

TER: Autour de l'homologie persistante

Moussa Doumbiya

Juillet 2022

Encadrant : Grégory Ginot.



Table des matières

1	Simplexe et complexe simplicial	4
1.1	Simplexe	4
1.1.1	Définition	4
1.2	Complexe simplicial	5
1.2.1	Définition	5
2	Du nuage de points à une structure simpliciale	6
2.0.1	Définition	7
2.1	Filtrations associées aux complexes de Čech et de Rips(Rips-Vietoris)	8
2.1.1	Définition	8
2.1.2	Définition	8
3	L'homologie persistante	10
3.1	Préliminaires	10
3.1.1	Définition (Chaines simpliciales)	10
3.2	Modules de persistance	12
3.2.1	Définition (Module de persistance) :	12
3.2.2	Définition (morphisme de module de persistance) :	13
3.2.3	Définition (somme directe de modules de persistance) :	13
3.2.4	Définition :	13
3.2.5	Définition (Module intervalle) :	14
3.2.6	Définition :	15
3.3	Algorithme de calcul du code barre d'un complexe simplicial	18
3.3.1	Algorithme :	21
3.3.2	Définition(matrice de bord) :	21
3.4	Stabilité	24
3.4.1	Définition (Accouplement partiel) :	24
3.4.2	Définition (Coût d'un accouplement partiel) :	25
3.4.3	Définition(distance Bottleneck)	27
3.4.4	Théorème de stabilité I	28
3.4.5	Théorème de stabilité II	29
3.4.6	Définition(<i>distance de Hausdorff</i>)	29
3.4.7	Définition(<i>distance de Gromov-Hausdorff</i>)	30

Introduction

Lorsque nous admirons une oeuvre d'art à travers une image pixelisée, nous voyons bien plus un grand nombre de pixels placés les uns à côté des autres. Face à une telle image, notre cerveau fait en sorte de relier les points (pixels) les uns aux autres d'une manière ou d'une autre si bien qu'une structure topologique est créée transformant ainsi une donnée discrète en un continuum.

Dans le cadre de l'analyse topologique des données on cherche à formaliser cette idée. Ainsi, lorsque nous disposons d'un nuage de points \mathbb{X} vu comme une discrétisation d'un espace topologique continu X inclus dans un espace euclidien, on cherche d'abord à trouver des techniques de triangulation de X (si on le suppose triangulable) à partir de \mathbb{X} de manière à ce qu'une approximation simpliciale de X soit créée.

Une première idée intuitive pour construire ce complexe simplicial à partir des données consiste à relier les points voisins mutuellement proches d'une certaine distance α définissant ainsi les sommets d'un simplexe pour avoir à la fin un complexe simplicial \mathbb{X}_α . Il s'agit là d'une définition que nous préciserons plus tard : Ce sont les complexes de Čech et de Rips-Vietoris sur les quels on va davantage se concentrer dans le cadre de ce TER. Il existe d'autres techniques de construction de complexe simplicial partant des données notamment les complexes de Delauney, les alpha-complexes que nous n'allons pas aborder ici. De manière plus générale, on peut définir une fonction continue f sur X et considérer les sous niveaux de f : $X_\alpha = f^{-1}(]-\infty, \alpha])$. Ainsi, en faisant varier la valeur α (lorsque celle-ci parcourt un sous-ensemble de \mathbb{N} ou de \mathbb{R}) on peut étudier l'évolution de la topologie de \mathbb{X}_α ou X_α en calculant son homologie pour tout α et de quelle manière celle-ci évolue : c'est le but de l'homologie persistante.

Dans ce TER, après avoir introduit comment construire un complexe simplicial associé aux données notamment ceux de Čech et de Rips-Vietoris, nous allons voir comment extraire de la topologie à partir des données avec l'homologie persistante et dans quelles mesures celle-ci permet d'analyser et de comparer des données sans oublier les algorithmes associés.

1 Simplexe et complexe simplicial

Pour définir l'homologie simpliciale (voir partie 3) d'un espace topologique nous aurons besoin de le découper en des parties plus "simples" que l'on appelle simplexe. On parle de triangulation de l'espace topologique. Il s'agit là que de simplexe et de complexe simplicial plongé dans un espace euclidien.

1.1 Simplexe

1.1.1 Définition

Un simplexe σ de dimension r plongé dans un espace euclidien \mathbb{R}^n est l'enveloppe convexe de $r+1$ points indépendants s_0, s_1, \dots, s_r .

Les s_i sont appelés sommets de σ . Les faces de σ sont les simplexes de dimension $r-1$ dont les sommets sont ceux de σ .

Une orientation de σ est le choix d'un ordre sur ses sommets modulo les permutations paires.

Un simplexe de dimension r sera appelé un r -simplexe.

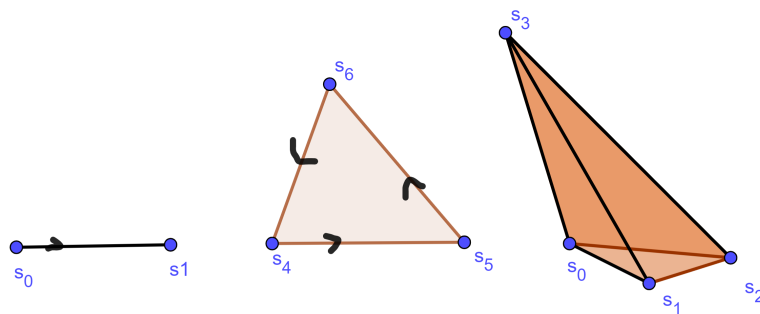


Figure 1 : De gauche à droite : Un 1-simplexe (ségment), un 2-simplexe (triangle plein), un 3-simplexe (tetraèdre plein)

1.2 Complexe simplicial

1.2.1 Définition

Un complexe simplicial K plongé dans un espace euclidien \mathbb{R}^n est une collection de simplexes dans \mathbb{R}^n vérifiant :

- 1) si un simplexe σ appartient à K alors toute face de σ appartient à K
- 2) si σ et τ appartiennent à K avec $\sigma \cap \tau \neq \emptyset$, alors $\sigma \cap \tau$ est le simplexe dont les sommets sont les sommets communs à σ et τ
- 3) Si $x \in \sigma \in K$, alors il existe un voisinage $U \subset \mathbb{R}^n$ de x qui ne rencontre qu'un nombre fini de simplexes de K .

On appelle sous complexe simplicial de K un sous ensemble de K qui est un

complexe simplicial.

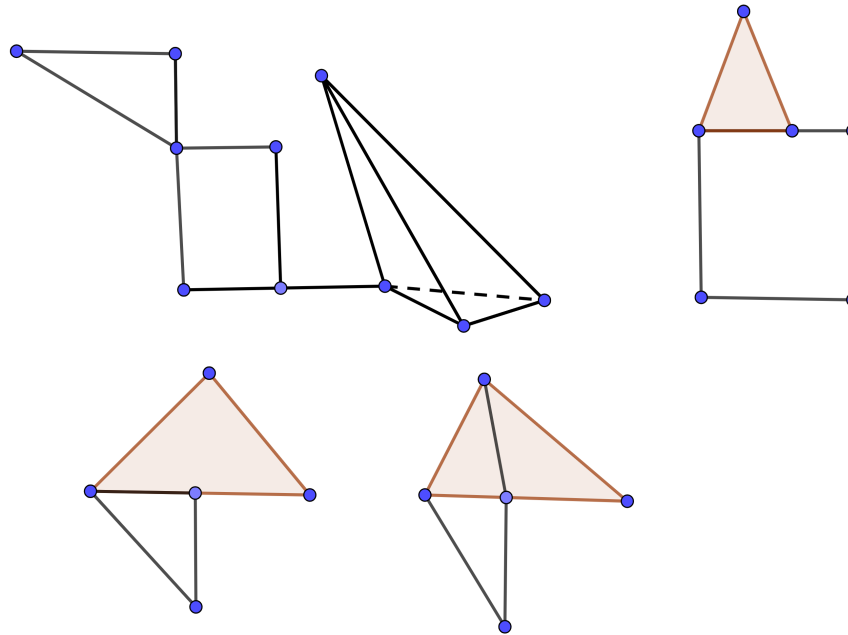


Figure 2 : En haut : à gauche un complexe simplicial et à droite une réunion de simplexe qui n'est pas simpliciale. En bas : à gauche une réunion de simplexes qui n'est pas simpliciale et à droite un complexe simplicial.

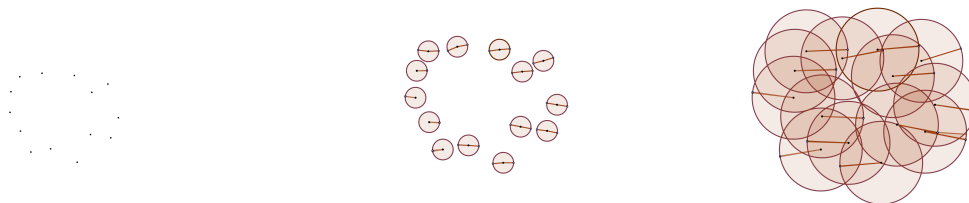
Remarque : Le bord d'un simplexe a une structure canonique de complexe simplicial.

2 Du nuage de points à une structure simpliciale

Face à un ensemble discret d'un jeu de données représenté par un nuage de points \mathbb{X} , la question qui se pose est comment construire un complexe simplicial associé aux données. Pour revenir sur l'une des constructions importantes qu'on a mentionnée en introduction : les complexes de Čech et

Rips-Vietoris. Il s'agit de décrire l'espace sous-jacent X comme union des boules centrées en les points de \mathbb{X} pour un rayon α choisi et en déduire une structure simpliciale. Un problème de choix de α se pose.

Par exemple si \mathbb{X} est un ensemble de points extrait autour d'un cercle. Lorsque α est très petit, alors \mathbb{X}_α n'est que la réunion d'un très grand nombre de petites boules disjointes et à donc autant de composantes connexes. Si α est très grand, \mathbb{X}_α a une composante connexe et homotope à un disque.



Pour répondre à ce problème et calculer fiablement l'homologie de X , on considère des filtrations de complexes ou d'espaces en faisant varier α .

2.0.1 Définition

Une filtration de complexes ou d'espaces topologiques indexée par une partie T de \mathbb{R} est une famille $(X_t)_{t \in T}$ de complexes ou d'espaces telle que $\forall s, t \in T$ avec $s \leq t$, on ait $X_s \subseteq X_t$

Exemples :

- Sous-niveaux d'une fonction : Soit $f : X \rightarrow \mathbb{R}$ une fonction continue. Alors $\forall t \in \mathbb{R}$, $X_t = f^{-1}(]-\infty, t])$ définit une filtration sur \mathbb{R} .
- Filtration simpliciale : Soit K un complexe simplicial et $f : K \rightarrow \mathbb{R}$ une fonction sur K telle que $f(\sigma) \leq f(\tau) \forall \sigma \subseteq \tau$. Alors $X_t = f^{-1}(]-\infty, t])$ définit une filtration sur \mathbb{R} .

2.1 Filtrations associées aux complexes de Cech et de Rips(Rips-Vietoris)

Comme nous l'avons évoqué dans l'introduction, il s'agit de l'une des techniques de construction de complexes les plus importantes utilisée en analyse topologique des données.

2.1.1 Définition

Etant donné un nuage de points fini $\mathbb{X} \subset \mathbb{R}^d$ de cardinal n ($n > d$ très souvent quand il s'agit de données réelles) et un réel α positif, on note $B(x_i, \alpha)$ pour tout $x_i \in \mathbb{X}$ la boule centrée en x_i et de rayon α

Le **complexe de Cech** (plongé dans \mathbb{R}^d) d'ensemble de sommets \mathbb{X} et de paramètre α que l'on note $\text{Cech}(\mathbb{X}, \alpha)$ est le complexe simplicial dont les k -simplexes σ avec $k \leq d$ sont définis comme suit :

$$\sigma = \{x_0, \dots, x_k\} \in \text{Cech}(\mathbb{X}, \alpha) \iff \bigcap_{i=0, \dots, k} B(x_i, \alpha) \neq \emptyset$$

Lorsque α parcourt \mathbb{R}_+ , on obtient une famille croissante de complexes simpliciaux : c'est la filtration associée aux complexes de Cech.

En effet on a : $\text{Cech}(\mathbb{X}, \alpha_1) \subseteq \text{Cech}(\mathbb{X}, \alpha_2)$ dès que $\alpha_1 \leq \alpha_2$.

2.1.2 Définition

Etant donné un nuage de points fini $\mathbb{X} \subset \mathbb{R}^d$ de cardinal n et un réel α positif, on note $B(x_i, \alpha)$ pour tout $x_i \in \mathbb{X}$ la boule centrée en x_i et de rayon α . On note $d(., .)$ la distance euclidienne.

Le **complexe de Rips** (plongé dans \mathbb{R}^d) d'ensembles de sommets \mathbb{X} et de paramètre α que l'on note $\text{Rips}(\mathbb{X}, \alpha)$ est le complexe simplicial dont les k -simplexes σ avec $k \leq d$ sont données par :

$$\sigma = \{x_0, \dots, x_k\} \in \text{Rips}(\mathbb{X}, \alpha) \iff B(x_i, \alpha) \cap B(x_j, \alpha) \neq \emptyset \iff d(x_i, x_j) \leq 2\alpha \quad \forall i, j \in \{0, \dots, k\}.$$

Comme pour le complexe de Cech, on obtient la filtration associée aux complexes de Rips lorsque α parcourt \mathbb{R}_+ .

Remarques :

$$\text{Cech}(\mathbb{X}, \alpha) \subseteq \text{Rips}(\mathbb{X}, \alpha) \text{ car } \bigcap_{i=0, \dots, k} B(x_i, \alpha) \neq \emptyset \implies B(x_i, \alpha) \cap B(x_j, \alpha) \neq \emptyset$$

Le complexe de Rips est plus facile à implémenter que celui de Cech puisqu'il revient juste à comparer deux nombres (la distance entre deux points et le paramètre α).

Définition(du Nerf) :

Soient $X \subseteq \mathbb{R}^d$. Le nerf d'un recouvrement $U = (U_i)_{i \in I}$ de X est le complexe simplicial $C(U)$ qui a U_i comme sommets et dont les simplexes sont définis comme suit :

$$\sigma = \{U_{i_0}, \dots, U_{i_k}\} \in C(U) \iff \bigcap_{j=0, \dots, k} U_{i_j} \neq \emptyset$$

Exemple : Le complexe de Cech est le nerf de l'union des boules $\bigcup_{x \in X} B(x, \alpha)$.

Théorème(du Nerf pour les convexes) :

Soient $X \subseteq \mathbb{R}^d$ une union finie d'ensembles fermés convexes $F = (F_i)_{i \in I}$ dans \mathbb{R}^d .

Alors X est homotope à $C(F)$.

Remarque : Par le théorème du Nerf, le complexe de Cech $C(X, \alpha)$ est homotope à l'union des boules $\bigcup_{x \in X} B(x, \alpha)$.

Ainsi il a un intérêt plus théorique par rapport à celui de Rips.



(1)

Figure 4 : [à gauche] Complexe de Cech dont les sommets sont extraits

autour d'un cercle pour un paramètre α et [à droite] complexe de Rips du même nuage de points que celui de Cech et pour le même paramètre α

3 L'homologie persistante

Une fois qu'on a défini une filtration de complexe simplicial sur les données notamment ceux de Rips (plus pratique) et de Cech (plus théorique) ou de façon plus abstraite une filtration d'espaces topologiques en définissant une fonction continue sur X et en considérant les sous-niveaux de celle-ci, cela nous permet de capturer de façon fiable la topologie de X en calculant l'homologie sur chaque échelle de la filtration et en étudiant ainsi l'évolution de celle-ci le long de la filtration.

3.1 Préliminaires

Dans toute la suite de ce document \mathbb{F} désignera le corps à deux éléments $\mathbb{Z}/2\mathbb{Z}$.

Soit K un complexe simplicial . On pose $K^{(i)} = \{\sigma \in K \mid \sigma \text{ est un } i\text{-simplexe}\}$

3.1.1 Définition (Chaines simpliciales)

On note $C_i(K)$ le \mathbb{F} -espace vectoriel des i -chaines sur K i.e librement engendré par K^i .

Une i -chaîne $c \in C_i(K)$ s'écrit donc comme suit : $c = \sum_{\sigma \in K^i} c_\sigma \sigma$ avec $c_\sigma \in \mathbb{F}$

On dispose d'une application bord naturelle linéaire. Par linéarité il suffit de la donner sur les i -simplexes qui engendrent $C_i(K)$:

$$\begin{array}{ccc} \partial : C_i(K) & \rightarrow & C_{i-1}(K) \\ \sigma & \mapsto & \sum \partial(\sigma) \end{array}$$

Lemme : $\partial \circ \partial = 0$.

On définit donc les notions suivantes :

1. l'espace vectoriel $Z_i(K)$ des i -cycles, i.e des i -chaines de bord nul : $Z_i(K) = \ker \partial$
2. l'espace vectoriel des i -bords $B_i(K) = \text{Im}(\partial : C_{i+1}(K) \rightarrow C_i(K))$

3. les i -èmes groupes d'homologie : $H_i(K) = Z_i(K)/B_i(K)$.

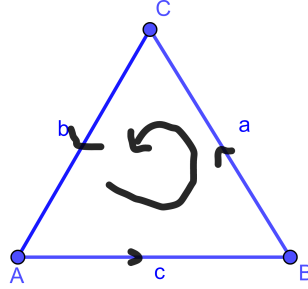


Figure 5(triangle non plein) : $C_{i \geq 2} = 0$ (il n'y a pas de i -simplexe pour $i \geq 2$), $C_1 = \mathbb{F}^3$ (on a trois 1-simplexes (a,b,c)), $C_0 = \mathbb{F}^3$ (on a trois 0-simplexes (A,B,C)).

$Z_0 = C_0 = \mathbb{F}^3$, $B_0 = \mathbb{F}^2$ (on a deux 1-simplexes par exemple a et b qui s'envoient sur le même bord C et un des trois 1-simplexe i.e c qui s'envoie sur les deux autres bords A et B) si bien que $H_0 = Z_0/B_0 = \mathbb{F}^3/\mathbb{F}^2 = \mathbb{F}$.

$Z_1 = \mathbb{F}$ (on a qu'un seul 1-cycle qui est le triangle $a+b+c$ (la seule 1-chaine qui est sans bord i.e dans le noyau)), $B_1 = 0$ (ya pas de 1-bord puisqu'il n'y a pas de 2-simplexe) si bien que $H_1 = Z_1/B_1 = \mathbb{F}/0 = \mathbb{F}$.

3.2 Modules de persistance

Une fois qu'on a une filtration, on peut à chaque echelle (ou temps) de la filtration calculer l'homologie de l'espace persistant correspondant et suivre ainsi l'évolution de celle-ci au cours du temps.

La notion de module de persistance est celle qui formalise algébriquement et encode cette évolution.

3.2.1 Définition (Module de persistance) :

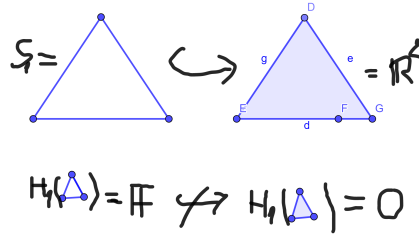
Un **module de persistance** sur une partie T de \mathbb{R} est la donnée d'une famille de \mathbb{F} -espaces vectoriels indexée par T $(M_t)_{t \in T}$ et des applications linéaires pour tout $s \leq t$, $m_s^t : M_s \longrightarrow M_t$ vérifiant

$$\begin{cases} m_s^s = id_{M_s} \quad \forall s \in T \\ \forall s \leq t \leq u : m_s^u = m_t^u \circ m_s^t \end{cases}$$

Exemples : Les n -ième groupes d'homologie des filtrations associées aux complexes de Rips $(H_n(R(\mathbb{X}_t)))_{t \in \mathbb{R}}$ ou de Čech $(H_n(C(\mathbb{X}_t)))_{t \in \mathbb{R}}$ et ceux des sous-niveaux d'une fonction continue sur X (comme nous l'avons défini dans la deuxième partie) $(H_n(X_t))_{t \in \mathbb{R}}$ sont des modules de persistance dont les morphismes sont induits par fonctorialité de l'homologie.

En effet $(R(\mathbb{X}_t))_{t \in \mathbb{R}}$ étant une filtration, on a : $R(\mathbb{X}_s) \subseteq R(\mathbb{X}_t)$ pour tout $s \leq t$ si bien qu'on a une application linéaire $m_s^t : H_n(R(\mathbb{X}_s)) \longrightarrow H_n(R(\mathbb{X}_t))$ vérifiant la structure d'un module persistant.

Remarque : l'inclusion entre les complexes ou espaces n'implique pas en général l'inclusion entre les groupes d'homologie.



3.2.2 Définition (morphisme de module de persistance) :

Soit M, N deux modules de persistance sur $T \subseteq \mathbb{R}$. Un morphisme $\varphi : M \rightarrow N$ est la donnée pour tout t , d'une application linéaire $\varphi_t : M_t \rightarrow N_t$ telle que $\forall s \leq t$, le diagramme suivant soit commutatif :

$$\begin{array}{ccc} M_s & \xrightarrow{m_s^t} & M_t \\ \varphi_s \downarrow & & \downarrow \varphi_t \\ N_s & \xrightarrow{n_s^t} & N_t \end{array}$$

i.e $\varphi_t \circ m_s^t = n_s^t \circ \varphi_s$.

On dira que $\varphi : M \rightarrow N$ est un isomorphisme si $\varphi_t : M_t \rightarrow N_t$ est un isomorphisme $\forall t \in T$.

3.2.3 Définition (somme directe de modules de persistance) :

Soit M et N deux modules de persistance sur $T \subseteq \mathbb{R}$. La somme directe $M \oplus N$ est définie comme suit : $\begin{cases} (M \oplus N)_t = M_t \oplus N_t, \forall t \in T \\ (m \oplus n)_s^t = m_s^t \oplus n_s^t, \forall s \leq t \end{cases}$

3.2.4 Définition :

On dit qu'un module de persistance M est décomposable s'il existe des modules $M_1, M_2 \neq 0$ tels que : $M \simeq M_1 \oplus M_2$ et indécomposable sinon.

Exemples :

- Soit M donné par $M(t) = \mathbb{F}^2 \forall t \in \mathbb{R}$ et $\forall s \leq t$, $m_s^t = \text{id} : \mathbb{F}^2 \rightarrow \mathbb{F}^2$.
Alors $M = M_1 \oplus M_2$ avec $M_1 = M_2$ donnée par $\forall t \in \mathbb{R}, M_1(t) = M_2(t) = \mathbb{F}$
et $\forall s \leq t$, $m_{1,s}^t = m_{2,s}^t = \text{id} : \mathbb{F} \rightarrow \mathbb{F}$.

- Soit M donné par $M(t) = \begin{cases} \mathbb{F} & \text{si } t < 0 \\ \mathbb{F}^2 & \text{si } t \in [0, 1[\\ \mathbb{F} & \text{si } t \geq 1 \end{cases}$ dont les morphismes sont
données par

$$m_s^t = \begin{cases} \text{id}_{\mathbb{F}} & \text{si } s \leq t < 0 \text{ ou } t \geq s \geq 1 \\ \text{id}_{\mathbb{F}^2} & \text{si } 0 \leq s \leq t < 1 \\ i : \mathbb{F} \hookrightarrow \mathbb{F}^2 & \text{si } s < 0 \text{ et } 0 \leq t < 1 \\ p : \mathbb{F}^2 \rightarrow \mathbb{F} & \text{si } 0 \leq s < 1 \text{ et } t \geq 1 \\ 0 : \mathbb{F} \rightarrow \mathbb{F} & \text{si } s < 0 \text{ et } t \geq 1 \end{cases} \quad \text{où } i \text{ est l'injection qui envoie } 1$$

sur $(1,0)$ et p est la projection qui envoie $(1,0)$ sur 0 . Alors $M = M_1 \oplus M_2$ où $M_1(t) = \mathbb{F}$ si $t \geq 0$ et 0 sinon et $\forall t \geq s \geq 0, m_{1,s}^t = \text{id}_{\mathbb{F}}$ et le morphisme nul sinon, et $M_2(t) = \text{id}_{\mathbb{F}}$ si $t < 1$ et $\forall s \leq t < 1, m_{2,s}^t = \text{id}_{\mathbb{F}}$ sinon le morphisme nul.

3.2.5 Définition (Module intervalle) :

Soit $I \subseteq \mathbb{R}$ un intervalle. On appelle module intervalle associé à I noté \mathbb{F}_I le module définit comme suit :

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

et dont les morphismes de structure sont $\text{id} : \mathbb{F}_I(s) \rightarrow \mathbb{F}_I(t)$ si $s \leq t$ et $s, t \in I$ sinon le morphisme nul.

Lemme : \mathbb{F}_I est indécomposable.

Preuve : S'il existe $M_1, M_2 \neq 0$ tels que : $\mathbb{F}_I = M_1 \oplus M_2$

Alors $\forall t \in I$, on a :

$$\mathbb{F} \simeq M_1(t) \oplus M_2(t)$$

Il suit donc que soit $M_1(t)$ est nul, soit $M_2(t)$ est nul.

Supposons $M_2(t)$ nul et montrons que $M_2(t')$ est nul $\forall t' \in I$

On prend $t' \in I$ avec $t \leq t'$ et on suppose $M_1(t') = 0$.

On regarde alors les morphismes de structure :

$$\text{id} : \mathbb{F} \rightarrow \mathbb{F} = 0 : M_1(t) \rightarrow 0 \oplus 0 : 0 \rightarrow M_2(t').$$

Ce qui est absurde !

Théorème(de structure des modules de persistance)[Gabriel,Auslander,Crawley-Boevey] :

Soit $M = (M_t)_{t \in T}$ un module de persistance ponctuellement de rang fini, i.e pour tout $t \in T$ M_t est de dimension finie. Alors M se décompose sous la

forme :

$$M \simeq \bigoplus_{I \in B} \mathbb{F}_I. (*)$$

où B est un multi-ensemble d'intervalle.

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

Remarque : Les modules de persistance associées aux filtrations des complexes de Cech et de Rips verifient les hypothèses du théorème : $\forall t \in \mathbb{R}$, $H_n(C(\mathbb{X}_t))$ et $H_n(R(\mathbb{X}_t))$ sont de dimension finie.

3.2.6 Définition :

Le multi-ensemble B est appelé **Code barre** de M que l'on note parfois $B(M)$.

Preuve de l'unicité de B dans $(*)$ dans le cas fini :

Supposons qu'il existe B et B' vérifiant $(*)$ avec $B \neq B'$. Il existe donc $I_0^n \in B$, $I_0^m \in B'$ (n et m désignant respectivement la multiplicité de I_0 dans B et B' comme multi-ensembles.) avec $n \neq m$.

On définit

$K = \{J \in B, I_0 \subset J \text{ (inclusion stricte)}\}$ et $K' = \{J' \in B', I_0 \subset J' \text{ (inclusion stricte)}\}$ et note k et k' leurs cardinaux respectifs.

Soit $t, t' \in I_0$ avec $t \leq t'$.

$$\text{Im}(m_t^{t'} : M_t \longrightarrow M_{t'})$$

$$\begin{aligned} &= \bigoplus_{t, t' \in I \in B} \mathbb{F}_I & &= \bigoplus_{t, t' \in I' \in B'} \mathbb{F}_{I'} \\ &= \bigoplus_{I_0 \in B} \mathbb{F}_{I_0} \bigoplus \bigoplus_{J \in K} \mathbb{F}_J & &= \bigoplus_{I_0 \in B'} \mathbb{F}_{I_0} \bigoplus \bigoplus_{J' \in K'} \mathbb{F}_{J'} \\ &= \mathbb{F}^n \bigoplus \mathbb{F}^k & &= \mathbb{F}^m \bigoplus \mathbb{F}^{k'}. \end{aligned}$$

Il suit donc que $n+k=n+k'$. comme $n \neq m$ alors $k \neq k'$.

Soient $t_1 \leq t \leq t_2$ avec $t \in I_0$, $t_1 \notin I_0$ et $t_1 \in J_-$ (intervalle qui contient strictement I_0 par valeurs inférieures indifféremment pris comme élément dans K ou K') et $t_2 \notin I_0$ et $t_2 \in J_+$ (intervalle qui contient strictement I_0 par valeurs supérieures indifféremment pris comme élément dans K ou K'). On note $K_+ \subseteq K$ le sous multi-ensemble qui ne contient que les intervalles J_+ (comme défini ci-dessus) et $K_- \subseteq K$ le sous multi-ensemble qui ne contient que les intervalles J_- (comme défini ci-dessus).

Et enfin on note K_{\mp} le sous multi-ensemble qui ne contient que les intervalles qui contiennent strictement I_0 par valeurs centrales.

Et on note k_+, k_-, k_{\mp} leurs cardinaux respectifs.

On définit de même K'_+, K'_-, K'_{\mp} pour K' ainsi que leurs cardinaux respectifs k'_+, k'_-, k'_{\mp} .

Remarque : $\{K_+, K_-, K_{\mp}\}$ est une partition de K tout comme $\{K'_+, K'_-, K'_{\mp}\}$ est une partition de K' . Donc on a $k = k_+ + k_- + k_{\mp}$ et $k' = k'_+ + k'_- + k'_{\mp}$.
On a :

$$\begin{aligned}
& \text{Im}(m_{t_1}^t : M_{t_1} \longrightarrow M_t) \oplus \text{Im}(m_t^{t_2} : M_t \longrightarrow M_{t_2}) \\
&= \bigoplus_{t_1, t \in I \in B} \mathbb{F}_I \oplus \bigoplus_{t, t_2 \in I \in B} \mathbb{F}_I = \bigoplus_{t_1, t \in I' \in B'} \mathbb{F}_{I'} \oplus \bigoplus_{t, t_2 \in I' \in B'} \mathbb{F}_{I'} \\
&= \bigoplus_{J \in K_-} \mathbb{F}_J \oplus \bigoplus_{J \in K_+} \mathbb{F}_J \oplus \bigoplus_{J \in K_{\mp}} \mathbb{F}_J = \bigoplus_{J' \in K'_-} \mathbb{F}_{J'} \oplus \bigoplus_{J' \in K'_+} \mathbb{F}_{J'} \oplus \bigoplus_{J' \in K'_{\mp}} \mathbb{F}_{J'} \\
&= \mathbb{F}^{k_-} \oplus \mathbb{F}^{k_+} \oplus \mathbb{F}^{k_{\mp}} = \mathbb{F}^{k'_-} \oplus \mathbb{F}^{k'_+} \oplus \mathbb{F}^{k'_{\mp}} \\
&= \mathbb{F}^k = \mathbb{F}^{k'} .
\end{aligned}$$

Il suit donc que $k=k'$: contradiction !

Exemples de calculs de codes de barres :

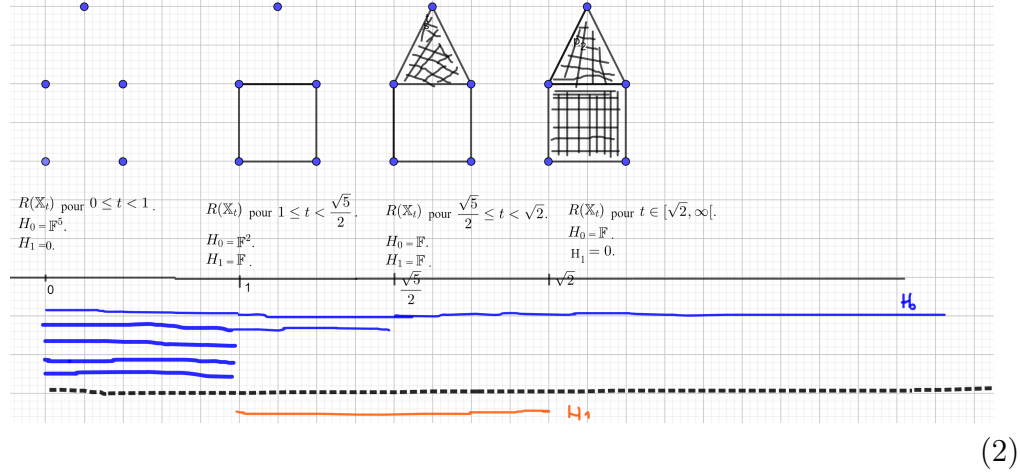


Figure 6 : Complexes de Rips associées à un nuage de cinq points pour un paramètre parcourant \mathbb{R}_+ et les groupes d'homologies associées en degré 0 et 1 à chaque échelle de la filtration ainsi que les codes barre associées :

- pour $t \in [0, 1[$, les boules ne se rencontrent pas. On a donc autant de composantes connexes qu'il ya de points donnés dans le nuage donc autant de classes d'homologie en degré 0 et aucune classe en degré 1. ce qui donne 4 barres en bleu pour H_0 et aucune barre pour H_1 .
- pour $t \in [1, \frac{\sqrt{5}}{2}[$, 4 boules se rencontrent (celles dont les centres forment le carré d'en bas) fusionnant ainsi leurs classes d'homologie en degré 0 en une seule (disons celle d'en bas à gauche) : autrement dit 3 classes sont tuées et 2 persistent. En même temps que des classes en degré 0 meurent, un 1-cycle i.e une classe d'homologie en degré 1 naît. D'où le prolongement des deux barres bleues (celles qui persistent) et une barre rouge apparaît.
- pour $t \in [\frac{\sqrt{5}}{2}, \sqrt{2}[$, toutes les boules se rencontrent donnant ainsi un complexe connexe. Cela se traduit en homologie de degré 0 par la persistance d'une seule classe : les deux qui ont survécu jusqu'ici se sont fusionnées en une seule (la même depuis le temps 0) et l'autre est tuée. Le 1-cycle est resté inchangé et aucun autre n'est créé donc la seule classe d'homologie en degré 1 qui a persisté jusqu'ici persiste encore. D'où une seule barre bleue continue et la barre rouge continue.
- pour $t \in [\sqrt{2}, \infty[$, rien a changé en homologie de degré 0 : le complexe une fois connexe va le rester pour toujours et à jamais le long de la

filtration. Ce qui donne une barre bleue infinie.

En même temps on a plus le 1-cycle i.e la seule classe d'homologie en degré 1 qui a persisté jusqu'ici est tuée ne laissant plus apparaître de barre rouge.

A la fin, comme indiqué sur la figure, on obtient les barres ;

Pour $H_0 : [0, \infty[$, $[0, 1[$ (3 fois) et $[0, \frac{\sqrt{5}}{2}[$.

Pour $H_1 : [1, \sqrt{2}[$

Donc on a pour cet exemple :

$$H_0 = \mathbb{F}_{[0,1[} \oplus \mathbb{F}_{[0,1[} \oplus \mathbb{F}_{[0,1[} \oplus \mathbb{F}_{[0, \frac{\sqrt{5}}{2}[} \oplus \mathbb{F}_{[0, \infty[}.$$

$$H_1 = \mathbb{F}_{[1, \sqrt{2}[}.$$

Remarque : Comme dans l'exemple précédent et de manière générale, le complexe de Čech et de Rips associé à un nuage fini représentant des données ne change (ajout de nouveaux simplexes) qu'un nombre fini de fois le long de la filtration. Le code barre est donc toujours fini dans ce cas. On peut donc écrire un algorithme pour calculer le code barre d'un complexe simplicial fini.

3.3 Algorithme de calcul du code barre d'un complexe simplicial

En pratique, la filtration $(F_t)_{t \in T \subseteq \mathbb{R}}$ de complexe simplicial associée aux données est une filtration qui ne change qu'un nombre fini de fois (les données bien qu'ils soient en grande quantité sont finies) et peut être renommée $(F_{t_k})_{k=1, \dots, n}$ où $t_1 < t_2 < \dots < t_n$ sont les échelles ou temps de changement (points où on a l'ajout de nouveaux simplexes)

L'algorithme prend donc en entrée un complexe simplicial fini filtré. Plusieurs simplexes peuvent venir à la fois, on va devoir les insérer un à un à la filtration pour pouvoir calculer son homologie. Ainsi on adapte la filtration en numérotant les simplexes de manière croissante de la façon suivante :

$\emptyset = F_0 \subseteq F_1 \subseteq \dots \subseteq F_n$ tel que :

— $F_{t+1} = F_t \cup \{\sigma_{t+1}\} \forall t = 0, \dots, N$. (on insère les simplexes un à un)

— $\sigma_s \subseteq \sigma_t \implies s \leq t$. (compatibilité avec la relation d'incidence)

On calcule l'homologie persistante du complexe au fur et à mesure qu'on ajoute les simplexes $\sigma_1, \dots, \sigma_N$ un à un. A la fin, on obtient des barres de la forme : $[i, j[$, où i correspond à l'indice d'un simplexe σ_i dont l'insertion crée une classe d'homologie (on dit qu'il est un simplexe positif) et j correspond

l'indice du simplexe σ_j dont l'insertion tue la classe de σ_i (on dit que c'est un simplexe négatif). Si la classe de σ_i n'est pas tuée, on obtient une barre infinie $[i, \infty[$. Pour revenir à la filtration de départ, on remplace i, j par son temps d'apparition. L'algorithme cherchera donc à mettre en pratique ce qui vient d'être décrit. Avant de l'écrire étudions l'exemple ci-après :

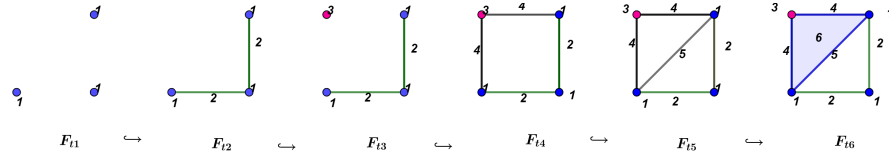


Figure 7 : Filtration d'un complexe simplicial fini avec les étapes sur lesquelles elle change (ajout de nouveaux simplexes). Comme on le voit sur la figure, au temps 1, la filtration change avec l'ajout de 3 simplexes de dimension 0, au temps 2 elle change avec l'ajout de 2 simplexes de dimension 1, au temps 3 elle change avec l'ajout d'un simplexe de dimension 0, au temps 4 elle change avec l'ajout de 2 simplexe de dimension 1, au temps 5 elle change avec l'ajout d'un simplexe de dimension 1 et enfin au temps 6 elle change avec l'ajout d'un simplexe de dimension 2. Donc plusieurs simplexes s'ajoutent en même temps. On va donc numéroter l'ensemble des simplexes de la filtration de façon croissante en respectant l'ordre d'arrivée et l'incidence. On obtient la filtration suivante :

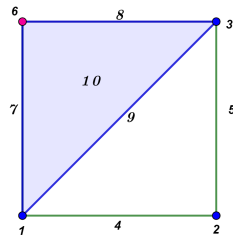


Figure 8 : Le complexe filtré de la figure 7 adapté par numéotation croissante des simplexes et en respectant la règle d'incidence.

Lorsque la numérotation croît de 1 à 10, on peut maintenant calculer son homologie persistance encodée par son code barre :

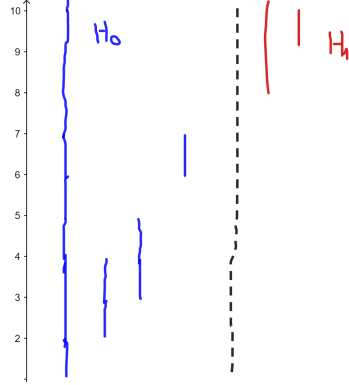


Figure 9 : Le code barre du complexe filtré de la figure 8.

En homologie de degré 0 : l'insertion successive des trois premiers simplexes crée chacun une classe d'homologie dont la deuxième et la troisième meurent en se fusionnent à la première respectivement à l'étape 4 et 5 pour ne laisser persister qu'une seule classe (la première) jusqu'à l'infini même si l'insertion du simplexe 6 crée à nouveau une classe d'homologie qui meurt à l'étape suivante.

En degré 1 : une classe d'homologie se crée seulement à partir de l'étape 8 et persiste jusqu'à l'étape suivante où une nouvelle classe se crée et meurt à son tour à l'étape suivante pour ne laisser survivre à l'infini que la première.

Ainsi on obtient les barres pour $H_0 : [1, \infty[, [2, 4[, [3, 5[, [6, 7[$ et pour $H_1 : [8, \infty[, [9, 10[$.

Pour finir, on revient à la filtration initiale en remplaçant le numéro de chaque simplexe de la filtration adaptée par son temps d'apparition dans la filtration initiale : $i \mapsto t_i$.

on a : $(1, 2, 3 \mapsto 1) ; (4, 5 \mapsto 2) ; (6 \mapsto 3) ; (7, 8 \mapsto 4) ; (9 \mapsto 5)$ et $(10 \mapsto 6)$

Les barres s'écrivent pour $H_0 : [1, \infty[, [1, 2[, [1, 2[, [3, 4[$ et pour $H_1 : [4, \infty[, [5, 6[$.

3.3.1 Algorithme :

Pour rendre effectif et implémentable l'algorithme on va encoder l'homologie persistante du complexe simplicial filtré adapté $(F_i)_{i=1,\dots,N}$ qu'il prend en entrée par sa matrice de bord. On calcule le bord du complexe au fur et à mesure qu'on insert un à un les simplexes de dimension variables.

3.3.2 Définition(matrice de bord) :

La matrice de bord de $(F_i)_{i=1,\dots,N}$ est la matrice $D_{i,j} \in M_N(\mathbb{F})$ lue dans la base des σ_i de l'opérateur de bord naturelle $\partial : C_*(F) \longrightarrow C_{*-1}(F)$

Remarques : σ_i est une face de σ_j ssi $D_{i,j} \neq 0$.

La matrice est triangulaire supérieure stricte car la numérotation croissante des simplexes respecte la relation d'incidence et que la face d'un simplexe est de dimension plus petite que celle-ci (exactement de co-dimension 1).

L'ajout d'un simplexe change potentiellement l'homologie du complexe (le simplexe crée ou tue une classe d'homologie). Il arrive que les simplexes se rejoignent par une même face (les bords sont liés). Ce qui se traduit par la présence des coefficients non nuls sur une même ligne. Il faut donc réduire la matrice pour pouvoir lire le code barre du complexe.

Définition : Pour toute matrice $A = (a_{i,j}) \in M_N$, on note $low_A(j)$ l'entier défini comme suit :

$$low_A(j) = \begin{cases} 0 & \text{si } \forall i, a_{i,j} = 0. \\ \max(i \mid a_{i,j} \neq 0) & \text{sinon.} \end{cases}.$$

L'algorithme va donc essentiellement reposer sur le calcul de la matrice de bord et sa réduction.

Algorithme :

Entrée : un complexe simplicial filtré fini adapté $(F_i)_{i=0,\dots,N}$ i.e $\forall i, F_{i+1} = F_i \cup \sigma_{i+1}$ (numérotation croissante des simplexes) et $\sigma_k \subseteq \sigma_l \implies k \leq l$ (préservation des règles d'incidence).

1) calculer $A=D_{i,j}$ la matrice de bord de F_i

2) pour j allant de 1 à N :

-tant qu'il existe $l < j$ avec $low_A(l) = low_A(j) \neq 0$

- réduire la matrice en remplaçant la j -eme colonne c_j de A par :

$$c_j - \frac{A_{low_A(j),j}}{A_{low_A(j),l}} c_l$$

3) nommer par \tilde{A} la matrice réduite obtenue à la fin du procédé 2) et par \tilde{c}_j ses colonnes.

4) extraire les j tels que $\tilde{c}_j = 0$

5) Pour tout j obtenu à la fin du procédé 4)

-s'il existe k tel que $j = low_{\tilde{A}}(k)$ alors retourner $[j, k[$

-sinon retourne $[j, \infty[$.

Théorème : *Le code barre de la filtration F_i est donnée par les j tels que $\tilde{c}_j = 0$ comme suit :*

- *s'il existe k tel que $j = low_{\tilde{A}}(k)$, on a la barre $[j, k[$ en degré $\dim \sigma_j$*
- *sinon on a la barre $[j, \infty[$ en degré $\dim \sigma_j$*

Preuve : La matrice de bord réduite encode la naissance et la mort de chaque classe d'homologie. Plus précisément si l'insertion du simplexe σ_j crée une classe j , cela se traduit sur la matrice réduite par la nullité de la colonne \tilde{c}_j .

Alors des deux choses, l'une : soit la classe j est tuée par l'insertion d'un simplexe σ_k (σ_j est une face de σ_k , donc $j \leq k$) i.e $j = low_{\tilde{A}}(k)$. Et le couple persistant (j, k) correspond à la naissance et la mort d'une classe en degré $\dim \sigma_j$, ce qui induit donc une barre $[j, k[$ finie en degré $\dim \sigma_j$.

soit la classe j n'est pas tuée le long de la filtration et cela se lit sur la matrice réduite par le fait qu'il n'existe pas de colonne \tilde{c}_k tel que $j = low_{\tilde{A}}(k)$. On a donc une barre $[j, \infty[$ infinie en degré $\dim \sigma_j$.

Exemple : On considère le complexe filtré suivant adapté pour l'algorithme

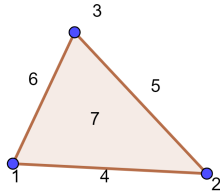


Figure 10 : triangle plein filtré pour l'algorithme.

On calcule la matrice de bord et on la réduit :

$$\begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ & & & 1 & & 1 & \\ & & & 1 & 1 & & \\ & & & & 1 & 1 & \\ & & & & & 1 & \\ & & & & & & 1 \\ & & & & & & 1 \end{pmatrix} \xrightarrow{\text{réduction}} \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ & & & 1 & & & \\ & & \boxed{1} & \boxed{1} & 1 & & \\ & & & \boxed{1} & & & \\ & & & & \boxed{1} & & \\ & & & & & 1 & \\ & & & & & & \boxed{1} \\ & & & & & & \boxed{1} \end{pmatrix}$$

Les barres sont données par les colonnes nulles 1,2,3 et 6 en degré de la dimension respective des simplexes correspondants.

Ainsi on a en dimension 0 :

- la colonne 1 et il n'existe pas de colonne k telle que $1 = low(k)$

ce qui donne donc la barre $[1, \infty[$.

-la colonne 2 et $2 = low(4)$ d'où la barre $[2, 4[$

- et enfin la colonne 3 et $3 = low(5)$ d'où la barre $[3, 5[$

En dimension 1, on a la colonne 6 et $6 = low(7)$ ce qui donne donc la barre $[6, 7[$.

Pour résumer les barres sont :

-pour $H_0 : [2, 4[; [3, 5[; [1, \infty[$.

-pour $H_1 : [6, 7[$.

3.4 Stabilité

Nous rappelons qu'on peut associer de façon unique à chaque module \mathbb{M} un code barre B . On peut ainsi construire l'application «code barre en degré n » :

$\mathbb{M} \longmapsto B_n(\mathbb{M})$ où ici $\mathbb{M} = (H_n(C(\mathbb{X}_t)))_t$ ou $\mathbb{M} = (H_n(R(\mathbb{X}_t)))_t$ ou $\mathbb{M} = (H_n(f^{-1}(\cdot - \infty, t)))_t$

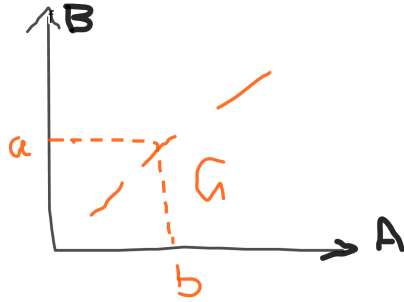
Nous allons définir une distance sur les codes barres et les comparer et voir que le code barre est un descripteur stable pour cette distance permettant de discriminer des données en ce sens qu'il encode entièrement et fidèlement l'homologie de l'espace sous-jacent aux données.

3.4.1 Définition (Accouplement partiel) :

Soient A et B deux multi-ensembles de barres. On appelle accouplement partiel de A vers B noté $A \xleftarrow{G} B$ un sous-ensemble $G \subset A \times B$ tel que les projections naturelles $G \longrightarrow A$ et $G \longrightarrow B$ soient injectives.

On appellera un élément $(a, b) \in G$ une paire accouplée et à contrario un élément $x \in A \sqcup B$ ne correspondant pas à un couple accouplé est un élément non-accouplé.

On peut voir G comme un graphe partiel qu'on l'on peut représenter comme suit :



Concrètement, on a :

$\forall a \in A, \exists$ au plus un $b \in B$ tel que $(a, b) \in G$.

$\forall b \in B, \exists$ au plus un $a \in A$ tel que $(a, b) \in G$.

3.4.2 Définition (Coût d'un accouplement partiel) :

Soit G un accouplement partiel de A vers B deux multi-ensembles de barres.

Pour toute paire accouplée $(a, b) \in G$ et tout élément $x \in A \sqcup B$ non accouplée. on appelle :

- coût de (a, b) noté $c(a, b) := \|a - b\|_\infty = \left\| \begin{pmatrix} a_0 - b_0 \\ a_1 - b_1 \end{pmatrix} \right\|_\infty = \max(|a_0 - b_0|, |a_1 - b_1|)$ où a_i, b_i sont les extrémités respectifs de a, b
- coût de x noté $c(x) := |x_0 - x_1|$ où x_i sont les extrémités de l'intervalle x .

— coût total de G noté $c(G) := \max(c(a, b), c(x), (a, b) \text{ paires accouplées}, x \text{ non-accouplées.})$

Remarque : $c(G) \in \mathbb{R}_+ \cup \{+\infty\}$.

Exemple :

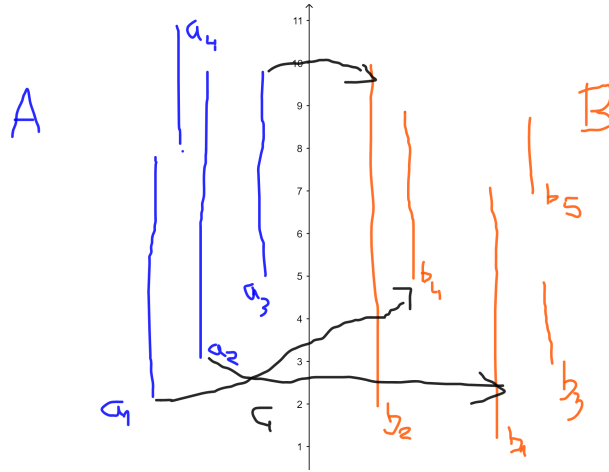


Figure 11 : Exemple d'accouplement partiel G entre deux codes barres A et B . Comme indiqué sur la figure $G = \{(a_1, b_4), (a_2, b_1), (a_3, b_2)\}$.

Pour calculer $c(G)$, on calcule d'abords :

$$\begin{aligned}
 - c(a_1, b_4) &= \left\| \begin{pmatrix} 8 - 9 \\ 2 - 5 \end{pmatrix} \right\|_{\infty} = 3; \\
 - c(a_2, b_1) &= \left\| \begin{pmatrix} 10 - 7 \\ 3 - 1 \end{pmatrix} \right\|_{\infty} = 3; \\
 - c(a_3, b_2) &= \left\| \begin{pmatrix} 10 - 10 \\ 5 - 2 \end{pmatrix} \right\|_{\infty} = 3; \\
 - c(a_4) &= |8 - 11| = 3; \\
 - c(b_3) &= |3 - 5| = 2; \\
 - c(b_5) &= |7 - 9| = 2.
 \end{aligned}$$

$$c(G) = \max(c(a_1, b_4), c(a_2, b_1), c(a_3, b_2), c(a_4), c(b_3), c(b_5)) = \max(3, 2) = 3.$$

3.4.3 Définition(distance Bottleneck)

Soient A et B deux multi-ensembles de barres. La distance Bottleneck notée $d_{\mathcal{B}}$ entre A et B est :

$$d_{\mathcal{B}}(A, B) := \inf_{\substack{A \xleftrightarrow{G} B \\ \text{partial matching}}} (c(G)).$$

Exemple : On considère les sous-niveaux de deux fonctions f et g sur un espace X et les codes barres associés à $H_0(f^{-1}(]-\infty, t]))_t$ et $H_0(g^{-1}(]-\infty, t]))_t$

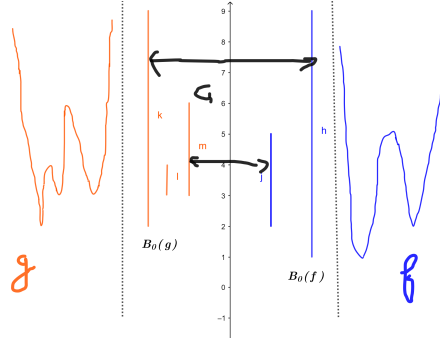


Figure 12 :Code barre d'une fonction f dont g est une perturbation (f et g sont "proches" dans la topologie de la norme infinie) et un accouplement partiel $B_0(g) \xleftrightarrow{G} B_0(f)$ avec $G = \{((k, h), (m, j))\}$ comme indiqué sur la figure.

Pour calculer $d_{\mathcal{B}}(B_0(g), B_0(f))$, on doit déterminer tous les accouplements partiels entre $B_0(g)$ et $B_0(f)$ et calculer leur coût. Or pour cet exemple, il est clair qu'il faut accoupler les deux barres infinies pour que le coût soit fini. En accouplant les deux barres infinies, on a que trois accouplements partiels possibles : $G; \{(k, h)\}; \{(k, h), (i, j)\}$ dont le coût minimum est réalisé par G .

D'où $d_{\mathcal{B}}(B_0(g), B_0(f)) = c(G) = 1$.

3.4.4 Théorème de stabilité I

Est ce que la distance Bottleneck traduit fidèlement la topologie de X ? Le théorème de stabilité apporte une réponse à cette question.

On dira qu'une fonction est ponctuellement de dimension finie si son homologie persistante associée (l'homologie associée aux sous-niveaux) est ponctuellement de dimension finie.

Remarque : Ce sera en pratique toujours le cas pour des fonctions associées à des données.

On va noter le code barre de l'homologie persistante en degré n associée aux sous-niveaux $B((H_n(f^{-1}[\cdot - \infty, t])))_t$ simplement par $B_n(f)$.

Théorème : Soient $f, g : X \rightarrow \mathbb{R}$ deux fonctions ponctuellement de dimension finie.

Pour tout entier n , on a :

$$d_{\mathcal{B}}(B_n(f), B_n(g)) \leq d(f, g).$$

où $d(f, g) := \|f - g\|_{\infty} = \sup_{x \in X} |f(x) - g(x)|$.

Les codes barres de l'homologie persistante de deux fonctions en tout degré n sont aussi proches que le sont ces deux fonctions dans la topologie de $C^0(X, \mathbb{R})$ muni de la $\|\cdot\|_{\infty}$.

Une idée de preuve : On admet les résultats de la théorie de l'homologie de Morse qui est une théorie beaucoup plus fondamentale et dépasse de loin le cadre de ce TER. On admet notamment que les points critiques localisent les lieux de naissance et de mort des classes d'homologie en ce sens que les valeurs critiques correspondent aux temps de naissance et de mort d'une classe.

On suppose X compacte (ce qui sera toujours le cas en pratique pour les données) et f, g continues.

f, g sont donc bornées et atteignent leurs bornes d'où la finitude de $d(f, g)$ d'une part et d'autre part on a que : $f^{-1}(\mathbb{R}) = g^{-1}(\mathbb{R}) = X$. Donc $B_n(f), B_n(g)$ encodent l'homologie du même espace X , leur distance Bottleneck ne peut

être donc infinie.

On peut donc dire que les barres de l'homologie persistante associée aux sous-niveaux de f sont de la forme $[f(s), f(t)[$ pour une barre finie et $[f(s), \infty[$ (ou $]\infty, f(s)]$) pour une barre infinie où s et t sont des points critiques de f . Il suit que :

le coût d'un accouplement partiel est de la forme $|f(s) - g(t)|$ (s point critique de f et t point critique de g) ou de la forme $|f(s_1) - f(s_2)|$ ou $|g(t_1) - g(t_2)|$ (s_i points critiques de f et t_i points critiques de g)

$$d_{\mathcal{B}}(B_n(f), B_n(g)) = \inf\{|f(s) - g(t)|, |f(s_1) - f(s_2)|, |g(t_1) - g(t_2)|\} \leq d(f, g)$$

3.4.5 Théorème de stabilité II

Cas des nuages de points :

On va devoir définir une métrique pour comparer des nuages de points qui à priori ne vivent pas dans un même espace.

3.4.6 Définition(*distance de Hausdorff*)

Soient K, L deux sous-espaces compacts d'un espace métrique (X, d) . La distance de Hausdorff est :

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

où $d(x, L) = \min_{y \in L} d(x, y)$.

On généralise cette définition au cas où K et L sont des espaces métriques abstraits en les plongeant isométriquement dans un même espace métrique Z .

On appelle plongement isométrique $\gamma : (X, d) \hookrightarrow (Y, d')$ une application continue isométrique qui est un homéomorphisme sur son image.

3.4.7 Définition (*distance de Gromov-Hausdorff*)

Soient X, Y deux espaces métriques. La distance de **Gromov-Hausdorff** est :

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

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

Exemple :

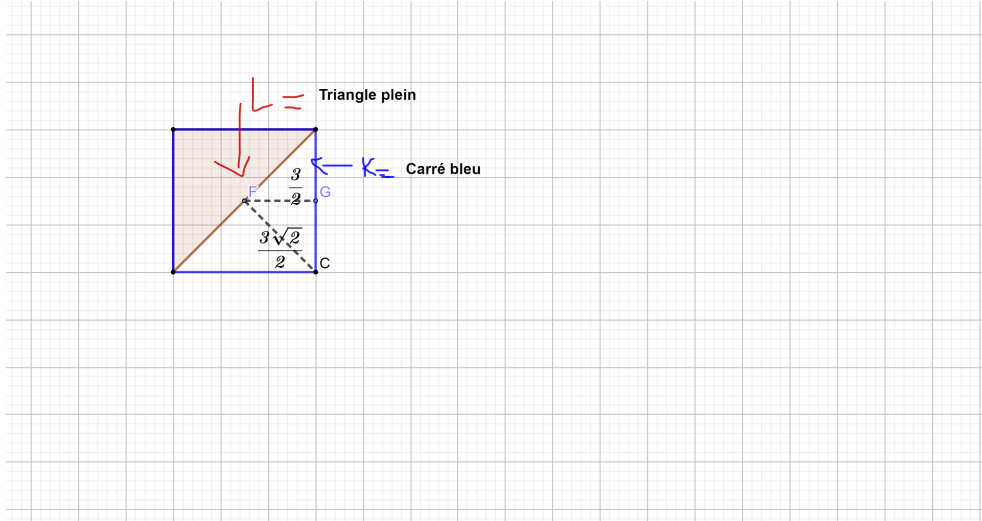


Figure 13 : On calcule la distance de Hausdorff entre deux compacts K et L .

$$\begin{aligned} - \max_{x \in K} (d(x, L)) &= d(C, F) = \frac{3\sqrt{2}}{2} \\ - \max_{y \in L} (d(y, K)) &= d(F, G) = \frac{3}{2}. \end{aligned}$$

$$d'où d_{\mathcal{H}}(K, L) = \max\left(\frac{3\sqrt{2}}{2}, \frac{3}{2}\right) = \frac{3\sqrt{2}}{2}.$$

Théorème : Soient \mathbb{X}, \mathbb{Y} deux nuages de points. On note $B_n(\mathbb{X})$ et $B_n(\mathbb{Y})$ les codes barres en degré n de l'homologie persistante associée respectivement aux filtrations de complexes de Rips $(R(\mathbb{X}_t))_t$ et $(R(\mathbb{Y}_t))_t$.

Pour tout entier $n \in \mathbb{N}$, on a :

$$d_{\mathcal{B}}(B_n(\mathbb{X}), B_n(\mathbb{Y})) \leq d_{\mathcal{G}, \mathcal{H}}(\mathbb{X}, \mathbb{Y}).$$

Des nuages de points qui discrétisent le même espace ont des homologies persistantes associées dont les codes barres sont proches. En d'autres termes, des objets "géométriquement proches" ont des codes barres proches.

Conclusion

Pour conclure, l'homologie persistante est une théorie très intéressante à plusieurs égards qui promet d'importantes innovations dans le domaine des sciences des données (et peut-être pas que) et que personnellement, nous la trouvons d'une beauté sidérante.

D'une part on ne peut être que stupéfait par l'intérêt pratique de la théorie. En partant seulement d'un nuage de points qui paraît informe, on en extrait une information pertinente par un descripteur assez puissant qu'est le code barre en se ramenant à de simples notions d'algèbre linéaire. On est d'autant plus content qu'on ait trouvé une métrique qui rend fiable ce descripteur et donc stable par perturbation. C'est ce que traduisent et garantissent les théorèmes de stabilité.

D'autre part on imagine assez facilement les champs d'application de la théorie. On peut penser à la reconnaissance et la classification de formes (en imagerie médicale par exemple), à l'étude (porosité par exemple) des nano-structures en physique des matériaux ou encore à la connectivité d'un réseau (informatique par exemple). Il ya là seulement quelques champs d'application pour les quels on a déjà appliqué l'homologie persistante mais la liste est plus longue.

Toutefois, il reste plusieurs perspectives de recherche à mener pour explorer d'avantage la théorie. L'échelle de filtration à la quelle on attient l'homologie de l'espace sous-jacent des données ou comprendre l'espace des codes barres en la munissant d'une structure algébrique et (ou) topologique et (ou) d'une mesure afin de faire des statistiques sont quelques exemples qui méritent d'être au moins étudiés.

Références

- [1] Grégory Ginot, Cours de Topologie Algébrique - Ecole Polytechnique. <https://www.math.univ-paris13.fr/ginot/MAT557TA/>
- [2] Grégory Ginot, Cours de TDA, Master 2 de Mathématiques des données - Université Sorbonne Paris Nord
- [3] Chazal Boissonnat Yvinec, Computational Topology Inference
- [4] Steve Oudot, Cours TDA-Ecole polytechnique. <https://geometrica.saclay.inria.fr/team/Steve.Oudot/>
- [5] Analysis Situs, Homologie : Une invitation à l'homologie persistante. <https://analysis-situs.math.cnrs.fr/Une-invitation-a-l-homologie-persistante.html>