



HAL
open science

Raconte-moi... la persistance topologique

Philippe Malbos

► **To cite this version:**

| Philippe Malbos. Raconte-moi... la persistance topologique. 2021, Vol. 169, pp. 44-50. hal-03345898

HAL Id: hal-03345898

<https://hal.science/hal-03345898v1>

Submitted on 15 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Raconte-moi... la persistance topologique

Philippe Malbos

1. Topologie et données

Avec le développement des outils numériques et l'augmentation prodigieuse des ressources en stockage et des puissances de calcul des processeurs, la production des données explose dans de très nombreux domaines des sciences, sciences humaines et sociales, en ingénierie et en santé. Ces données sont de natures très différentes, mais, souvent volumineuses et produites sous forme de vecteurs de très grande dimension. Il est ainsi très difficile de les visualiser. Généralement, très peu de coordonnées sont pertinentes par rapport aux questions que l'on se pose sur ces données. Aussi, il est nécessaire d'éliminer les informations inutiles, alors même que l'on ne sait pas toujours quelles sont les coordonnées pertinentes à observer. Il arrive également que les données produites soient « bruitées » avec des informations manquantes. L'enjeu est alors d'analyser qualitativement et quantitativement ces données en tenant compte de toutes ces problématiques des processus de production des données.

Des méthodes d'analyse statistique, probabiliste, topologique et géométrique ont été développées afin de construire des algorithmes efficaces et robustes pour l'exploration et l'analyse des données de grande dimension. Nous expliquons ici comment une méthode topologique, appelée *l'homologie persistante*, issue de la topologie algébrique et de la théorie de Morse peut être appliquée pour analyser des jeux de données représentés par des nuages de points équipés d'une fonction distance. Ces nuages de points doivent être considérés comme des échantillons finis et bruités, prélevés sur un objet géométrique. L'objectif est alors d'étudier les propriétés topologiques

de l'objet géométrique en éliminant les caractéristiques topologiques de petites tailles correspondant au bruit. Nous allons montrer comment l'homologie persistante permet d'extraire l'information utile au sein de ces données et comment fonder la comparaison des jeux de données sur cette information utile. L'analyse de données par persistance topologique est ainsi une approche descriptive. Il existe de nombreuses applications pratiques, notamment dans les problèmes de reconnaissance de formes. Parmi les belles applications, citons la découverte d'un sous type de cancer du sein avec un taux de survie à 100% avec la méthode de persistance appliquée à des données biologiques d'expression des gènes [11], ou encore l'étude de certaines caractéristiques moléculaires de la structure de protéine [14].

2. La persistance topologique

La *persistance topologique* est une méthode de la topologie algorithmique qui consiste à éliminer l'information inutile dans les données. Il s'agit d'encoder l'évolution de la topologie selon une décomposition en sous-niveaux d'une filtration et de mesurer l'importance relative des caractéristiques topologiques de la filtration. Cette approche émergea simultanément dans plusieurs travaux à la fin des années 1990, [1, 8, 13, 7].

Commençons par une illustration et examinons la topologie des sous-niveaux d'une fonction d'une seule variable réelle $f : I \rightarrow \mathbb{R}$ définie sur un intervalle réel I dont le graphe est donné par la figure 1. Cela revient à étudier les espaces $K_a := f^{-1}((-\infty, a])$ selon la croissance du réel a . Si $a < x_1$, alors K_a est vide et

lorsque $x_1 \leq a < x_2$, alors l'espace K_a est formé d'une seule composante connexe. Pour $x_2 \leq a < x_3$, K_a possède deux composantes connexes et pour $x_3 \leq a < x_4$, l'espace K_a ne possède plus qu'une seule composante connexe issue de la fusion des deux précédentes composantes connexes.

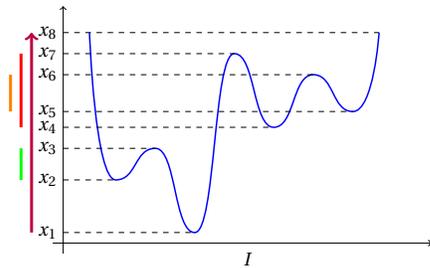


FIGURE 1 – Graphe de la fonction f

Au passage de $a = x_4$, une deuxième composante connexe est à nouveau créée, puis une troisième au passage de $a = x_5$, laquelle disparaît quand $a = x_6$ et K_a ne possède plus qu'une seule composante connexe après $a = x_7$. La deuxième composante créée pour $a = x_2$ disparaît lorsque $a = x_3$. On dit alors que l'on *apparie* les valeurs x_2 et x_3 . Cet appariement est représenté par un point dans le *diagramme de persistance* de la figure 2; on parle de *paire de persistance*. L'intervalle de persistance associé est indiqué en vert sur la figure 2. Cet intervalle décrit la durée de vie de la composante connexe créée en x_2 . Il y a deux autres intervalles de persistance pour des composantes connexes créées en x_4 , puis en x_5 . Toute cette information est résumée dans le *code barre* qui encode la naissance et la mort des composantes connexes.

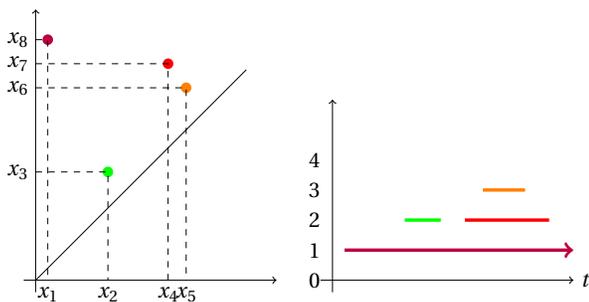


FIGURE 2 – Diagramme de persistance et code barre de la fonction f

Considérons un bruitage de la fonction f décrit par le graphe de la figure 3. Ce graphe présente la même forme globale que la fonction f , mais les variations topologiques sont beaucoup plus nombreuses. Aussi, le diagramme de persistance de la fonction représentée par la figure 4 contient beaucoup plus de points que celui de la fonction f . Notons que ce bruitage de la fonction induit des perturbations topologiques très locales, créant ainsi des appariements de points qui se situent proches de la diagonale du diagramme de persistance.

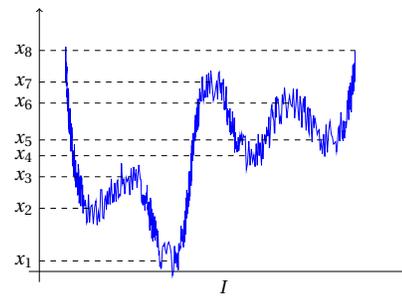


FIGURE 3 – Fonction f bruitée

Les propriétés topologiques de la forme globale sont ainsi caractérisées par une longue persistance, alors que celles induites par les perturbations topologiques très locales sont de courte persistance.

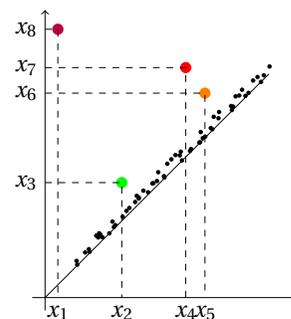


FIGURE 4 – Diagramme de persistance de la fonction f bruitée

Cet exemple porte sur une fonction sur un espace de dimension un, mais en général on s'intéresse à des données qui sont représentées par des fonctions définies sur des espaces topologiques de grande dimension. L'évolution de la topologie des sous-

niveaux est alors caractérisée par celui du nombre de composantes connexes, de boucles, de cavités... Par exemple, la figure 5 représente le code barre d'une variété de dimension 2 et la persistance de ses caractéristiques topologiques en dimension 0 et 1. On crée alors des diagrammes de persistance pour chacune de ces dimensions.

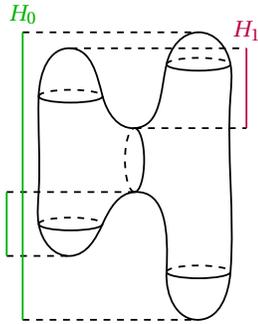


FIGURE 5 – Code barre d'une variété de dimension 2

La nature topologique des modifications est ici plus complexe qu'en dimension un. Pour modéliser cela, nous utiliserons la structure de *complexe simplicial* et nous décrirons les nombres de composantes connexes, boucles, cavités par des invariants homologiques de ces complexes, appelés les *nombres de Betti*. Nous montrerons ensuite comment filtrer ce complexe simplicial pour révéler ses propriétés topologiques persistantes.

3. Les nombres de Betti

L'*homologie* consiste à associer, à un objet géométrique, une suite de groupes abéliens dont les dimensions déterminent respectivement ses nombres de composantes connexes, boucles, cavités... La définition de l'homologie est basée sur une description combinatoire de l'espace. Nous présentons ici une version simple d'homologie où l'espace est décrit par un *complexe simplicial*.

3.1. Les complexes simpliciaux

La notion de *complexe simplicial* permet de décrire certains espaces topologiques à partir de don-

nées combinatoires, en généralisant le procédé de triangulation d'une surface. Nous présentons ici cette structure simpliciale dans sa forme algébrique, mais elle peut également être introduite de façon plus abstraite.

Considérons un ensemble $\mathcal{P} = \{u_0, \dots, u_k\}$ de points de l'espace \mathbb{R}^d . Ces points sont dits *affinement indépendants* si, lorsque deux combinaisons affines $\sum_i \lambda_i u_i$ et $\sum_i \mu_i u_i$ sont égales, alors $\lambda_i = \mu_i$ pour tout $0 \leq i \leq k$. Une combinaison affine $\sum_i \lambda_i u_i$ est dite *convexe* lorsque tous les scalaires λ_i sont positifs et que $\sum_i \lambda_i = 1$. L'ensemble des combinaisons convexes forme l'*enveloppe convexe* de \mathcal{P} .

Un *k-simplexe* est l'enveloppe convexe de $k + 1$ points affinement indépendants u_0, \dots, u_k de \mathbb{R}^d . On note alors $\sigma = \text{conv}(u_0, \dots, u_k)$ et on dit que σ est de *dimension k*. La figure 6 représente les simplexes dans les premières dimensions.

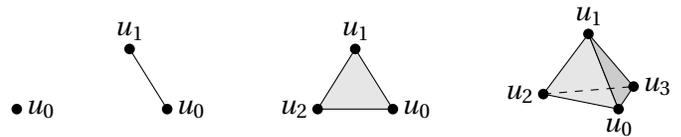


FIGURE 6 – Simplexes en dimension 0, 1, 2, 3

Une *face* du *k-simplexe* σ est l'enveloppe convexe d'un sous-ensemble non-vide de $\{u_0, \dots, u_k\}$. Lorsque τ est une face de σ , on note $\tau \leq \sigma$. On appelle *complexe simplicial* (fini) une famille finie K de simplexes vérifiant les deux propriétés suivantes

- i) si $\sigma \in K$ et $\tau \leq \sigma$, alors $\tau \in K$;
- ii) si $\sigma, \sigma' \in K$, alors l'intersection $\sigma \cap \sigma'$ est soit vide, soit exactement une face commune à σ et σ' .

La *dimension* de K est la dimension maximale de ses simplexes. Les points u_i qui engendrent les simplexes de K sont appelés les *sommets* de K .

3.2. Les groupes d'homologie

Donnons-nous un complexe simplicial K de dimension n et considérons $\mathbb{Z}/2\mathbb{Z}$ comme corps de

base. Pour $0 \leq k \leq n$, une k -chaîne est une combinaison linéaire

$$c = \sum_{i=1}^{n_k} \lambda_i \sigma_i,$$

où $\{\sigma_1, \dots, \sigma_{n_k}\}$ est l'ensemble des k -simplexes de K et les λ_i sont égaux à 0 ou 1. L'ensemble des k -chaînes de K forme un espace vectoriel, que nous noterons $C_k(K)$. On définit le bord d'un k -simplexe $\sigma = \text{conv}(u_0, \dots, u_k)$ comme la $(k-1)$ -chaîne, notée $\partial(\sigma)$, formée de toutes ses $(k-1)$ -faces, i.e.,

$$\partial(\sigma) = \sum_{j=0}^k \text{conv}(u_0, \dots, \check{u}_j, \dots, u_k),$$

où $\text{conv}(u_0, \dots, \check{u}_j, \dots, u_k)$ désigne la $(k-1)$ -face de σ ne contenant pas le sommet u_j . Par linéarité, on étend l'opération ∂ aux k -chaînes de K et on définit ainsi l'opérateur bord comme l'application linéaire

$$\partial : C_k(K) \rightarrow C_{k-1}(K)$$

qui associe à une k -chaîne $c = \sum \lambda_i \sigma_i$ la somme de des bords de ses simplexes : $\partial(c) = \sum \lambda_i \partial(\sigma_i)$.

Une propriété remarquable de cet opérateur est qu'il satisfait la relation $\partial\partial = 0$. Ainsi, au complexe simplicial K on associe un *complexe de chaînes* défini par la suite d'applications linéaires suivantes :

$$0 \rightarrow C_n(K) \xrightarrow{\partial} C_{n-1}(K) \xrightarrow{\partial} \dots \xrightarrow{\partial} C_1(K) \xrightarrow{\partial} C_0(K) \xrightarrow{\partial} 0.$$

Pour $0 \leq k \leq n$, on appelle k -cycle de K les éléments du noyau

$$Z_k(K) = \ker(\partial : C_k(K) \rightarrow C_{k-1}(K)),$$

en posant $C_{-1}(K) = 0$. Les k -bords sont les éléments de l'image

$$B_k(K) = \text{im}(\partial : C_{k+1}(K) \rightarrow C_k(K)).$$

De l'égalité $\partial\partial = 0$, on déduit que chaque k -bord est un k -cycle, autrement dit, on a une inclusion de sous-espaces vectoriels $B_k(K) \subseteq Z_k(K)$. Nous pouvons ainsi considérer l'espace vectoriel quotient

$$H_k(K) = Z_k(K) / B_k(K),$$

appelé le k -ème groupe d'homologie de K . Ses éléments s'écrivent $\bar{c} = c + B_k(K)$, où c est un k -cycle, et sont appelés les *classes d'homologie du complexe K* .

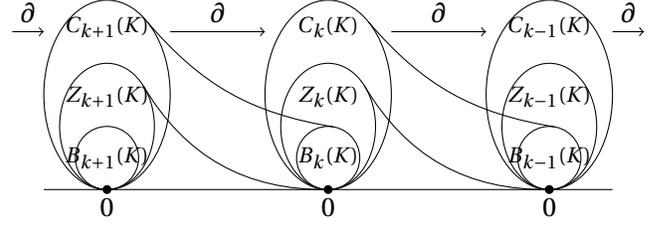


FIGURE 7 – Espaces des cycles et des bords reliés par les opérateurs bords.

La dimension de l'espace vectoriel $H_k(K)$ est appelée le k -ème nombre de Betti de K et notée $\beta_k(K)$. On montre que $\beta_0(K)$ correspond au nombre de composantes connexes du complexe K , $\beta_1(K)$ au nombre de cycles, $\beta_2(K)$ au nombre de cavités... On montre également que les nombres de Betti sont des invariants topologiques, c'est-à-dire que si K et L sont deux complexes simpliciaux dont les espaces qui les réalisent sont homéomorphes, alors $\beta_k(K) = \beta_k(L)$, pour tout $k \geq 0$.

La terminologie *nombre de Betti* a été introduite par Henri Poincaré dans [12, §6.] en l'hommage à son ami Enrico Betti.

3.3. Calcul de l'homologie

Le calcul effectif des nombres de Betti d'un complexe simplicial K peut se faire via la méthode de réduction par *forme normale de Smith* des matrices associées aux opérateurs bords ∂ , comme dans [10, 4]. La k -ème *matrice bord* $[\partial_k]$ encode la décomposition du bord de chaque k -simplexe en fonction de ses $(k-1)$ -faces. De façon explicite, $[\partial_k]_i^j$ est égal à 1, si le i -ème $(k-1)$ -simplexe est une face du j -ème k -simplexe, et 0 sinon. On montre alors que le k -ième nombre de Betti du complexe K s'obtient par la formule suivante :

$$\beta_k(K) = \text{rang}(C_k(K)) - \text{rang}([\partial_k]) - \text{rang}([\partial_{k+1}]).$$

Le calcul de ces rangs peut s'effectuer par réduction à la forme normale de Smith des matrices bords de K .

4. La persistance topologique selon une filtration

4.1. Filtration d'un complexe simplicial

La persistance d'un complexe simplicial K se calcule selon une *filtration* du complexe, qui consiste à ordonner l'ensemble de ses simplexes afin de le décomposer en sous-niveaux, comme nous l'avons fait précédemment avec la fonction f de la figure 1. Formellement, une filtration de K est définie par une suite $(K_i)_i$ de sous-complexes de K , ordonnés par l'inclusion :

$$\emptyset = K_0 \subset K_1 \subset \dots \subset K_p = K,$$

et telle que pour tout $0 \leq i < p$, $K_{i+1} = K_i \cup \{\sigma\}$, où σ est un simplexe de K n'appartenant pas à K_i .

La *persistance topologique* sur le complexe K selon cette filtration consiste en l'étude de l'évolution de la topologie des sous-complexes K_i . Le point essentiel dans le calcul de la persistance réside donc dans la méthode de filtration choisie. Par exemple, pour une fonction f à valeurs réelles, définie sur les sommets de K , on associe à tout k -simplexe σ de sommets u_0, \dots, u_k la valeur

$$f(\sigma) = \max_{0 \leq i \leq k} f(u_i).$$

On ordonne ainsi tous les simplexes de K en fonction des valeurs de f et on en déduit une filtration de K par les sous-niveaux de $K_a := f^{-1}((-\infty, a])$ de f .

4.2. Homologie persistante

Donnons-nous une filtration $(K_i)_{0 \leq i \leq p}$ d'un complexe simplicial K . Pour tous $j \leq \ell$, l'inclusion $K_j \subset K_\ell$ induit une application linéaire $h_k^{j,\ell} : H_k(K_j) \rightarrow H_k(K_\ell)$ pour tout $k \geq 0$. Ainsi, la filtration $(K_i)_i$ induit une suite d'applications linéaires

$$H_k(K_0) \rightarrow H_k(K_1) \rightarrow \dots \rightarrow H_k(K_p) = H_k(K),$$

pour chaque dimension $k \geq 0$.

En notant respectivement $Z_k^j(K)$ et $B_k^j(K)$ les sous-espaces vectoriels formés des k -cycles et des k -bords du sous-complexe K_j , on définit le k -ème *espace d'homologie persistante* de K comme le quotient

$$H_k^{j,\ell}(K) = Z_k^j(K) / (B_k^\ell(K) \cap Z_k^j(K)),$$

qui consiste en les classes d'homologie du sous-complexe K_j qui sont encore en vie dans le complexe K_ℓ . La dimension de cet espace homologique est appelé le k -ème *nombre de Betti persistant* et noté $\beta_k^{j,\ell}$.

Une classe d'homologie \bar{c} de $H_k(K_j)$ est dite *naître* en K_j lorsque $\bar{c} \notin H_k^{j-1,j}(K)$. On dit qu'elle *disparaît* dans K_ℓ si elle s'annule dans $H_k(K_\ell)$ sans être nulle dans $H_k(K_{\ell-1})$, c'est-à-dire $h_k^{j,\ell-1}(\bar{c}) \notin H_k^{j-1,\ell-1}(K)$ et $h_k^{j,\ell}(\bar{c}) \in H_k^{j-1,\ell}(K)$, comme illustré par la figure 8. La différence d'indice $\ell - j$ est alors appelée l'*indice de persistance* de la classe d'homologie \bar{c} .

4.3. Diagrammes de persistance

On peut représenter l'information de la persistance avec des diagrammes de dimension 2, comme nous l'avons fait dans notre premier exemple avec les diagrammes de persistance des figures 2 et 4. On montre que le nombre $\mu_k^{j,\ell}$ de classes d'homologie en dimension k qui naissent en K_j et disparaissent en K_ℓ est donné par la formule

$$\mu_k^{j,\ell} = (\beta_k^{j,\ell-1} - \beta_k^{j,\ell}) - (\beta_k^{j-1,\ell-1} - \beta_k^{j-1,\ell}).$$

Le *diagramme de k -ème persistance* de la filtration $(K_i)_i$ est un multi-ensemble obtenu en traçant dans le plan réel étendu chaque point (a_j, a_ℓ) correspondant à un couple de classes d'homologies appareillées, avec sa multiplicité $\mu_k^{j,\ell}$. Certains points peuvent être confondus et d'autres peuvent avoir des coordonnées infinies. Par définition, tous les points sont au-dessus de la diagonale et la distance d'un point à la diagonale représente la persistance des classes homologiques correspondantes.

4.4. Calcul de la persistance

Les nombres de Betti persistants peuvent être calculés avec la méthode de la forme normale de Smith

comme dans la section précédente. Les paires de persistance peuvent être également calculées à partir d'un algorithme similaire. Considérons un complexe simplicial K constitué de n simplexes $\sigma_1, \dots, \sigma_n$ et fixons une filtration $(K_i)_i$ de K . On ordonne les simplexes selon la filtration de telle façon que si τ est une face d'un simplexe σ , alors $\tau < \sigma$. Définissons la *matrice bord* associée à la filtration du complexe, comme la matrice $[\partial]$ de $\mathcal{M}_n(\mathbb{Z}/2\mathbb{Z})$, dont la j -ème colonne représente les faces de codimension 1 du simplexe σ_j , i.e., $[\partial]_i^j = 1$, si σ_i est une face de codimension 1 de σ_j et 0 sinon. Notons $b(j)$ le numéro de la ligne contenant le coefficient 1 le plus bas dans la colonne j . Si la colonne j est nulle, alors $b(j)$ est non défini. La matrice bord est dite *réduite* si $b(j) \neq b(j_0)$ pour toutes colonnes non nulles $j \neq j_0$.

Par construction, la matrice $[\partial]$ est triangulaire strictement supérieure, mais n'est pas réduite en général. L'algorithme suivant transforme la matrice $[\partial]$ en une matrice réduite R .

```

R ← [∂]
pour (j = 1 à n) faire
  tant que (∃ j₀ < j | b(j₀) = b(j)) faire
    j ← j + j₀
  fin tant que
fin pour

```

On montre que le k -ème nombre de Betti du complexe K est obtenu par la formule

$$\beta_k(K) = Z_k(R) - B_k(R),$$

où $Z_k(R)$ est le nombre de colonnes nulles et $B_k(R)$ celui de colonnes non nulles, correspondantes à des k -simplexes du complexe K .

On peut alors extraire de la matrice R les paires de persistance. En effet, si $b(j) = i > 0$, alors le simplexe σ_j va s'appareiller avec le simplexe σ_i pour faire mourir une classe d'homologie. Si $b(j) = 0$, alors le simplexe σ_j va créer un nouveau cycle et donc donner naissance à une nouvelle classe d'homologie. Dans

ce cas, s'il n'existe pas de k tel que $b(k) = j$, alors la classe d'homologie ne disparaît pas avec la filtration, on parle de *cycle essentiel*. Les paires de persistance sont ainsi les couples $(\sigma_{b(i)}, \sigma_i)$, où i est tel que $b(i) \neq 0$.

Le coût de cet algorithme est cubique en le nombre de simplexes. Dans les applications pratiques, le nombre de simplexes peut être très grand et la matrice bord être une matrice creuse. Le premier algorithme rapide de calcul de l'homologie persistante a été introduit dans [7] avec une implémentation par des matrices creuses.

4.5. Stabilité des diagrammes de persistance

Nous terminons par un résultat de stabilité obtenu dans [3] pour des fonctions définies sur un complexe simplicial. Considérons deux fonctions f et g définies sur les sommets d'un complexe simplicial K et les deux filtrations de K associées. Soient D_f et D_g les diagrammes de persistance de K associés à ces filtrations. On définit la distance *bottleneck* entre ces diagrammes en posant :

$$d_B(D_f, D_g) = \inf_{\gamma} \sup_{x \in D_f} \|x - \gamma(x)\|_{\infty},$$

où $\gamma : D_f \rightarrow D_g$ parcourt l'ensemble des bijections de multi-ensembles. Le résultat de stabilité suivant, montré dans [3], établit que

$$d_B(D_f, D_g) \leq \sup_x |f(x) - g(x)|,$$

où x parcourt l'ensemble des sommets du complexe K . Ce résultat montre que de petites variations sur les fonctions f et g impliquent de petites variations sur les diagrammes de persistance associés relativement à la distance bottleneck.

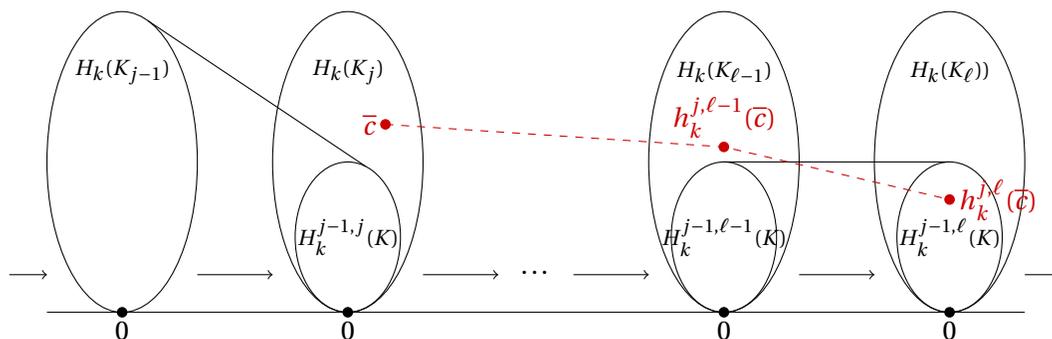


FIGURE 8 – La classe d’homologie \bar{c} née dans K_j et disparaît dans K_{ℓ} d’indice de persistance $\ell - j$.

Pour en savoir plus

Le livre d’Edelsbrunner et Harer [6] présente les principales méthodes de la topologie algorithmique. Les constructions esquissées dans cette note y sont très largement développées. Les articles [5] et [2] constituent également de très bonnes introductions au sujet et présentent de nombreux exemples applicatifs. Nous renvoyons le lecteur qui voudrait en savoir plus sur les notions de complexes simpliciaux et de constructions homologiques à [10] et [9]. Les procédés classiques de filtration d’un complexe simplicial associé à un nuage de points sont présentés dans [6]. La notion d’appariement homologique apparaît déjà dans les travaux de Barannikov dans [1], où la notion de code barre relativement à une filtration est introduite sous la terminologie de *forme canonique*.

References

- [1] S. A. Barannikov. The framed morse complex and its invariants. In *Singularities and bifurcations*, pages 93–115. Providence, RI : American Mathematical Society, 1994.
- [2] Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2) :255–308, 2009.
- [3] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1) :103–120, 2007.
- [4] Jean-Guillaume Dumas, Frank Heckenbach, David Saunders, and Volkmar Welker. Computing simplicial homology based on efficient Smith normal form algorithms. In *Algebra, geometry, and software systems*, pages 177–206. Springer, Berlin, 2003.
- [5] Herbert Edelsbrunner and John Harer. Persistent homology—a survey. In *Surveys on discrete and computational geometry*, volume 453 of *Contemp. Math.*, pages 257–282. Amer. Math. Soc., Providence, RI, 2008.
- [6] Herbert Edelsbrunner and John L. Harer. *Computational topology*. American Mathematical Society, Providence, RI, 2010.
- [7] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. volume 28, pages 511–533. 2002. *Discrete and computational geometry and graph drawing* (Columbia, SC, 2001).
- [8] Patrizio Frosini and Claudia Landi. Size theory as a topological tool for computer vision. *Pattern Recognition and Image Analysis*, 9(4) :596–603, 1999.
- [9] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.
- [10] James R. Munkres. *Elements of algebraic topology*. Addison-Wesley Publishing Company, Menlo Park, CA, 1984.
- [11] Monica Nicolau, Arnold J Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17) :7265–7270, 2011.

- [12] Henri Poincaré. Analysis situs. *Journal de l'École Polytechnique*, t. 1 :1–121, 1895.
- [13] V. Robins. Towards computing homology from finite approximations. In *Proceedings of the 14th Summer Conference on General Topology and its Applications (Brookville, NY, 1999)*, volume 24, pages 503–532 (2001), 1999.
- [14] Kelin Xia and Guo-Wei Wei. Persistent homology analysis of protein structure, flexibility, and folding. *Int J Numer Method Biomed Eng*, 30(8) :814–44, Aug 2014.

L'auteur remercie Isaac Ren, François Laudenbach et Romain Joly pour leurs relectures et leurs suggestions avisées.