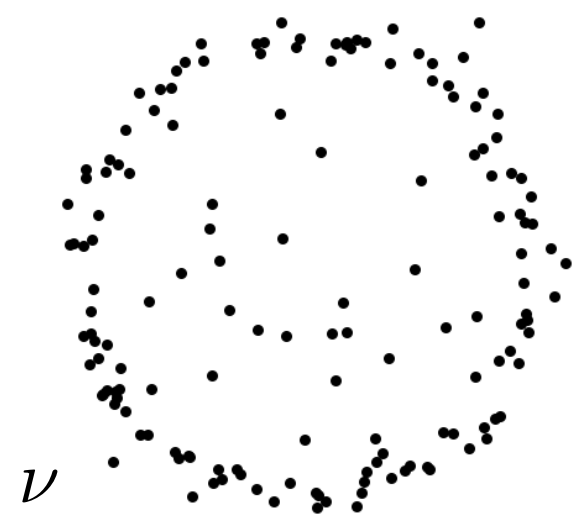
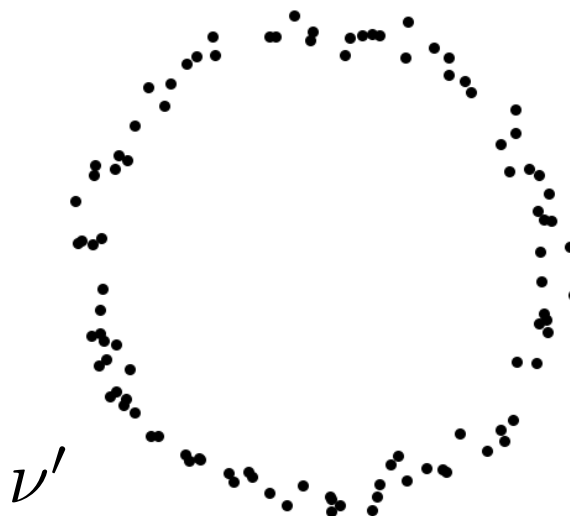
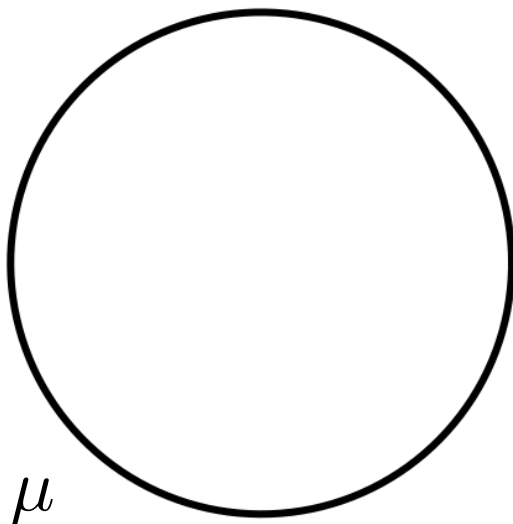
La quantité  $c$  est faible si la mesure  $\mu$  est proche de la mesure de Hausdorff sur un sous-variété.

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

Soient  $\mu, \nu$  des mesures de probabilité. Soit  $\nu'$  une mesure de probabilité à support compact inclus dans  $\text{supp}(\nu)$ .

La distance bottleneck entre les filtrations-DTM  $W[\mu, m]$  et  $W[\nu, m]$  est bornée par :

$$m^{-\frac{1}{2}} W_2(\mu, \nu') + m^{-\frac{1}{2}} W_2(\nu', \nu) + c(\mu, m, p) + c(\nu', m, p)$$



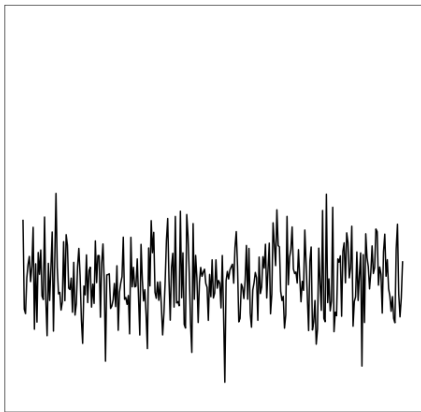
# Exemple d'application

20/21 (1/3)

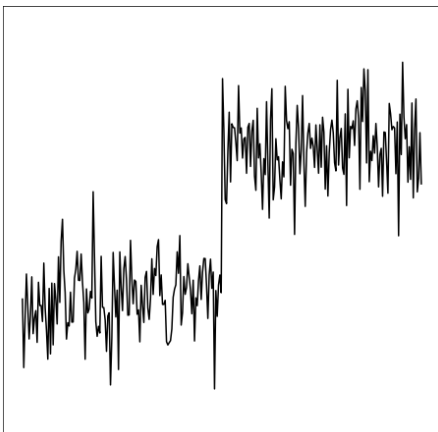
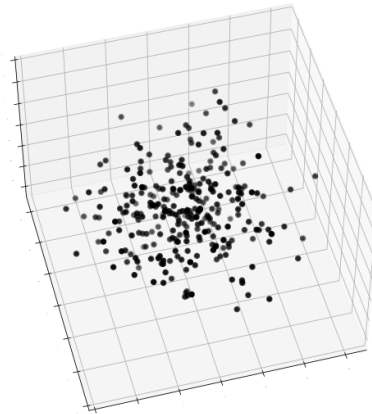
On s'intéresse à l'usure des bâtiments.

On obtient des séries temporelles (mesures de capteurs inertiels). On transforme une série temporelle  $(x_1, x_2, \dots, x_n)$  en le nuage de points

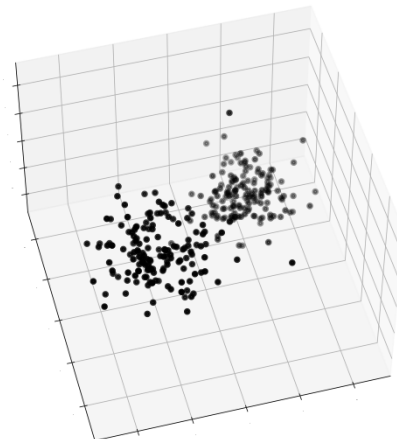
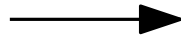
$$\{(x_1, x_2, x_3), (x_2, x_3, x_4), \dots, (x_{n-2}, x_{n-1}, x_n)\}.$$



bâtiment en bon état



bâtiment abimé



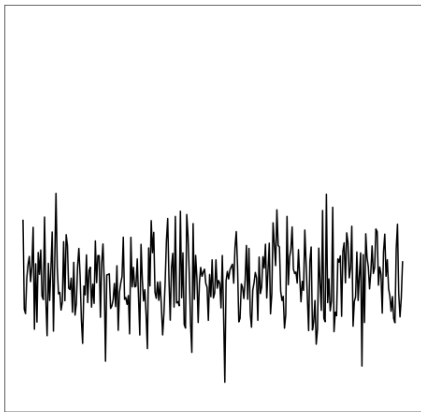
# Exemple d'application

20/21 (2/3)

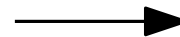
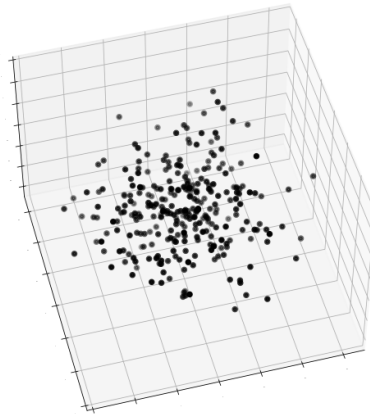
On s'intéresse à l'usure des bâtiments.

On obtient des séries temporelles (mesures de capteurs inertiels). On transforme une série temporelle  $(x_1, x_2, \dots, x_n)$  en le nuage de points

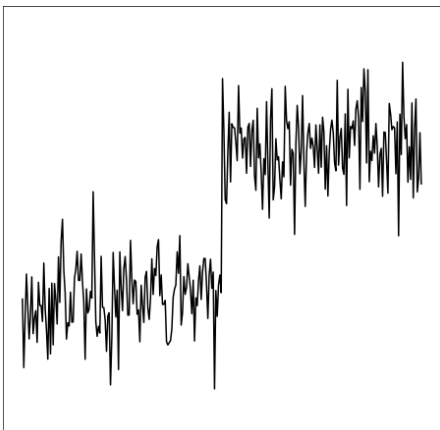
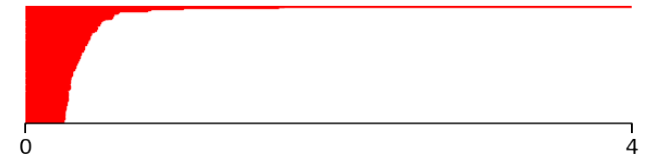
$$\{(x_1, x_2, x_3), (x_2, x_3, x_4), \dots, (x_{n-2}, x_{n-1}, x_n)\}.$$



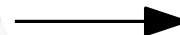
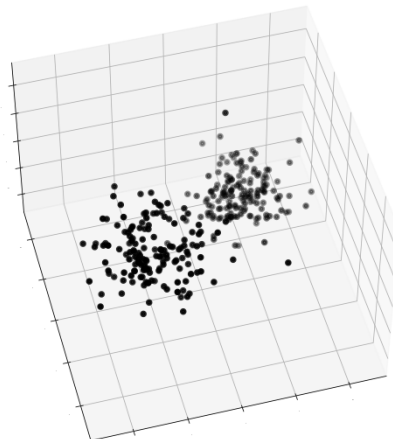
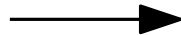
bâtiment en bon état



Codes-barres en  $H_0$  de la filtration de Čech



bâtiment abimé



codes-barres semblables...

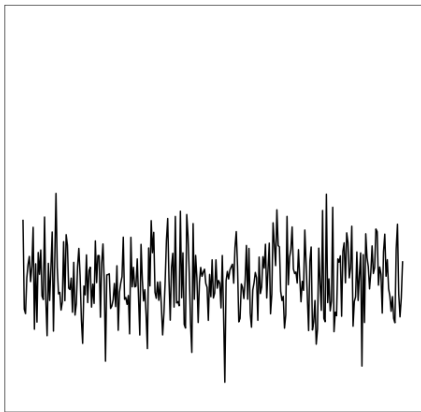
# Exemple d'application

20/21 (3/3)

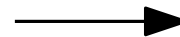
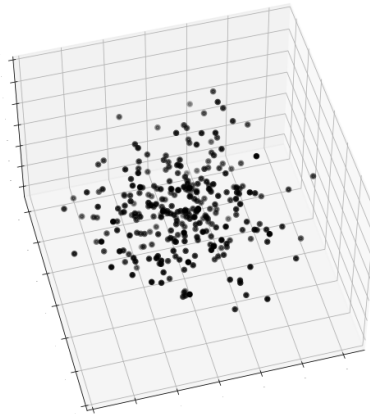
On s'intéresse à l'usure des bâtiments.

On obtient des séries temporelles (mesures de capteurs inertiels). On transforme une série temporelle  $(x_1, x_2, \dots, x_n)$  en le nuage de points

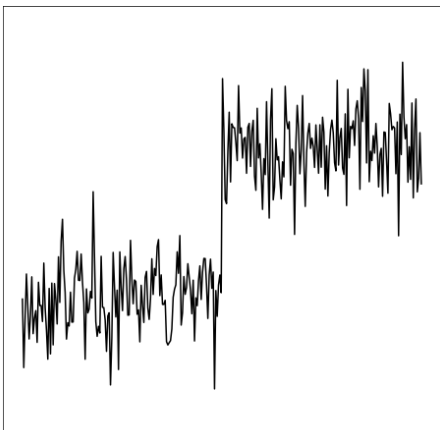
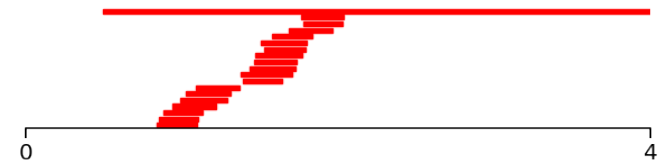
$$\{(x_1, x_2, x_3), (x_2, x_3, x_4), \dots, (x_{n-2}, x_{n-1}, x_n)\}.$$



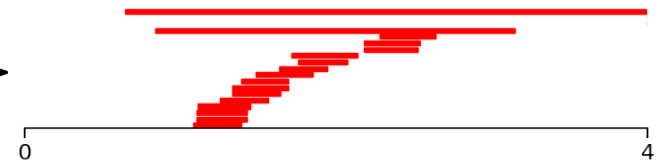
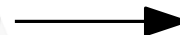
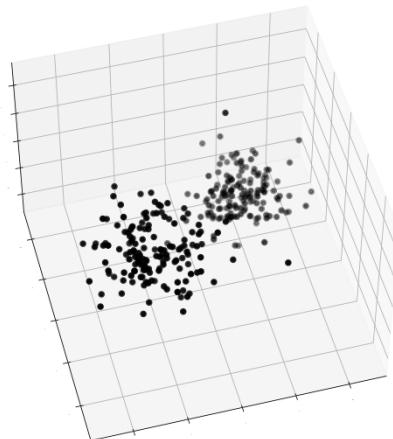
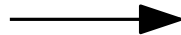
bâtiment en bon état



Codes-barres en  $H_0$  de la filtration-DTM



bâtiment abimé



codes-barres différenciables !

# Conclusion

- L'homologie persistante permet d'estimer l'homologie des objets géométriques sous-jacents aux jeux de données.
- Repose sur l'hypothèse qu'un tel objet géométrique sous-jacent existe.
- On peut réparer la théorie pour le cas où le jeu de données contient des points aberrants.

# Conclusion

- L'homologie persistante permet d'estimer l'homologie des objets géométriques sous-jacents aux jeux de données.
- Repose sur l'hypothèse qu'un tel objet géométrique sous-jacent existe.
- On peut réparer la théorie pour le cas où le jeu de données contient des points aberrants.

Merci !