

UN APERÇU DES MODULES DE PERSISTANCE ET DE LEURS APPLICATIONS

par Grégory Ginot

Introduction

La *persistance homologique* a en grande partie été développée pour être un des outils principaux de l'étude géométrique des « données », aussi appelée *analyse topologique des données*. Cette dernière est un domaine assez récent et très actif qui trouve de nombreuses applications dans divers domaines scientifiques, notamment via les méthodes issues de la théorie de la persistance homologique, qui a des racines anciennes en géométrie différentielle et topologie algébrique.

D'un point de vue purement mathématique, la théorie de la persistance peut se voir comme l'étude d'*espaces filtrés* $(X_i)_i$ via des invariants algébriques (aussi effectifs que possible) et des propriétés métriques naturelles qu'on peut leur associer (ce dernier point étant l'un des apports récents cruciaux). Et c'est essentiellement ce qui sera développé dans ces notes. Il s'agit le plus souvent d'espaces topologiques, variétés ou complexes simpliciaux (mais cela a du sens dans d'autres cadres).

Il est important de noter qu'un des points clés de la théorie de *la persistance est qu'elle vise à étudier la filtration en elle-même* et pas simplement à la voir comme un outil pour calculer l'espace colimite $\bigcup_{i \in I} X_i$.

Avant de détailler un peu plus la persistance et les modules persistants, on va essayer de préciser pourquoi elle intervient en analyse topologique des données. L'idée de cette branche est d'étudier des données (complexes et possiblement en grande dimension) en essayant d'évaluer leur géométrie et/ou leur topologie, avec comme objectif des tâches de type classification des données, visualisation, régression ou reconnaissance de patterns communs. Souvent, ces données apparaissent comme un espace métrique (\mathbb{X}, d) , typiquement un grand sous-ensemble fini d'un espace euclidien — qu'on qualifie souvent de nuage de points.

Par « information topologique » extraite d'un tel ensemble, on entend la forme que prennent ces données dans l'espace ambiant (possiblement de grande dimension). Ce qu'on appelle la *manifold hypothesis* prescrit que (après avoir éventuellement partitionné les données en groupes) ces données s'accumulent sur une sous-variété X de l'espace ambiant. On peut alors penser les données comme une *discrétisation* de cet objet continu X , qui

peut être *a priori* inconnu de l’observateur. L’analyse topologique des données a pour but d’utiliser (et d’approximer donc) la topologie de X pour comprendre ou organiser les données.

La persistance homologique s’appuie sur les méthodes de la topologie algébrique classique *adaptées aux espaces filtrés* pour estimer la topologie des données. Reprenons l’exemple d’un sous-ensemble \mathbb{X} d’un espace disons euclidien (muni de la distance induite), vu comme une approximation d’une sous-variété X . À tout point x de notre ensemble discret \mathbb{X} , on peut associer la boule $B(x, r)$ (c’est-à-dire épaissir notre point x) et la réunion $\mathbb{X}(r) = \bigcup_{x \in \mathbb{X}} B(x, r)$ de ces boules. Si \mathbb{X} est une bonne approximation de X , alors cette réunion de boules représente un espace topologique qui, pour un r ni trop petit (sinon on ne voit que les points) ni trop grand, va être un épaississement de X qui lui sera *homotope*. En pratique, on remarque que cette famille $(\mathbb{X}(r))_{r \geq 0}$ est filtrée au sens où $r < r'$ implique $\mathbb{X}(r) \subset \mathbb{X}(r')$. On obtient donc ici un espace topologique filtré $\mathbb{X}(r)$ dont la filtration est paramétrée par l’ensemble \mathbb{R}_+ . En pratique, si \mathbb{X} est fini, il n’y a qu’un nombre fini de valeurs de r où la topologie de $\mathbb{X}(r)$ change, et donc notre espace est en fait équivalent à un espace filtré par un sous-ensemble fini de \mathbb{R} (que l’on ne connaît pas forcément *a priori*). Ceci constitue un premier exemple d’espace filtré apparaissant en analyse topologique de données.

Une autre famille standard (et très proche) d’exemples est donnée par les sous-niveaux d’une fonction $f: X \rightarrow \mathbb{R}$, c’est-à-dire la collection des $\{x \in X, f(x) < t\}_{t \in \mathbb{R}}$ qui est là encore naturellement paramétrée par \mathbb{R} .

Un dernier exemple standard est donné par le cas d’un graphe (avec la distance de plus court chemin). On a alors des informations topologiques évidentes comme le nombre de composantes connexes du graphe (qui correspondent à partitionner) et le nombre de cycles dans chaque composante. Notons que ces informations topologiques sont là aussi « à homotopie près », au sens où le nombre de composantes connexes rend un arbre et un point équivalents, et où le nombre de cycles correspond au groupe fondamental (de la composante connexe considérée) du graphe. On peut aussi extraire des informations de plus grande dimension, ce qui revient à construire un complexe simplicial (par exemple le complexe de clique du graphe) qui tient compte des connectivités d’ordre supérieur dans le graphe.

La persistance homologique a pour objet d’étudier les espaces filtrés (par un sous-ensemble de \mathbb{R} dans le cadre originel, mais il y a des développements importants en direction du cadre à plusieurs paramètres, *cf.* partie 5) via des invariants topologiques. Un principe, du point de vue des applications en analyse topologique des données, est donc que l’on souhaite avoir trois propriétés fondamentales :

- ils doivent être invariants par homotopie (de façon à ne pas être sensibles à de petites déformations, comme dans l’exemple des réunions de boules ci-dessus) ;
- ils doivent être calculables sur machine (de façon à avoir des applications) ;

- ils doivent être comparables pour divers jeux de données, ce qui va se faire par l’intermédiaire de (pseudo-)métriques si possible là aussi implémentables sur machine.

Les invariants qui vont réaliser cela seront encodés par la notion de *module de persistance* associée à un objet filtré dans la section 1. Ces modules sont essentiellement la structure obtenue en passant à l’homologie dans un espace filtré. Cette structure a de très bonnes propriétés dans le cadre algébrique, notamment le fait que l’on peut décomposer les modules de persistance en indécomposables simples caractérisés par un intervalle de \mathbb{R} . La donnée de ces intervalles est un objet combinatoire, appelé son *code-barres*, qui permet de le manipuler informatiquement et combinatoirement facilement. Ces propriétés seront le cœur de la partie 2.

Comme évoqué plus haut, une idée clé de la persistance est le fait que les modules de persistance ont une (pseudo)-métrique naturelle, appelée *distance d’entrelacement*, qui a une traduction combinatoire sur le code-barres appelée distance bottleneck, c’est-à-dire du « goulot de bouteille » en français). C’est l’outil essentiel pour comparer des objets persistants et obtenir des estimées intéressantes dans les applications. Ces distances seront définies et comparées dans la partie 3.

Dans les applications, un point important est que les distances que l’on a obtenues pour comparer les modules persistants (ou leurs code-barres associés à des données) reflètent bien la géométrie des données. Ceci est établi par ce que l’on appelle les théorèmes de stabilité énoncés en partie 4, qui essentiellement garantissent que l’homologie persistante fournit des estimateurs consistants (au sens statistique du terme) de la topologie des données.

Dans la section 5, nous évoquerons des généralisations et raffinements importants de l’homologie persistante, notamment le passage à plusieurs paramètres qui est un enjeu important non seulement dans la théorie mais aussi dans les applications concrètes à d’autres sciences.

Enfin, dans la partie 6, nous parlerons du problème de l’inférence, c’est-à-dire que nous reviendrons au problème initial de déterminer la topologie d’un espace X à partir d’une discrétisation et expliciterons ce qu’apportent les outils introduits dans les sections précédentes dans ce contexte.

Notons que du point de vue de l’étude des espaces filtrés, il n’est évidemment pas nécessaire que ces derniers proviennent de données. La persistance homologique a un intérêt mathématique propre. Elle a de fait trouvé des applications en géométrie symplectique (via les filtrations naturelles dans les complexes d’homologie de Floer : L. Polterovich, Rosen, Samvelyan et Zhang, 2020), a été étudiée en lien avec la théorie de Morse avant même l’analyse topologique des données, et a également des applications en géométrie que nous évoquerons brièvement dans la partie 6.

Notations et conventions

- Dans toute la suite \mathbb{F} sera un corps. Dans les applications pratiques de la persistance homologique, ce corps est en général un corps fini (notamment pour éviter d’avoir des problèmes d’arrondi dans les calculs sur machine).
- Si on ne précise rien, les groupes d’homologie seront les groupes d’homologie singulière (pour un espace topologique) ou les groupes d’homologie simpliciale (pour un complexe simplicial) qui calculent précisément l’homologie singulière de l’espace topologique dont le complexe simplicial est une triangulation.
- Si on ne précise rien, les groupes d’homologie ou espaces vectoriels considérés seront sur le corps \mathbb{F} . On notera $\mathbb{F}\text{-Vect}$ la catégorie des \mathbb{F} -espaces vectoriels.

On renvoie le lecteur à Chazal, Silva, Glisse et Oudot (2016), Oudot (2015), Baranikov (1994), L. Polterovich, Rosen, Samvelyan et Zhang (2020) et Botnan et Lesnick (2023) pour des références détaillées sur la persistance homologique et ses applications en géométrie symplectique, et des preuves détaillées de résultats des sections 2, 4 notamment.

1. Modules de persistance

Nous introduisons dans cette partie la notion de module de persistance, qui est le miroir algébrique des espaces filtrés. Cette notion peut se généraliser pour des structures autres que des filtrations par des sous-ensembles de \mathbb{R} et pour des structures autres que les espaces vectoriels, comme nous le nous verrons ci-dessous.

DÉFINITION 1.1 (Module de persistance). — *Un module de persistance est la donnée d’une collection $(M(t))_{t \in \mathbb{R}}$ de \mathbb{F} -espaces vectoriels et pour tout $s \leq t$, d’applications linéaires $\iota_{s \leq t}^M : M(s) \rightarrow M(t)$, vérifiant :*

- $\iota_{s \leq s}^M = \text{id}_{M(s)}$;
- pour tout $s \leq t \leq u$, on a $\iota_{t < u}^M \circ \iota_{s < t}^M = \iota_{s < u}^M$.

On notera simplement $M = (M(s), \iota_{s \leq t}^M)_{s,t}$ cette donnée.

Si M et N sont deux modules de persistance, un morphisme $f : M \rightarrow N$ est la donnée d’une collection $f_t : M(t) \rightarrow N(t)$ ($t \in \mathbb{R}$) vérifiant la condition suivante : pour tout $s < t$, le diagramme

$$\begin{array}{ccc} M(s) & \xrightarrow{\iota_{s < t}^M} & M(t) \\ f_s \downarrow & & \downarrow f_t \\ N(s) & \xrightarrow{\iota_{s < t}^N} & N(t) \end{array}$$

est commutatif.

La composition des morphismes se fait point par point $((g \circ f)_t = g_t \circ f_t)$ et l'on obtient ainsi une catégorie.

Exemple 1.2 (Module à support sur un intervalle). — Soit E un \mathbb{F} -espace vectoriel et soit I un intervalle de \mathbb{R} . On dispose d'un module de persistance associé défini comme suit : pour tout $t \in \mathbb{R}$, on a

$$E_I(t) = \begin{cases} E & \text{si } t \in I \\ \{0\} & \text{sinon,} \end{cases}$$

et les morphismes structuraux sont donnés, pour tout $s \leq t$, par $\iota_{s \leq t}^{E_I} = \text{id}_E$ si $t, s \in I$ et (nécessairement) le morphisme nul sinon.

Exemple 1.3. — Pour tout réel t , on pose $\tilde{\mathbb{F}}(t) = \mathbb{F}$ et, pour $s < t$,

$$\tilde{\iota}_{s \leq t}^{\mathbb{F}} = \begin{cases} \text{id}_{\mathbb{F}} & \text{si } t \leq 0 \\ 0 & \text{si } s \leq 0 \text{ et } t > 0 \\ \text{id}_{\mathbb{F}} & \text{si } s > 0 \end{cases}$$

On vérifie que cette structure définit bien un module de persistance tel que pour tout t , on ait $\tilde{\mathbb{F}}(t) = \mathbb{F}_{\mathbb{R}}(t)$, mais dont les morphismes structuraux diffèrent. En particulier, $\tilde{\mathbb{F}}$ n'est pas isomorphe à $\mathbb{F}_{\mathbb{R}}$ comme module de persistance (ce que l'on peut vérifier élémentairement et qui découle aussi de la proposition 2.3 ci-dessous).

La notion de module de persistance que nous avons donnée est un cas particulier associé à la structure d'ensemble ordonné (\mathbb{R}, \leq) . Plus généralement, soit $(S, <)$ un ensemble partiellement ordonné. On lui associe la catégorie $S^{<}$ dont les objets sont les éléments de S et les ensembles de morphismes

$$\text{Hom}_{S^{<}}(s, t) = \begin{cases} \{*\} & \text{si } s = t \text{ ou } s < t; \\ \emptyset & \text{sinon.} \end{cases}$$

Autrement dit, il y a un unique morphisme de s vers t si $s \leq t$ et aucun sinon. L'unicité garantit que l'on obtient bien une catégorie.

Rappelons que si \mathcal{C} est une petite catégorie et \mathcal{D} une catégorie, la catégorie $\text{Fun}(\mathcal{C}, \mathcal{D})$ est la catégorie dont les objets sont les foncteurs de \mathcal{C} vers \mathcal{D} et les morphismes les transformations naturelles entre foncteurs.

On peut réécrire la notion de module de persistance comme suit.

LEMME 1.4. — *La catégorie des modules de persistance de la définition 1.1 est la catégorie $\text{Fun}(\mathbb{R}^{<}, \mathbb{F}\text{-Vect})$ des foncteurs de la catégorie associée à (\mathbb{R}, \leq) vers celle des \mathbb{F} -espaces vectoriels.*

On peut également remplacer la catégorie des \mathbb{F} -espaces vectoriels par toute catégorie. Ainsi on a la notion générale suivante.

DÉFINITION 1.5. — Soit $(S, <)$ un ensemble partiellement ordonné. La catégorie des modules de persistance modelés sur $(S, <)$ est celle des foncteurs de $S^<$ vers $\mathbb{F}\text{-Vect}$.

Si \mathcal{C} est une catégorie, on appelle catégorie des \mathcal{C} -objets de persistance modelés sur $(S, <)$ la catégorie des foncteurs de $S^<$ vers \mathcal{C} (simplement appelés objets persistants si $S^< = \mathbb{R}^<$).

En particulier, tout sous-ensemble de \mathbb{R} héritant d'une structure d'ensemble partiellement ordonné, nous avons immédiatement une notion de modules et objets de persistance pour tout sous-ensemble de \mathbb{R} , en particulier \mathbb{R}_+ , un ensemble dénombrable discret, un sous-ensemble fini...). Et par ailleurs, tout objet persistant sur \mathbb{R} induit une telle structure par restriction. Notons que si $S \subset \mathbb{R}$ est discret, on peut étendre tout module de persistance F sur S en un module sur \mathbb{R} en prenant pour $t \in [s, s'[$, avec s, s' deux éléments consécutifs de S , $F(t) = F(s)$ et en étendant les morphismes de structure par l'identité sur $[s, s'[$.

Remarque 1.6. — Dans les applications pratiques associées à un nuage de points, on a en pratique un sous-ensemble fini de \mathbb{R} sur lequel se concentrent les seules évolutions du module de persistance associé. Il n'est en général pas évident de déterminer a priori quel sera ce sous-ensemble. De plus, la notion d'objet de persistance sur \mathbb{R} est utile pour comprendre la variété que l'on souhaite approximer.

Exemple 1.7 (Espaces de persistance). — Soit $(X(t))_{t \in \mathbb{R}}$ une famille d'espaces topologiques tels que pour $s < t$ on ait $X(s) \subset X(t)$. Alors cette collection $(X(s), X(s) \hookrightarrow X(t))_{s < t}$ définit un espace topologique de persistance. On pourra remarquer que les exemples de sous-niveaux de fonction et de réunion de boules de l'introduction sont précisément de cette nature. Et que tout espace filtré au sens usuel l'est également (pour le sous-ensemble \mathbb{N} de \mathbb{R}).

De même, un complexe simplicial filtré $F_0X \subset \dots \subset F_nX \subset \dots$ est un complexe simplicial de persistance modelé sur \mathbb{N} .

On va maintenant donner deux exemples cruciaux à la base du développement de la théorie et provenant de l'analyse topologique de données. Ils sont basés sur la remarque élémentaire suivante :

Exemple 1.8 (Module de persistance associé à un espace de persistance)

Soit encore $(X(t))_{t \in \mathbb{R}}$ une famille d'espaces topologiques telle que pour $s < t$ on ait $X(s) \subset X(t)$. Notons $\iota_{s \leq t}: X(s) \hookrightarrow X(t)$ les inclusions. Alors, en prenant les groupes d'homologie en degré i , on obtient des applications linéaires $H_i(\iota_{s \leq t}, \mathbb{F}): H_i(X(s), \mathbb{F}) \rightarrow H_i(X(t), \mathbb{F})$ qui font de

$$\left(H_i(X(t), \mathbb{F}), H_i(\iota_{s \leq t}, \mathbb{F}) \right)_{s, t}$$

un module de persistance (car prendre les groupes d'homologie est un foncteur). On notera que bien que $\iota_{s \leq t}$ soit une inclusion, on n'a aucune propriété particulière pour les morphismes induits en homologie.

On appelle parfois ce module de persistance l'homologie persistante de $(X(t))$, mais cette terminologie peut aussi souvent servir à désigner les images des morphismes de structure (indexés par $s \leq t$). Nous préférons donc éviter d'utiliser la terminologie « homologie persistante » dans ces notes pour éviter des confusions.

Notons qu'une construction similaire à l'exemple précédent fonctionne pour tout espace topologique de persistance (autrement dit un objet de persistance dans la catégorie des espaces topologiques), ou pour tout complexe de chaînes de persistance.

Remarque 1.9. — La remarque naïve est ici que tout foncteur $\mathcal{C} \rightarrow \mathcal{D}$ induit un foncteur des \mathcal{C} -objets de persistance vers les \mathcal{D} -objets de persistance, par simple composition de foncteurs.

Voici maintenant deux exemples fondamentaux déjà évoqués dans l'introduction.

Exemple 1.10 (Sous-niveaux). — Soit X un espace topologique et $f: X \rightarrow \mathbb{R}$ une application continue. Le sous-niveau (ouvert) de hauteur t est la préimage $X^f(t) := f^{-1}(]-\infty, t])$. On a en particulier que $X^f(s) \subset X^f(t)$ si $s \leq t$. D'après l'exemple 1.8, on dispose alors pour tout $i \in \mathbb{N}$ du module de persistance

$$H_i(X^f(t), \mathbb{F})$$

appelé module de persistance en degré i de f .

Donnons maintenant un exemple clé en lien avec l'approximation discrète d'un sous-espace.

Exemple 1.11 (Module de Čech). — Soit \mathbb{X} un sous-ensemble d'un espace métrique (Y, d) et soit $r > 0$. On construit un complexe simplicial (abstrait) $C(\mathbb{X}_r)$ dont les sommets sont les points de \mathbb{X} .

L'ensemble $\Delta^n(\mathbb{X}_t)$ des n -simplexes de $C(\mathbb{X}_t)$ est l'ensemble des $n + 1$ -uplets $\{x_0, \dots, x_n\}$ de points de \mathbb{X} tels que

$$\bigcap_{i=0 \dots n} B(x_i, r) \neq \emptyset$$

où $B(x_i, r)$ est la boule ouverte de rayon r centrée en x_i .

Par convention on pose $C(\mathbb{X}_r) = \emptyset$ pour $r \leq 0$.

Par construction, pour tout $r < r'$, l'ensemble $C(\mathbb{X}_r)$ est un sous-complexe simplicial de $C(\mathbb{X}_{r'})$. En passant aux groupes d'homologie simpliciale, on obtient pour tout entier naturel i un module de persistance

$$H_i C(\mathbb{X}) = \left(H_i(C(\mathbb{X}_r), \mathbb{F}) \right)_r$$

dont les morphismes structuraux sont induits par les inclusions.

On peut remarquer que l’homologie simpliciale du complexe de Čech est la même chose que l’homologie singulière de sa réalisation topologique, c’est-à-dire la réunion $\bigcup_{i \in \mathbb{N}} \left(\bigcup_{\sigma \in \Delta^i(\mathbb{X}_t)} \langle \sigma \rangle \right) / \sim$ de tous les simplexes de $C(\mathbb{X}_r)$ quotientée par la relation identifiant les faces communes des simplexes.

Cet espace a une autre interprétation. Il s’agit en fait d’une abstraction combinatoire de la réunion des boules de rayon r dont les centres sont les points de \mathbb{X} .

PROPOSITION 1.12 (Corollaire du théorème du nerf). — *Soit \mathbb{X} un sous-ensemble de \mathbb{R}^m . Alors la réalisation géométrique du complexe de Čech $|C(\mathbb{X}_r)|$ est homotope à $\bigcup_{x \in \mathbb{X}} B(x, r)$ et le diagramme suivant est commutatif à homotopie près :*

$$\begin{array}{ccc} |C(\mathbb{X}_r)| & \hookrightarrow & |C(\mathbb{X}_{r'})| \\ \simeq \downarrow & & \simeq \downarrow \\ \bigcup_{x \in \mathbb{X}} B(x, r) & \hookrightarrow & \bigcup_{x \in \mathbb{X}} B(x, r') \end{array}$$

En particulier les modules de persistance $H_i C(\mathbb{X})$ sont les mêmes que ceux de l’espace topologique de persistance donné par les réunions de boules $\bigcup_{x \in \mathbb{X}} B(x, r)$ (cf. l’introduction).

Notons que si \mathbb{X} est fini, alors pour r petit, l’homologie persistante est concentrée en degré 0 et y est donnée par l’espace vectoriel engendré par les points de \mathbb{X} ; et que de plus, pour r grand, le complexe de Čech est homotope à une boule et donc contractile. L’information intéressante capturée par le module de persistance réside dans les transformations qui se présentent entre ces deux extrêmes.

Remarque 1.13. — Dans les applications pratiques aux sciences des données, le complexe de Čech d’un ensemble fini \mathbb{X} est souvent trop lourd à calculer et à stocker, car il nécessite de regarder des intersections de nombreux disques. On utilise des variantes plus simples. Par exemple le complexe de Rips défini, pour tout $t > 0$, comme le complexe simplicial abstrait dont les sommets sont les points de \mathbb{X} et l’ensemble $\mathcal{R}_n(\mathbb{X})(t)$ des n -simplexes de $\mathcal{R}(\mathbb{X})(t)$ est l’ensemble des $n + 1$ -uplets

$\{x_0, \dots, x_n\}$ de points de \mathbb{X} tels que

$$\forall i, j, \quad d(x_i, x_j) < t.$$

On appelle module de persistance de Rips en degré i , le module de persistance obtenu en prenant l’homologie en degré i de $\mathcal{R}(\mathbb{X})(t)$ (comme dans l’exemple 1.8).

Notons que l’on a des inclusions évidentes

$$\mathcal{R}(\mathbb{X}(r)) \subset C(\mathbb{X}_r) \subset \mathcal{R}(\mathbb{X}(2r)).$$

Ceci permet, via un changement logarithmique de paramétrisation de \mathbb{R} de relier précisément les modules de persistance de Čech et de Rips, et en particulier de contrôler la différence entre leurs code-barres (voir 2.5) respectifs

2. Structure des modules de persistance

Un intérêt majeur des modules de persistance (notamment par rapport aux espaces de persistance) provient de la structure algébrique supplémentaire dont on dispose lorsqu'on est à valeurs dans un corps — structure qui découle de la théorie des représentations d'algèbres.

Étant donnée une catégorie de foncteurs à valeurs dans les espaces vectoriels, la catégorie des modules de persistance hérite de limites et colimites (donc de produits et coproduits). En particulier nous avons la notion de **somme directe** $M \oplus N$ de deux modules de persistance M, N ; cette somme directe est définie, pour tout $t \in \mathbb{R}$ et tout $s \leq t$, par

$$(M \oplus N)(t) = M(t) \oplus N(t), \quad (M \oplus N)(s) \xrightarrow{(\iota_s^M, \iota_s^N)} (M \oplus N)(t).$$

Nous notons simplement 0 le module de persistance qui vaut 0 en tout point.

DÉFINITION 2.1. — *On dit qu'un module persistant M est décomposable s'il est isomorphe à une somme directe $N_1 \oplus N_2$ avec $N_1, N_2 \neq 0$. Sinon on dit qu'il est indécomposable.*

Exemple 2.2. — Le module de persistance $\tilde{\mathbb{F}}$ de l'exemple 1.3 se décompose sous la forme

$$\tilde{\mathbb{F}} = \mathbb{F}]_{-\infty, 0] \oplus \mathbb{F}]_{0, +\infty[.$$

Si la notion précédente de décomposabilité a un sens très général, le fait de travailler sur un corps nous donne la caractérisation essentielle suivante des indécomposables. Rappelons la construction des modules intervalles 1.2.

PROPOSITION 2.3. — *Les modules de persistance indécomposables sont les modules intervalles de rang 1, c'est-à-dire les \mathbb{F}_I où I est un intervalle quelconque de \mathbb{R} .*

Remarque 2.4. — Les modules indécomposables sont en général réductibles. De fait, les irréductibles non triviaux sont les modules de persistance de la forme $\mathbb{F}_{\{x\}}$, c'est-à-dire ceux de rang 1 supportés par un point.

Le théorème principal des modules de persistance énonce que ces derniers ont une décomposition très simple en indécomposables.

THÉORÈME 2.5 (de structure des modules persistants : Cochoy et Oudot, 2020 ; Botnan et Crawley-Boevey, 2020)

Soit M un module de persistance ponctuellement de rang fini, c'est-à-dire tel que pour tout $t \in \mathbb{R}$, $M(t)$ soit de dimension finie. Alors M se décompose sous la forme

$$(1) \quad M \cong \bigoplus_{I \in \mathbf{B}(M)} \mathbb{F}_I$$

où $\mathbf{B}(M)$ est un multi-ensemble d'intervalles.

La décomposition est unique à isomorphisme près et à l'ordre près des termes dans la somme directe.

Par un multi-ensemble on entend un ensemble avec répétition possible de chaque intervalle.

DÉFINITION 2.6. — *Le code-barre d'un module de persistance M est le multi-ensemble d'intervalles $\mathbf{B}(M)$ donné par le théorème de décomposition.*

Le code-barre est une donnée combinatoire très simple qui caractérise complètement tout module de persistance ponctuellement de type fini. En particulier, cette donnée est facilement stockable informatiquement.

Exemple 2.7. — Le module de persistance $\tilde{\mathbb{F}}$ de l'exemple 1.3 a pour code-barre $\{]-\infty, 0],]0, +\infty[\}$.

Exemple 2.8. — Si X est une variété compacte et f est une fonction de Morse, alors le code-barre $\mathbf{B}_i(X^f)$ des sous-niveaux de f (exemple 1.10) est constitué d'un nombre fini de barres de la forme $]a_q, b_q]$ et de barres infinies $]c_p, \infty[$. Ces dernières sont en nombre exactement égal au nombre de Betti $\beta_i(X) = \dim H_i(X, \mathbb{F})$. Les bornes des barres finies détectent des points critiques de la fonction f .

Remarque 2.9. — Les hypothèses du théorème de structure 2.5 sont toujours vérifiées pour les modules de persistance associés à un complexe de Čech (ou de Rips) d'un ensemble fini.

Le théorème de structure tel que nous l'avons énoncé est un cas particulier d'un théorème plus général pour tout module de persistance défini sur un ensemble totalement ordonné $(S, <)$. En particulier, il s'applique par restriction à tout sous-ensemble de \mathbb{R} et donc aux homologies d'espaces filtrés au sens usuel. Avant d'énoncer ledit théorème général, précisons la notion suivante.

DÉFINITION 2.10 (Module intervalle, cas général). — *Un intervalle d'un ensemble partiellement ordonné $(S, <)$ est un sous-ensemble I de S qui est connexe (c'est-à-dire que deux points de l'ensemble sont toujours comparables) et convexe au sens de l'ordre, c'est-à-dire tel que pour tout $a \leq b$ dans I , si $t \in S$ et $a \leq t \leq b$, alors $t \in I$.*

On dispose des modules intervalles \mathbb{F}_I , définis comme dans l'exemple 1.2 de la manière suivante : pour tout $s \in S$,

$$\mathbb{F}_I(s) = \begin{cases} \mathbb{F} & \text{si } s \in I \\ \{0\} & \text{sinon.} \end{cases}$$

et les morphismes structuraux sont donnés, pour tout $s < t$, par $\mathbb{F}_I(s) \rightarrow \mathbb{F}_I(t) = \text{id}_{\mathbb{F}}$ si $t, s \in I$ et (nécessairement) le morphisme nul sinon.

LEMME 2.11. — *Soit $(S, <)$ un ensemble partiellement ordonné. Pour tout intervalle I de S , le module \mathbb{F}_I est indécomposable.*

En général, il existe d'autres modules indécomposables ; cependant, dans le cas totalement ordonné, on a l'analogie du théorème 2.5 de structure. Rappelons qu'un module de persistance modelé sur S est dit ponctuellement de rang fini si pour tout $s \in S$, l'espace $M(t)$ est de dimension finie.

THÉORÈME 2.12 (CB ; Botnan et Crawley-Boevey, 2020). — *Soit M un module de persistance modelé sur un ensemble totalement ordonné $(S, <)$ ponctuellement de rang fini. Alors M se décompose sous la forme*

$$(2) \quad M \cong \bigoplus_{I \in \mathbf{B}(M)} \mathbb{F}_I$$

où $\mathbf{B}(M)$ est un multi-ensemble d'intervalles de S . La décomposition est unique à isomorphisme près et à l'ordre près des termes dans la somme directe.

3. Distance entre modules de persistance

Dans les applications, notamment pour l'analyse des données, un point fondamental est de pouvoir comparer à quel point des modules de persistance sont proches l'un de l'autre ; et en particulier si par exemple le module de persistance associé à une discrétisation d'un espace est proche de celui de l'espace en question.

La structure persistante permet de définir une notion de (pseudo-)distance naturelle entre les objets persistants. Par ailleurs, cette structure peut se transférer sur le code-barre, permettant de la calculer d'une manière algorithmique.

La distance induite par la structure persistante a un sens pour les espaces vectoriels et plus généralement pour des objets de persistance dans toute catégorie. Elle est basée sur la notion de décalage.

DÉFINITION 3.1. — *Soit $F: \mathbb{R}^{\leq} \rightarrow \mathcal{C}$ un objet persistant dans \mathcal{C} et soit $\varepsilon > 0$. On note $F[\varepsilon]: \mathbb{R}^{\leq} \rightarrow \mathcal{C}$ l'objet persistant défini comme suit : $F[\varepsilon](t) = F(t + \varepsilon)$ et, pour $s \leq t$, le morphisme structurel $\iota_{s \leq t}^{F[\varepsilon]}: F[\varepsilon](s) \rightarrow F[\varepsilon](t)$ est donné par*

$$\iota_{s \leq t}^{F[\varepsilon]} = \iota_{s+\varepsilon \leq t+\varepsilon}^F: F[\varepsilon](s) = F(s + \varepsilon) \longrightarrow F(t + \varepsilon) = F[\varepsilon](t).$$

On appelle $F[\varepsilon]$ le décalage de longueur ε de F .

De façon similaire, si $f: F \rightarrow G$ est un morphisme de modules de persistance, on pose

$$f[\varepsilon] = \left(F[\varepsilon](t) = F(\varepsilon + t) \xrightarrow{f_{t+\varepsilon}} G(\varepsilon + t) = G[\varepsilon](t) \right)_{t \in \mathbb{R}}.$$

On énumère maintenant les propriétés élémentaires du décalage.

LEMME 3.2. — *Pour tout $\varepsilon > 0$, on a les résultats suivants.*

— $F[\varepsilon]$ est un objet de persistance de \mathcal{C} et $(F \mapsto F[\varepsilon], f \mapsto f[\varepsilon])$ est un endofoncteur de la catégorie des objets de persistance de \mathcal{C} .

- De plus, la collection de morphismes $\left(\iota_{t \leq t+\varepsilon}^F : F(t) \longrightarrow F(t+\varepsilon) \right)_{t \in \mathbb{R}}$ est un morphisme de modules de persistance $F \rightarrow F[\varepsilon]$, que l'on notera $\tau^F[\varepsilon]$ et appellera morphisme de décalage de longueur ε .
- Pour tout $\varepsilon' > 0$, on a $\tau^F[\varepsilon] \circ \tau^F[\varepsilon'] = \tau^F[\varepsilon + \varepsilon']$.

DÉFINITION 3.3 (Distance d'entrelacement). — Soit \mathcal{C} une catégorie (par exemple celle des \mathbb{F} -espaces vectoriels). Soient $F, G : \mathbb{R}^{\leq} \rightarrow \mathcal{C}$ des objets persistants dans \mathcal{C} . On dit que les objets persistants F et G sont ε -entrelacés s'il existe deux morphismes de modules de persistance

$$\mu : F \longrightarrow G[\varepsilon] \quad \text{et} \quad \nu : G \longrightarrow F[\varepsilon]$$

tels que les diagrammes suivants

$$\begin{array}{ccccc} & & \tau^F[2\varepsilon] & & \\ & \curvearrowright & & \curvearrowleft & \\ F & \xrightarrow{\mu} & G[\varepsilon] & \xrightarrow{\nu[\varepsilon]} & F[2\varepsilon] \\ & & & & \\ G & \xrightarrow{\nu} & F[\varepsilon] & \xrightarrow{\mu[\varepsilon]} & G[2\varepsilon] \\ & & \tau^G[2\varepsilon] & & \end{array}$$

soient commutatifs.

On appelle distance d'entrelacement entre F et G l'infimum

$$d(F, G) := \inf \{ \varepsilon \geq 0 \mid F \text{ et } G \text{ sont } \varepsilon\text{-entrelacés} \} \in \mathbb{R}_+ \cup \{ \infty \}.$$

Autrement dit, deux modules de persistance sont ε -entrelacés si l'on peut factoriser simultanément les morphismes de décalage de 2ε de l'un au travers de l'autre.

On peut vérifier le résultat suivant.

LEMME 3.4. — La distance d'entrelacement est une pseudo-distance :

$$d(F, G) \geq 0 \quad \text{et} \quad d(F, H) \leq d(F, G) + d(G, H).$$

On appelle *modules éphémères* les modules de persistance à distance nulle de 0 : par exemple, un module supporté sur un point $\mathbb{F}_{\{x\}}$ est éphémère.

Notons que la distance d'entrelacement a du sens pour n'importe quelle catégorie \mathcal{C} à l'arrivée (mais peut y être très difficilement calculable). Tout foncteur de \mathcal{C} dans \mathcal{D} est automatiquement 1-lipschitzien pour cette distance.

LEMME 3.5. — Si $\Psi : \mathcal{C} \rightarrow \mathcal{D}$ est un foncteur, alors le foncteur induit $\text{Pers}(\mathcal{C}) \rightarrow \text{Pers}(\mathcal{D})$ est 1-lipschitzien : pour toute paire $(F, G) : \mathbb{R}^{\leq} \rightarrow \mathcal{C}$ d'objets \mathcal{C} -persistants, on a

$$d(\Psi \circ F, \Psi \circ G) \leq d(F, G).$$

Le lemme s'applique en particulier pour les foncteurs groupes d'homologie $H_i(-, \mathbb{F})$.

Exemple 3.6. — Soient E un \mathbb{F} -espace vectoriel et $I = [a, b]$ un intervalle. Comme on a $t + \varepsilon \in [a, b]$ si et seulement si $t \in [a - \varepsilon, b - \varepsilon]$, on obtient que le décalage par ε du module intervalle E_I est

$$E_I[\varepsilon] = E_{[a-\varepsilon, b-\varepsilon]}.$$

C'est donc le même module, mais porté par la translation de ε sur la droite de l'intervalle I . Ici on peut remplacer l'intervalle fermé par des intervalles ouverts ou semi-ouverts bien sûr.

Soit $I = \langle a, b \rangle$ un intervalle (où on utilise \langle, \rangle pour ne pas préciser si les bornes sont ouvertes ou fermées). L'exemple précédent montre que le morphisme de décalage naturel $\mathbb{F}_I \rightarrow \mathbb{F}_I[\varepsilon]$ est nul si $\varepsilon > b - a$ et non nul si $\varepsilon < b - a$. On peut en déduire le résultat suivant

LEMME 3.7. — *On a*

$$d(\mathbb{F}_{\langle a, b \rangle}, 0) = \frac{b - a}{2} \quad \text{et} \quad d(\mathbb{F}_{\langle a, b \rangle}, \mathbb{F}_J) = 0$$

pour tout intervalle J avec les mêmes extrémités que I . De plus

$$d(\mathbb{F}_{\langle a, b \rangle}, \mathbb{F}_J) = +\infty$$

si J est non borné.

Exemple 3.8. — On peut évaluer la distance entre les modules intervalles $\mathbb{F}_{[0,2]}$ et $\mathbb{F}_{[1,3]}$ par un argument similaire. Pour tout $\varepsilon > 1$, les morphismes de décalage $\mathbb{F}_{[0,2]} \rightarrow \mathbb{F}_{[0,2]}[2\varepsilon]$ et $\mathbb{F}_{[1,3]} \rightarrow \mathbb{F}_{[1,3]}[2\varepsilon]$ sont nuls. Par conséquent, $\mathbb{F}_{[0,2]}$ et $\mathbb{F}_{[1,3]}$ sont ε -entrelacés. En revanche, si $\varepsilon < 1$, le morphisme de décalage $\tau^{\mathbb{F}_{[0,2]}}[2\varepsilon]: \mathbb{F}_{[0,2]} \rightarrow \mathbb{F}_{[0,2]}[2\varepsilon]$ évalué en $t = 0$ est l'identité de \mathbb{F} . Mais comme $\mathbb{F}_{[1,3]}[\varepsilon](0) = 0$, on en déduit que $\mathbb{F}_{[0,2]}$ et $\mathbb{F}_{[1,3]}$ ne sont pas ε -entrelacés.

On en conclut que $d(\mathbb{F}_{[0,2]}, \mathbb{F}_{[1,3]}) = 1$. Par un raisonnement similaire, on peut vérifier par exemple que $d(\mathbb{F}_{[0,5]}, \mathbb{F}_{[4,7]}) = 2$ ou que $d(\mathbb{F}_{[0,2]}, \mathbb{F}_{[4,7]}) = 1.5$.

On peut généraliser les exemples précédents pour calculer « à la main » la distance entre deux modules de persistance indécomposables, et vérifier que le résultat dépend de la taille des intervalles et de la comparaison entre les points initiaux et finaux respectifs des intervalles. La théorie de structure 2.5 suggère alors de transférer la distance entre modules de persistance vers leurs code-barres. C'est ce que nous allons faire maintenant.

Soient A, B des (multi)-ensembles de barres, c'est-à-dire des multi-ensembles d'intervalles de \mathbb{R} . Si I est un intervalle, on notera $s_I \leq t_I$ ses extrémités.

DÉFINITION 3.9 (Correspondance partielle). — *Une correspondance partielle entre A et B , notée $A \xleftarrow{G} B$, est la donnée d'un sous-ensemble $G \subset A \times B$ tel que les projections naturelles $G \rightarrow A$ et $G \rightarrow B$ soient injectives.*

On dira qu'un élément $(I, J) \in G$ est une paire accouplée et a contrario qu'un élément $I \in A \amalg B$ qui ne correspond pas à une paire accouplée est un élément non accouplé.

À toute correspondance partielle $G \subset A \times B$, on associe son coût :

- si $(I, J) \in G$, le coût de (I, J) est $c(I, J) = \|(s_I, t_I) - (s_J, t_J)\|_\infty$;
- si $K \in A \amalg B$ est une barre non accouplée, le coût de K est $c(K) = \frac{|s_K - t_K|}{2}$.
- Le coût total de G est

$$c(G) := \max\left(c(I, J), c(K), \text{ sur } (I, J) \text{ paires accouplées, } K \text{ non accouplée}\right).$$

L'infimum du coût des accouplements partiels permet de définir la pseudo-distance suivante :

DÉFINITION 3.10 (Distance bottleneck). — Soient A, B des code-barres. La distance bottleneck entre A et B est

$$d_{\text{bn}}(A, B) := \inf_{A \xrightarrow{G} B} (c(G)),$$

la borne inférieure étant prise sur les correspondances partielles entre A et B .

Le théorème suivant permet de calculer la distance d'entrelacement en termes de la distance bottleneck.

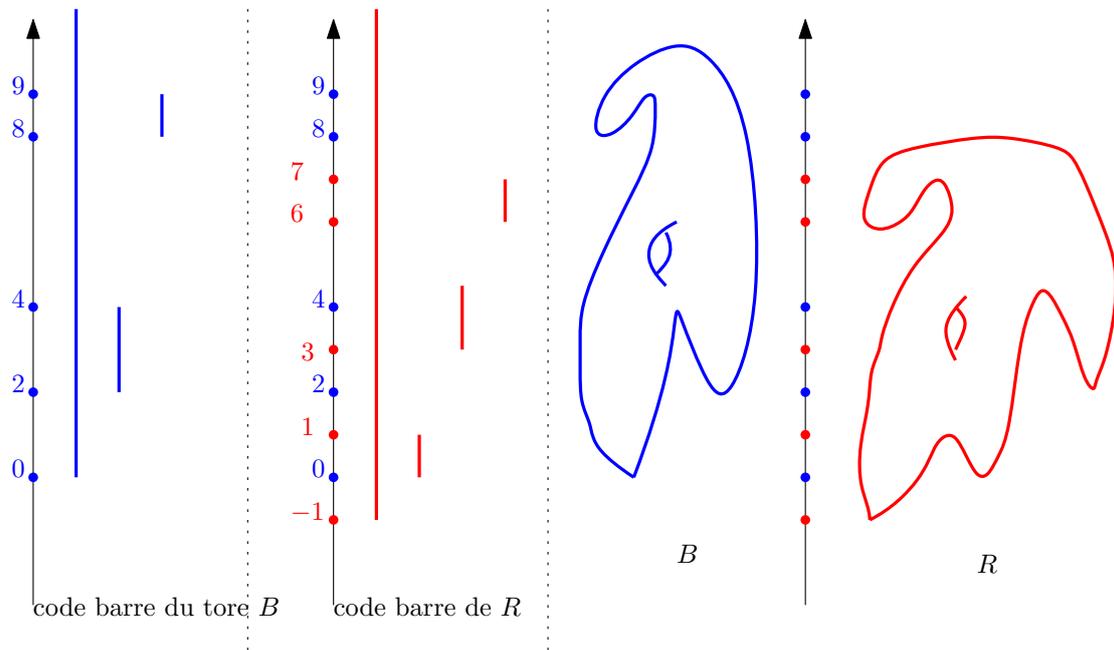
THÉORÈME 3.11 (Isométrie, Bauer et Lesnick, 2014). — Soient M, N deux modules de persistance ponctuellement de type fini ; alors on a l'égalité

$$d(M, N) = d_{\text{bn}}(\mathbf{B}(M), \mathbf{B}(N))$$

entre la distance d'entrelacement des modules et la distance bottleneck entre les code-barres associés.

Un point clé dans les applications est que la distance bottleneck peut se calculer algorithmiquement via des algorithmes efficaces d'appariement.

Exemple 3.12 (Projections de 2-tores). — On considère un tore (rouge) R et un tore (bleu) B avec deux projections différentes sur \mathbb{R} données par la fonction hauteur. On peut leur associer leurs groupes d'homologie en degré 0, comme décrit à gauche dans la figure ci-dessous.



On peut vérifier que la distance bottleneck entre les (code-barres en degré 0 des) deux tores est

$$d_{\text{bn}}(\mathbf{B}(H_0(R^{p_z})), \mathbf{B}(H_0(B^{p_z}))) = 1.$$

Évidemment, dans la pratique, on a en général un *très* grand nombre de barres à comparer.

Notons que la notion de coût (et donc de distance bottleneck) utilise de la norme infinie pour définir la distance entre des barres. On peut évidemment utiliser les normes p pour obtenir des distances très utilisées en pratique.

DÉFINITION 3.13 (distance de Wasserstein). — *Soit G une correspondance partielle et soit $p \in \mathbb{N}^*$. Le p -coût de Wasserstein de G est*

$$c_p(G) = \left(\sum_{(I,J) \in G} \|(s_I, t_I) - (s_J, t_J)\|_\infty^p + \sum_{K \text{ non accouplé}} \left(\frac{|s_K - t_K|}{2} \right)^p \right)^{1/p}.$$

La p -distance de Wasserstein entre deux code-barres \mathbb{A} , \mathbb{B} est

$$d_{p\text{-Was}}(\mathbb{A}, \mathbb{B}) := \inf_{\mathbb{A} \xleftrightarrow{G} \mathbb{B}} \left(c_p(G) \right),$$

la borne inférieure étant prise sur les correspondances partielles entre \mathbb{A} et \mathbb{B} .

En particulier, la distance bottleneck est la limite quand $p \rightarrow \infty$ des p -distances de Wasserstein. Contrairement à cette dernière, les p -distances de Wasserstein tiennent (aussi) compte du nombre de petites barres.

4. Stabilité

Une question élémentaire qui se pose est de savoir si la notion de distance d'entrelacement entre modules de persistance associée à des objets géométriques est un invariant de la géométrie, c'est-à-dire est « continue » en un sens à préciser.

La réponse est donnée par les théorèmes suivants, dits « de stabilité », qui justifient l'utilisation des modules de persistance dans les applications associées aux exemples 1.11 et 1.10.

On commence par les modules de persistance associés aux sous-niveaux d'une fonction.

On dira qu'une fonction $f: X \rightarrow \mathbb{R}$ est ponctuellement de dimension finie si son module de persistance associé $H_i(X^f)$ est ponctuellement de dimension finie.

Remarque 4.1. — C'est toujours le cas en pratique pour des fonctions associées à des données finies.

Notons que si $\|f - g\|_\infty \leq \varepsilon$, alors les espaces persistants X^f et X^g sont ε -entrelacés puisque $f(x) \leq t$ implique $g(x) \leq x + \varepsilon$ et symétriquement. Le lemme 3.5 et le théorème d'isométrie impliquent alors le résultat suivant.

THÉORÈME 4.2 (Théorème de Stabilité I). — *Soient $f, g: X \rightarrow \mathbb{R}$ des fonctions ponctuellement de dimension finie. On note $\mathbf{B}_n(f)$ et $\mathbf{B}_n(g)$ les code-barres associés à l'homologie persistante en degré n des sous-niveaux de f et g .*

Pour tout entier n , on a

$$d_{\text{bn}}(\mathbf{B}(f), \mathbf{B}(g)) \leq \|f - g\|_\infty.$$

Passons au cas des complexes de Čech 1.11, autrement dit aux modules de persistance associés à un nuage de points.

Rappelons que si K, L sont deux sous-espaces compacts d'un espace métrique (Y, d) , alors leur distance de Hausdorff est

$$d_{\mathcal{H}}(K, L) = \max\left(\max_{x \in K} (d(x, L)), \max_{y \in L} (d(y, K))\right).$$

Si maintenant K et L sont des espaces métriques compacts abstraits, on a la généralisation suivante. Appelons *plongement isométrique* $(X, d) \rightarrow (Y, d')$ une application continue entre espaces métriques qui est un homéomorphisme sur son image et une isométrie.

DÉFINITION 4.3. — *Soit K, L deux espaces métriques compacts. Leur distance de Gromov–Hausdorff est*

$$d_{\mathcal{GH}}(K, L) = \inf_{Z, \gamma_K, \gamma_L} d_{\mathcal{H}}(\gamma_K(K), \gamma_L(L))$$

où γ_K, γ_L sont des plongements isométriques de K et L dans un même espace métrique Z .

La distance de Gromov–Hausdorff que nous donnons ici diffère de certaines conventions par un facteur de 2. Le deuxième théorème de stabilité énonce :

THÉORÈME 4.4 (Théorème de Stabilité II : Chazal, De Silva et Oudot, 2014)

Soient \mathbb{X}, \mathbb{Y} deux espaces métriques compacts. On note $\mathbf{B}_n(\mathbb{X})$ et $\mathbf{B}_n(\mathbb{Y})$ les code-barres associées à l’homologie persistante en degré n des complexes de Čech associés.

Pour tout entier n , on a

$$d_{\text{bn}}(\mathbf{B}(\mathbb{X}), \mathbf{B}(\mathbb{Y})) \leq d_{\mathcal{GH}}(\mathbb{X}, \mathbb{Y}).$$

Les théorèmes de stabilité suggèrent que les petites barres dans un code-barre représentent du bruit, au sens où les changements de topologie (manifestés par la distance de la convergence uniforme ou la distance de Gromov–Hausdorff) ne sont perceptibles que par les grandes barres *a priori*.

Les théorèmes de stabilité ne s’étendent pas *stricto sensu* aux distances de Wasserstein, mais il existe des variantes.

Remarque 4.5. — Le deuxième théorème de stabilité est vrai pour les complexes de Rips 1.13 et toutes les variantes utilisées à la place de Čech également.

5. Persistance « supérieure »

Un enjeu *théorique* mais aussi *pratique* de la théorie de la persistance est dans les versions avec plus de structure. Ici, on veut évoquer deux possibilités, et nous développerons succinctement seulement la deuxième.

- D’une part, on peut raffiner l’homologie en introduisant des structures homotopiques supérieures. Plus précisément, la problématique est que de nombreux espaces ne peuvent pas être discriminés par leurs simples groupes d’homologie. Même si les modules persistants sont en pratique bien plus efficaces, il existe des discrétisations naturelles d’espaces non homotopes qui ne sont pas discernables via leurs code-barres. En topologie algébrique, il est bien connu que l’on peut raffiner les groupes d’homologie en rajoutant des structures algébriques (éventuellement à homotopie près) supplémentaires. Par exemple la structure d’algèbre graduée de la cohomologie, les structures A_∞ , de modules sur l’algèbre de Steenrod... On peut raffiner l’homologie persistante pour travailler dans des catégories d’objets munis de ces structures. On peut même considérer un type d’homotopie persistant pour les espaces topologiques persistants. Évidemment, dans ce cadre les distances d’entrelacement sont beaucoup plus difficiles à calculer, mais raffinent celles des modules persistants sous-jacents. Une large littérature et implémentation d’algorithmes est consacrée à ces développements.

- Les approches ci-dessus permettent de considérer les modules de persistance $CH_i(\mathbb{X})$ en considérant tous les degrés en même temps et autorisent des interactions non triviales entre degrés différents qui discriminent davantage les distances. Une variante consiste à interpréter de manière faisceautique les modules de persistance, cf. Kashiwara et Schapira (2018) et Berkouk et Petit (2021).
- On peut considérer des modules de persistance à *plusieurs* paramètres. Ceci n'est pas seulement motivé par une question théorique évidente mais aussi par de nombreuses motivations pratiques en sciences expérimentales. Il est en effet parfois important de considérer des objets filtrés par *plusieurs* paramètres qui peuvent venir de plusieurs données étudiées simultanément (il y a des exemples pratiques en physique, astronomie ou biologie notamment). Un autre cas pratique très important est donné par l'étude des sous-niveaux de la fonction distance $d_X: \mathbb{R}^n \rightarrow \mathbb{R}$ à un espace X (cf. les applications à l'inférence dans la section 6) couplée à une fonction densité $g: \mathbb{R}^d \rightarrow \mathbb{R}$ estimée à partir de X qui permet de réduire l'effet de possibles points aberrants. L'idée est précisément ici de filtrer l'espace \mathbb{R}^n par le couple (d_X, g) . Un dernier exemple naturel est donné par les niveaux d'une fonction comme dans l'exemple 5.3. Dans tous les cas, on obtient des espaces *persistants multiparamétriques* dont les groupes d'homologie ont une structure persistante à plusieurs paramètres.

Nous allons évoquer un peu plus en détail la problématique des modules de persistance à plusieurs paramètres.

Le produit $S \times T$ d'ensembles partiellement ordonnés $(S, <_S)$, $(T, <_T)$ hérite d'un ordre naturel donné par « $(s, t) <_{S \times T} (s', t')$ si et seulement si $s <_S s'$ et $t <_T t'$ ». Ce produit n'est que partiellement ordonné même si S et T le sont totalement (sauf cas trivial).

DÉFINITION 5.1 (Persistance à d paramètres). — *La catégorie des objets de persistance à d paramètres à valeurs dans une catégorie \mathcal{C} est la catégorie des foncteurs $(\mathbb{R}, <)^d \rightarrow \mathcal{C}$ où $(\mathbb{R}, <)^d$ est la catégorie associée à \mathbb{R}^d muni de la structure d'ordre produit.*

Pour $\mathcal{C} = \mathbb{F}\text{-Vect}$, on dira simplement module de persistance à d paramètres ou module de d -persistance.

On définit de même les notions d'objets de d -persistance indexés par \mathbb{Z}^d (ou un produit $I_1 \times \cdots \times I_d$ de sous-ensembles de \mathbb{R}) plutôt que par \mathbb{R}^d .

La distance d'entrelacement pour les modules de persistance à d paramètres se définit de manière analogue au cas à 1 paramètre. On remarque que pour tout vecteur $\alpha \in \mathbb{R}_+^d$ à coordonnées positives, on a un endofoncteur de décalage $F \mapsto F[\alpha]$ des objets de d -persistance où

$$F[\alpha]\left((t_1, \dots, t_d)\right) = F(t_1 + \alpha_1, \dots, t_d + \alpha_d)$$

est défini comme le décalage de la définition 3.1 (appliqué dans chaque coordonnée).

DÉFINITION 5.2. — Soient $F, G: (\mathbb{R}, <)^d \rightarrow \mathcal{C}$ des objets persistants dans \mathcal{C} et soit $\varepsilon \geq 0$. On dit que F et G sont ε -entrelacés s'il existe deux morphismes de modules de persistance

$$\mu: F \longrightarrow G[\vec{\varepsilon}] \quad \text{et} \quad \nu: G \longrightarrow F[\vec{\varepsilon}],$$

où $\vec{\varepsilon} = (\varepsilon, \dots, \varepsilon)$, tels que les diagrammes suivants

$$\begin{array}{ccc} & \xrightarrow{\tau^F[2\vec{\varepsilon}]} & \\ & \curvearrowright & \\ F & \xrightarrow{\mu} G[\vec{\varepsilon}] \xrightarrow{\nu[\vec{\varepsilon}]} & F[2\vec{\varepsilon}] \\ & & \\ G & \xrightarrow{\nu} F[\vec{\varepsilon}] \xrightarrow{\mu[\vec{\varepsilon}]} & G[2\vec{\varepsilon}] \\ & \curvearrowleft & \\ & \xrightarrow{\tau^G[2\vec{\varepsilon}]} & \end{array}$$

soient commutatifs. La distance d'entrelacement entre F et G est l'infimum

$$d(F, G) := \inf \{ \varepsilon \geq 0 \mid F \text{ et } G \text{ sont } \varepsilon\text{-entrelacés} \} \in \mathbb{R}_+ \cup \{ \infty \}.$$

Exemple 5.3 (Niveaux d'une fonction). — Soit $f: X \rightarrow \mathbb{R}$ une fonction. On peut raffiner l'étude de ses sous-niveaux en considérant plutôt les *niveaux*. Précisément, pour tout $(x, y) \in \mathbb{R}^2$, on pose

$$N_X^f((x, y)) = \begin{cases} f^{-1}(] - x, y]) & \text{si } x + y > 0 \\ \emptyset & \text{sinon.} \end{cases}$$

On a alors que, pour $x' + y' > 0$, $(x, y) \leq (x', y')$ est équivalent à $] - x, y[\subset] - x', y'[$. En particulier, $N_X^f((x, y)) \subset N_X^f((x', y'))$ lorsque $(x, y) \leq (x', y')$, ce qui fait de N_X^f un espace topologique de 2-persistance. En prenant ses groupes d'homologie on obtient donc des modules de persistance à 2 paramètres

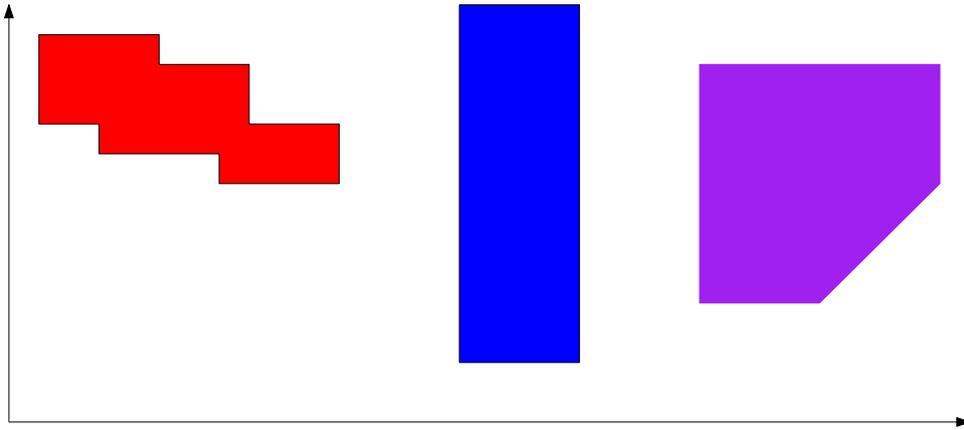
$$\left(H_i(N_X^f), H_i\left(f^{-1}(] - x, y[) \hookrightarrow f^{-1}(] - x', y'[) \right) \right)_{(x, y) \leq (x', y')}$$

dont les morphismes de structure sont donnés par l'application du foncteur homologie aux inclusions canoniques. On l'appelle *module de 2-persistance des niveaux* de f . Il encode toutes les informations locales de la topologie de X vue au travers de f .

Les modules de persistance associés aux niveaux d'une fonction sont par définition nuls en tout point $(x, y) \in \mathbb{R}^2$ qui est situé en-dessous (au sens large) de l'antidiagonale $\Delta = \{(-x, x), x \in \mathbb{R}\}$.

On notera $\text{Per}(\Delta_+)$ la sous-catégorie des modules de persistance de $(\mathbb{R}, <)^2$ qui valent 0 en tout point sur ou en-dessous de l'antidiagonale.

Par ailleurs, le théorème de stabilité 4.2 s'adapte de manière similaire au cas des sous-niveaux :

FIGURE 1. Des exemples d'intervalles de $(\mathbb{R}, >)^2$

PROPOSITION 5.4. — Soient $f, g: X \rightarrow \mathbb{R}$ des fonctions ponctuellement de dimension finie. On note $\mathbf{B}_i(N_X^f)$ et $\mathbf{B}_i(N_X^g)$ les code-barres associés aux modules de persistance à 2 paramètres en degré i des niveaux de f et g . Alors, on a

$$d(\mathbf{B}_i(N_X^f), \mathbf{B}_i(N_X^g)) \leq \|f - g\|_\infty.$$

Exemple 5.5. — Les groupes d'homologie d'un complexe simplicial bifiltré $(F_{p,q})_{p,q \in \mathbb{N}}$ ont naturellement une structure de module de persistance à 2 paramètres dont les morphismes de structure sont induits par le passage à l'homologie dans les inclusions $F_{p,q} \hookrightarrow F_{p',q} \hookrightarrow F_{p',q'}$ si $(p, q) \leq (p', q')$.

Exemple 5.6. — Soient $f, g: X \rightarrow \mathbb{R}$ deux fonctions continues. On a alors un espace topologique de persistance à 2 paramètres

$$X^{f,g}(s, t) = \{x \in X, f(x) \leq s \text{ et } g(x) \leq t\}$$

dont les groupes d'homologie fournissent un module de persistance à 2 paramètres.

Une problématique majeure qui survient pour la persistance à plusieurs paramètres est que l'ordre n'est plus total et en particulier le théorème de structure 2.12 ne s'applique pas.

On dispose encore de *modules intervalles* comme dans la définition 2.10, mais ils ne représentent pas tous les indécomposables et on ne dispose pas de classification élégante de ces derniers. La problématique n'est pas sans rappeler celle du passage de l'étude des polynômes à une variable à l'étude des polynômes à plusieurs variables. Et on sait que dès que l'on a au moins deux paramètres, la classification des indécomposables est *sauvage* au sens de la théorie des représentations, voir Botnan et Lesnick (2023).

En particulier, les indécomposables ne sont pas caractérisés par leur support, comme c'est le cas pour les modules de persistance ponctuellement de rang fini modélés sur un ordre total.

Exemple 5.7. — On considère le module de persistance M sur $\{0, 1, 2\} \times \{0, 1\} \subset \mathbb{R}^2$ donné par le diagramme suivant, où les flèches représentent les morphismes de structure $\iota_{(a,x) \leq (b,y)}^M$:

$$\begin{array}{ccccc} \mathbb{F} & \xrightarrow{\delta} & \mathbb{F} \oplus \mathbb{F} & \xrightarrow{\pi_1} & \mathbb{F} \\ \uparrow & & \uparrow i_1 & & \uparrow \text{id} \\ 0 & \longrightarrow & \mathbb{F} & \xrightarrow{\text{id}} & \mathbb{F} \end{array}$$

où δ est le morphisme diagonal $\lambda \mapsto (\lambda, \lambda)$, π_1 et i_1 sont respectivement la projection et l'inclusion dans le premier facteur de $\mathbb{F} \oplus \mathbb{F}$.

Le module M est *indécomposable* de support l'intervalle $I = \{(0, 1), (1, 1), (2, 1), (1, 0), (2, 0)\}$. Le module intervalle \mathbb{F}_I a par définition le même support mais n'est pas isomorphe à M (comme on le voit immédiatement en considérant leurs valeurs en $(1, 1)$).

Si la classification complète est hors de portée, on peut cependant se poser la question de l'existence de décompositions simples pour certaines classes de modules persistants multi-paramétriques, et de celle de déterminer le cas échéant cette décomposition.

Soit \mathcal{I} une famille d'intervalles de \mathbb{R}^d . On dira qu'un module de persistance à d paramètres M admet une décomposition en \mathcal{I} -modules intervalles si $M = \bigoplus_{\mathbf{B}(M)} \mathbb{F}_I$ où $\mathbf{B}(M)$ est un multi-ensemble d'intervalles de \mathcal{I} et que cette décomposition est unique à isomorphisme et ré-indexage près.

Voici une famille d'exemples dans \mathbb{R}^2 .

Remarque 5.8. — Dans la suite de cette section, ce que nous faisons avec \mathbb{R}^2 peut être généralisé assez facilement au produit de deux ensembles totalement ordonnés.

On commence par définir une famille simple d'intervalles : les rectangles.

DÉFINITION 5.9 (Rectangles et blocs). — Un rectangle de $(\mathbb{R}, >)^2$ est un produit $I \times J$ où I, J sont des intervalles de \mathbb{R} . On notera $\mathcal{R}\text{ec}$ l'ensemble des rectangles et on appellera module rectangle un module \mathbb{F}_R où R est un rectangle.

Un bloc est un rectangle $I \times J$ d'une des formes suivantes :

- bloc naissant : I et J ont $+\infty$ comme extrémité ;
- bloc mourant : I et J ont $-\infty$ comme extrémité ;
- bloc horizontal : si $I = \mathbb{R}$;
- bloc vertical : si $J = \mathbb{R}$.

On notera $\mathcal{B}\text{loc}$ l'ensemble des blocs et on appellera module bloc un module \mathbb{F}_S où S est un bloc.

On étend les notions précédentes au cas de $\text{Per}(\Delta_+)$, les modules de persistance nuls en-dessous de la diagonale, et les modules blocs et rectangles sont étendus à ce cadre en modifiant $\mathbb{F}_R((x, y))$ de sorte qu'il vaille 0 si $x + y \leq 0$. On notera $\mathcal{B}\text{loc}_+$ l'ensemble des blocs qui intersectent le demi-plan au-dessus de l'anti-diagonale, c'est-à-dire dont les modules blocs associés sont non nuls.

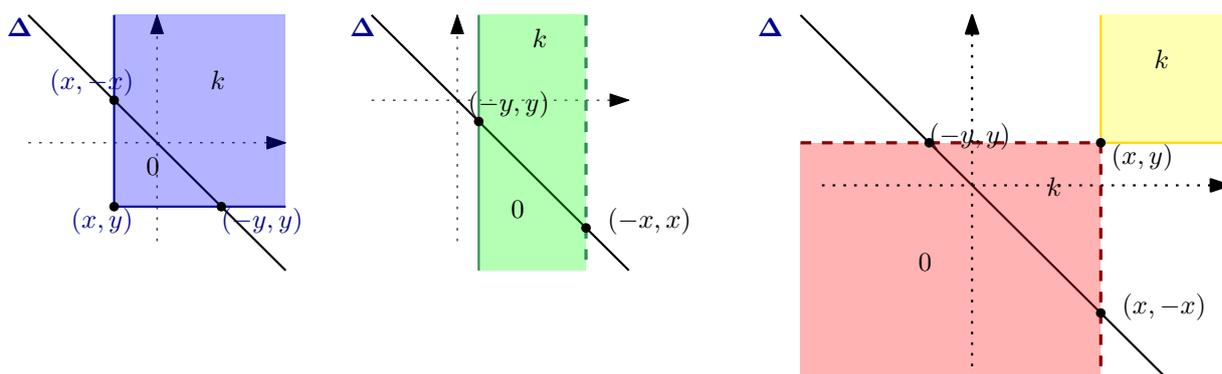


FIGURE 2. Des blocs (et leur modules intervalles associés sur Δ_+) : un bloc naissant à gauche, un bloc mourant à droite (en rouge) et son bloc naissant dual (en jaune) et un bloc vertical au milieu.

On notera que \mathbb{R}^2 (et quelques autres cas particuliers) sont des blocs plusieurs types différents, mais cela ne posera pas de problème particulier.

Remarque 5.10. — Il y a une dualité entre les blocs naissants et les blocs mourants (distincts de \mathbb{R}^2) : le dual d’un bloc naissant de la forme $]a, +\infty[\times]b, +\infty[$ est le bloc mourant $] - \infty, a] \times] - \infty, b]$ et plus généralement à $\langle a, +\infty[\times \langle b, +\infty[$ on associe le bloc mourant $] - \infty, a \rangle^{op} \times] - \infty, b \rangle^{op}$ où la notation \rangle^{op} signifie que l’on inverse fermé et ouvert : $]^{op} = [$ et $[^{op} =]$. On notera B^\vee le bloc dual d’un bloc naissant (resp. mourant) ; on a $(B^\vee)^\vee = B$.

Si \mathcal{I} est une famille d’intervalles de \mathbb{R}^2 , on peut restreindre cette famille à une sous-partie de \mathbb{R}^2 . En particulier, on appelle grille carrée de \mathbb{R}^2 la donnée de 4 points de \mathbb{R}^2 comme ci-dessous :

$$\begin{array}{ccc} (a, y) & \xrightarrow{\leq} & (x, y) . \\ \leq \uparrow & & \uparrow \leq \\ (a, b) & \xrightarrow{\leq} & (b, x) \end{array}$$

Pour tout module de persistance M sur \mathbb{R}^2 et tout carré comme ci-dessus on obtient un diagramme commutatif

$$\begin{array}{ccc} M(a, y) & \xrightarrow{\iota_{(a,y) \leq (x,y)}^M} & M(x, y) \\ \iota_{(a,b) \leq (a,y)}^M \uparrow & & \uparrow \iota_{(b,x) \leq (x,y)}^M \\ M(a, b) & \xrightarrow{\iota_{(a,b) \leq (b,x)}^M} & (b, x) . \end{array}$$

La notion de décomposition en rectangles ou blocs se restreint aux grilles carrées (et y a des formes très simples laissées au lecteur). On a les théorèmes de type local vers global remarquables suivants :

THÉORÈME 5.11 (Botnan et Crawley-Boevey, 2020 ; Cochoy et Oudot, 2020)

Un module de persistance ponctuellement de rang fini sur \mathbb{R}^2 admet une décomposition en blocs si et seulement si sa restriction à toute grille carrée admet une décomposition en blocs.

THÉORÈME 5.12 (Botnan, Lebovici et Oudot, 2022). — *Un module de persistance ponctuellement de rang fini sur \mathbb{R}^2 admet une décomposition en rectangles si et seulement si sa restriction à toute grille carrée admet une décomposition en rectangles.*

Le cas des blocs et rectangles dans ces théorèmes est spécifique. On peut montrer que les rectangles sont maximaux pour la propriété d’être testable localement sur les carrés, voir Botnan, Lebovici et Oudot (2022).

Ces résultats permettent de donner des algorithmes pour tester si un module de persistance admet une décomposition en blocs ou rectangles et le cas échéant de donner la décomposition. Par exemple

PROPOSITION 5.13 (Botnan, Lebovici et Oudot, 2022). — *Il existe un algorithme en temps $O(n^{\omega+2})$ qui teste si un module de persistance associé à une bifiltration d’un complexe simplicial avec N simplexes est décomposable en rectangles. Ici ω est l’exposant du coût de la multiplication matricielle. (On sait que $\omega \leq 2.373$.)*

Les modules de persistance associés aux niveaux d’une fonction continue (cf. l’exemple 5.3) forment une classe de modules de persistance admettant une décomposition spécifique en blocs. On rappelle que la notion de bloc dual est définie dans la remarque 5.10.

PROPOSITION 5.14 (Décomposition des niveaux, Berkouk, Ginot et Oudot, 2019)

Soit $f : X \rightarrow \mathbb{R}$ une fonction continue, dont les modules persistants associés $H_i(N_X^f)$ sont ponctuellement de type fini.

Alors, pour tout $i \in \mathbb{N}$, $H_i(N_X^f)$ est décomposable en blocs (non inclus sous l’anti-diagonale). Plus précisément, on a des décompositions uniques (à isomorphisme et ré-indexage près)

$$H_i(N_X^f) = \bigoplus_{B \in \mathbf{B}_i(N_X^f)} \mathbb{F}_B$$

où les $\mathbf{B}_i(N_X^f) \subset \mathcal{B}\text{loc}_+$ sont des multi-ensembles de blocs vérifiant que pour tout bloc mourant $B \in \mathbf{B}_i(N_X^f)$ alors son bloc naissant dual B^\vee est dans $\mathbf{B}_{i+1}(N_X^f)$ et, réciproquement, pour tout bloc naissant \tilde{B} dans $\mathbf{B}_j(N_X^f)$ dont le bloc dual \tilde{B}^\vee est dans $\mathcal{B}\text{loc}_+$ alors ce dual est dans $\mathbf{B}_{j-1}(N_X^f)$.

La proposition précédente montre que non seulement les niveaux de fonctions admettent des décompositions en blocs simples, mais que l’on observe une interaction globale entre les différents degrés des modules de persistance. Autrement dit, les blocs mourants et leurs blocs naissants duaux apparaissent en paires avec des degrés décalés de 1. Ce phénomène provient du fait que les modules de persistance associés aux niveaux

d'une fonction ont une interprétation cofaisceautique qui produit une interaction de type suite exacte de Mayer-Vietoris entre leurs évaluations en les points le plus petit et le plus grand d'une grille carrée.

Exemple 5.15. — On considère le cercle unité $S^1 = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 = 1\}$ et les fonctions $p, q: S^1 \rightarrow \mathbb{R}$ données respectivement par la projection verticale $p: (x, y) \mapsto x$ et l'application constante $q: (x, y) \mapsto 0$.

On peut alors vérifier que l'on a les isomorphismes

$$H_i \left(q^{-1} \left(] - x, y[\right) \right) = \mathbb{F} \text{ si } -x < 0 < y \text{ et } i = 0, 1, \text{ et } 0 \text{ sinon,}$$

alors que $H_i \left(p^{-1} \left(] - x, y[\right) \right) = 0$ si $] - x, y[\cap [-1, 1] = \emptyset$ et $H_{\geq 2} \left(p^{-1} \left(] - x, y[\right) \right) = 0$. Enfin on a

$$H_1 \left(p^{-1} \left(] - x, y[\right) \right) = \mathbb{F} \quad \text{si } (-x, y) \supset (-1, 1), \text{ et } 0 \text{ sinon ;}$$

$$H_0 \left(p^{-1} \left(] - x, y[\right) \right) = \mathbb{F}^2 \quad \text{si }] - x, y[\subset] - 1, 1[$$

et

$$H_0 \left(p^{-1} \left(] - x, y[\right) \right) = k \quad \text{si }] - x, y[\cap [-1, 1] \neq \emptyset \text{ mais }] - x, y[\not\subset] - 1, 1[$$

On peut en déduire la décomposition en blocs $H_0(N_{S^1}^p) = \mathbb{F}_R \oplus \mathbb{F}_B$ et $H_1(N_{S^1}^p) = \mathbb{F}_O$ où R , B et O sont les blocs rouge, bleu et orange de la figure 3. On notera que les blocs orange et rouge sont duaux. Enfin $H_i(N_{S^1}^q) = \mathbb{F}_J$ est donné par les blocs jaunes de la figure 3 en degrés 0 et 1 (qui sont des blocs naissants dont le dual n'est pas dans $\mathcal{B}loc_+$).

En utilisant un raisonnement comme dans l'exemple 3.6 on peut montrer que, pour $i = 0, 1$, les modules de persistance à 2 paramètres $H_i(N_{S^1}^p)$ et $H_i(N_{S^1}^q)$ ne sont pas ε -entrelacés pour $\varepsilon < 1$. La propriété de stabilité 5.4 implique alors que les distances d'entrelacement entre ces modules sont données, pour $i = 0, 1$, par

$$d(H_i(N_{S^1}^p), H_i(N_{S^1}^q)) = 1.$$

6. Applications

6.1. Inférence

Les modules de persistance ont initialement été développés pour leurs applications en analyse topologique des données, et en particulier autour du problème de l'*inférence*. La problématique est, étant donné un nuage de points \mathbb{X} (c'est-à-dire un sous-ensemble discret d'un espace métrique (Y, d)), de déterminer les propriétés topologico-géométriques de l'espace géométrique dont il est une approximation. Plus généralement, l'analyse topologique des données étudie les structures topologiques sous-jacentes à des données.

Les théorèmes de stabilité de la partie 4 permettent de construire des estimateurs consistants en statistiques à partir d'échantillonnages qui se sont révélés très utiles en pratique. Ils sont en partie à la base des applications en sciences dures de la persistance.

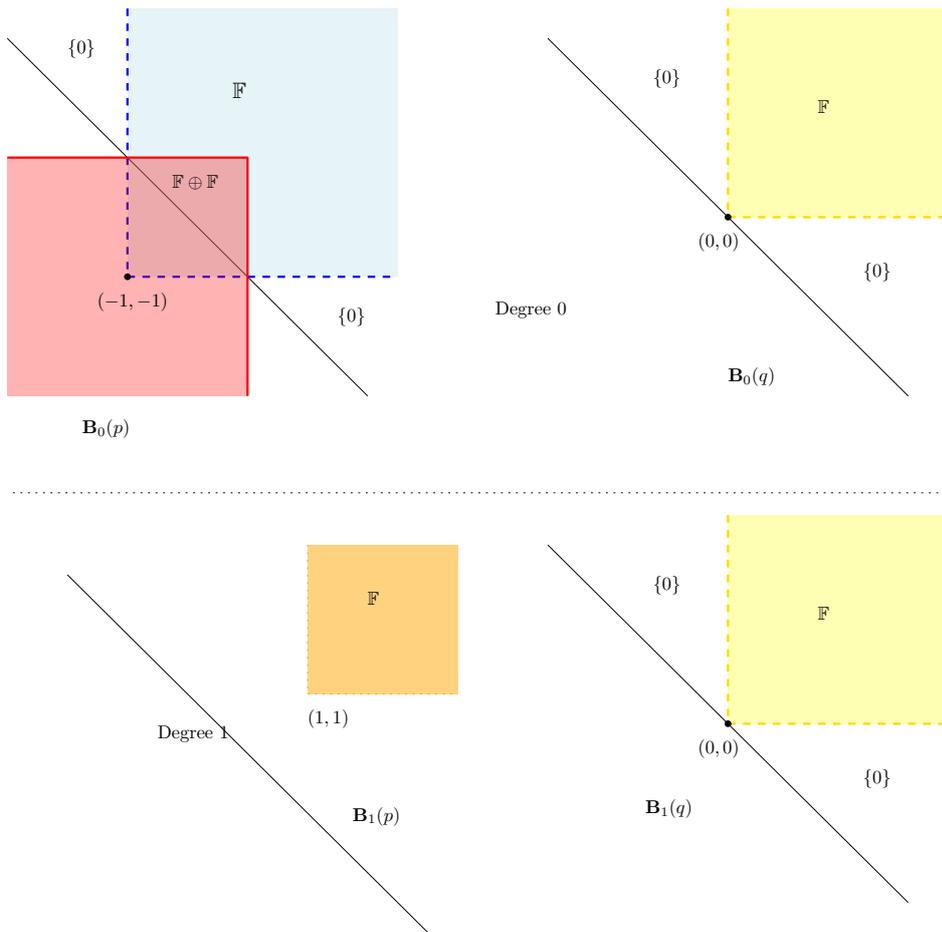


FIGURE 3. À gauche les blocs de $H_i(N_{S^1}^p)$ en degré 0 et 1 et à droite ceux de $H_i(N_{S^1}^q)$.

Expliquons maintenant la problématique de l'inférence en pratique. Soit X un sous-espace compact d'un espace métrique Y . On note

$$d_X: \begin{cases} Y & \longrightarrow \mathbb{R} \\ y & \longmapsto d_X(y) = \inf_{x \in X} d(y, x) \end{cases}$$

la fonction distance à X .

DÉFINITION 6.1. — Pour tout $i \in \mathbb{N}$, on note $\mathcal{HP}_i(X)$ le module de persistance $H_i(X^{d_X}, \mathbb{F})$ associé à l'espace topologique persistant $X^{d_X}: t \mapsto d_X^{-1}(] - \infty, t])$ (cf. l'exemple 1.10).

En particulier, on a $\mathcal{HP}_i(X)(0) = H_i(X)$.

Soit maintenant \mathbb{X} un sous-espace fini de X que l'on veut voir comme une approximation discrète de X . Les théorèmes de stabilité 4.4 nous donnent

$$d(\mathcal{HP}_i(X), \mathcal{HP}_i(\mathbb{X})) \leq d_{GH}(X, \mathbb{X}).$$

En particulier, si \mathbb{X} est proche de X , alors les modules de persistance associés sont proches et en particulier leurs code-barres sont proches.

Dans les applications, on préfère remplacer la fonction $d_{\mathbb{X}}$ par un modèle purement combinatoire qui est implémentable sur un ordinateur. Ceci est possible grâce au corollaire du théorème du nerf (proposition 1.12), via le complexe de Čech de \mathbb{X} . On déduit de ce corollaire :

LEMME 6.2. — *Il existe des isomorphismes de modules de persistance $\mathcal{HP}_i(\mathbb{X}) \cong H_iC(\mathbb{X})$ et en particulier*

$$d(\mathcal{HP}_i(X), H_iC(\mathbb{X})) \leq d_{\mathcal{GH}}(X, \mathbb{X}).$$

Ainsi, le module de persistance associé à \mathbb{X} est complètement calculé par $H_iC(\mathbb{X})$, qui est un module de persistance associé à un complexe cellulaire fini et donc « efficacement » implémentable (et implémenté) sur machine.

Si la distance de Gromov–Hausdorff de \mathbb{X} est proche de celle de X , alors les code-barres de $H_iC(\mathbb{X})$ sont proches de ceux de X et permettent donc d’approximer les invariants topologiques de X .

Pour simplifier, on suppose maintenant que $Y = \mathbb{R}^d$, et on suppose toujours que X est un sous-ensemble compact quelconque de Y . Commençons par définir, pour tout $z \in \mathbb{R}^d$,

$$(3) \quad L_X^{\min}(z) := \{x \in X, d_X(z) = \|x - z\|\}$$

le sous-espace (non vide) de X des points réalisant la distance de z à X . Évidemment, $L_X^{\min}(z)$ est un singleton si $z \in X$; de manière plus générale, si $L_X^{\min}(z)$ est un singleton, alors son unique élément est le projeté (orthogonal) de z sur X .

DÉFINITION 6.3 (Axe médian et amplitude). — *L’axe médian, noté M_X , de X est l’ensemble des $z \in \mathbb{R}^d$ tels que $\text{card}(L_X^{\min}(z)) > 1$.*

L’amplitude de X est

$$\text{rch}(X) := \inf_{x \in X, z \in M_X} (\|x - z\|).$$

Par définition, on a une application bien définie $p_X: \mathbb{R}^d \setminus M_X \rightarrow X$ qui à un point z associe le point $p_X(z)$ réalisant la distance $d_X(z)$ de z à X . De plus, cette application est continue.

Lorsque X est une sous-variété lisse, l’amplitude est non nulle en vertu du lemme suivant.

LEMME 6.4. — *Si X est une sous-variété C^1 et est compacte, alors $\text{rch}(X) > 0$.*

Exemple 6.5. — Si X est convexe dans \mathbb{R}^d , alors $M_X = \emptyset$ (et $\text{rch}(X) = \infty$).

— Si X est une sphère S de rayon R , alors M_S est son centre et $\text{rch}(S) = R$.

— Si X est un tore de révolution T obtenu comme surface de révolution de rayon R d’un cercle de rayon $r < R$, alors M_T est l’âme du tore (c’est-à-dire le cercle de rayon R orthogonal à l’axe) et $\text{rch}(T) = r$.

Notons que si X est une variété compacte dans \mathbb{R}^d telle que $\text{rch}(X) > 0$, alors, pour tout $t \in [0, \text{rch}(X))$, le t -ème sous-niveau $d_X^{-1}([-\infty, t])$ est homotope à X et plus précisément l'inclusion $X \subset d_X^{-1}([-\infty, t])$ est un rétracte par déformation.

L'invariance homotopique de l'homologie nous donne alors que, pour tous $0 \leq s \leq t < \text{rch}(X)$, l'application $HP_i(X)(s) \rightarrow HP_i(X)(t)$ est l'identité. On en déduit que, en tout degré $i \in \mathbb{N}$, le code-barre de X a précisément $\beta_i(X) = \dim(H_i(X))$ barres contenant l'intervalle $[0, \text{rch}(X)[$. Le lemme 6.2 permet alors de démontrer le théorème suivant.

THÉORÈME 6.6 (Chazal, Silva, Glisse et Oudot, 2016). — *Soient X un sous-espace compact de \mathbb{R}^d avec $\text{rch}(X) > 0$ et \mathbb{X} un sous-ensemble fini de X tel que $d_{G\mathcal{H}}(X, \mathbb{X}) < \varepsilon < \frac{\text{rch}(X)}{2}$. Alors, pour tout entier $i \in \mathbb{N}$, l'ensemble des barres du code-barre $\mathbf{B}_i(C(\mathbb{X}))$ qui traversent l'intervalle $[\varepsilon, \text{rch}(X) - \varepsilon]$ est égal à la dimension de $H_n(X)$. Toutes les barres de $\mathbf{B}(\mathbb{X})$ qui sont dans $[\varepsilon, \text{rch}(X) - \varepsilon]$ sont de longueur strictement plus petite que 2ε .*

Ce résultat exprime que lorsque la discrétisation est proche de l'espace X , alors on peut retrouver les invariants de X à partir du code-barre (calculable algorithmiquement). Évidemment, en pratique, on ne connaît pas l'amplitude de X .

Plus généralement, étant données des données abstraites, les résultats précédents suggèrent que si les code-barres de deux jeux de données sont à des distances suffisamment proches l'un de l'autre, c'est que les données ont des propriétés topologiques proches. Ceci peut être utilisé pour déterminer si deux données sont proches ou remarquer des spécificités. Mais évidemment, là aussi, déterminer ce que veut dire «proche» n'est pas évident a priori. Les barres en degré i témoignent souvent d'interactions complexes entre les paramètres/données.

Dans les applications pratiques, il faut faire des analyses statistiques qui peuvent dépendre d'une connaissance/collaboration avec des experts qui permet d'avoir des évaluations de bons ε . Ou bien on utilise des noyaux et des algorithmes d'apprentissage pour déterminer des groupes de données proches ou non.

Remarque 6.7. — Dans les applications pratiques, on utilise souvent les distances de Wasserstein à la place de la distance d'entrelacement/bottleneck. Bien que l'on ne dispose pas de théorèmes de stabilité généraux dans ce cas-là, les distances de Wasserstein tiennent compte du nombre de petites barres à la différence de la distance d'entrelacement, et en pratique discriminent parfois mieux (sans nécessairement que l'on ne comprenne bien ce que l'on calcule).

En termes de statistiques, on peut donner un intervalle de confiance pour la distance de Hausdorff entre une discrétisation et un espace topologique, et ainsi appliquer le théorème 6.6.

Par exemple, soit μ une mesure de probabilité de support un compact $X \subset \mathbb{R}^d$ qui est (a, b) -standard, où $a, b > 0$; ceci signifie que, pour tous $x \in X$ et $r > 0$, on a

$$\mu(B(x, r)) \leq \min(1, ar^b).$$

On a alors le résultat suivant, qui donne un intervalle de confiance pour les code-barres (via la stabilité).

PROPOSITION 6.8. — *Si \mathbb{X} est une discrétisation à N points de X et $\varepsilon > 0$, alors la probabilité $\mathbb{P}(d_{\mathcal{H}}(X, \mathbb{X}) > \varepsilon)$ que la distance entre X et \mathbb{X} soit plus grande que ε admet la majoration suivante :*

$$\mathbb{P}(d_{\mathcal{H}}(X, \mathbb{X}) > \varepsilon) \leq \frac{4^b}{a\varepsilon^b} \exp\left(-N \frac{a\varepsilon^b}{2^b}\right).$$

Remarque 6.9. — Ces notions d’inférence et plus généralement la stabilité des code-barres ont de nombreuses applications, on renvoie aux références suivantes :

- Bobrowski, Mukherjee et Taylor (2017), pour des exemples d’applications à la construction d’estimateurs consistants ;
- Aukerman, Carrière, Chen, Gardner, Rabadán et Vanguri (2021), pour un (parmi de nombreux) exemple d’applications à la biologie médicale, en l’occurrence ici l’étude de tissus cancéreux et de prédictions d’évolution de tumeurs ;
- Lee, Barthel, Dłotko, Moosavi, Hess et Smit (2017), pour un (parmi une multitude) exemple d’applications en science des matériaux ;

La persistance à 2 paramètres a également trouvé des applications concrètes. On renvoie par exemple à Carrière et Blumberg (2020) pour des exemples en lien avec l’apprentissage.

6.2. Géométrie

Nous allons maintenant brièvement mentionner des applications géométriques récentes de la persistance déduites une nouvelle fois des propriétés de stabilité, liées à des comptes grossiers de valeurs critiques.

Un premier exemple est en géométrie analytique affine. Soit $f: \mathbb{C}^n \rightarrow \mathbb{C}^n$ une fonction analytique. Son *module maximum* est la fonction

$$\mu_f(r) = \max_{z \in \overline{B}(0,r)} \|f(z)\|$$

donnée par le maximum sur une boule fermée de rayon $r > 0$ de son module. En particulier, on note $\overline{B}(a, r)$ la boule fermée de centre a et rayon r .

Soit $\delta > 0$. On peut alors considérer le sous-ensemble $Z^f(\delta, r) = f^{-1}(\overline{B}(0, \delta)) \cap \overline{B}(0, r) \subset \mathbb{C}^n$ des éléments de la boule de rayon r qui sont à distance au plus δ d’un zéro.

Notons que $Z^f(\delta, r)$ est en général non connexe et que ses composantes connexes peuvent contenir aucun ou plusieurs zéros de f .

DÉFINITION 6.10. — *On appelle épaissement δ -grossier d’un zéro de f une composante connexe de $Z^f(\delta, r)$ qui contient au moins un zéro de f , et on note $Z_{\text{coarse}}^f(\delta, r)$ réunion de ces composantes connexes.*

On appelle nombre de zéros δ -grossiers de f la quantité $\mathfrak{z}_0^f(r, \delta)$ définie comme le cardinal de l'ensemble des épaississements δ -grossiers de zéros de f . Autrement dit, c'est le nombre de composantes connexes de $Z^f(\delta, r)$ qui contiennent un zéro.

Le théorème suivant borne le nombre de zéros grossiers de f en fonction du module maximum.

THÉORÈME 6.11 (Buhovsky, I. Polterovich, L. Polterovich, Shelukhin et Stojisavljević, 2023)

Soit $f: \mathbb{C}^n \rightarrow \mathbb{C}^n$ une fonction analytique. Pour tous $a > 1$, $r > 0$ et $\delta < \frac{\mu_f(ar)}{2}$, on a

$$\mathfrak{z}_0^f(r, \delta) \leq C \left(\log \left(\frac{\mu_f(ar)}{\delta} \right) \right)^{2n-1}$$

où C est une constante qui ne dépend que de a et n .

Remarque 6.12. — Le module maximum a des propriétés qui évoquent la notion de degré d'un polynôme. Par exemple, si $\limsup \frac{\log(\mu_f(r))}{\log(r)}$ vérifie $\limsup \frac{\log(\mu_f(r))}{\log(r)} < k + 1$, alors f est un polynôme de degré k . Pour $n = 1$, l'inégalité de Jensen permet de prouver que si $f(0) \neq 0$, alors le nombre de zéros dans $\overline{B}(0, r)$ est borné par $C \log(\mu_f(r))$. Un contre-exemple célèbre de Cornalba et Shiffman (1972) montre qu'aucun résultat similaire n'est vrai en général dès que $n = 2$ (cette question est connue sous le nom de problème de Bézout transcendant). Le théorème 6.11 établit ces résultats pour les zéros grossiers : le point en est que ce compte grossier permet d'éviter les nombres de zéros produits par des fonctions très oscillantes.

Nous allons juste donner un aperçu de la manière dont les modules persistants interviennent dans la preuve en introduisant ainsi une idée fructueuse en géométrie.

Avant cela, nous avons besoin d'une notation nouvelle. Pour $f: X \rightarrow \mathbb{R}$ ponctuellement de dimension finie, notons $N_{i, \delta}(f)$ le nombre de barres dans $\mathbf{B}_i(X^f)$ qui sont de longueur $> \delta$. Le théorème de stabilité 4.2 implique le corollaire suivant :

COROLLAIRE 6.13. — Soient $0 < \beta < \alpha$. Si $f, g: X \rightarrow \mathbb{R}$ sont des fonctions continues sur une variété compacte et $\|f - g\|_\infty \leq \alpha - \beta$, alors pour tout $i \in \mathbb{N}$, on a

$$N_{i, 2\alpha}(f) \leq N_{i, 2\beta}(g).$$

Quelques éléments de la preuve du théorème 6.11. — Notons que $\mathfrak{z}_0^f(r, \delta)$ s'identifie au nombre de Betti $\beta_0(Z_{\text{coarse}}^f(\delta, r)) = \dim(H_0(Z_{\text{coarse}}^f(\delta, r)))$. On peut réinterpréter cela en termes du module de persistance en degré 0 associé à la fonction $|f|: \mathbb{C}^n \rightarrow \mathbb{R}$.

En effet le nombre de zéros δ -grossiers de f est par définition la dimension de l'image du morphisme de structure $\iota_{0, \delta}^{X^{|f|}}: H_0(X^f)(0) \rightarrow H_0(X^f)(\delta)$ (puisque les groupes d'homologie en degré 0 ont les composantes connexes comme base canonique et que $f = 0$ est équivalent à $|f| \leq 0$). Les composantes connexes qui indexent la base canonique de

cette image correspondent, par définition du code barre, à des barres de longueur au moins δ . On en déduit que

$$(4) \quad \mathfrak{z}_0^f(r, \delta) \leq N_{0,\delta}(|f|).$$

Dès lors, pour démontrer le théorème 6.11, on est ramené à borner le nombre de barres en degré 0 de longueur au moins δ . Ce n'est a priori pas évident. La preuve de Buhovsky, I. Polterovich, L. Polterovich, Shelukhin et Stojisavljević (2023) est basée sur les deux résultats (non triviaux) suivants :

— Soient $P_1, \dots, P_n: \mathbb{C}^n \rightarrow \mathbb{C}$ des fonctions polynomiales de degré au plus k . On note $P = (P_1, \dots, P_n): \mathbb{C}^n \rightarrow \mathbb{C}^n$. Alors il existe une constante C (ne dépendant que de n) telle que

$$(5) \quad N_{0,\delta}(|P|) \leq C k^{2n-1}.$$

— Pour tout entier $k \in \mathbb{N}$, il existe une fonction polynomiale $P = (P_1, \dots, P_n): \mathbb{C}^n \rightarrow \mathbb{C}^n$ de degré au plus $k - 1$ telle que

$$(6) \quad \sup_{x \in \overline{B}(0,r)} |f(x) - P(x)| \leq \frac{a^{1-k}}{a-1} \mu_f(ar).$$

En prenant $k = \left\lceil \frac{\log(\mu_f(ar)\delta)}{\log(a)} \right\rceil$, on déduit de (6) que sur la boule de rayon r , on a $\| |f| - |P| \|_\infty \leq C_a \delta$ où C_a est une constante qui ne dépend que de a . On déduit de (5) et de (4) que

$$N_{0,2C_a\delta}(|f|) \leq N_{0,\varepsilon}(|P|) \leq C \left(1 + 1/\log(a)\right)^{2n-1} \left(\log \left(\frac{\mu_f(ar)}{\delta} \right) \right)^{2n-1}$$

pour ε suffisamment petit.

En faisant un changement de variables, on obtient la formule voulue

$$N_{0,\delta}(|f|) \leq C' \left(\log \left(\frac{\mu_f(ar)}{\delta} \right) \right)^{2n-1}$$

où C' ne dépend que de n et de a . □

Les idées de compte grossier introduites dans le théorème 6.11 permettent de démontrer d'autres résultats. Notons par exemple le remarquable théorème ci-dessous.

Soit $f: \Omega \rightarrow \mathbb{R}$. On appelle domaine nodal de f une composante connexe de $\Omega \setminus f^{-1}(0)$. Un domaine nodal S est dit δ -profond si $\sup_{x \in S} |f(x)| > \delta$. On note $\mathcal{M}_0(f, \delta)$ le cardinal de l'ensemble des domaines nodaux δ -profonds. Plus généralement, pour $i \geq 0$, on note

$$\mathcal{M}_i(f, \delta) = \text{rang} \left(H_i(\{x \setminus |f(x)| > \delta\}) \rightarrow H_i(\Omega \setminus f^{-1}(0)) \right).$$

On constate que pour $i = 0$ les deux définitions coïncident.

On dispose de bornes sur les nombres de domaines modaux en tout degré similaires au théorème 6.11 dues à Lev Buhovsky, Jordan Payette, Iosif Polterovich, Leonid Polterovich, Egor Shelukhin et Vukašin Stojisavljević.

THÉORÈME 6.14 (Buhovsky, I. Polterovich, L. Polterovich, Shelukhin et Stojisavljević, 2023)

Soit P un opérateur elliptique défini sur un domaine $\Omega \subset \mathbb{R}^n$ et soit E_λ^P l'espace engendré par les fonctions propres de P de valeurs propres inférieures ou égales à λ . Pour tous entiers $i \geq 0$ et $k > n/2$, il existe une constante C telle que pour toute fonction $f \in E_\lambda^P$ de norme 1, on a

$$\mathcal{M}_i(f, \delta) \leq C \frac{(1 + \lambda)^{n/2}}{\delta^{n/k}}.$$

La preuve est assez similaire à celle de 6.11, au sens où, après avoir réinterprété l'énoncé de 6.14 comme donnant une borne supérieure du nombre de domaines nodaux en termes de norme de Sobolev d'une fonction lisse, on utilise de manière cruciale le corollaire 6.13 après avoir approximé f par un polynôme sur un cube, pour lequel on a des bornes calculables explicitement.

Références

- Andrew Aukerman, Mathieu Carrière, Chao Chen, Kevin Gardner, Raúl Rabadán et Rami Vanguri (2021). « Persistent homology based characterization of the breast cancer immune microenvironment : a feasibility study », *J. Comput. Geom.* **12** (2), p. 183-206.
- Serguei Barannikov (1994). *The framed Morse complex and its invariants*. Providence, RI : American Mathematical Society, p. 93-115.
- Ulrich Bauer et Michael Lesnick (2014). « Induced matchings of barcodes and the algebraic stability of persistence », in : *Proceedings of the 30th annual symposium on computational geometry, SoCG '14, Kyoto, Japan, June 8–11, 2014*. New York, NY : Association for Computing Machinery (ACM), p. 355-364.
- Nicolas Berkouk, Grégory Ginot et Steve Oudot (2019). *Level-sets persistence and sheaf theory*. arXiv : [1907.09759](https://arxiv.org/abs/1907.09759) [math.AT].
- Nicolas Berkouk et François Petit (2021). « Ephemeral persistence modules and distance comparison », *Algebr. Geom. Topol.* **21** (1), p. 247-277.
- Omer Bobrowski, Sayan Mukherjee et Jonathan E. Taylor (2017). « Topological consistency via kernel estimation », *Bernoulli* **23** (1), p. 288-328.
- Magnus Bakke Botnan et William Crawley-Boevey (2020). « Decomposition of persistence modules », *Proc. Am. Math. Soc.* **148** (11), p. 4581-4596.
- Magnus Bakke Botnan, Vadim Lebovici et Steve Oudot (2022). « On rectangle-decomposable 2-parameter persistence modules », *Discrete Comput. Geom.* **68** (4), p. 1078-1101.

- Magnus Bakke Botnan et Michael Lesnick (2023). « An introduction to multiparameter persistence », in : *Representations of algebras and related structures. International conference on representations of algebras, ICRA 2020, 9–25 November 2020*. Berlin : European Mathematical Society (EMS), p. 77-150.
- Lev Buhovsky, Iosif Polterovich, Leonid Polterovich, Egor Shelukhin et Vukašin Stojisavljević (2023). *Persistent transcendental Bézout theorems*. arXiv : [2307.02937](https://arxiv.org/abs/2307.02937) [[math.CV](https://arxiv.org/abs/2307.02937)].
- Mathieu Carrière et Andrew Blumberg (2020). « Multiparameter persistence images for topological machine learning », *Advances in Neural Information Processing Systems* **33**.
- Frédéric Chazal, Vin De Silva et Steve Oudot (2014). « Persistence stability for geometric complexes », *Geom. Dedicata* **173**, p. 193-214.
- Frédéric Chazal, Vin de Silva, Marc Glisse et Steve Oudot (2016). *The structure and stability of persistence modules*. SpringerBriefs Math. Cham : Springer.
- Jérémy Cochoy et Steve Oudot (2020). « Decomposition of exact pfd persistence bimodules », *Discrete Comput. Geom.* **63** (2), p. 255-293.
- Maurizio Cornalba et Bernard Shiffman (1972). « A counterexample to the "transcendental Bezout problem" », *Ann. Math. (2)* **96**, p. 402-406.
- Masaki Kashiwara et Pierre Schapira (2018). « Persistent homology and microlocal sheaf theory », *J. Appl. Comput. Topol.* **2** (1-2), p. 83-113.
- Yongjin Lee, Senja Barthel, Paweł Dłotko, Mohamad Moosavi, Kathryn Hess et Berend Smit (mai 2017). « Quantifying similarity of pore-geometry in nanoporous materials », **8** (1).
- Steve Oudot (2015). *Persistence theory. From quiver representations to data analysis*. T. 209. Math. Surv. Monogr. Providence, RI : American Mathematical Society (AMS).
- Leonid Polterovich, Daniel Rosen, Karina Samvelyan et Jun Zhang (2020). *Topological persistence in geometry and analysis*. T. 74. Univ. Lect. Ser. Providence, RI : American Mathematical Society (AMS).

Grégory Ginot

LAGA, UMR 7539

Université Sorbonne Paris Nord

99 av. J.-B. Clément,

93430 Villetaneuse

E-mail : ginot@math.univ-paris13.fr