

Chapitre 4

Persistance homologique

4.1 Motivation

La notion de persistance homologique apparaît dans les articles de Robins [Rob99] puis d'Edelsbrunner et al. [ELZ00] dans le cadre de la théorie de l'approximation. Cette notion a donné suite à de nombreux développements [ZC05, CSEM06, EMP06, CSEH07, CSEH08, EH08]. Le principe est d'encoder un processus d'évolution d'un objet plutôt que l'objet lui-même. Pour commencer, on se donne une filtration d'un complexe simplicial K :

$$K_1 \subset K_2 \subset \dots \subset K_n = K$$

Cette filtration induit d'après la section 3.4 une suite de morphismes en homologie :

$$H_*(K_1) \rightarrow \dots \rightarrow H_*(K_n)$$

où chaque flèche est induite par l'inclusion. Clairement, cette suite de morphismes est un invariant plus riche pour la filtration que la simple collection des groupes d'homologie. On pourra considérer par exemple deux filtrations de longueur 2 du tore commençant par un cycle contractile ou non.

En travaillant avec des coefficients dans un corps on obtient ainsi une chaîne d'applications linéaires. On définit de manière naturelle une notion de morphismes entre deux telles chaînes : c'est une suite d'applications linéaires entre les espaces des deux chaînes telle que le diagramme global commute. La question se pose alors de classier les chaînes à isomorphisme près. On va montrer que la notion d'intervalle de persistance fournit un invariant complet.

4.2 Classification des chaînes d'applications linéaires

4.2.1 Décomposition canonique

On se donne une chaîne d'applications linéaires de longueur n entre des espaces vectoriels E_i de dimensions finies sur un corps \mathbb{F} donné :

$$E_1 \xrightarrow{\phi_1} \dots \xrightarrow{\phi_n} E_{n+1} \quad (4.1)$$

Pour tout $(i, j) \in [1, n+1]^2$ on définit $\phi_{i,j} : E_i \rightarrow E_j$ comme la composition des ϕ_k entre E_i et E_j . Plus précisément, on pose

- $\forall i \in [1, n] : \phi_{i,i} = \text{Id}_{E_i}$
- $\forall 1 \leq i < j \leq n+1 : \phi_{i,j} = \phi_{i+1,j} \circ \phi_i$ et $\phi_{j,i} = 0$.

Exemple 4.2.1 On note $\text{Id}_{\mathbb{F}}(i, j)$ la chaîne

$$\underbrace{0 \rightarrow \dots \rightarrow 0}_{(i-1) \times 0} \rightarrow \underbrace{\mathbb{F} \xrightarrow{\text{Id}} \dots \xrightarrow{\text{Id}} \mathbb{F}}_{(j-i+1) \times \mathbb{F}} \rightarrow \underbrace{0 \rightarrow \dots \rightarrow 0}_{(n-j+1) \times 0}$$

C'est-à-dire avec les notations ci-dessus : E_k est l'espace de dimension 1 sur \mathbb{F} pour $k \in [i, j]$ et 0 sinon, tandis que ϕ_k est l'identité pour $k \in [i, j[$ et l'application nulle sinon.

Le résultat principal de cette section est une décomposition canonique des chaînes d'applications linéaires.

Théorème 4.2.2 Pour toute chaîne $(\phi_i)_{1 \leq i \leq n}$ de longueur n de la forme (4.1), il existe un unique multi-ensemble d'intervalles, I , tel que $(\phi_i)_{1 \leq i \leq n}$ est isomorphe à la somme directe

$$\bigoplus_{[i,j] \in I} \text{Id}_{\mathbb{F}}(i, j) \quad (4.2)$$

où chaque intervalle $[i, j]$ apparaît dans cette somme avec sa multiplicité dans I . Les intervalles de I sont appelés intervalles de persistance de $(\phi_i)_{1 \leq i \leq n}$.

Dans cette somme directe l'espace d'indice k est la somme directe des espaces – ici 0 ou \mathbb{F} – de même indice dans les $\text{Id}_{\mathbb{F}}(i, j)$. Les applications linéaires sont également obtenus par somme directe des applications – ici 0 ou l'identité – de même indice.

Soit I le multi-ensemble d'intervalles de persistances du théorème. On note $n_{i,j}$ la multiplicité de l'intervalle de persistance $[i, j]$ dans I . Pour tout $1 \leq i \leq j \leq n$ on pose $\beta_{i,j} = \text{rang}(\phi_{i,j})$. On vérifie aisément que

$$\beta_{i,j}(\text{Id}_{\mathbb{F}}(k, l)) = \begin{cases} 1 & \text{si } [i, j] \subset [k, l] \\ 0 & \text{sinon.} \end{cases}$$

et que si $(\phi_k)_{1 \leq k \leq n}$ et $(\psi_k)_{1 \leq k \leq n}$ sont deux chaînes d'applications :

$$\beta_{i,j}((\phi_k)_k \oplus (\psi_k)_k) = \beta_{i,j}((\phi_k)_k) + \beta_{i,j}((\psi_k)_k)$$

Lemme 4.2.3 Pour tout $1 \leq i < j \leq n+1 : n_{i,j} = \beta_{i,j} - \beta_{i-1,j} - (\beta_{i,j+1} - \beta_{i-1,j+1})$.

Preuve : D'après ce qui précède, chaque $\text{Id}_{\mathbb{F}}(k, \ell)$ de la décomposition contribue pour 1 dans $\beta_{i,j}$ si et seulement si $[i, j] \subset [k, \ell]$. Dit autrement, $\beta_{i,j}$ compte le nombre d'intervalles de persistance contenant $[i, j]$. Par suite, $\delta_{i,j} := \beta_{i,j} - \beta_{i-1,j}$ compte le nombre d'intervalles de persistance de la forme $[i, \ell]$ avec $\ell \geq j$. Et finalement $n_{i,j} = \delta_{i,j} - \delta_{i,j+1}$. \square

Le lemme montre l'unicité du multi-ensemble d'intervalles dans le théorème. L'existence, montrée ci-dessous, repose sur la notion de base compatible.

Exercice 4.2.4 *On dit qu'une chaîne est décomposable si elle est somme directe de deux chaînes non-triviales (différentes de $0 \rightarrow 0 \rightarrow \dots$). Dans le cas contraire, la chaîne est dite indécomposable.*

1. *Montrer par une récurrence sur $\sum \dim E_i$ que toute chaîne du type (4.1) est somme directe d'indécomposables.*
2. *Montrer que pour tout i, j , $\text{Id}_{\mathbb{F}}(i, j)$ est indécomposable.*
3. *Montrer qu'une chaîne qui n'est pas isomorphe à une $\text{Id}_{\mathbb{F}}(i, j)$ est décomposable.*
4. *En déduire que toute chaîne admet une décomposition de la forme (4.2).*

4.2.2 Bases compatibles

Pour $x \in E_i$, on pose

$$x(j) = \phi_{i,j}(x)$$

Définition 4.2.5 *Une base compatible pour la chaîne (4.1) est la donnée d'une famille $X \subset \bigcup_i E_i$, l'union étant considérée comme disjointe, telle que pour $1 \leq i \leq n+1$ la famille*

$$X(i) := \{x(i) \mid (x \in X) \wedge (x(i) \neq 0)\}$$

est une base de E_i et pour x, y distincts dans X , $x(i) \neq 0 \implies x(i) \neq y(i)$.

Définition 4.2.6 *À tout élément x d'une base compatible, on associe son intervalle de persistance $I_x := \{i \mid x(i) \neq 0\}$.*

Lemme 4.2.7 *Les intervalles de persistance des éléments d'une base compatible coïncident avec les intervalles de persistance du théorème 4.2.2.*

Preuve : Il suffit d'adapter la preuve du lemme 4.2.3 en définissant cette fois n_{ij} comme la multiplicité de l'intervalle $[i, j]$ parmi les intervalles de persistance des éléments d'une base compatible X et en remarquant que $\beta_{i,j} = |\{x \in X \mid [i, j] \subset I_x\}|$. \square

Par la suite on utilise par commodité la fonction d'activation $a : X \rightarrow [1, n+1]$ telle que $\forall x \in X : x \in E_{a(x)}$.

Proposition 4.2.8 *Toute chaîne du type (4.1) admet une base compatible.*

Preuve : On procède par récurrence sur la longueur de la chaîne. Pour une longueur nulle (E_1 seul est donné), une base quelconque de E_1 fournit une base compatible. On considère une chaîne du type (4.1) de longueur $n > 0$ et on suppose que X est une base compatible pour la sous-chaîne de longueur $n - 1$:

$$E_1 \xrightarrow{\phi_1} \dots \xrightarrow{\phi_{n-1}} E_n \quad (4.3)$$

Soit $k = |X|$ le nombre d'éléments de X . On définit de manière récursive des bases compatibles $X_1 = X, X_2, \dots, X_k$ pour (4.3). Le but est d'obtenir une base compatible X_k dont les éléments s'envoient dans E_{n+1} soit sur 0 soit sur une sous-famille libre de E_{n+1} . Pour cela on commence par ordonner les éléments x_1, x_2, \dots, x_k de X de manière non-décroissante pour l'activation, i.e. $a(x_j) \leq a(x_{j+1})$. On suppose avoir construit $X_{i-1} = \{y_1, y_2, \dots, y_k\}$ pour un $i > 1$, l'ordre des y_j étant non-décroissant pour l'activation, de sorte que les images non-nulles dans E_{n+1} de $\{y_1, y_2, \dots, y_{i-1}\}$ forment une famille libre.

- Si $y_i(n+1)$ ($= \phi_n(y_i(n))$) est nul ou linéairement indépendant de $\{y_j(n+1)\}_{j < i}$ alors on pose $X_i = X_{i-1}$,
- sinon, on peut écrire $y_i(n+1) = \sum_{j < i} \lambda_j y_j(n+1)$ et on pose $y'_i = y_i - \sum_{j < i} \lambda_j y_j(i)$ et

$$X_i = X_{i-1} \setminus \{y_i\} \cup \{y'_i\}$$

Clairement X_i est une base compatible pour (4.3) : pour $1 \leq j \leq n$, la famille $X_i(j)$ (cf. définition 4.2.5), obtenue par changement de bases élémentaire à partir de la famille $X_{i-1}(j)$, est une base de E_j . De plus, par construction, les images non-nulles dans E_{n+1} des i premiers éléments de X_i forment une famille libre.

On obtient par récurrence une base compatible X_k de (4.3) telle que les éléments non nuls de $\{x(n+1)\}_{x \in X_k}$ forment une famille libre dans E_{n+1} .

Il reste à compléter X_k par une base d'un supplémentaire de $\phi_n(E_n)$ pour obtenir une base compatible pour (4.1). \square

Preuve de l'existence de la décomposition canonique du théorème 4.2.2 : Notons tout d'abord qu'une décomposition canonique de la forme (4.2) admet une base compatible formée, pour chaque $\text{Id}_{\mathbb{F}}(i, j)$ de la décomposition, d'un générateur de \mathbb{F}^1 dans l'espace d'indice i . L'intervalle de persistance d'un tel générateur est précisément $[i, j]$. On note X_I une base compatible de (4.2) ainsi formée.

Soit une chaîne de la forme (4.1). Par la proposition 4.2.8, on peut choisir une base compatible X pour cette chaîne. Soit I le multi-ensemble des intervalles de persistance des éléments de X . On note F_i l'espace d'indice i dans la décomposition canonique et on considère une bijection $f : X \rightarrow X_I$ qui respecte les intervalles de persistance (cf. lemme 4.2.7). Cette bijection induit pour chaque i un morphisme $f_i : E_i \rightarrow F_i$ obtenu en envoyant la base de E_i associée à X (cf. définition 4.2.5) sur la base correspondante de F_i via $f : f_i(x(i)) = f(x)(i)$. Clairement les f_i forment une application de chaîne et l'inverse de f induit l'application inverse. \square

4.2.3 Persistance des sous-chaînes

On considère une chaîne $(\phi_i)_{1 \leq i \leq n}$:

$$E_1 \xrightarrow{\phi_1} \dots \xrightarrow{\phi_n} E_{n+1}$$

et une injection croissante $\kappa : [1, m+1] \rightarrow [1, n+1]$. On définit une *sous-chaîne* $(\phi'_j)_{1 \leq j \leq m}$:

$$E'_1 \xrightarrow{\phi'_1} \dots \xrightarrow{\phi'_m} E'_{m+1}$$

de la chaîne $(\phi_i)_{1 \leq i \leq n}$ en posant $E'_j = E_{\kappa(j)}$, pour $j \in [1, m+1]$, et en posant, avec les notations de la section 4.2.1, $\phi'_j = \phi_{\kappa(j), \kappa(j+1)}$, pour $j \in [1, m]$. On pose de plus

$$\begin{aligned} \lambda : [1, \kappa(m+1)] &\rightarrow [1, m+1] \\ i &\mapsto \min\{j \in [1, m+1] \mid \kappa(j) \geq i\} \end{aligned}$$

et

$$\begin{aligned} \mu : [\kappa(1), n+1] &\rightarrow [1, m+1] \\ i &\mapsto \max\{j \in [1, m+1] \mid \kappa(j) \leq i\} \end{aligned}$$

où l'on convient que $\min \emptyset = m+1$ et $\max \emptyset = 1$.

Lemme 4.2.9 *Soit I le multi-ensemble des intervalles de persistance de la chaîne $(\phi_i)_{1 \leq i \leq n}$. Alors le multi-ensemble des intervalles de persistance de $(\phi'_j)_{1 \leq j \leq m}$ est $\{[\lambda(i), \mu(j)] \mid [i, j] \in I \text{ et } \lambda(i) \leq \mu(j)\}$.*

Preuve : Il suffit de partir d'une décomposition canonique de $(\phi_i)_{1 \leq i \leq n}$ de la forme (4.2) et de remarquer qu'une composition d'applications identité ou nulle reste de l'un de ces deux types. Nous laissons le lecteur combler les détails. Une autre preuve consiste à partir d'une base compatible X pour $(\phi_i)_{1 \leq i \leq n}$. On note $a : X \rightarrow [1, n+1]$ la fonction d'activation telle que $\forall x \in X : x \in E_{a(x)}$ et on pose $a'_x = \kappa \circ \lambda(a(x))$ si $a(x) \leq \kappa(m+1)$. Alors, avec les notations de la section 4.2.2, il est immédiat que la famille $X' = \{x(a'_x) \mid x \in X \text{ et } a(x) \leq \kappa(m+1) \text{ et } x(a'_x) \neq 0\}$ est une base compatible pour $(\phi'_j)_{1 \leq j \leq m}$. Notons que $x(a'_x) \in E_{a'_x} = E'_{\kappa^{-1}(a'_x)} = E'_{\lambda(a(x))}$. De plus si l'intervalle de persistance (cf. définition 4.2.6) de $x \in X$ est $[a_x, b_x]$, celui de $x(a'_x) \in X'$ est $[\lambda(a_x), \mu(b_x)]$. \square

4.3 Persistance des filtrations de complexes simpliciaux

Soit K un complexe simplicial et \mathcal{K} une filtration :

$$K_1 \subset K_2 \subset \dots \subset K_n = K \tag{4.4}$$

Les inclusions de cette filtration induisent une chaîne de morphismes entre les groupes d'homologie de chaque complexe de la filtration :

$$H(K_1) \rightarrow H(K_2) \rightarrow \dots \rightarrow H(K_n) \tag{4.5}$$

On considère l'homologie sur un corps \mathbb{F} , de sorte que les $H(K_i)$ sont des espaces vectoriels. Selon l'usage on note $C(K_i)$ (resp. $Z(K_i)$, resp. $B(K_i)$) l'espace des chaînes (resp. cycles, resp. des bords) de K_i . On omet volontairement la dimension pour ne pas alourdir les notations. On peut soit la rétablir partout où elle est utile, soit considérer que les espaces $C(K_i), Z(K_i), B(K_i), H(K_i)$ sont des sommes directes de ces espaces usuellement pris dans chaque dimension.

Définition 4.3.1 *On appelle intervalles de persistance de la filtration \mathcal{K} les intervalles de persistance de la chaîne induite (4.5). On note $I(\mathcal{K})$ le multi-ensemble des intervalles de persistance de \mathcal{K} .*

4.3.1 Filtrations simples

Pour des raisons algorithmiques on s'intéresse aux filtrations obtenues en ajoutant un à un les simplexes de K . La filtration \mathcal{K} est dite *simple* si K_1 est réduit à un sommet $\sigma_1 \in K$ et si $K_i, 2 \leq i \leq n$, se déduit de K_{i-1} par l'ajout d'un unique simplexe $\sigma_i \in K$. Une filtration simple s'identifie donc à un ordre sur les simplexes de K dont les préfixes sont des sous-complexes de K . On supposera \mathcal{K} simple dans ce qui suit.

En considérant l'application bord comme un endomorphisme de $C(K_i)$, on obtient par la formule du rang d'une application linéaire :

$$\dim C(K_i) = \dim Z(K_i) + \dim B(K_i)$$

On en déduit en retranchant à cette équation celle pour K_{i-1} , et en notant que $\dim C(K_i)$ est le nombre de simplexes de K_i :

$$(\dim Z(K_i) - \dim Z(K_{i-1})) + (\dim B(K_i) - \dim B(K_{i-1})) = 1$$

Les deux termes entre parenthèses étant positifs ou nuls on a pour tout i ,

1. ou bien $\dim Z(K_i) = \dim Z(K_{i-1}) + 1$ et $B(K_i) = B(K_{i-1})$,
2. ou bien $Z(K_i) = Z(K_{i-1})$ et $\dim B(K_i) = \dim B(K_{i-1}) + 1$.

Définition 4.3.2 *On dira que l'indice i (ou le simplexe σ_i) dans la filtration simple \mathcal{K} est positif dans le premier cas, et négatif dans le second. L'ensemble des indices positifs et négatifs de \mathcal{K} sont respectivement notés $\mathcal{P}(\mathcal{K})$ et $\mathcal{N}(\mathcal{K})$.*

Lemme 4.3.3 *Les conditions suivantes sont équivalentes*

- σ_i est positif,
- σ_i est dans le support d'un cycle $z \in Z(K_i)$ et $Z(K_i) = Z(K_{i-1}) \oplus \mathbb{F}z$,
- $\partial\sigma_i \in B(K_{i-1})$,

La preuve est laissée à titre d'exercice. Notons que dans tous les cas on a

$$B(K_i) = B(K_{i-1}) + \mathbb{F}\partial\sigma_i, \tag{4.6}$$

la somme étant directe si et seulement si σ_i est négatif.

Lemme 4.3.4 *Tout intervalle de $I(\mathcal{K})$ est de la forme $[i, j]$ avec*

$$(i, j + 1) \in \mathcal{P}(\mathcal{K}) \times (\mathcal{N}(\mathcal{K}) \cup \{n + 1\}).$$

De plus,

- *tout indice positif est la borne inférieure d'un unique intervalle de $I(\mathcal{K})$ de multiplicité un, et*
- *tout indice négatif j est tel que $j - 1$ est la borne supérieure d'un unique intervalle de $I(\mathcal{K})$ de multiplicité un.*

On notera que $n + 1$ n'est pas un indice de la filtration et que n peut être la borne supérieure de plusieurs intervalles de persistance.

Preuve : On note n_i (resp. m_i) le nombre d'intervalles de persistance de (4.5) commençant (resp. se terminant) en i . On note également β_i le rang du morphisme $H(K_i) \rightarrow H(K_{i+1})$ induit par l'inclusion $K_i \subset K_{i+1}$. D'après la forme canonique (4.2), la dimension de $H(K_i)$ est le nombre d'intervalles de persistance contenant i . En partitionnant ces intervalles en ceux qui commencent en i et les autres, on obtient :

$$\dim H(K_i) = n_i + \beta_{i-1}$$

de même, en séparant ces intervalles en ceux qui se terminent en i et les autres, on obtient :

$$\dim H(K_i) = m_i + \beta_i$$

Puisque $B(K_i) \subset B(K_{i+1}) \subset Z(K_i) \subset Z(K_{i+1})$, l'image de $H(K_i) = Z(K_i)/B(K_i)$ dans $H(K_{i+1}) = Z(K_{i+1})/B(K_{i+1})$ est isomorphe à $Z(K_i)/B(K_{i+1})$, d'où

$$\beta_i = \dim Z(K_i) - \dim B(K_{i+1})$$

On en déduit, avec $\dim H(K_i) = \dim Z(K_i) - \dim B(K_i)$, que

$$n_i = \dim Z(K_i) - \dim Z(K_{i-1}) \quad \text{et} \quad m_i = \dim B(K_{i+1}) - \dim B(K_i)$$

Ces deux égalités permettent de conclure, compte tenu de la définition des indices positifs et négatifs. \square

Définition 4.3.5 *Le lemme précédent permet de définir les applications naissance et mort*

$$N : \mathcal{N}(\mathcal{K}) \rightarrow \mathcal{P}(\mathcal{K}) \quad \text{et} \quad M : \mathcal{P}(\mathcal{K}) \rightarrow \mathcal{N}(\mathcal{K}) \cup \{n + 1\}$$

telles que pour tout indice i positif et tout indice j négatif les intervalles $[i, M(i) - 1]$ et $[N(j), j - 1]$ sont des intervalles de $I(\mathcal{K})$.

Corollaire 4.3.6

$$\begin{aligned} I(\mathcal{K}) &= \{[i, M(i) - 1]\}_{i \in \mathcal{P}(\mathcal{K})} \\ &= \{[N(j), j - 1]\}_{j \in \mathcal{N}(\mathcal{K})} \cup \{[i, n + 1]\}_{i \in \mathcal{P}(\mathcal{K}) \setminus \text{Im} N} \end{aligned}$$

4.3.2 Bases compatibles des bords

Notre objectif est de calculer les signes des simplexes et les intervalles de persistance d'une filtration simple $\mathcal{K} : K_1 \subset K_2 \subset \dots \subset K_n = K$ de manière efficace. On introduit pour cela la notion de base compatible des bords.

Définition 4.3.7 Une base compatible des bords pour \mathcal{K} est une famille $\{x_j\}_{j \in J}$, $J \subset [1, n]$, de cycles de K vérifiant les conditions :

1. $\forall i \in [1, n]$, $\{x_j \mid (j \in J) \wedge (j \leq i)\}$ est une base de $B(K_i)$,
2. L'application $\nu : J \rightarrow [1, n]$ qui associe à tout j l'indice maximal $\nu(j)$ des simplexes du support de x_j est injective.

Lemme 4.3.8 Pour toute base compatible, la fonction ν définie ci-dessus coïncide avec la fonction N de la définition 4.3.5.

Preuve : La condition 1 ci-dessus et la note suivant l'équation (4.6) montrent que $J = \mathcal{N}(\mathcal{K})$. Le lemme 4.3.3 montre également que pour tout $j \in J$, $\nu(j) \in \mathcal{P}(\mathcal{K})$. On définit des cycles $z_i \in Z(K_i)$ pour tout $i \in \mathcal{P}(\mathcal{K})$ comme suit :

- Si $i \in \mathcal{P}(\mathcal{K}) \setminus \nu(J)$, on choisit z_i tel que $Z(K_i) = Z(K_{i-1}) \oplus \mathbb{F}z_i$ (cf. lemme 4.3.3).
- Sinon, $i = \nu(j)$ pour un certain $j \in J$ et on pose $z_{\nu(j)} = x_j$.

Remarquons que le simplexe d'indice maximal du support de z_i est σ_i et donc (cf. lemme 4.3.3) $\{z_j\}_{j \in \mathcal{P}(\mathcal{K}), j \leq i}$ forme une base de $Z(K_i)$. On note $[z]_i$ la classe d'homologie du cycle z dans $H(K_i)$. Il reste à vérifier que les $[z_i]_i$ forment une base compatible pour (4.5) au sens de la définition 4.2.5 et que les intervalles de persistance des $[z_{\nu(j)}]_{\nu(j)}$ sont de la forme $[\nu(j), j - 1]$. Vérifions que

$Z(i) = \{[z_j]_i\}_{(j \leq i) \wedge (j \notin \nu(J))} \cup \{[z_{\nu(j)}]_i\}_{(j \in J) \wedge (\nu(j) \leq i) \wedge (j > i)}$ est une base de $H(K_i)$. Puisque $[z_{\nu(j)}]_j = 0$, on sait déjà par la remarque ci-dessus que $Z(i)$ est une famille génératrice. Reste à voir qu'elle est libre : une combinaison linéaire des classes de $Z(i)$ est nulle si la combinaison des cycles correspondant est dans $B(K_i)$, c'est-à-dire égale à une combinaison de $\{x_j \mid (j \in J) \wedge (j \leq i)\}$. On déduit aisément la nullité des coefficients d'une telle combinaison du fait que les z_j (et x_j) sont échelonés en tant que chaînes exprimés sur la base des simplexes. Il reste à remarquer que l'intervalle de persistance d'un $[z_i]_i$ est $\{j \mid [z_i]_j \in Z(j)\}$. \square

Proposition 4.3.9 Toute filtration simple admet une base compatible des bords.

La preuve est donnée par l'algorithme de la section suivante.

Exercice 4.3.10 Montrer que $N(i + 1) = i$ implique que σ_i est une face de σ_{i+1} .

4.3.3 Algorithme

Le lemme 4.3.8 ci-dessus et le corollaire 4.3.6 indiquent qu'il suffit de construire une base compatible des bords pour calculer le signe de chaque simplexe et les intervalles de persistance d'une filtration simple.

La construction se fait par récurrence sur la taille n de la filtration. Le cas $n = 1$ est trivial puisque l'unique simplexe de la filtration est nécessairement positif. Supposons donc avoir construit une base compatible des bords $\{x_j\}_{j \in J}$ pour la sous-filtration \mathcal{K}' :

$$K_1 \subset K_2 \subset \dots \subset K_{n-1}$$

On désigne par $N : J \rightarrow \mathcal{P}(\mathcal{K}')$ la fonction de naissance associée. Supposons pouvoir calculer une décomposition du bord de σ_n sous la forme

$$\partial\sigma_n = \sum_{j \in J} \alpha_j x_j + x_n \quad (4.7)$$

de sorte que

1. ou bien x_n est nul,
2. ou bien l'indice maximal, k_n , des simplexes du support de x_n n'est pas dans $N(J)$.

Si x_n est nul alors $B(K_n) = B(K_{n-1})$ (cf. (4.6)) et $\{x_j\}_{j \in J}$ reste une base compatible des bords pour \mathcal{K} . Sinon, c'est le cas pour $\{x_j\}_{j \in J} \cup \{x_n\}$ et on a de plus $N(n) = k_n$.

Notons que par la condition 2 de la définition 4.3.7 toute base compatible des bords forme une base échelonnée de vecteurs sur la base canonique des simplexes. La décomposition (4.7) peut donc s'obtenir par la méthode du pivot de Gauss décrite par le pseudo-code suivant :

```

y := ∂σn
i := indice maximal des simplexes du support de y
Tant que ( (y ≠ 0) ∧ (i ∈ N(J)) )
  j := N-1(i)
  α := coefficient de σi dans y
  β := coefficient de σi dans xj
  y := y - (α/β)xj
  i := indice maximal des simplexes du support de y
Fin tant que
\* y contient xn à la sortie de la boucle *

```

Analyse de la complexité On stocke chaque cycle x_j par la liste des coefficients de ses simplexes dans un tableau indexé par les n simplexes de K . On représente la fonction N par un tableau de taille n ; la case j contient $N(j)$ si σ_j est négatif et 0 sinon. On stocke également la fonction inverse N^{-1} dans un tableau de taille n . Le calcul de x_n dans l'algorithme précédent prend un temps $O(n^2)$. On en déduit

Proposition 4.3.11 *Le calcul d'une base compatible des bords et de la fonction N pour une filtration de longueur n requiert un temps $O(n^3)$ sur une machine \mathbb{F} -RAM (i.e. les opérations élémentaires $+$, $-$, $*$, \div sur \mathbb{F} prennent un temps constant). En particulier, on peut calculer les intervalles de persistance de la filtration dans le même temps.*

4.4 Diagramme de persistance

Se donner une filtration d'un complexe K revient à se donner une fonction $f : K \rightarrow \mathbb{R}$ qui soit non-décroissante, c'est-à-dire telle que :

$$\forall \sigma, \tau \in K : \sigma \prec \tau \implies f(\sigma) \leq f(\tau)$$

où \prec est la relation "être face de". En effet, si $f_1 < f_2 < \dots < f_n$ est la suite des *valeurs* de f alors la suite de complexes

$$f^{-1}([-\infty, f_1]) \subset f^{-1}([-\infty, f_2]) \subset \dots \subset f^{-1}([-\infty, f_n]) \quad (4.8)$$

fournit une filtration de K que l'on note \mathcal{K}_f . Inversement, toute filtration $K_1 \subset K_2 \subset \dots \subset K_n = K$ est de la forme \mathcal{K}_f pour la fonction f définie par $f(\sigma) = i$ pour $\sigma \in K_i \setminus K_{i-1}$.

Soit $f : K \rightarrow \mathbb{R}$ une fonction non-décroissante et $I = I(\mathcal{K}_f)$ le multi-ensemble d'intervalles de persistance de la filtration \mathcal{K}_f associée. Dit autrement, I est le multi-ensemble d'intervalles de persistance de la chaîne :

$$H(K_1) \rightarrow H(K_2) \rightarrow \dots \rightarrow H(K_n)$$

où l'on a posé $K_i = f^{-1}([-\infty, f_i])$.

Définition 4.4.1 *Le diagramme de persistance de f , noté $D(f)$, est la réunion dans le plan étendu $(\mathbb{R} \cup \{-\infty, +\infty\})^2$*

- *du multi-ensemble de points $\{(f_i, f_{j+1})\}_{[i,j] \in I}$, où l'on a posé $f_{n+1} = +\infty$, et*
- *des points de la diagonale du plan $\{x = y\}$ comptés chacun avec multiplicité infinie (dénombrable).*

On note Δ^∞ le multi-ensemble des points de la diagonale avec multiplicité infinie.

On dira qu'une filtration $\mathcal{K} : K_1 \subset K_2 \subset \dots \subset K_m = K$ est *compatible* avec la fonction $f : K \rightarrow \mathbb{R}$ si \mathcal{K}_f est une sous-filtration de \mathcal{K} . Autrement dit, si f est constante sur chaque $K_i \setminus K_{i-1}$ et si la fonction $f_{\mathcal{K}} : [1, m] \rightarrow \mathbb{R}$, $i \mapsto f(K_i \setminus K_{i-1})$ est non-décroissante.

Définition 4.4.2 *le diagramme de persistance d'une fonction non-décroissante $f : K \rightarrow \mathbb{R}$ relativement à une filtration \mathcal{K} compatible avec f est le multi-ensemble de points*

$$D(f, \mathcal{K}) = \Delta^\infty \cup \{(f_{\mathcal{K}}(i), f_{\mathcal{K}}(j+1))\}_{[i,j] \in I(\mathcal{K})}$$

où l'on a posé $f_{\mathcal{K}}(m+1) = +\infty$, m étant la longueur \mathcal{K} .

En particulier, $D(f) = D(f, \mathcal{K}_f)$.

Lemme 4.4.3 *$D(f) = D(f, \mathcal{K})$ pour toute filtration \mathcal{K} compatible avec f .*

Preuve : Soit $f_1 < f_2 < \dots < f_n$ la suite des valeurs de f sur K . On pose $\kappa : [1, n] \rightarrow [1, m], i \mapsto \max\{j \mid f_{\mathcal{K}}(j) = f_i\}$. Clairement, $f^{-1}([-\infty, f_i]) = K_{\kappa(i)}$ et la chaîne d'homologie induite par \mathcal{K}_f définit, avec l'injection κ , une sous-chaîne de celle induite par \mathcal{K} selon la terminologie de la section 4.2.3. On en déduit par le lemme 4.2.9 que $I(\mathcal{K}_f) = \{[\lambda(i), \mu(j)] \mid [i, j] \in I(\mathcal{K}) \text{ et } \lambda(i) \leq \mu(j)\}$, avec λ et μ comme dans le lemme. D'où

$$D(f) = D(f, \mathcal{K}_f) = \Delta^\infty \cup \{(f_{\lambda(i)}, f_{\mu(j)+1}) \mid [i, j] \in I(\mathcal{K}) \text{ et } \lambda(i) \leq \mu(j)\}$$

On vérifie aisément avec les définitions de κ, λ et μ que $f_{\lambda(i)} = f_{\mathcal{K}}(i)$ et $f_{\mu(j)+1} = f_{\mathcal{K}}(j+1)$. Soit encore

$$D(f) = \Delta^\infty \cup \{(f_{\mathcal{K}}(i), f_{\mathcal{K}}(j+1)) \mid [i, j] \in I(\mathcal{K}) \text{ et } \lambda(i) \leq \mu(j)\}$$

Mais si $\lambda(i) > \mu(j)$ pour un certain intervalle $[i, j] \in I(\mathcal{K})$ alors $f_{\mathcal{K}}(i) = f_{\mathcal{K}}(j+1)$ et le point correspondant $(f_{\mathcal{K}}(i), f_{\mathcal{K}}(i))$ est "absorbé" par la diagonale Δ^∞ . On conclut finalement

$$D(f) = \Delta^\infty \cup \{(f_{\mathcal{K}}(i), f_{\mathcal{K}}(j+1))\}_{[i,j] \in I(\mathcal{K})} = D(f, \mathcal{K}). \quad \square$$

Ainsi le diagramme de persistance de f peut être calculé à partir de n'importe quelle filtration compatible avec f .

4.4.1 Stabilité du diagramme de persistance

Soient $f, g : K \rightarrow \mathbb{R}$ deux fonctions non-décroissantes. On note

$$\|f - g\|_\infty = \sup_{\sigma \in K} |f(\sigma) - g(\sigma)|$$

leur distance L_∞ .

On définit également une (pseudo) distance¹ d entre diagrammes de persistance par

$$d(D, D') = \inf_{\phi} \sup_{p \in D} \|p - \phi(p)\|_\infty$$

où $\phi : D \rightarrow D'$ décrit l'ensemble des bijections entre D et D' et $\|p - q\|_\infty = \max\{|x_p - x_q|, |y_p - y_q|\}$ (on pose $|+\infty - x| = +\infty$ si $x \neq +\infty$ et 0 sinon). Notons que D et D' ont tous les deux même cardinal (du continu) du fait de la présence de la diagonale Δ^∞ .

Proposition 4.4.4 ([CSEH07, CSEM06]) $d(D(f), D(g)) \leq \|f - g\|_\infty$

Preuve : On pose $f_t = f + t(g - f)$. Notons que f_t est non-décroissante sur K . Pour toute paire de simplexes $\sigma, \tau \in K$, il existe un réel $u \in [0, 1]$ tel que le signe (au sens large) de $f_t(\sigma) - f_t(\tau)$ reste constant pour $t \in [0, u]$ et de même pour $t \in [u, 1]$. Il existe donc une partition finie $0 = t_0 < t_1 < \dots < t_r = 1$ de² $[0, 1]$ telle que l'ordre relatif des

¹Il ne s'agit pas d'une véritable distance car d peut prendre une valeur infinie.

² $r \leq \binom{m}{2} + 1$ où m est le nombre de simplexes de K

f_t -valeurs des simplexes soit indépendant de t sur chaque intervalle $[t_i, t_{i+1}]$. On en déduit pour chaque $i \in [0, r - 1]$ une filtration simple \mathcal{K}_i compatible avec toutes les fonctions f_t pour $t \in [t_i, t_{i+1}]$. Par le lemme 4.4.3, on a

$$D(f_t) = D(f_t, \mathcal{K}_i) = \Delta^\infty \cup \{(f_t(\sigma_a), f_t(\sigma_{b+1}))\}_{[a,b] \in I(\mathcal{K}_i)}$$

où σ_a est le a -ième simplexe de la filtration simple \mathcal{K}_i . En considérant la bijection évidente entre $D(f_{t_i})$ et $D(f_{t_{i+1}})$ qui vaut l'identité sur Δ^∞ et envoie $(f_{t_i}(\sigma_a), f_{t_i}(\sigma_{b+1}))$ sur $(f_{t_{i+1}}(\sigma_a), f_{t_{i+1}}(\sigma_{b+1}))$, on déduit $d(D(f_{t_i}), D(f_{t_{i+1}})) \leq (t_{i+1} - t_i) \|f - g\|_\infty$. Par application de l'inégalité triangulaire on obtient finalement

$$d(D(f), D(g)) \leq \sum_i (t_{i+1} - t_i) \|f - g\|_\infty = \|f - g\|_\infty$$

□