

---

# Introduction à l'homologie persistante

---

Raphaël TINARRAGE

Datashape, Inria Saclay

30 mai 2018

## Table des matières

|   |                      |   |
|---|----------------------|---|
| 1 | Exemple introductif  | 1 |
| 2 | Algèbre persistante  | 2 |
| 3 | En présence de bruit | 7 |
|   | Références           | 9 |

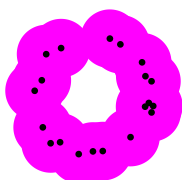
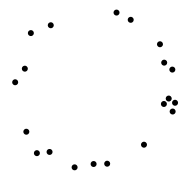
## 1 Exemple introductif

Dans le domaine de l'analyse de données, nous avons souvent affaire à des ensembles de données discrètes, que l'on peut interpréter comme des nuages de points dans des espaces euclidiens de grande dimension. Pour les traiter, l'analyse topologique des données repose sur le principe qu'il existe des informations de nature topologique et géométrique à extraire de ces nuages. Par les contraintes mises en jeu, nous pouvons présumer que ces données se concentrent sur des sous-espaces de petite dimension. Ainsi, si le nuage est suffisamment dense, il est envisageable de les estimer.

Parmi les outils mathématiques à disposition pour l'étude des espaces topologiques, l'homologie se révèle ici être d'une pertinence particulière : elle donne des informations sur ces espaces, et prend des formes implémentables du point de vue algorithmique.


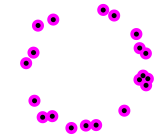
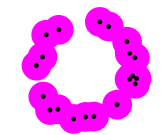
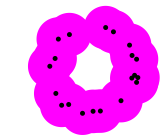
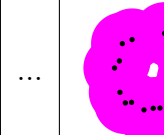
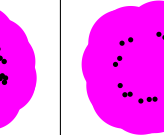
**Le problème de l'échelle.** Soit  $X$  un nuage de point, qui peut être un sous-ensemble d'un espace euclidien, un sous-ensemble d'une variété, ou un espace métrique discret. En tant qu'espace topologique,  $X$  n'apporte pas plus d'information que le nombre de points qui le compose. Pour étudier l'espace continu sur lequel  $X$  repose, il faut interpoler entre ces points.

Considérons par exemple que  $X$  soit un sous-ensemble de  $\mathbb{R}^2$  tiré proche du cercle  $\mathbb{S}^1$ . On aimerait reconnaître le cercle à partir de  $X$ . Posons alors sur chacun des points de  $X$  une boule de rayon  $t$ . L'espace obtenu est  $X^t := \bigcup_{x \in X} B(x, t)$ , le voisinage tubulaire de  $X$ . Si le rayon est bien choisi, on voit que  $\mathbb{S}^1$  est un rétract par déformation de  $X^t$ . Autrement dit,  $X^t$  est du même type d'homotopie que le cercle sous-jacent.



Pour chaque valeur de  $t$ , nous pouvons considérer l'homologie singulière de  $X^t$ . Les groupes d'homologie singulière à coefficients dans un corps  $k$ , notés  $H_*(X^t)$ , donnent des informations sur la topologie de cet ensemble. En pratique, le problème se pose lorsque nous ne savons pas choisir a priori le paramètre  $t$ . Dans notre exemple, s'il est trop petit, le voisinage tubulaire est composé de plusieurs parties connexes ; s'il est trop grand, il est

homéomorphe à un disque. L'idée de la persistance est de calculer l'homologie de ces voisinages tubulaires  $X^t$  pour toutes les valeurs de  $t \in \mathbb{R}^+$ , et de suivre au fur et à mesure leur évolution :

| $t$        | 0   | 0,1   | 0,3   | 0,5  | ... | 0,8   | 1   |
|------------|---|---|---|--|-----|---|---|
| $X^t$      |  |  |  |  | ... |  |  |
| $H_0(X^t)$ | $k^{20}$  | $k^{14}$  | $k^2$   | $k$  | ... | $k$   | $k$   |
| $H_1(X^t)$ | 0   | 0   | 0   | $k$  | ... | $k$   | 0   |

Nous avons ici affaire à une suite croissante de sous-espaces topologiques  $\{X^t\}_{t \in \mathbb{R}^+}$ , avec des inclusions :

$$\dots \subset X^r \subset X^s \subset X^t \subset \dots$$

En appliquant l'homologie singulière, on obtient par naturalité une suite d'espaces vectoriels :

$$\dots \rightarrow H_*(X^r) \rightarrow H_*(X^s) \rightarrow H_*(X^t) \rightarrow \dots,$$

où la commutativité du diagramme au niveau des inclusions se traduit en une commutativité au niveau homologique :

$$\begin{array}{ccc} X^r & \hookrightarrow & X^s & \hookrightarrow & X^t \\ & \searrow & & \searrow & \\ & & & & \end{array} \quad \text{donne} \quad \begin{array}{ccc} H_*(X^r) & \longrightarrow & H_*(X^s) & \longrightarrow & H_*(X^t) \\ & \searrow & & \searrow & \\ & & & & \end{array}$$

La structure algébrique que nous venons de construire s'appelle un module de persistance. Dans cette structure, nous sommes capables de suivre l'évolution des cycles de  $X^t$ , leur naissance et leur mort.

La quantité suivante apparaît déjà en 1992 dans les travaux de Frosini [1] :

$$r_s^t := \text{rang}(H_n(X^s) \rightarrow H_n(X^t)).$$

Elle s'interprète comme le nombre de  $n$ -cycles de  $X^s$  qui ont survécu jusque dans  $X^t$ .

## 2 Algèbre persistante

**Module de persistance.** Les considérations précédentes se formulent en toute généralité dans la structure algébrique de module de persistance. La première définition apparaît dans l'article [2] de Zomorodian et Carlsson, publié en 2005. On considère pour le moment que les indices sont dans  $T = \mathbb{N}$  ou  $\mathbb{R}$ .

**Définition 1.** Un module de persistance (indexé par  $T$ ) est la donnée d'une famille de  $k$ -espaces vectoriels  $\mathbb{V} = (V_t)_{t \in T}$  et d'applications linéaires  $(v_s^t : V_s \rightarrow V_t)_{s \leq t \in T}$  tels que

- $\forall t \in T, v_t^t$  est la fonction identité sur  $V_t$ , et
- $\forall s \leq t \leq u \in T, v_t^u \circ v_s^t = v_s^u$ .

La dernière condition se traduit par le diagramme commutatif suivant, pour tout  $\forall s, t, u \in T$  tels que  $s \leq t \leq u$  :

$$\begin{array}{ccc} V_r & \xrightarrow{v_r^t} & V_t \\ & \searrow v_r^s & \nearrow v_s^t \\ & & V_s \end{array}$$

Si  $T = \mathbb{N}$ , les applications  $v_s^t$  sont entièrement déterminées par les  $v_n^{n+1}, n \in T$ .

Les modules de persistance forment une catégorie abélienne. On y définit les sommes de modules de persistance, et les morphismes.

**Définition 2.** La somme de deux modules de persistance est le module de persistance obtenu en sommant terme à terme les espaces vectoriels (somme directe) et en concaténant les applications linéaires.

**Définition 3.** Pour  $\epsilon \geq 0$ , un  $\epsilon$ -morphisme entre des modules de persistance  $\mathbb{U}$  et  $\mathbb{V}$  est une suite d'applications linéaires  $\Phi = (\phi_t : U_t \rightarrow V_{t+\epsilon})_{t \in T}$  telle que le diagramme suivant commute en tout  $s, t \in T$  tels que  $s \leq t$  :

$$\begin{array}{ccc} U_s & \xrightarrow{u_s^t} & U_t \\ \downarrow \phi_s & & \downarrow \phi_t \\ V_{s+\epsilon} & \xrightarrow{v_{s+\epsilon}^{t+\epsilon}} & V_{t+\epsilon} \end{array}$$

Quand  $\epsilon = 0$  et  $\Phi$  est une famille d'isomorphismes, on dit que les modules sont isomorphes. En pratique, cette notion est trop rigide, et on a recours à la définition suivante :

**Définition 4.** Soient  $\mathbb{U}$  et  $\mathbb{V}$  deux modules de persistance, et  $1_{2\epsilon}^{\mathbb{U}}, 1_{2\epsilon}^{\mathbb{V}}$  les  $2\epsilon$ -endomorphismes de modules de persistance donnés par la structure de module de persistance :  $(1_{2\epsilon}^{\mathbb{U}})_t = u_t^{t+\epsilon}$  et  $(1_{2\epsilon}^{\mathbb{V}})_t = v_t^{t+\epsilon}$ .

On dit que  $\mathbb{U}$  et  $\mathbb{V}$  sont  $\epsilon$ -entrelacés s'il existe deux morphismes de degré  $\epsilon$ ,  $\Phi : \mathbb{U} \rightarrow \mathbb{V}$  et  $\Psi : \mathbb{V} \rightarrow \mathbb{U}$ , tels que  $\Psi \circ \Phi = 1_{2\epsilon}^{\mathbb{U}}$  et  $\Phi \circ \Psi = 1_{2\epsilon}^{\mathbb{V}}$ .

Si les modules de persistance en question sont issus de l'homologie d'espaces topologiques, on interprète la définition précédente ainsi : les temps de naissance et de mort des cycles non-triviaux de deux modules  $\epsilon$ -entrelacés ne peuvent différer que de  $\epsilon$ .

Enfin, étant donnés deux modules de persistance  $\mathbb{U}$  et  $\mathbb{V}$ , on définit la distance d'entrelacement :

$$d_i(\mathbb{U}, \mathbb{V}) = \inf\{\epsilon \geq 0 \text{ tel que } \mathbb{U} \text{ et } \mathbb{V} \text{ soient } \epsilon\text{-entrelacés}\},$$

avec pour convention  $\inf \emptyset = +\infty$ .

La distance d'entrelacement est une pseudo-distance entre les modules de persistance. Elle nous permet de les comparer.

**Exemples de modules de persistance.** La première construction est celle des voisinages tubulaires d'un sous-ensemble de points  $X \subset \mathbb{R}^n$ . Les groupes d'homologie singulière  $V_t = H_*(X^t)$ , munis des applications linéaires induites par les inclusions, forment un module de persistance.

Plus généralement, si  $E$  un espace topologique et  $f : E \rightarrow \mathbb{R}$  une fonction continue, les sous-niveaux  $f^t = f^{-1}(] - \infty, t])$  de  $f$  forment une famille croissante de sous-espaces topologiques indexée par  $\mathbb{R}$ . Pour tout  $r, s, t \in \mathbb{R}$  tels que  $r \leq s \leq t$ ,

$$f^r \subset f^s \subset f^t.$$

Les groupes d'homologie singulière  $V_t = H_*(f^t)$  donnent lieu à un module de persistance indexé par  $\mathbb{R}$ . On retrouve la première construction en prenant pour  $f$  la fonction distance au compact  $X$ .

Un autre exemple est donné par les filtrations de complexes simpliciaux. Si  $K$  est un complexe simplicial abstrait (éventuellement infini), on appelle filtration de  $K$  une suite croissante de sous-complexes

$$\emptyset = K^0 \subset K^1 \subset \dots \subset K.$$

L'homologie simpliciale de ces complexes donne aussi lieu à une suite d'espaces vectoriels  $V_i = H_*(K^i)$ . Les inclusions entre les sous-complexes simpliciaux sont des applications simpliciales qui induisent des morphismes  $v_i^j : V_i \rightarrow V_j$ . On obtient une structure de module de persistance indexé par  $\mathbb{N}$ .

**Décomposition des modules de persistance.** Les modules-intervalles sont des modules de persistance particulièrement importants. Soit toujours  $k$  un corps. Si  $I \subset T$  est un intervalle, le module intervalle  $\mathbb{I}(I)$  est défini par :

- les espaces vectoriels, pour tout  $t \in T$ ,  $\begin{cases} V_t = k & \text{si } t \in I \\ = 0 & \text{sinon.} \end{cases}$
- les morphismes, pour  $s \leq t \in T$ ,  $\begin{cases} v_s^t = \text{id} & \text{si } s, t \in I \\ = 0 & \text{sinon.} \end{cases}$

On dit qu'un module de persistance  $\mathbb{V}$  est décomposable en modules-intervalle s'il est isomorphe à une somme de modules intervalles :

$$\mathbb{V} = \bigoplus_{I \in L} \mathbb{I}(I).$$

Dans [2], les auteurs étudient la décomposabilité des modules de persistance indexés par  $T = \mathbb{N}$ . Dans ce cadre, ils établissent une correspondance avec les modules gradués sur l'anneau des polynômes  $k[X]$ . Ils obtiennent :

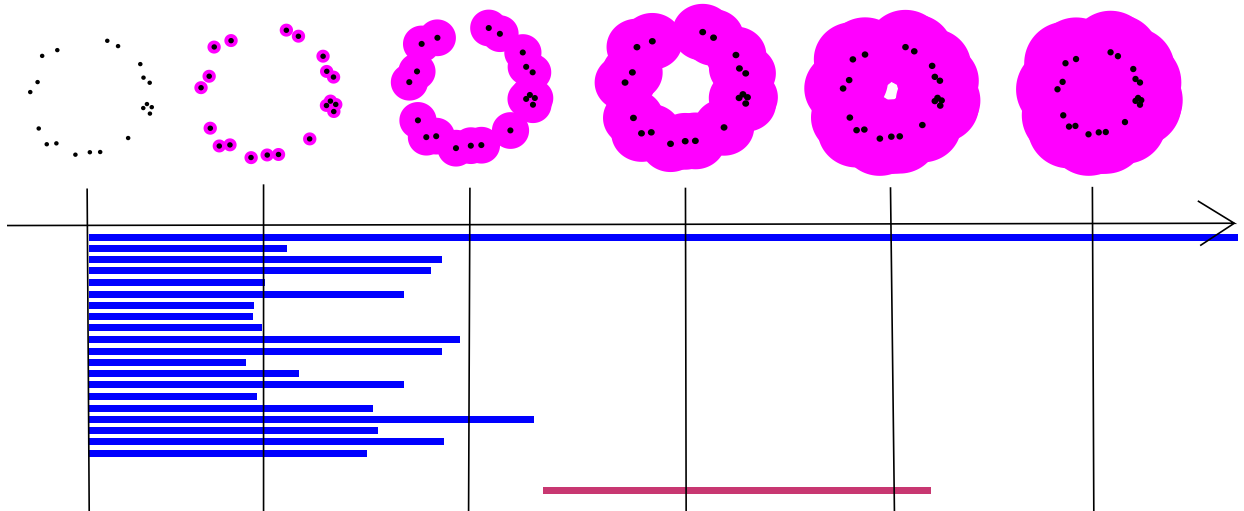
**Théorème 5.** Soit  $\mathbb{V}$  un module de persistance tel que chacun de ses espaces vectoriels  $V_i$  soit de dimension finie, et que les applications  $v_i^j$  soient des isomorphismes pour  $i$  assez grand. Alors  $\mathbb{V}$  est décomposable en modules-intervalle.

Lorsque  $T = \mathbb{R}$ , la tâche se complique. On peut obtenir la décomposabilité des modules lorsqu'ils n'ont qu'un nombre fini de valeurs critiques, c'est à dire de valeurs  $t \in \mathbb{R}$  telles que pour tout  $\epsilon > 0$  assez petit, l'application  $v_{t-\epsilon}^{t+\epsilon}$  ne soit pas un isomorphisme.

En 2014, dans [3], Crawley-Boevey montre une condition de finitude suffisante :

**Théorème 6.** Soit  $\mathbb{V}$  un module de persistance sur  $\mathbb{R}$  tel que chaque espace  $V_t$  soit de dimension finie. Alors  $\mathbb{V}$  se décompose en somme (éventuellement infinie) de modules-intervalles.

La figure suivante reprend l'exemple introductif :



La décomposition du module de persistance  $(H_0(X^t))_{t \geq 0}$  est représentée en bleu, et celle du module de persistance  $(H_1(X^t))_{t \geq 0}$  en rouge. Chaque barre correspond à un module-intervalle dans la décomposition.

**Diagramme de persistance.** Edelsbrunner, Letscher et Zomorodian [4] introduisent en 2002 une façon commode de représenter la décomposition d'un module de persistance. Considérons un module décomposé :

$$\mathbb{V} = \bigoplus_{I \in L} \mathbb{I}(a_i^*, b_i^*),$$

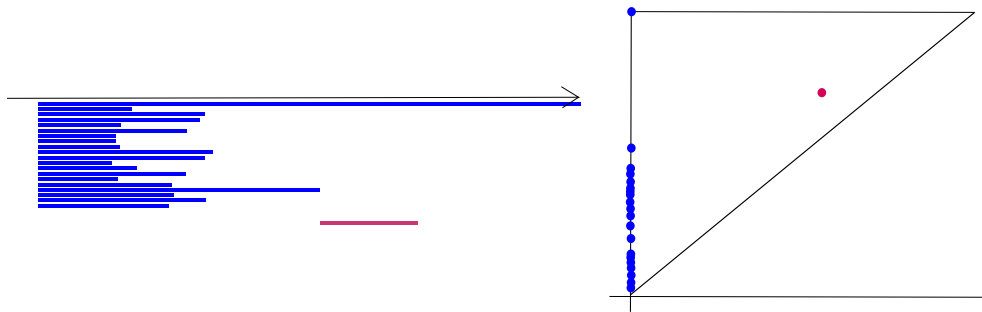
où l'étoile en exposant signifie que l'intervalle est ouvert ou fermé de ce côté de l'intervalle.

**Définition 7.** Le diagramme de persistance de  $\mathbb{V}$  est le multi-ensemble

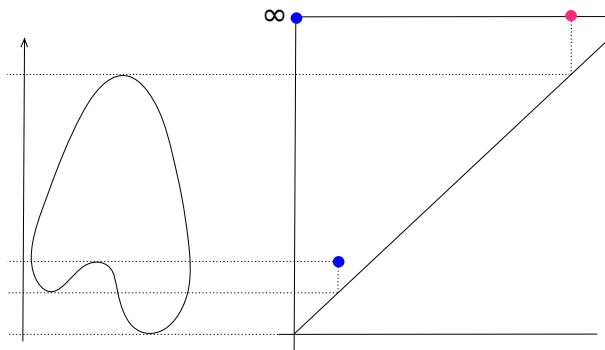
$$\text{dgm}(\mathbb{V}) := \{(a_l, b_l), l \in L\} \setminus \Delta \subset \bar{\mathbb{R}}^2,$$

où les points sont comptés avec multiplicité. On a retiré du diagramme les points  $(a_l, b_l)$  de la diagonale  $\Delta = \{(x, x), x \in \mathbb{R}\}$ .

La figure suivante représente le diagramme de persistance du module de persistance considéré dans l'exemple introductif. La point bleu de coordonnées  $(0, +\infty)$  signifie la présence d'un module-intervalle  $\mathbb{I}([0, +\infty[)$  dans la décomposition en modules-intervalle de ce module de persistance.



L'exemple suivant est le diagramme de persistance d'une courbe de  $\mathbb{R}^2$  filtrée par la projection sur l'axe vertical. Les sous-niveaux de cette fonction sont vides, ou sont des courbes à une ou deux composantes connexes. On pose  $H_*(\emptyset) = 0$ . On voit apparaître sur le diagramme de persistance une composante connexe qui persiste jusqu'à l'infini (en bleu, de coordonnées  $(0, +\infty)$ ), et une composante connexe qui persiste un temps fini (en bleu, près de la diagonale). Un cycle non-trivial n'apparaît dans  $H_1(X^t)$  que lorsque  $f(\cdot - \infty, t]$  contient la courbe en entière. Ce cycle persiste ensuite jusqu'à l'infini.



Étant donné deux diagrammes de persistance, on dispose d'une autre distance : la distance bottleneck. Deux diagrammes sont proches en distance bottleneck si on peut appairer leurs points sans trop les déplacer. On formalise cette idée ainsi :

- Un appariement partiel entre deux multi-ensembles  $A, B \subset \mathbb{R}$  est un sous-multi-ensemble  $M \subset A \times B$  tel que :

- pour tout  $a \in A$ , il y a au plus un  $b \in B$  tel que  $(a, b) \in M$ ,
- pour tout  $b \in B$ , il y a au plus un  $a \in A$  tel que  $(a, b) \in M$ .

De tels points  $a \in A$  et  $b \in B$  tels que  $(a, b) \in M$  sont dit appariés. On dit que les autres points sont non-appariés, ou de manière équivalente appariés avec la diagonale  $\Delta$ .

- Pour  $\delta \geq 0$ , on dit que l'appariement  $M$  est un  $\delta$ -appariement si :

- si  $(a, b) \in M$ , alors  $d_\infty(a, b) \leq \delta$ ,
- si  $a \in A$  n'est pas apparié, alors  $d_\infty(a, \Delta) \leq \delta$ ,
- si  $b \in B$  n'est pas apparié, alors  $d_\infty(b, \Delta) \leq \delta$ .

( $d_\infty$  représente la distance euclidienne)

**Définition 8.** La distance bottleneck entre  $\mathbb{U}$  et  $\mathbb{V}$  est définie comme la borne inférieure des valeurs  $\delta \geq 0$  telles qu'il existe un  $\delta$ -appariement entre  $\text{dgm}(\mathbb{U})$  et  $\text{dgm}(\mathbb{V})$ . Elle est notée  $d_b(\mathbb{U}, \mathbb{V})$ .

On montre aussi que la distance bottleneck est une pseudo-distance entre les modules de persistance décomposables.

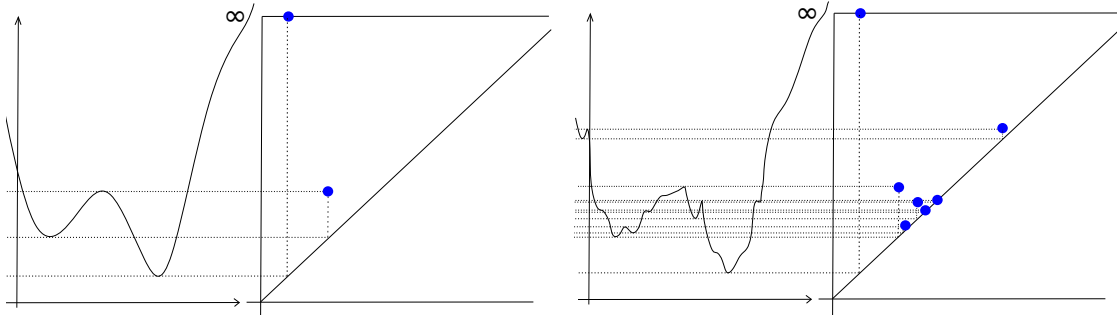
**Théorème de stabilité.** Les diagrammes de persistance sont intensément utilisés en analyse topologique des données. Leur stabilité est une notion capitale : si deux données initiales sont proches, qu'en est-il de leurs diagrammes ?

La première instance du théorème de stabilité a été montrée en 2007 par Cohen-Steiner, Edelsbrunner et Harer [5].

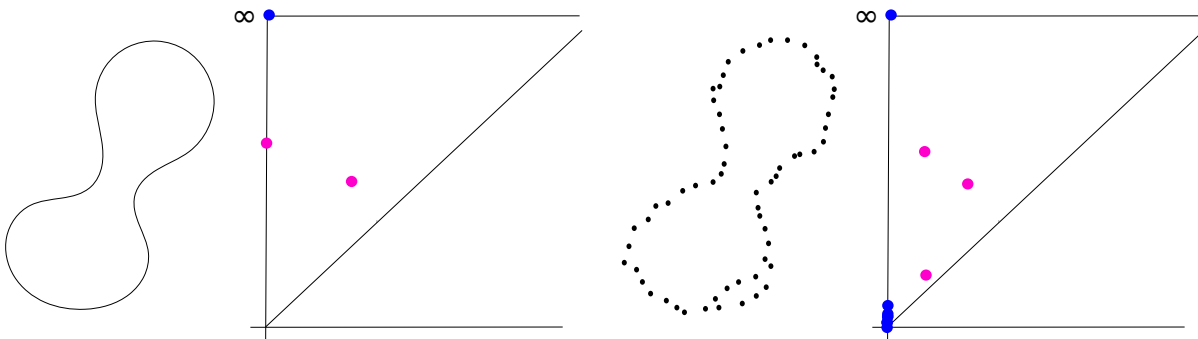
**Théorème 9.** Soit  $\mathbb{X}$  un espace topologique triangulable, et  $f, g : \mathbb{X} \rightarrow \mathbb{R}$  deux fonctions continues. Soient  $\mathbb{V}_f$  et  $\mathbb{V}_g$  les modules de persistance des sous-niveaux de ces fonctions. On suppose que ces modules de persistance ne possèdent qu'un nombre fini de valeurs critiques. Alors ces modules sont décomposables en modules-intervalle, et :

$$d_b(\mathbb{V}_f, \mathbb{V}_g) \leq \|f - g\|_\infty.$$

Pour illustrer cette propriété, nous représentons le diagramme de persistance des sous-niveaux d'une fonction de  $\mathbb{R}$  dans  $\mathbb{R}$ , ainsi que son diagramme lorsqu'elle est soumise à de petites perturbations en norme infinie. On voit apparaître des points proches de la diagonale.



Soient  $K, L \subset \mathbb{R}^d$  deux sous-ensembles finis. Si ces compacts sont proches en distance de Hausdorff, alors c'est aussi le cas pour les fonctions distances correspondantes en norme infinie. On obtient alors que les modules de persistance des voisinages tubulaires sont proches :



On observe l'apparition de points bleu et rose proches de la diagonale, et déplacement de deux points roses.

Remarquons que le résultat de stabilité précédent se démontre presque immédiatement si l'on remplace la distance bottleneck par la distance d'entrelacement :  $d_i(\mathbb{V}_f, \mathbb{V}_g) \leq \|f - g\|_\infty$ . En fait, le théorème de stabilité est un corollaire direct du théorème d'isométrie suivant, mis au propre en 2016 par Chazal, de Silva, Glisse et Oudot [6] :

**Théorème 10.** Soient  $\mathbb{U}$  et  $\mathbb{V}$  deux modules de persistance indexés par  $\mathbb{R}$  et décomposables en modules intervalles. Alors leurs distances bottleneck et d'entrelacement coïncident.

**Représentations de carquois.** Dans [7], publié en 2015, Oudot raconte le lien entre les modules de persistance et les représentations de carquois. Cela permet de généraliser les résultats précédents à d'autres notions de modules de persistance.

Pour commencer, si  $T$  est un ensemble partiellement ordonné, on peut considérer la catégorie associée, où les objets sont les éléments de  $T$  et où pour tout  $a, b \in T$ , une unique flèche  $a \rightarrow b$  est présente lorsque  $a \leq b$ . Un module de persistance sur  $T$  est alors défini comme un foncteur de la catégorie associée à  $T$  vers la catégorie des espaces vectoriels. Si  $T = \mathbb{N}$  ou  $\mathbb{R}$ , cette notion coïncide avec les modules de persistance précédemment définis. Si  $T = \mathbb{R}^2$  pour l'ordre partiel

$$(a, b) < (c, d) \iff a < c \text{ et } b < d,$$

alors l'on parle de multipersistance. Les modules de persistance indexés par  $\mathbb{R}^2$  ne jouissent pas de propriétés de décomposabilité comparables à ceux indexés par  $\mathbb{R}$ , et sont un sujet de recherche actuel.

Plus généralement, si  $Q$  est un carquois, on peut définir un module de persistance sur  $Q$  comme une représentation (linéaire) de  $Q$  (voir [7]). Un cas particulièrement étudié est celui du carquois  $A_n$  :

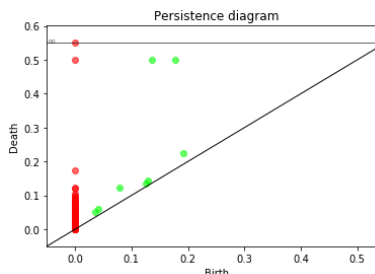
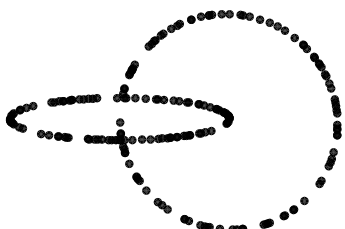
$$\bullet \longleftrightarrow \bullet \longleftrightarrow \dots \longleftrightarrow \bullet \longleftrightarrow \bullet,$$

où le sens des flèches est arbitraire. Ceci donne lieu à la notion de persistance zigzag (voir [8]).

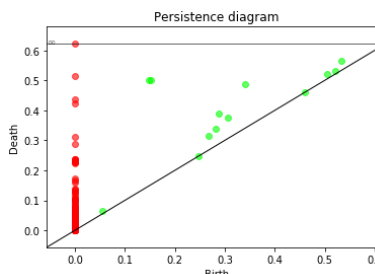
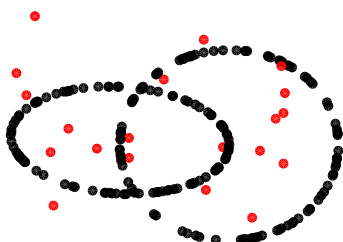
### 3 En présence de bruit

**Le problème des points aberrants.** Lorsque les données présentent quelques points aberrants, la fonction distance à un nuage de point ne permet plus d'estimer le diagramme de persistance de l'espace sous-jacent.

En guise d'exemple, considérons un échantillon  $X$  de 200 points tirés sur deux cercles entrelacés dans  $\mathbb{R}^3$ . L'homologie singulière sur  $k = \mathbb{Z}/2\mathbb{Z}$  des voisinages tubulaires  $(X^t)_{t \geq 0}$  de ce nuage de points définissent un module de persistance décomposable en modules-intervalle, dont on trace le diagramme de persistance :



Les points du diagramme du module de persistance  $(H_0(X^t))_{t \geq 0}$  sont en rouge et ceux de  $(H_1(X^t))_{t \geq 0}$  en vert. Nous ajoutons ensuite un échantillon de 20 points uniformément distribués dans le pavé :



Le diagramme de persistance est fortement perturbé. Pour remédier à ce problème, il faut remplacer la distance  $d_X$  par une fonction qui serait robuste au bruit. C'est le dessein de la distance à la mesure, une fonction qui tient compte de la densité de la distribution de points  $X$ . Elle a été introduite en 2011 par Chazal, Cohen-Steiner et Mériqot [9].

**Distance à la mesure.** On considère une mesure de probabilité  $\mu$  sur  $\mathbb{R}^d$ , et un paramètre  $m \in [0, 1]$ . Pour tout  $x \in \mathbb{R}^d$ , on définit la quantité :

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

La distance à la mesure  $\mu$  se définit ensuite comme une moyenne quadratique des fonctions  $\delta_{\mu,m}$ .

**Définition 11.** Soit  $m_0 \in [0, 1[$ . La distance à la mesure  $\mu$  de paramètre de masse  $m_0$  est la fonction :

$$d_{\mu,m_0} : \mathbb{R}^d \longrightarrow \mathbb{R} \\ x \longmapsto \sqrt{\frac{1}{m_0} \int_0^{m_0} \delta_{\mu,m}^2(x) dm}$$

La distance à la mesure admet une reformulation élégante. On dit qu'une mesure  $\nu$  est une sous-mesure de  $\mu$  si on a  $\nu \leq \mu$  en tant que fonctions. L'ensemble des sous-mesures de  $\mu$  est noté  $\text{Sub}(\mu)$ . Si  $m_0 \in \mathbb{R}$ , on note  $\text{Sub}_{m_0}(\mu)$  l'ensemble des sous-mesures de  $\mu$  de masse  $m_0$ . Mérigot montre dans [10] en 2009 que la distance à la mesure s'écrit comme un problème de transport entre sous-mesures, à l'aide de la distance de Wasserstein (voir [11] page 105).

**Théorème 12.** Pour tout  $x \in \mathbb{R}^d$ ,

$$d_{\mu,m_0}(x) = \frac{1}{\sqrt{m_0}} \inf\{W_2(m_0\delta_x, \nu), \nu \in \text{Sub}_{m_0}(\mu)\},$$

et cette borne inférieure est réalisée par une mesure  $\nu \in \text{Sub}_{m_0}(\mu)$ .

L'auteur y montre aussi deux propriétés de régularité fondamentales :

**Proposition 13.** La distance à la mesure  $d_{\mu,m}$  est 1-lipschitzienne, et la fonction

$$f : x \in \mathbb{R}^d \mapsto \|x\|^2 - d_{\mu,m}^2(x)$$

est convexe.

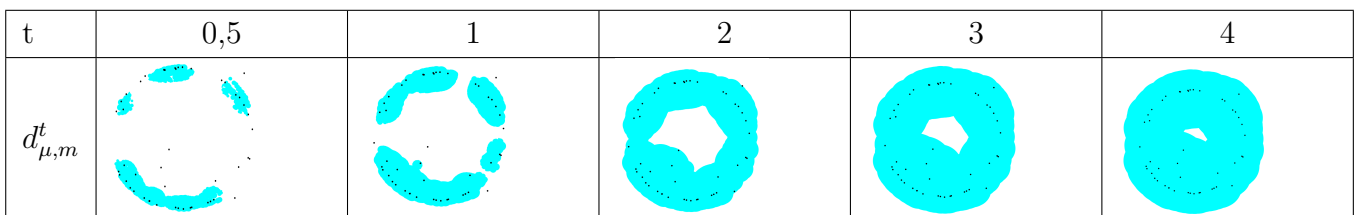
Enfin, la distance à la mesure est continue en la mesure :

**Proposition 14.** Soient  $\mu$  et  $\nu$  deux mesures de probabilité sur  $\mathbb{R}^d$ . Alors pour tout  $m \in [0, 1[$ ,

$$\|d_{\mu,m} - d_{\nu,m}\|_\infty \leq W_2(\mu, \nu).$$

**Estimation robuste.** Si  $X$  est un nuage de points, on peut considérer la mesure empirique  $\mu$  sur  $X$ . La dernière proposition montre que la distance à la mesure  $d_{\mu,m}$  est une fonction stable de  $X$  au sens suivant : si l'on rajoute peu de points aberrants à  $X$ , alors la distance à la mesure varie peu.

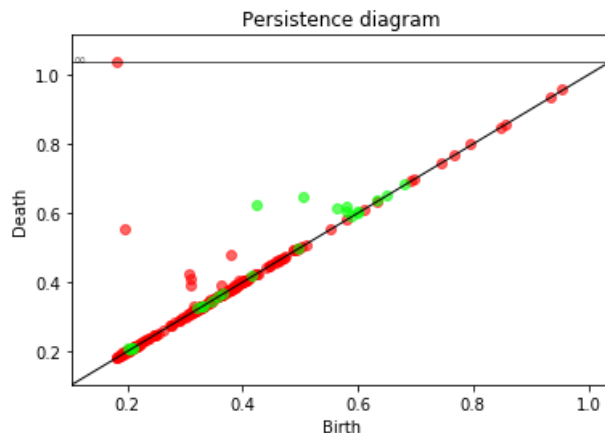
En guise d'exemple, nous traçons quelques sous-niveaux de  $d_{\mu,m}$  pour  $m = 0.1$  et  $\mu$  la mesure empirique d'un nuage de points bruité qui échantillonne le cercle :



Ces sous-niveaux sont proches de ceux du cercle lui-même.



Le module de persistance des sous-niveaux de la fonction  $d_{\mu,m}$  est robuste aux points aberrants. La figure suivante représente le diagramme de persistance correspondant à l'exemple de deux cercles entrelacés et plongés dans du bruit, lorsque l'espace est filtré par la distance à la mesure ( $m = 0.1$ ) :



Sur ce diagramme, on distingue deux points rouges et deux points verts éloignés de la diagonale : des propriétés topologiques de l'espace initial ont été retrouvées.

## Références

- [1] P. Frosini. Discrete computation of size functions. *Journal of combinatorics, information and system sciences*, 1992.
- [2] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 2005.
- [3] W. Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules. *arXiv :1210.0819v3*, 2014.
- [4] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Computational Geometry*, 2002.
- [5] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 2007.
- [6] F. Chazal, V. de Silva, M. Glisse, and S. Y. Oudot. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics, 2016.
- [7] S. Y. Oudot. *Persistence Theory : From Quiver Representations to Data Analysis*. American Mathematical Society, 2015.
- [8] G. Carlsson and V. de Silva. Zigzag persistence. *Foundations of Computational Mathematics*, 2010.
- [9] F. Chazal, D. Cohen-Steiner, and Q. Mérigot. Geometric inference for probability measures. *Foundations of Computational Mathematics*, 2011.
- [10] Q. Mérigot. Détection de structure géométrique dans les nuages de points. <https://tel.archives-ouvertes.fr/tel-00443038v2>, 2010.
- [11] C. Villani. *Optimal transport, old and new*. Springer, 2008.