



CaMIUS

Volume 3 Cahiers mathématiques de l'Université de Sherbrooke

| | |
|--|----|
| M. Ethier | 1 |
| <i>Introduction à l'homologie persistante avec application à la suspension topologique</i> | |
| X. Peng and J. Zhang | 19 |
| <i>Cluster algebras and Markoff numbers</i> | |
| J.-F. Marceau | 27 |
| <i>Pavages additifs</i> | |
| D. Boulet-St-Jacques, M. Chabot et G. Douville | 39 |
| <i>Groupes de frise des répétitions des carquois de type Dynkin</i> | |
| A. Salvail-Bérard | 54 |
| <i>Réseaux de neurones</i> | |
| N. Bureau | 70 |
| <i>Exemples d'extensions galoisiennes de degré 24 sur \mathbb{Q}</i> | |

INTRODUCTION À L'HOMOLOGIE PERSISTANTE AVEC APPLICATION À LA SUSPENSION TOPOLOGIQUE

MARC ETHIER, AUTEUR INVITÉ

RÉSUMÉ L'homologie persistante, qui étudie la durée de la persistance de propriétés topologiques le long d'une filtration d'espaces, est fréquemment appliquée à de nombreuses disciplines, notamment à la comparaison de formes en informatique. Cependant, une question se pose : pour comparer deux objets, vaut-il mieux utiliser un grand nombre de fonctions de mesure pour extraire leurs propriétés, ou plutôt considérer un nombre réduit de fonctions et utiliser tout le pouvoir de la persistance pour comparer leurs propriétés en dimensions plus élevées ? Cela demeure une question ouverte, mais cet article y jettera un coup d'œil en plus de fournir une procédure se servant du concept topologique de suspension pour aider à construire des exemples numériques.

Marc Ethier est un étudiant au doctorat en mathématiques à l'Université de Sherbrooke. Son article a été écrit à la demande du comité de rédaction de CaMUS.

1 Introduction

La topologie algébrique est une branche des mathématiques qui consiste à étudier les propriétés d'espaces topologiques en se servant d'outils algébriques. En d'autres termes, à un espace topologique on peut associer un objet algébrique (groupe, module, espace vectoriel, etc.) qui, à travers sa structure, donne de l'information sur certains invariants topologiques de l'espace de départ. La topologie algébrique possède de plus en plus d'applications dans d'autres domaines du savoir, comme l'informatique, l'ingénierie, les sciences médicales, etc., étant donné que les objets algébriques en question peuvent fréquemment être calculés de façon algorithmique. En l'occurrence, le problème de la reconnaissance de formes en informatique bénéficie d'être abordé à l'aide de l'homologie persistante, une discipline qui consiste, compte tenu un objet (espace topologique) que l'on traverse avec une fonction continue en considérant, pour chaque valeur u , le sous-espace où la fonction prend des valeurs inférieures à u , à étudier combien de temps les invariants topologiques de ce sous-espace persistent. Cependant,

J'aimerais remercier le professeur Tomasz Kaczynski pour m'avoir suggéré le projet qui a mené aux principaux résultats décrits dans cet article. Je voudrais aussi remercier Claudia Landi de l'Università di Modena e Reggio Emilia pour m'avoir fourni l'image de la figure 2.

une question se pose : vaut-il mieux, pour améliorer le pouvoir de discrimination des méthodes algorithmiques de comparaison de formes, utiliser plus qu'une seule fonction continue, ou plutôt considérer des invariants topologiques en plus grande dimension ?

Cet article vise à entamer de répondre à cette question en introduisant une méthode basée sur le concept topologique de la suspension. La suspension est une opération topologique qui augmente de un la dimension des propriétés d'un espace. Cette méthode pourra entre autres permettre la construction de modèles ayant une topologie complexe en dimension quelconque, qui pourront servir à tester et comparer des algorithmes de comparaison de formes. Un autre but de cet article est de fournir au lecteur une idée intuitive de la topologie algorithmique et des applications qu'elle possède dans les sciences.

Nous commencerons, à la section 2, par décrire l'homologie telle qu'elle est utilisée en topologie algébrique, ainsi que ce domaine de recherche plus récent qu'est l'homologie persistante. À la section 3, nous introduirons le concept de suspension topologique, ainsi qu'une fonction que nous pouvons définir sur la suspension d'un espace et qui a les propriétés que nous recherchons. Nous décrivons ensuite les effets que cette opération a sur la persistance. Finalement, à la section 4, nous validerons les résultats obtenus grâce à des tests numériques.

2 Homologie persistante

2.1 Homologie

Le calcul d'homologie est une procédure fréquemment utilisée en topologie algébrique pour établir une correspondance entre des espaces topologiques et des objets algébriques tels que des modules ou des espaces vectoriels. De façon générale, ceci veut dire que deux espaces topologiques « identiques » (c'est-à-dire homéomorphes) auront la même structure homologique. Il existe une variété de théories de l'homologie définies sous des hypothèses différentes, mais les espaces qui seront considérés dans les cas d'application sont suffisamment simples pour que toutes ces théories coïncident. Une théorie de l'homologie relativement aisée à comprendre est l'*homologie simpliciale*, définie sur les complexes simpliciaux, et qui peut être étendue aux espaces triangulables, c'est-à-dire aux espaces topologiques homéomorphes à un complexe simplicial. Cette théorie, et d'autres théories de l'homologie, sont décrites dans [Mun84]. Si nous travaillons sur un corps \mathbb{K} , l'homologie de l'espace topologique X , dénotée $H_*(X; \mathbb{K}) = \{H_q(X; \mathbb{K})\}_{q \in \mathbb{Z}}$ ou simplement $H_*(X)$, sera une suite d'espaces vectoriels isomorphes à \mathbb{K}^{n_q} , qui sont tous entièrement déterminés par leur dimension n_q . L'information recueillie dans l'homologie de X peut donc être entièrement décrite par la dimension des espaces d'homologie, que nous nommons *nombres de Betti* en l'honneur du topologue italien du 19^e siècle Enrico Betti.

Sans avoir à entrer trop profondément dans les théories sous-tendant l'homologie, nous pouvons dire que le q^e nombre de Betti de l'espace X , dénoté $\beta_q(X)$, correspond au nombre de « trous de dimension q » de X . Par exemple,

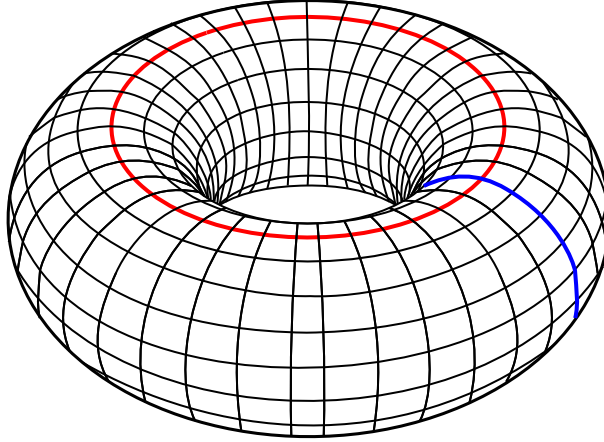


FIGURE 1 – Le tore, avec ses deux cycles d'ordre 1 indiqués. Version éditée d'une image produite par l'utilisateur de Wikipedia *YassineMrabet*.

si nous nous en tenons aux dimensions les plus fréquemment considérées dans les applications, $\beta_0(X)$ correspond au nombre de composantes connexes de X , $\beta_1(X)$ au nombre de trous, au sens ordinaire, dans X , et $\beta_2(X)$ au nombre de cavités enclavées dans X . À titre d'exemple, considérons la sphère topologique de dimension k , soit S^k . (S^0 se compose de deux points disjoints, S^1 est un cercle, S^2 est une sphère ordinaire, etc.) On remarque que pour tout $k > 0$, $\beta_q(S^k) = 1$ si $q = 0$ ou $q = k$, alors que $\beta_q(S^k) = 0$ pour les autres valeurs de q , tandis que $\beta_q(S^0) = 2$ si $q = 0$ et $\beta_q(S^0) = 0$ pour les autres valeurs de q . Cette apparente asymétrie a entraîné l'introduction du concept d'*homologie réduite* $\tilde{H}_*(X)$, dont la différence avec l'homologie ordinaire, en ce qui concerne les nombres de Betti, est que le nombre de Betti réduit d'ordre 0, $\tilde{\beta}_0(X)$, égale $\beta_0(X) - 1$ en supposant que X est non vide, tandis que $\tilde{\beta}_q(X) = \beta_q(X)$ pour $q > 0$. En d'autres mots, on peut interpréter $\tilde{\beta}_0(X)$ comme le nombre d'espaces vides entre les composantes connexes de X . On obtient donc que $\tilde{\beta}_k(S^k) = 1$, et $\tilde{\beta}_q(S^k) = 0$ si $0 \leq q \neq k$.

Exemple 2.1. Considérons le tore T . Ses nombres de Betti sont $\beta_0(T) = 1$, $\beta_1(T) = 2$ et $\beta_2(T) = 1$. En effet, le tore comprend une composante connexe et une cavité connexe. De plus, il possède deux « trous », ce que vous pouvez vérifier de la façon suivante : placez une corde autour de l'ouverture du tore. Il vous est impossible, même en étirant et déplaçant cette corde sur le tore, de la réduire à un point sans sortir de la surface. De même, si vous passez une extrémité de la corde dans l'ouverture et la joignez ensuite à l'autre extrémité, vous obtenez une boucle que vous ne pouvez pas réduire à un point. On vérifie aisément que ce *cycle d'ordre 1* est différent du précédent. Ces deux cycles sont illustrés sur la figure 1. De même, les nombres de Betti réduits du tore sont $\tilde{\beta}_0(T) = 0$, $\tilde{\beta}_1(T) = 2$, $\tilde{\beta}_2(T) = 1$.

2.2 Persistence

L'homologie persistante est apparue comme sujet de recherche au tournant des années 1990, d'abord comme *théorie de taille* (*size theory*) dans les travaux de Frosini *et al.*, puis de façon plus générale et dans des termes plus homologiques grâce à Edelsbrunner *et al.* Ce sujet se base sur le concept d'une *filtration* d'espaces topologiques, c'est-à-dire une collection indexée d'espaces topologiques X_u où la variable d'indice u appartient à un ensemble (complètement ou partiellement) ordonné, de telle sorte que $X_u \subseteq X_v$ si $u \leq v$. L'inclusion $j^{(u,v)} : X_u \hookrightarrow X_v$ induit une application linéaire $H_*(j^{(u,v)})$ entre les homologies $H_*(X_u)$ et $H_*(X_v)$. Formellement, dans le langage de la théorie des catégories, H_* est un *foncteur covariant* entre la catégorie des espaces topologiques, dont les morphismes sont des fonctions continues, et la catégorie des suites d'espaces vectoriels, dont les morphismes sont des suites d'applications linéaires. L'image de $H_q(j^{(u,v)})$, informellement parlant, correspond aux composantes topologiques de dimension q qui existent dans X_u et existent encore dans X_v . Dans le langage de la persistance, nous disons que ces composantes sont « nées » au temps u ou avant, et « meurent » plus tard que v .

Un exemple typique d'une filtration que nous voyons fréquemment dans les applications est celui de la filtration de sous-niveau pour une fonction continue $f : X \rightarrow \mathbb{R}$, dite *fonction de mesure* ou de *filtration*. Dans ce cas, on définit X_u comme $\{x \in X \mid f(x) \leq u\}$ pour $u \in \mathbb{R}$, et on note l'inclusion de X_u dans X_v , pour $u \leq v$, par $j_f^{(u,v)}$. Le choix de la fonction de mesure f permet d'extraire des informations sur la géométrie de l'objet X , informations qui peuvent être caractéristiques. C'est pourquoi cette théorie peut être appliquée à la reconnaissance de formes en informatique. Tout comme dans la sous-section 2.1, nous travaillons sur un corps \mathbb{K} , et l'image de $H_q(j_f^{(u,v)})$ est donc un espace vectoriel. Soit $\Delta^+ = \{(u, v) \in \mathbb{R}^2 \mid u < v\}$, et définissons le *nombre de Betti persistant* d'ordre q pour la filtration associée à f comme

$$\beta_{f,q} = \dim \operatorname{im} H_q(j^{(u,v)}).$$

Pour des fins de simplicité de la notation, la dimension q pourra dans certains cas être omise, et dans d'autres cas être placée en exposant plutôt qu'en indice (soit β_f^q). Nous pouvons de manière similaire définir les *nombre de Betti persistants réduits*, qui seront notés $\tilde{\beta}_{f,q}$. La note précédente est aussi valide pour ceux-ci.

Les nombres de Betti persistants peuvent être décrits de façon simple par leur *diagramme de persistance*.

Définition 2.2. Soit $p = (u, v) \in \Delta^+$. Nous définissons sa *multiplicité* $\mu(p)$ comme le minimum, pour tout $\epsilon > 0$ tel que $u + \epsilon < v - \epsilon$, de

$$\beta_f(u + \epsilon, v - \epsilon) - \beta_f(u - \epsilon, v - \epsilon) - \beta_f(u + \epsilon, v + \epsilon) + \beta_f(u - \epsilon, v + \epsilon).$$

Si $\mu(p) > 0$, alors on appelle p un *point propre* pour β_f . Les points propres de β_f correspondent à des composantes topologiques nées en u et qui meurent en

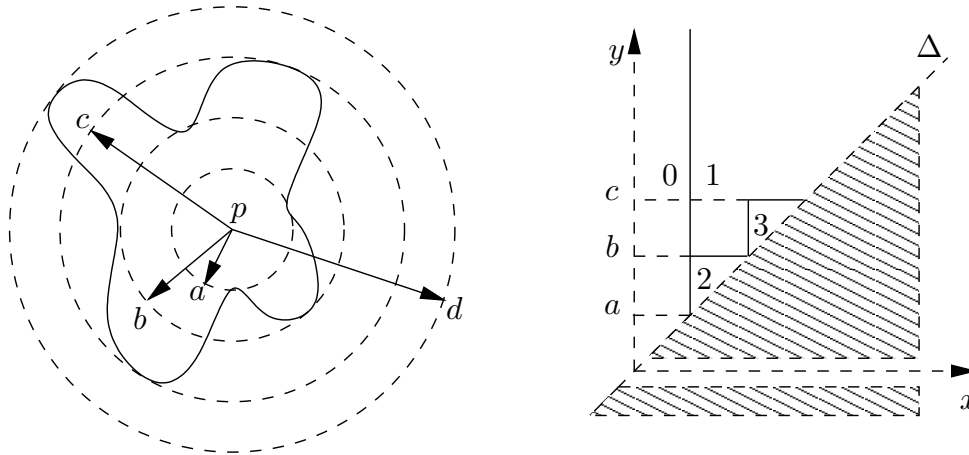


FIGURE 2 – À gauche : un modèle X , muni de la fonction f représentant la distance entre chaque point de X et le point dénoté p . À droite : le diagramme de persistance pour β_f^0 . On y voit un point à l’infini (d’abscisse a), un point propre de multiplicité 1 (aux coordonnées (a, b)) et un point propre de multiplicité 2 (aux coordonnées (b, c)).

v . Additionnellement, pour une droite verticale r d’équation $x = u$, on définit sa *multiplicité* $\mu(r)$ comme le minimum, pour tout $\epsilon > 0$ tel que $u + \epsilon < 1/\epsilon$, de

$$\beta_f(u + \epsilon, 1/\epsilon) - \beta_f(u - \epsilon, 1/\epsilon).$$

Si $\mu(r) > 0$, alors r est appelé un *point à l’infini* pour β_f , que nous identifions avec la paire (u, ∞) . Les points à l’infini de β_f correspondent à des composantes topologiques nées en u et qui ne meurent jamais, donc qui sont inhérentes à l’espace X . Finalement, le *diagramme de persistance* de β_f est un *multi-ensemble* (c’est-à-dire une généralisation d’un ensemble dans lequel les éléments peuvent apparaître plus d’une fois) composé des points propres et points à l’infini de β_f , chacun apparaissant le nombre de fois correspondant à sa multiplicité.

La figure 2 montre un modèle X muni d’une fonction de mesure continue f , ainsi que les diagrammes de persistance pour son nombre de Betti persistant $\beta_f^0(u, v)$.

Dans les applications, l’espace des diagrammes de persistance peut être muni d’une distance. L’intérêt de procéder ainsi est d’avoir une mesure calculable de la différence entre deux diagrammes, qui puisse potentiellement nous fournir de l’information sur la différence entre les modèles, c’est-à-dire les espaces munis de fonctions de mesure, qu’ils représentent. Une telle distance est la *distance d’appariement* (*matching distance*, [dFL06]). Soient $f : X \rightarrow \mathbb{R}$ et $g : Y \rightarrow \mathbb{R}$ deux fonctions de mesure continues définies sur deux modèles X et Y , et β_f, β_g les nombres de Betti persistants (d’un ordre quelconque) pour les fonctions f et

g . Pour éviter les technicalités, la définition de la distance d'appariement d n'a pas été incluse ici, mais intuitivement parlant, elle calcule le coût de la bijection optimale qui déplace les points du diagramme de β_f sur ceux du diagramme de β_g , gardant à l'esprit que si le point (u, v) est tel que $v - u$ est faible, il correspond à une composante de faible persistance et peut être éliminé sans grand coût. La valeur d fournit une borne inférieure pour la *pseudo-distance naturelle* entre les paires (X, f) et (Y, g) définie comme

$$\inf_h \max_{p \in X} |f(p) - g(h(p))|$$

où h varie parmi les homéomorphismes entre X et Y .

2.3 Persistence multidimensionnelle

Au cours des dernières années l'*homologie persistante multidimensionnelle* a gagné en importance comme sujet de recherche. Cette théorie s'intéresse au cas où la filtration d'espaces dépend de plus d'une variable ; en particulier, on pense à la filtration de sous-niveau pour une fonction de mesure sur X qui prend ses valeurs dans \mathbb{R}^k . Elle est alors définie à partir de l'ordre partiel sur \mathbb{R}^k : $X_u = \{x \in X \mid f(x) \preceq u\}$ pour $u \in \mathbb{R}^k$, où $x \preceq y$ signifie que $x_i \leq y_i$ pour $i = 1, \dots, k$. L'avantage de cette théorie est que la persistance multidimensionnelle permet de soutirer plus d'information sur la géométrie de X que la persistance ordinaire, mais son inconvénient est qu'il n'existe alors pas de descripteur pour les nombres de Betti persistants β_f aussi simple que les diagrammes de persistance. Il est par contre possible de réduire le calcul de l'homologie persistante multidimensionnelle au calcul de l'homologie persistante ordinaire pour une famille paramétrée de fonctions à valeur dans \mathbb{R} [BCF⁺08], soit :

$$\text{red}_{(\vec{l}, \vec{b})}^f(x) = \min_{i=1, \dots, k} l_i \max_{i=1, \dots, k} \frac{f_i(x) - b_i}{l_i}$$

où (\vec{l}, \vec{b}) est choisi dans l'ensemble des paires admissibles

$$\text{Adm}_k = \left\{ (\vec{l}, \vec{b}) \in \mathbb{R}^k \times \mathbb{R}^k \mid \forall i \ l_i > 0, \sum_i l_i = 1, \sum_i b_i = 0 \right\}.$$

Une *distance d'appariement k -dimensionnelle* $D(\beta_f, \beta_g)$ entre les nombres de Betti persistants β_f pour $f : X \rightarrow \mathbb{R}^k$ et β_g pour $g : Y \rightarrow \mathbb{R}^k$ peut ensuite être définie comme le maximum sur Adm_k de la distance d'appariement $d\left(\beta_{\text{red}_{(\vec{l}, \vec{b})}^f}, \beta_{\text{red}_{(\vec{l}, \vec{b})}^g}\right)$. Grâce à cette formulation réduite, nous pouvons utiliser la persistance multidimensionnelle dans les applications.

3 Suspension

3.1 Définition et propriétés

La *suspension* de l'espace topologique X est définie comme l'espace quotient $SX = X \times [-1, 1] / \sim$, où \sim est l'identification $(x, 1) \sim (y, 1)$ et $(x, -1) \sim (y, -1)$

pour tous $x, y \in X$ [Mun84, §33]. Nous pouvons visualiser cet espace comme un double cône sur X dont les sommets se trouvent aux valeurs de paramètre $s = \pm 1$. L'opération S est un foncteur qui décale la dimension de l'homologie réduite par 1 [Mun84, Th. 33.2], c'est-à-dire

$$\tilde{H}_{q+1}(SX) \cong \tilde{H}_q(X), \quad q \geq -1. \quad (1)$$

Nous pouvons réécrire cette équation en termes d'homologie ordinaire :

$$H_{q+1}(SX) \cong H_q(X), \quad \text{pour } q \geq 1, \quad (2)$$

$$H_1(SX) \oplus \mathbb{K} \cong H_0(X), \quad \text{et } H_0(SX) \cong \mathbb{K}. \quad (3)$$

Soit maintenant $x_0 \in X$. La *suspension réduite* de l'espace pointé (X, x_0) (un espace pointé est un espace topologique dont un point est spécifié comme point de base) est l'espace pointé $(\Sigma X, (x_0, 0))$, où ΣX est le quotient de SX obtenu par l'identification additionnelle $(x_0, s) \sim (x_0, 0)$ pour tout $s \in [-1, 1]$. Si X est triangulable, alors ΣX est homotopiquement équivalent à SX [Hat02, §4.3]. La figure 3 montre un exemple d'un espace topologique et de sa suspension.

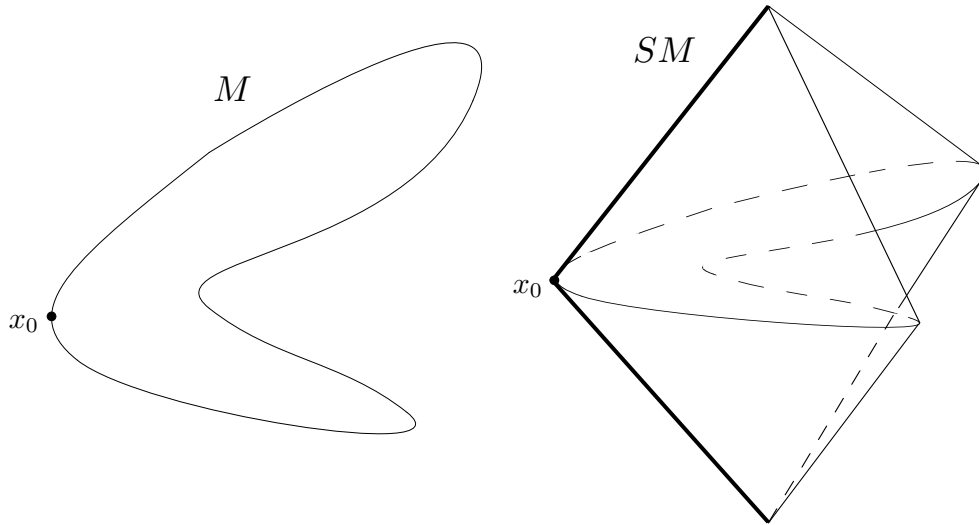


FIGURE 3 – À gauche : une variété M homéomorphe à S^1 , avec un point de base x_0 , qui peut être, par exemple, le minimum global de la première fonction de coordonnées. À droite : sa suspension SM . Dans la suspension réduite ΣM , tout le segment $\{x_0\} \times [-1, 1]$ (trait plus épais) serait réduit à un point.

Considérons maintenant des espaces topologiques munis de fonctions de mesure. Pour fixer les idées, supposons que $X = M$ soit une variété compacte lisse, et que $f : M \rightarrow \mathbb{R}$ soit une *fonction de Morse*, c'est-à-dire une fonction de classe C^2 dont les points critiques sont non dégénérés et possèdent des valeurs critiques distinctes. Soit m_0 le minimum global de f , atteint en x_0 .

Définissons $Ef : M \times [-1, 1] \rightarrow \mathbb{R}$ comme

$$Ef(x, s) = s^2 m_0 + (1 - s^2) f(x). \quad (4)$$

On note que Ef atteint son minimum global sur l'union de $\{x_0\} \times [-1, 1]$ avec $M \times \{-1, 1\}$. Le long de chaque segment $s \mapsto (x, s)$, elle atteint son maximum en $(x, 0)$. La proposition suivante est une simple application du calcul différentiel. Rappelons que l'*indice de Morse* d'un point critique p de f est le nombre de directions linéairement indépendantes dans lesquelles f décroît à partir de p . Par exemple, un minimum est d'indice de Morse 0, un maximum est d'indice de Morse $\dim M$, et un col a un indice de Morse entre ces deux valeurs.

Proposition 3.1. *Le point (p, s) est un point critique de Ef si et seulement si l'une ou l'autre des conditions suivantes sont vérifiées : en premier lieu, p est un point critique de f autre que x_0 et $s = 0$, ou alors, en second lieu, $p = x_0$ ou $s = \pm 1$. De plus, dans le premier cas, p est d'indice de Morse λ si et seulement si $(p, 0)$ est d'indice de Morse $\lambda + 1$.*

Démonstration. Selon (4), nous obtenons

$$\nabla Ef(x, s) = ((1 - s^2)\nabla f(x), 2sm_0 - 2sf(x)).$$

Les composantes en x de ∇Ef sont nulles si et seulement si $\nabla f = 0$ ou $s = \pm 1$, tandis que la composante en s est nulle si et seulement si $s = 0$ ou $f(x) = m_0$ (cette dernière condition équivalant à $x = x_0$). Nous pouvons donc déceler deux cas où $\nabla Ef(p, s) = 0$: celui où $\nabla f(p) = 0$ et $s = 0$, avec $p \neq x_0$, ainsi que celui, composé de points non isolés où Ef atteint son minimum global de m_0 , où $p = x_0$ ou $s = \pm 1$.

Dans le premier cas, on note que p est un point critique d'indice de Morse λ de f si et seulement si il existe λ directions linéairement indépendantes dans M suivant lesquelles f décroît à partir de p . Ceci équivaut à dire qu'il existe λ directions de décroissance pour Ef dans $M \times \{0\}$, en plus de la direction correspondant à la variable s où Ef diminue aussi puisque $f(p) > m_0$. L'indice de Morse de $(p, 0)$ est donc de $\lambda + 1$. \square

Définissons les applications $S_{\mathbb{R}}f : SM \rightarrow \mathbb{R}$ et $\Sigma_{\mathbb{R}}f : \Sigma M \rightarrow \mathbb{R}$ comme compositions de Ef avec l'application quotient, respectivement, de $M \times [-1, 1]$ dans SM , et dans ΣM . L'énoncé de la proposition 3.1 s'applique aussi à $S_{\mathbb{R}}f$, incluant le fait que les points sur $\{x_0\} \times [-1, 1]$, tous critiques avec valeur critique m_0 , sont dégénérés. Par contre, en considérant $\Sigma_{\mathbb{R}}f$, on vérifie la condition d'isolation pour la classe d'équivalence de $(x_0, 0)$, qui devient l'unique point critique d'indice de Morse 0 de $\Sigma_{\mathbb{R}}f$. Le fait que SM n'est pas lisse aux deux sommets du cône, et que ΣM n'est pas lisse à son point de base, n'affecte pas les invariants homologiques étant donné que la fonction de mesure y atteint alors son minimum global m_0 .

Pour tout $u \in \mathbb{R}$, nous pouvons considérer les ensembles de sous-niveau

$$M_u = \{x \in M \mid f(x) \leq u\}$$

et

$$(SM)_u = \{(x, s) \in SM \mid S_{\mathbb{R}}f(x, s) \leq u\}.$$

Il est aisé de vérifier que $S(M_u) \subset (SM)_u$, mais l'inclusion inverse n'est cependant pas vraie. Nous pouvons néanmoins prouver le résultat suivant.

Lemme 3.2. *L'ensemble $S(M_u)$ est homotopiquement équivalent à $(SM)_u$.*

Démonstration. Si $u = m_0$ est le minimum absolu de f , alors il est évident que $S(M_u) = (SM)_u$.

Soit $EM = M \times [-1, 1]$, et notons $(EM)_u$ l'ensemble de sous-niveau pour Ef . Supposons premièrement que u est une valeur régulière, et construisons une déformation forte $H : X \times [0, 1] \rightarrow X$ de l'ensemble

$$X = (EM)_u \cup (M \times \{-1, 1\})$$

sur

$$A = E(M_u) \cup (M \times \{-1, 1\}).$$

Ces ensembles sont illustrés sur la figure 4. Puisque, pour tout $(x, s) \in A$, $H((x, s), \lambda)$ doit être constant pour $\lambda \in [0, 1]$, il suffira de définir H de façon continue sur $\overline{X \setminus A} \times [0, 1]$. Pour chaque $(x, s) \in X \setminus A$, $f(x) > u$. Rappelons que $\nabla Ef(x, s)$ égale 0 si et seulement si ou bien $s = \pm 1$ et $f(x) = m_0$, ou alors $\nabla f(x) = 0$. Le cas $f(x) = m_0$ a déjà été vérifié en début de démonstration. On note que $\{x_0\} \times [-1, 1]$ est disjoint de $\overline{X \setminus A}$. Suivant l'hypothèse que u est une valeur régulière, le champ de vecteurs gradient $-\nabla Ef$ définit un flot $t \mapsto \varphi((x, s), t)$ sur $\overline{X \setminus A}$ sur les trajectoires duquel Ef diminue. En l'absence de points critiques, tout $(x, s) \in \overline{X \setminus A}$ est envoyé par φ sur A en un temps fini $t(x, s)$. En effet, le fait que $t(x, s) < \infty$ peut être vérifié par l'équation différentielle pour la seconde coordonnée de $-\nabla Ef$,

$$\frac{ds}{dt} = 2sf(x) - 2sm_0,$$

dont la solution est $s(t) = s_0 e^{2t(f(x) - m_0)}$. Si la trajectoire de (x, s) n'atteint pas $E(M_u)$ en premier, alors la seconde coordonnée s doit atteindre 1 ou -1 . Par des arguments classiques de la théorie des systèmes dynamiques (cf. la preuve de continuité de τ dans [KMM04, Th. 10.37]), $t(x, s)$ est une fonction continue de $(x, s) \in X$.

Définissons $H : \overline{X \setminus A} \times [0, 1] \rightarrow \overline{X \setminus A}$ par la formule suivante :

$$H((x, s), \lambda) = \varphi((x, s), \lambda t(x, s)).$$

Selon les propriétés du flot, $H((x, s), 0) = (x, s)$, et selon le choix de $t(x, s)$, $H((x, s), 1) \in A$, pour tout $(x, s) \in \overline{X \setminus A}$. Si $(x, s) \in A$, alors $t(x, s) = 0$ et ainsi $H((x, s), \lambda) = (x, s)$ pour tout $\lambda \in [0, 1]$. On en conclut que H est une déformation forte de X sur A .

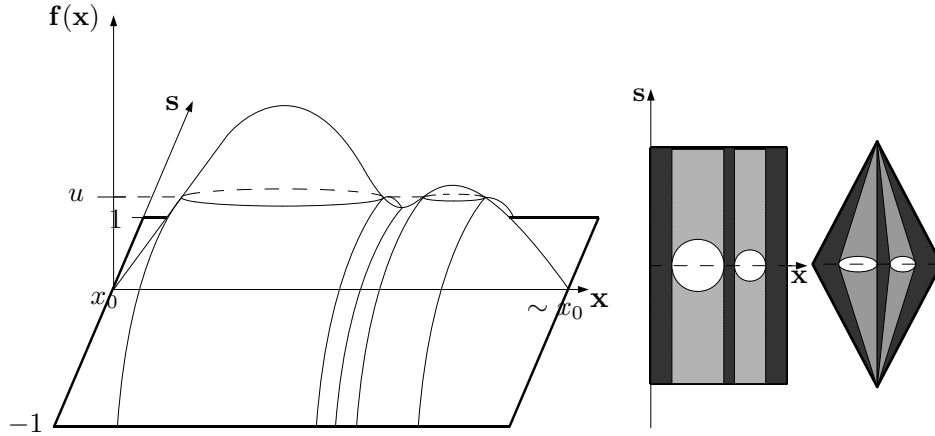


FIGURE 4 – À gauche : la fonction $Ef : M \times [-1, 1] \rightarrow \mathbb{R}$ pour une variété M avec fonction de mesure $f : M \rightarrow \mathbb{R}$, qui atteint son minimum global en x_0 . La variété a été « déconnectée » en x_0 pour faciliter la visualisation. On voit que Ef atteint son minimum global sur l'union de $\{x_0\} \times [-1, 1]$ avec $M \times \{-1, 1\}$ (trait plus épais). Au milieu : Ef vue de haut. En gris foncé, l'ensemble $E(M_u)$; en gris pâle et foncé, l'ensemble $(EM)_u$ (lemme 3.2). À droite : $S(M_u)$ en gris foncé, $(SM)_u$ en gris pâle et foncé. $M \times \{-1\}$ et $M \times \{1\}$ sont tous deux réduits à un point.

Puisque, pour tout λ , $H(\cdot, \lambda)$ est l'application identité sur $M \times \{-1, 1\} \subset A$, la déformation $\tilde{H} : (SM)_u \times [0, 1] \rightarrow (SM)_u$ est bien définie par la formule

$$\tilde{H}(q(x, s), \lambda) = q(H((x, s), \lambda)),$$

où q désigne l'application quotient de EM dans SM . Ceci démontre que pour une valeur régulière u , il existe une équivalence homotopique $S(M_u) \cong (SM)_u$.

Soit maintenant u une valeur critique autre que le minimum absolu m_0 que nous avons déjà vérifié. Puisque u est isolée, il existe $\epsilon > 0$ tel que toutes les valeurs dans $(u, u + \epsilon]$ sont régulières. Dans la théorie de Morse classique, il est connu que M_u est un rétract de déformation forte de $M_{u+\epsilon}$, et donc $M_u \cong M_{u+\epsilon}$. Ainsi $S(M_u) \cong S(M_{u+\epsilon})$. Par un argument équivalent, $(SM)_u \cong (SM)_{u+\epsilon}$. Nous avons déjà montré que $S(M_{u+\epsilon}) \cong (SM)_{u+\epsilon}$, et donc, par transitivité, $S(M_u) \cong (SM)_u$. \square

3.2 Suspension et nombres de Betti persistants

Nous discuterons maintenant de l'effet de l'opération de suspension sur les nombres de Betti persistants pour une fonction de mesure f . Nous savons que M est triangulable, et nous supposons que dans les expérimentations numériques, nous utiliserons une triangulation ayant comme propriété que les points critiques de f (et en particulier x_0) soient des sommets. En fait, f sera linéaire par morceaux sur cette triangulation. Voici notre premier résultat.

Théorème 3.3. *Soit $f : M \rightarrow \mathbb{R}$ une fonction de Morse sur une variété M compacte et connexe. Pour tout $(u, v) \in \Delta^+$,*

$$\tilde{\beta}_{S_{\mathbb{R}}f}^{q+1}(u, v) = \tilde{\beta}_f^q(u, v), \quad q \geq -1. \quad (5)$$

Pour les nombres de Betti persistants non réduits, nous avons

$$\beta_{S_{\mathbb{R}}f}^{q+1}(u, v) = \beta_f^q(u, v), \quad q \geq 1, \quad (6)$$

$$\beta_{S_{\mathbb{R}}f}^1(u, v) = \begin{cases} \beta_f^0(u, v) & \text{si } u < m_0, \\ \beta_f^0(u, v) - 1 & \text{sinon} \end{cases}, \quad (7)$$

et

$$\beta_{S_{\mathbb{R}}f}^0(u, v) = \begin{cases} 0 & \text{si } u < m_0, \\ 1 & \text{sinon} \end{cases}. \quad (8)$$

Ces formules sont aussi valides pour $\Sigma_{\mathbb{R}}f$.

Démonstration. L'équation (5) est une conséquence de (1), du lemme 3.2, et des propriétés fonctorielles de l'opération de suspension. L'équation (6) en découle directement. L'équation (8) provient de (3) et du fait que m_0 est le minimum absolu, et qu'en cette valeur naît donc une composante dont la persistance est infinie. Pour prouver l'équation (7), les arguments précédents prouvent la commutativité du diagramme suivant :

$$\begin{array}{ccc} H_0(M_u) & \xrightarrow{H_0(j_f^{(u,v)})} & H_0(M_v) \\ \downarrow \cong & & \downarrow \cong \\ H_1(SM_u) \oplus \mathbb{K} & \xrightarrow{H_1(j_{S_{\mathbb{R}}f}^{(u,v)})} & H_1(SM_v) \oplus \mathbb{K} \end{array}$$

Soient x_0, v_1, \dots, v_n des sommets représentatifs d'une base des générateurs de l'homologie H_0 pour M_u , et soient x_0, w_1, \dots, w_m ceux d'une base de l'homologie H_0 pour M_v . Ce que ceci veut dire est que l'espace M_u est composé de $n + 1$ composantes connexes, une d'entre elles associée à x_0 (il s'agit de celle dont la persistance est infinie) et n d'entre elles associées à des sommets v_i de notre triangulation. L'homologie $H_0(M_u)$ est un espace vectoriel de dimension $n + 1$ sur \mathbb{K} , que nous pouvons écrire

$$H_0(M_u) = \bigoplus_{k=1}^n \mathbb{K}[v_k - x_0] \oplus \mathbb{K}[x_0] = \tilde{H}_0(M_u) \oplus \mathbb{K}[x_0].$$

Il en va de même pour $H_0(M_v)$, remplaçant les v_k par les w_k et n par m . Ainsi $H_0(j_f^{(u,v)})$ se sépare en (ϕ_1, ϕ_2) , où

$$\phi_1 : \tilde{H}_0(M_u) \rightarrow \tilde{H}_0(M_v)$$

et $\phi_2 = \text{id}_{\mathbb{K}[x_0]}$. L'isomorphisme indiqué par la flèche verticale à la gauche du diagramme précédent se sépare en (ψ_1, ψ_2) où $\psi_1 : \tilde{H}_0(M_u) \rightarrow H_1(SM_u)$ et $\psi_2([x_0]) = 1_{\mathbb{K}}$. Le fait que la suspension du cycle d'ordre 0, $v_k - x_0$, est un cycle d'ordre 1 montre la validité de cet argument. Il en est de même pour l'isomorphisme associé à la flèche verticale de droite. Nous obtenons

$$\text{rang } H_1(j_{S_{\mathbb{R}}f}^{(u,v)}) = \text{rang } \phi_1 = \beta_f^0(u, v) - 1.$$

Puisque ΣM_u est homotopiquement équivalent à SM_u pour chaque u , les arguments sont aussi valides pour $\Sigma_{\mathbb{R}}f$. \square

Nous utiliserons le théorème 3.3 pour prouver des résultats concernant la distance d'appariement entre les nombres de Betti persistants.

Théorème 3.4. *Soient M, N deux variétés compactes, connexes et homéomorphes l'une avec l'autre. Soient aussi leurs fonctions de mesure, $f : M \rightarrow \mathbb{R}$ et $g : N \rightarrow \mathbb{R}$ deux fonctions de Morse. On notera m_0 et n_0 les minima respectifs de f sur M et de g sur N . Alors*

$$d(\tilde{\beta}_{S_{\mathbb{R}}f}^{q+1}, \tilde{\beta}_{S_{\mathbb{R}}g}^{q+1}) = d(\tilde{\beta}_f^q, \tilde{\beta}_g^q), \quad q \geq -1. \quad (9)$$

Pour la distance d'appariement entre nombres de Betti persistants non réduits :

$$d(\beta_{S_{\mathbb{R}}f}^{q+1}, \beta_{S_{\mathbb{R}}g}^{q+1}) = d(\beta_f^q, \beta_g^q), \quad q \geq 1, \quad (10)$$

$$d(\beta_{S_{\mathbb{R}}f}^1, \beta_{S_{\mathbb{R}}g}^1) \leq d(\beta_f^0, \beta_g^0), \quad (11)$$

et

$$d(\beta_{S_{\mathbb{R}}f}^0, \beta_{S_{\mathbb{R}}g}^0) = |m_0 - n_0|. \quad (12)$$

De plus, si $m_0 = n_0$, nous avons que

$$d(\beta_{S_{\mathbb{R}}f}^{q+1}, \beta_{S_{\mathbb{R}}g}^{q+1}) = \begin{cases} d(\beta_f^q, \beta_g^q) & \text{si } q \geq 0, \\ 0 & \text{si } q = -1 \end{cases}. \quad (13)$$

Ces formules sont aussi valides pour $\Sigma_{\mathbb{R}}f$ and $\Sigma_{\mathbb{R}}g$.

Démonstration. Les équations (9) et (10) sont des corollaires directs, respectivement, de (5) et (6). L'équation (12), quant à elle, est dérivée de (8), cette dernière équation signifiant que le diagramme de persistance de $\beta_{S_{\mathbb{R}}f}^0$ possède un seul point, soit un point à l'infini avec m_0 comme abscisse. De même, le diagramme de $\beta_{S_{\mathbb{R}}g}^0$ possède seulement un point à l'infini avec abscisse n_0 . Il en suit que $d(\beta_{S_{\mathbb{R}}f}^0, \beta_{S_{\mathbb{R}}g}^0) = |m_0 - n_0|$.

Pour prouver l'équation (11), notez qu'en vertu de (7), les diagrammes de β_f^0 and $\beta_{S_{\mathbb{R}}f}^1$ ne diffèrent qu'en ceci que le premier a un point à l'infini avec abscisse m_0 que le second n'a pas. De même, le diagramme de β_g^0 possède un point à l'infini avec abscisse n_0 non présent dans celui de $\beta_{S_{\mathbb{R}}g}^1$. Ainsi, le calcul de $d(\beta_{S_{\mathbb{R}}f}^1, \beta_{S_{\mathbb{R}}g}^1)$ se fait sur des sous-ensembles propres des ensembles de points utilisés dans le calcul de $d(\beta_f^0, \beta_g^0)$. L'équation (11) en découle directement. Par

contre, si $m_0 = n_0$, ces deux points à l'infini, étant distants l'un de l'autre de 0, seront appariés dans la bijection optimale entre les diagrammes de β_f^0 et β_g^0 et n'influenceront pas la valeur de $d(\beta_f^0, \beta_g^0)$. De cet argument, ainsi que des équations (10) et (12), on conclut à la validité de (13). \square

Pour ce qui est de la distance d'appariement multidimensionnelle, le résultat suivant devrait être vrai.

Conjecture 3.5. *Soient M, N deux variétés compactes, connexes et homéomorphes l'une avec l'autre. Soient aussi leurs fonctions de mesure, $f : M \rightarrow \mathbb{R}^k$ et $g : N \rightarrow \mathbb{R}^k$ deux fonctions dont les composantes sont des fonctions de Morse. Alors*

$$D(\tilde{\beta}_{S_{\mathbb{R}f}}^{q+1}, \tilde{\beta}_{S_{\mathbb{R}g}}^{q+1}) = D(\tilde{\beta}_f^q, \tilde{\beta}_g^q), \quad q \geq 0. \quad (14)$$

Pour la distance d'appariement k -dimensionnelle entre les nombres de Betti persistants non réduits :

$$D(\beta_{S_{\mathbb{R}f}}^{q+1}, \beta_{S_{\mathbb{R}g}}^{q+1}) = D(\beta_f^q, \beta_g^q), \quad q \geq 1, \quad (15)$$

et

$$D(\beta_{S_{\mathbb{R}f}}^1, \beta_{S_{\mathbb{R}g}}^1) \leq D(\beta_f^0, \beta_g^0). \quad (16)$$

La conjecture 3.5, bien qu'étant probablement vraie, ne semble pas pouvoir être démontrée de manière aisée. Par contre, une preuve existe dans le cas où f et g sont des *interpolations par axes* [CEF⁺] de données définies sur les sommets de triangulations. Le lecteur est invité à consulter [EK] pour plus de détails.

4 Exemples numériques

Les exemples numériques décrits dans cette section ont été obtenus avec des fonctions de mesure bidimensionnelles. Nous notons que dans ce cas l'ensemble des paires admissibles Adm_2 est composé des couples $((a, 1-a), (b, -b))$ tels que $a \in (0, 1)$ et $b \in \mathbb{R}$. Pour un complexe simplicial \mathcal{K} , nous noterons $\mathcal{V}(\mathcal{K})$ l'ensemble de ses sommets, et nous considérerons sa *suspension simpliciale* $S(\mathcal{K})$ dont l'ensemble des sommets est $\mathcal{V}(\mathcal{K})$ union deux points disjoints w_0 et w_1 , et dont les simplexes sont ceux de \mathcal{K} auxquels, pour tout simplexe $[v_0, \dots, v_d]$ dans \mathcal{K} , on ajoute $[v_0, \dots, v_d, w_0]$ et $[v_0, \dots, v_d, w_1]$. Il est connu que si \mathcal{K} est une triangulation de la variété M , alors $S(\mathcal{K})$ est une triangulation de SM . Pour une fonction de mesure $\varphi : \mathcal{V}(\mathcal{K}) \rightarrow \mathbb{R}^2$, nous noterons $S_{\mathbb{R}\varphi} = (S_{\mathbb{R}\varphi_1}, S_{\mathbb{R}\varphi_2}) : \mathcal{V}(S(\mathcal{K})) \rightarrow \mathbb{R}^2$ la fonction telle que $S_{\mathbb{R}\varphi_i}(v) = \varphi_i(v)$ pour tout sommet v de \mathcal{K} et $S_{\mathbb{R}\varphi_i}(w_j) = m_{i,0}$, $j = 0, 1$, $m_{i,0}$ étant le minimum absolu de φ_i sur $\mathcal{V}(\mathcal{K})$. Nous utiliserons aussi les notations $\Phi_{(a,b)}$ pour $\text{red}_{((a,1-a),(b,-b))}^\varphi$ et $S_{\mathbb{R}}\Phi_{(a,b)}$ pour $\text{red}_{((a,1-a),(b,-b))}^{S_{\mathbb{R}\varphi}}$.

Nos exemples sont destinés à procurer un support expérimental à la théorie décrite dans la section 3. Nous avons utilisé deux paires de modèles, chacun d'entre eux étant un complexe simplicial en dimension 1 ou 2, homéomorphe à une variété compacte et connexe. Nos résultats ont été obtenus à l'aide d'une

implantation de l'algorithme décrit dans [BCFG11] servant à calculer des bornes inférieures pour la distance d'appariement bidimensionnelle entre des nombres de Betti persistants avec une précision croissante. Cet algorithme, prenant en entrée des paires (K, φ) et (L, ψ) , calcule $d(\beta_{\Phi_{(a,b)}}, \beta_{\Psi_{(a,b)}})$, où (a, b) est choisi sur des grilles de plus en plus fines jusqu'à ce que la tolérance requise soit atteinte. Notre implantation (codée en Java avec l'aide de N. Cavazza, co-auteur de [CEF⁺]) intègre le logiciel d'homologie persistante JPLex [dSCVJ] dans l'étape du calcul des nombres de Betti persistants.

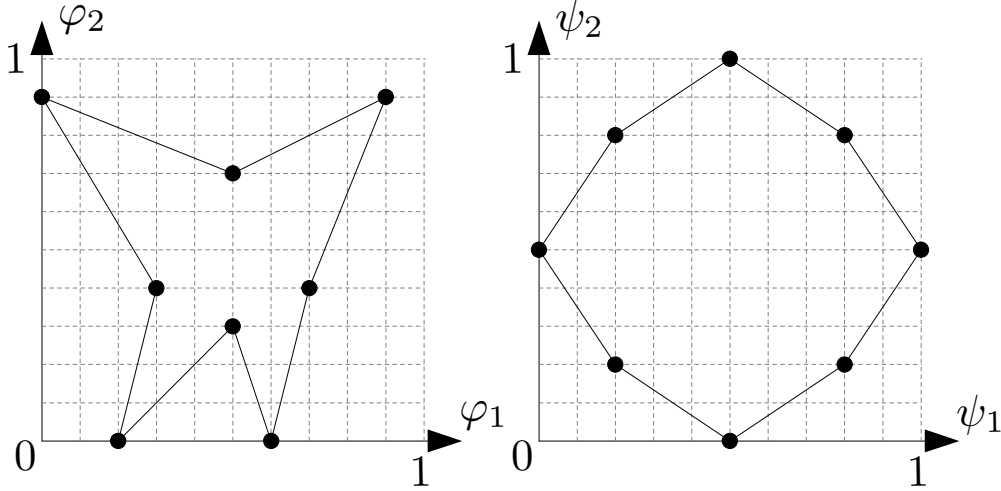


FIGURE 5 – Paire de modèles pour l'exemple 1D. Les fonctions de mesure sont les fonctions de coordonnées x et y .

Exemple 4.1. L'exemple 1D est la paire de modèles (K, φ) et (L, ψ) montrés à la figure 5. Tel qu'illustré, φ_1 et ψ_1 , ainsi que φ_2 et ψ_2 , atteignent le même minimum global (dans les deux cas 0), alors nous nous attendons à ce que l'équation (13) soit vérifiée. Les distances d'appariement 2D sont calculées avec une tolérance $\epsilon = 0,01$, mais vu la simplicité de l'exemple, la probabilité est très forte qu'elles soient en fait exactes. Les résultats de nos calculs peuvent être consultés dans les deux tableaux suivants :

| | β | $\tilde{\beta}$ | | $q = 0$ | $q = 1$ |
|--|---------|-----------------|--|---------|---------|
| $d(\beta_{\varphi_1}^0, \beta_{\psi_1}^0)$ | 0,05 | 0,05 | $d(\beta_{S_{\mathbb{R}}\varphi_1}^q, \beta_{S_{\mathbb{R}}\psi_1}^q)$ | 0 | 0,05 |
| $d(\beta_{\varphi_2}^0, \beta_{\psi_2}^0)$ | 0,15 | 0,15 | $d(\beta_{S_{\mathbb{R}}\varphi_2}^q, \beta_{S_{\mathbb{R}}\psi_2}^q)$ | 0 | 0,15 |
| $D(\beta_{\varphi}^0, \beta_{\psi}^0)$ | 0,2 | 0,15 | $D(\beta_{S_{\mathbb{R}}\varphi}^q, \beta_{S_{\mathbb{R}}\psi}^q)$ | 0,1 | 0,15 |

Tel que montré dans l'exemple, il n'est pas garanti que $D(\beta_{S_{\mathbb{R}}\varphi}^0, \beta_{S_{\mathbb{R}}\psi}^0)$ égale 0 même si φ_i et ψ_i atteignent le même minimum global pour $i = 1, 2$. Les résultats pour les nombres de Betti persistants réduits dans le cas des modèles suspendus n'ont pas été inclus dans le tableau, puisqu'ils égalent ceux pour les nombres de Betti persistants non réduits. Ceci n'est cependant pas nécessairement vrai en

général. La figure 6 montre additionnellement les graphiques de $d(\beta_{\Phi(a,b)}^q, \beta_{\Psi(a,b)}^q)$ et de $d(\beta_{\mathbb{S}_{\mathbb{R}}\Phi(a,b)}^q, \beta_{\mathbb{S}_{\mathbb{R}}\Psi(a,b)}^q)$, considérés comme fonctions de a et b , pour quelques valeurs intéressantes de q .

Exemple 4.2. Le modèle 2D est illustré à la figure 7. La fonction de mesure $\varphi : \mathcal{V}(\mathcal{K}) \rightarrow \mathbb{R}^2$ est définie de la manière suivante : supposons que $\mathcal{V}(\mathcal{K}) = \{v_1, \dots, v_n\}$ et que c est le centre de masse de K défini comme moyenne pondérée des barycentres de chaque triangle. Nous définissons un vecteur principal

$$\vec{w} = \frac{\sum_{i=1}^n (v_i - c) \|v_i - c\|_2}{\sum_{i=1}^n \|v_i - c\|_2^2}$$

où $\|\cdot\|_2$ est la norme euclidienne. Soit maintenant d la droite passant par c et ayant \vec{w} comme vecteur directeur, et π le plan passant par c et ayant \vec{w} comme vecteur normal. La fonction φ est définie comme

$$\varphi_1(v_i) = 1 - \frac{\text{dist}(v_i, d)}{\max_{j=1, \dots, n} \text{dist}(v_j, d)}$$

et

$$\varphi_2(v_i) = 1 - \frac{\text{dist}(v_i, \pi)}{\max_{j=1, \dots, n} \text{dist}(v_j, \pi)},$$

où dist est la distance euclidienne entre un point et la droite d ou le plan π . La fonction de mesure ψ est définie de manière équivalente sur l'ensemble des sommets $\mathcal{V}(\mathcal{L})$. Une conséquence directe de ces définitions est que les minima globaux de φ_1 , φ_2 , ψ_1 and ψ_2 seront tous 0. La table 1 montre les résultats de nos calculs. Toutes les bornes inférieures calculées pour la distance d'appariement 2D sont présentées, dans le but d'illustrer la comparabilité des résultats entre le modèle original et le modèle suspendu à chaque étape de l'algorithme.

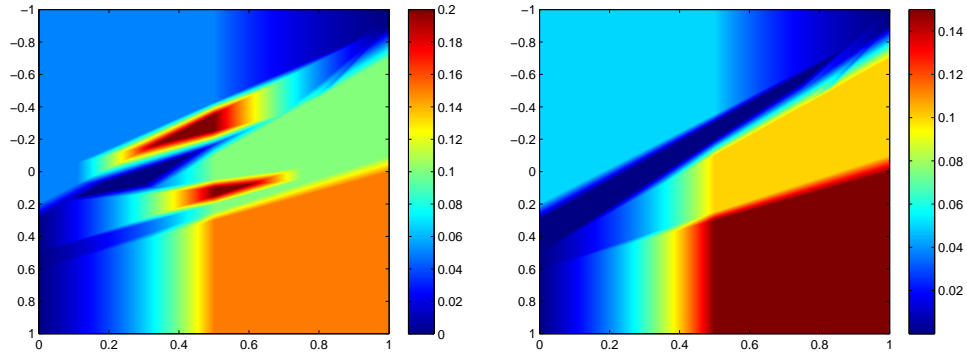


FIGURE 6 – Première image : graphique de $d(\beta_{\Phi(a,b)}^0, \beta_{\Psi(a,b)}^0)$. Deuxième image : graphique de $d(\tilde{\beta}_{\Phi(a,b)}^0, \tilde{\beta}_{\Psi(a,b)}^0)$, qui est identique à celui de $d(\beta_{\mathbb{S}_{\mathbb{R}}\Phi(a,b)}^1, \beta_{\mathbb{S}_{\mathbb{R}}\Psi(a,b)}^1)$, tel qu'attendu selon l'équation (9). L'axe horizontal correspond à a , l'axe vertical à b .

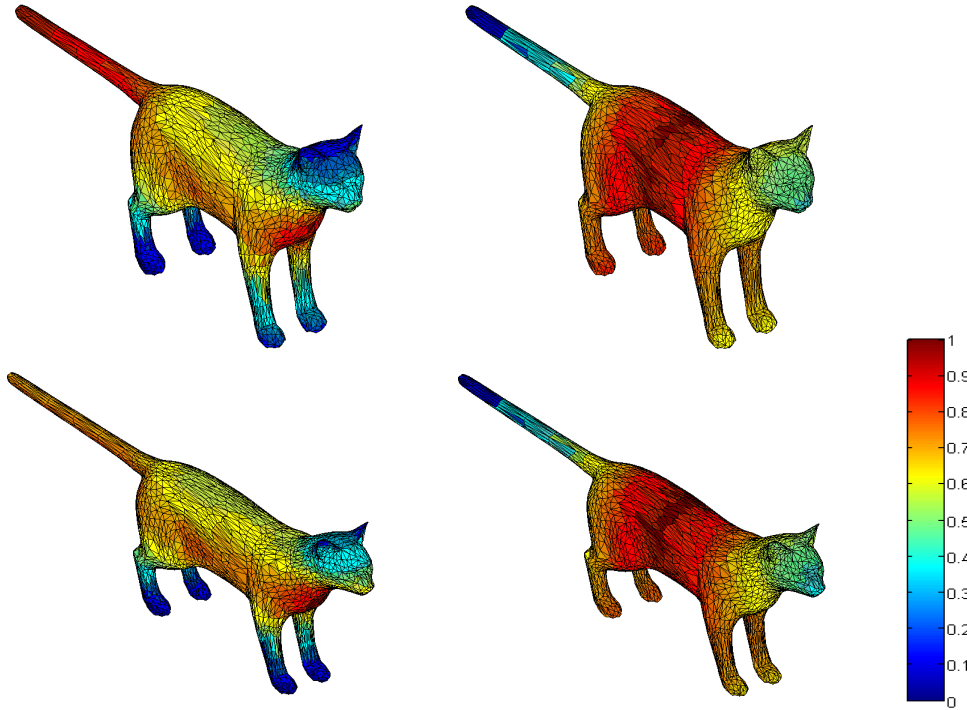


FIGURE 7 – Première rangée : modèle K et fonctions de mesure φ_1 et φ_2 . Deuxième rangée : modèle L et fonctions de mesure ψ_1 et ψ_2 . Remerciements aux auteurs de [BCFG11] pour nous avoir fourni les modèles.

| | ϵ | | | ϵ | |
|--|------------|-----------|--|------------|-----------|
| $d(\beta_{\varphi_1}^0, \beta_{\psi_1}^0)$ | | 0,118 165 | $d(\tilde{\beta}_{\varphi_1}^0, \tilde{\beta}_{\psi_1}^0)$ | | 0,118 165 |
| $d(\beta_{\varphi_2}^0, \beta_{\psi_2}^0)$ | | 0,032 043 | $d(\tilde{\beta}_{\varphi_2}^0, \tilde{\beta}_{\psi_2}^0)$ | | 0,032 043 |
| $D(\beta_{\varphi}^0, \beta_{\psi}^0)$ | 9/8 | 0,194 217 | $D(\tilde{\beta}_{\varphi}^0, \tilde{\beta}_{\psi}^0)$ | 9/8 | 0,118 165 |
| | 9/16 | 0,224 227 | | 9/16 | 0,127 301 |
| | 9/32 | 0,225 394 | | 9/32 | 0,135 530 |
| | 9/64 | 0,225 394 | | 9/64 | 0,144 274 |

| | ϵ | | | ϵ | |
|--|------------|-----------|--|------------|-----------|
| $d(\beta_{\varphi_1}^1, \beta_{\psi_1}^1)$ | | 0,031 129 | $d(\beta_{S_{\mathbb{R}}\varphi_1}^1, \beta_{S_{\mathbb{R}}\psi_1}^1)$ | | 0,118 165 |
| $d(\beta_{\varphi_2}^1, \beta_{\psi_2}^1)$ | | 0,039 497 | $d(\beta_{S_{\mathbb{R}}\varphi_2}^1, \beta_{S_{\mathbb{R}}\psi_2}^1)$ | | 0,032 043 |
| $D(\beta_{\varphi}^1, \beta_{\psi}^1)$ | 9/8 | 0,039 497 | $D(\beta_{S_{\mathbb{R}}\varphi}^1, \beta_{S_{\mathbb{R}}\psi}^1)$ | 9/8 | 0,118 165 |
| | 9/16 | 0,046 150 | | 9/16 | 0,127 301 |
| | 9/32 | 0,046 150 | | 9/32 | 0,135 530 |
| | 9/64 | 0,046 150 | | 9/64 | 0,144 274 |

| | ϵ | | | ϵ | |
|--|------------|-----------|--|------------|-----------|
| $d(\beta_{S_{\mathbb{R}}\varphi_1}^2, \beta_{S_{\mathbb{R}}\psi_1}^2)$ | | 0,031 129 | $d(\beta_{S_{\mathbb{R}}\varphi_1}^0, \beta_{S_{\mathbb{R}}\psi_1}^0)$ | | 0,000 000 |
| $d(\beta_{S_{\mathbb{R}}\varphi_2}^2, \beta_{S_{\mathbb{R}}\psi_2}^2)$ | | 0,039 497 | $d(\beta_{S_{\mathbb{R}}\varphi_2}^0, \beta_{S_{\mathbb{R}}\psi_2}^0)$ | | 0,000 000 |
| $D(\beta_{S_{\mathbb{R}}\varphi}^2, \beta_{S_{\mathbb{R}}\psi}^2)$ | 9/8 | 0,039 497 | $D(\beta_{S_{\mathbb{R}}\varphi}^0, \beta_{S_{\mathbb{R}}\psi}^0)$ | 9/8 | 0,155 527 |
| | 9/16 | 0,046 150 | | 9/16 | 0,171 368 |
| | 9/32 | 0,046 150 | | 9/32 | 0,179 821 |
| | 9/64 | 0,046 150 | | 9/64 | 0,194 217 |

TABLE 1 – Deux premiers tableaux : résultats pour modèles originaux avec β_0 et $\tilde{\beta}_0$. Troisième tableau : résultats pour modèles originaux avec β_1 . Trois derniers tableaux : résultats pour modèles suspendus avec β_1 , β_2 et β_0 . Résultats pour modèles suspendus avec $\tilde{\beta}_0$ identiques au sixième tableau.

Références

- [BCF⁺08] S. BIASOTTI, A. CERRI, P. FROSINI, D. GIORGI et C. LANDI : Multidimensional size functions for shape comparison. *J. Math. Imaging Vision*, 32(2):161–179, 2008.
- [BCFG11] S. BIASOTTI, A. CERRI, P. FROSINI et D. GIORGI : A new algorithm for computing the 2-dimensional matching distance between size functions. *Pattern Recognition Letters*, 32:1735–1746, 2011.
- [CEF⁺] N. CAVAZZA, M. ETHIER, P. FROSINI, C. LANDI et T. KACZYNSKI : Comparison of persistent homologies for vector functions : from continuous to discrete and back. <http://arxiv.org/abs/1201.3217>.
- [dFL06] M. D’AMICO, P. FROSINI et C. LANDI : Using matching distance in size theory : a survey. *Int. J. Imaging Syst. Technol.*, 16(5):154–161, 2006.
- [dSCVJ] V. de SILVA, G. CARLSSON et M. VEJDEMO JOHANSSON : Plex : A system for computational homology. <http://comp-top.stanford.edu/u/programs/jplex/>.
- [EK] M. ETHIER et T. KACZYNSKI : Suspension of a measuring function. Prépublication septembre 2012.
- [Hat02] A. HATCHER : *Algebraic Topology*. Cambridge University Press, 2002.
- [KMM04] T. KACZYNSKI, K. MISCHAIKOW et M. MROZEK : *Computational Homology*. Numéro 157 de Applied Mathematical Sciences. Springer-Verlag, 1^{re} édition, 2004.
- [Mun84] J. MUNKRES : *Elements of Algebraic Topology*. Addison-Wesley, 1984.

MARC ETHIER
DÉPARTEMENT DE MATHÉMATIQUES, UNIVERSITÉ DE SHERBROOKE
Courriel: marc.ethier@usherbrooke.ca

CLUSTER ALGEBRAS AND MARKOFF NUMBERS

XUEYUAN PENG AND JIE ZHANG

ABSTRACT We introduce Markoff numbers and reveal their connection to the cluster algebra associated to the once-punctured torus.

1 Introduction

Cluster algebras were first introduced and studied in [FZ02] by Fomin and Zelevinsky. The theory of cluster algebras has connections with various areas of mathematics like quiver representations, Lie theory, combinatorics, Teichmüller theory and Poisson geometry.

Markoff numbers, which have been introduced in the work of A. A. Markoff [Mar79, Mar80], are numbers that satisfy the Diophantine Equation

$$a^2 + b^2 + c^2 = 3abc.$$

The study of Markoff numbers is an important subject in combinatorics and plays a role in the theory of rational approximation of irrational numbers.

In this note, we introduce Markoff numbers in the next section. In the third section, we study the cluster algebra arising from the once-punctured torus. The last section is devoted to the connections between Markoff numbers and the cluster algebra.

2 Markoff numbers and Markoff triples

Definition 2.1. A triple (a, b, c) of positive integers is a **Markoff triple** if (a, b, c) is a solution to the Diophantine Equation

$$a^2 + b^2 + c^2 = 3abc.$$

The numbers a, b, c occurring in a Markoff triple are called **Markoff numbers**.

Example 2.2. $(1, 1, 1), (1, 1, 2), (1, 2, 5), (1, 13, 5), (2, 5, 29)$ are Markoff triples, therefore 1, 2, 5, 13, 29 are Markoff numbers.

Exercice 2.3. Is $(13, 2, 5)$ a Markoff triple?

The authors are grateful to their supervisor Professor Thomas Brüstle for his guidance and help.

Each Markoff triple can produce other Markoff triples by means of the following proposition.

Proposition 2.4. *Given a Markoff triple (a, b, c) , set*

$$a' = 3bc - a, \quad b' = 3ac - b, \quad c' = 3ab - c.$$

Then (a', b, c) , (a, b', c) and (a, b, c') are also Markoff triples.

Proof. We only show that (a', b, c) is a Markoff triple. It suffices to verify that

$$a'^2 + b^2 + c^2 = 3a'bc.$$

$$\begin{aligned} a'^2 + b^2 + c^2 &= (3bc - a)^2 + b^2 + c^2 = 9b^2c^2 + a^2 - 6abc + a^2 + b^2 \\ &= 9b^2c^2 + 3abc - 6abc = 9b^2c^2 - 3abc = 3(3bc - a)bc = 3a'bc. \end{aligned}$$

Similarly, one can check that (a, b', c) and (a, b, c') are also Markoff triples. \square

Therefore one can obtain three new Markoff triples from a given Markoff triple (a, b, c) by **mutations** at each position which we denote as follows:

$$(a, b, c) \xrightarrow{\mu_1} (a', b, c), (a, b, c) \xrightarrow{\mu_2} (a, b', c), (a, b, c) \xrightarrow{\mu_3} (a, b, c').$$

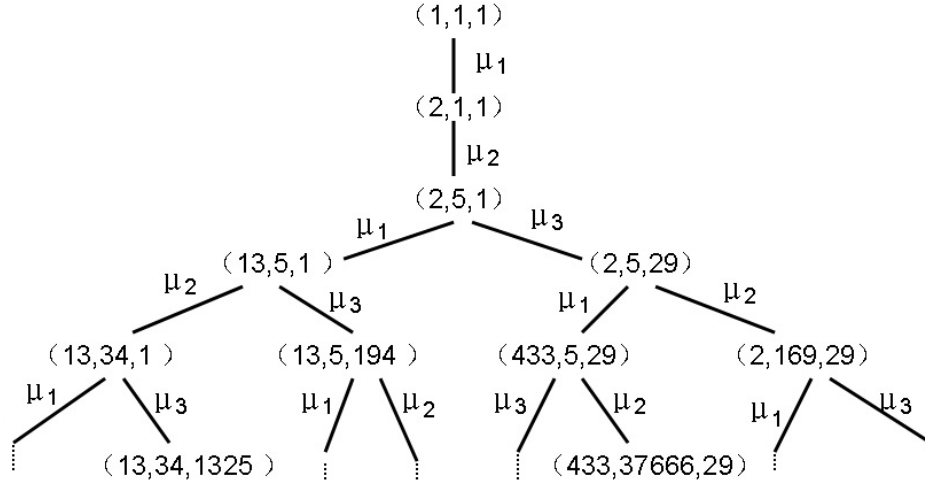
Remark 2.5. The mutation is an involution which means $\mu_i^2 = \text{Identity}$, where $i = 1, 2, 3$. For example,

$$\mu_1^2(a, b, c) = \mu_1(3bc - a, b, c) = (3bc - (3bc - a), b, c) = (a, b, c).$$

A. A. Markoff found in [Mar79] that all Markoff triples are connected by mutations:

Theorem 2.6. *All Markoff triples are obtained from $(1, 1, 1)$ by iterated mutations.*

Note that all permutations of entries of a Markoff triple (a, b, c) are still Markoff triples. For example, $(1, 2, 5)$, $(1, 5, 2)$, $(2, 1, 5)$, $(2, 5, 1)$, $(5, 1, 2)$, $(5, 2, 1)$ are all Markoff triples. In the following, we consider Markoff triples up to permutation and draw a **Markoff tree** with its vertices given by Markoff triples and edges given by mutations:



From the above Markoff tree, one can see that $(433, 37666, 29)$ and $(1, 1, 1)$ are connected as follows:

$$(1, 1, 1) \xrightarrow{\mu_1} (2, 1, 1) \xrightarrow{\mu_2} (2, 5, 1) \xrightarrow{\mu_3} (2, 5, 29) \xrightarrow{\mu_1} (433, 5, 29) \xrightarrow{\mu_2} (433, 37666, 29)$$

which implies

$$\mu_2 \mu_1 \mu_3 \mu_2 \mu_1 (1, 1, 1) = (433, 37666, 29).$$

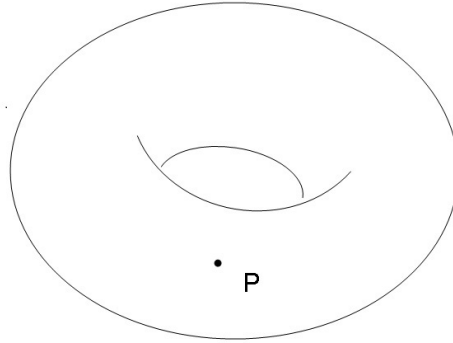
Exercise 2.7. Obtain $(14701, 169, 29)$ from $(1, 1, 1)$.

Frobenius [Fro13] claims that every Markoff number appears uniquely as the largest element of a Markoff triple (up to permutation). In fact, this is still a conjecture:

Conjecture 2.8. *The Markoff triple is determined by its largest entry (up to permutation).*

3 The once-punctured torus

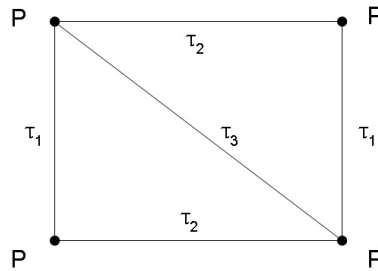
The cluster algebras associated to compact connected oriented Riemann surfaces with marked points have been studied in [SFT08] by Fomin, Shapiro and Thurston. They associated a quiver Q_Γ to each triangulation Γ of the surface, then the mutations of the quivers correspond to the flips of the triangulations. In this section, we consider the cluster algebra arising from the once punctured surface T as follows :



where P is the only puncture on the surface T .

3.1 Quiver associated to the once-punctured torus

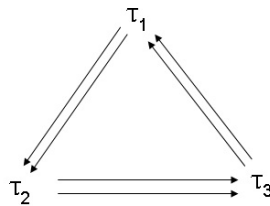
To get the quiver associated to the once punctured torus T , we first need to find a triangulation Γ of T . We cut the torus from P along the vertical line τ_1 and horizontal line τ_2 , then we get the following square S :



By adding one more arc τ_3 as in the above picture, one obtains a triangulation $\Gamma = \{\tau_1, \tau_2, \tau_3\}$ of T . Recall from [SFT08, LF09, ABCJP10] that a quiver $Q_\Gamma = (Q_0, Q_1)$ is associated to each triangulation Γ of the marked surface by the following rule:

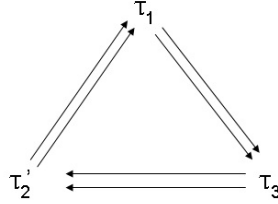
- Q_0 is given by the internal arcs of Γ .
- The set of arrows Q_1 is defined as follows: Whenever there is a triangle Δ in Γ containing two internal arcs a and b , then there is an arrow $\rho : a \rightarrow b$ in Q_1 if a is a predecessor of b with respect to clockwise orientation at the joint vertex of a and b in Δ .

According to the above rule, the quiver Q_Γ of the once-punctured torus T can be easily obtained from the square S above as follows:

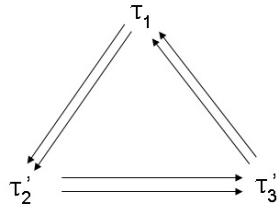


Recall in [BSJ12, FZ02] that the cluster algebra $\mathcal{A} = \mathcal{A}(\mathbf{x}, Q)$ with initial seed $\mathbf{x} = \{x_1, x_2, x_3\}$, $Q = Q_\Gamma$ is a subalgebra of $\mathbb{Q}(x_1, x_2, x_3)$ generated by cluster variables which are obtained by mutations. In order to understand \mathcal{A} , we first want to see what happens if we perform mutations at each vertex of Q , see [Dup10] for more details about quiver mutations.

- We try to mutate τ_2 first, then the quiver $\mu_2(Q) := \mu_{\tau_2}(Q) = Q^{op}$ is as follows

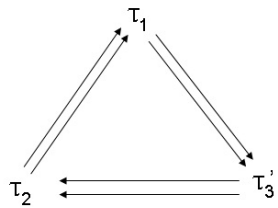


with a new cluster (x_1, x'_2, x_3) where $x'_2 = \frac{x_1^2 + x_3^2}{x_1}$. By mutating at τ_3 again, we get $\mu_3(Q^{op}) := \mu_{\tau_3}(Q^{op}) = \mu_{\tau_3}\mu_{\tau_2}(Q) = Q$ as follows:

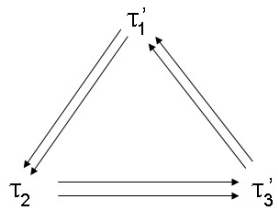


with a new cluster (x_1, x_2', x'_3) where $x'_3 = \frac{x_1^2 + x_2'^2}{x_3}$.

- If we mutate τ_3 first, then $\mu_3(Q) := \mu_{\tau_3}(Q) = Q^{op}$ is as follows:



with a new cluster (x_1, x_2, x'_3) where $x'_3 = \frac{x_1^2 + x_2^2}{x_3}$. Similarly, if we continue mutating at τ_1 , then $\mu_1(Q^{op}) := \mu_{\tau_1}(Q^{op}) = \mu_{\tau_1}\mu_{\tau_3}(Q) = Q$ as follows:



with a new cluster (x'_1, x_2, x'_3) where $x'_1 = \frac{x_2^2 + x_3^2}{x_1}$.

The above procedure implies that the quiver Q_Γ is stable under mutations:

Lemma 3.1. $\mu_i(Q) = Q^{op}$ and $\mu_i(Q^{op}) = Q$ where $i = 1, 2, 3$.

By Fomin-Zelevinsky's work in [FZ02], each cluster variable $f(x_1, x_2, x_3)$ in the cluster algebra \mathcal{A} can be written as

$$f := f(x_1, x_2, x_3) = \frac{g(x_1, x_2, x_3)}{x_1^{d_1} x_2^{d_2} x_3^{d_3}}$$

where $g \in \mathbb{Z}[x_1, x_2, x_3]$ and $d_i \geq 0$ for $i = 1, 2, 3$. Combining the mutation rule of cluster variables with Lemma 3.1, we get the following lemma:

Lemma 3.2. *Let (f_1, f_2, f_3) be a cluster of \mathcal{A} , then*

$$\begin{aligned} \mu_1(f_1, f_2, f_3) &= \left(\frac{f_2^2 + f_3^2}{f_1}, f_2, f_3 \right) \\ \mu_2(f_1, f_2, f_3) &= \left(f_1, \frac{f_1^2 + f_3^2}{f_2}, f_3 \right) \\ \mu_3(f_1, f_2, f_3) &= \left(f_1, f_2, \frac{f_1^2 + f_2^2}{f_3} \right). \end{aligned}$$

Proof. We only prove $\mu_1(f_1, f_2, f_3) = \left(\frac{f_2^2 + f_3^2}{f_1}, f_2, f_3 \right)$. By Lemma 3.1, the quiver associated to the cluster (f_1, f_2, f_3) must be Q or Q^{op} . Without loss of generality, we assume it is Q . Therefore, by definition of mutations of clusters

$$f'_1 = \frac{1}{f_1} \left(\prod_{i \rightarrow 1 \text{ in } Q} f_i + \prod_{1 \rightarrow j \text{ in } Q} f_j \right) = \frac{1}{f_1} (f_2^2 + f_3^2).$$

This completes the proof. □

4 Conclusion

We show in this section the connection between Markoff numbers and the cluster algebra \mathcal{A} . The following is our main theorem:

Theorem 4.1. *Let (f_1, f_2, f_3) be a cluster in the cluster algebra \mathcal{A} . Then $(f_1(1, 1, 1), f_2(1, 1, 1), f_3(1, 1, 1))$ is a Markoff triple.*

Proof. We first consider the initial seed $(f_1, f_2, f_3) = (x_1, x_2, x_3)$, then it is obvious that $(f_1(1, 1, 1), f_2(1, 1, 1), f_3(1, 1, 1)) = (1, 1, 1)$ is a Markoff triple. Therefore the theorem is true for the initial seed.

Assume (f_1, f_2, f_3) is a cluster of \mathcal{A} such that $(f_1(1, 1, 1), f_2(1, 1, 1), f_3(1, 1, 1))$ is a Markoff triple. It suffices to prove that $(f'_1(1, 1, 1), f_2(1, 1, 1), f_3(1, 1, 1))$ is also a Markoff triple where $\mu_1(f_1, f_2, f_3) = (f'_1, f_2, f_3)$ since each cluster is obtained by iterated mutations from the initial seed.

Lemma 3.2 implies that $f'_1 = \frac{f_2^2 + f_3^2}{f_1}$, hence

$$\begin{aligned} f'_1(1, 1, 1) &= \frac{f_2^2(1, 1, 1) + f_3^2(1, 1, 1)}{f_1(1, 1, 1)} \\ &= \frac{3f_1(1, 1, 1)f_2(1, 1, 1)f_3(1, 1, 1) - f_1^2(1, 1, 1)}{f_1(1, 1, 1)} \\ &= 3f_2(1, 1, 1)f_3(1, 1, 1) - f_1(1, 1, 1). \end{aligned}$$

By Proposition 2.4, $(f'_1(1, 1, 1), f_2(1, 1, 1), f_3(1, 1, 1))$ is also a Markoff triple. Similarly, one can prove that $(f_1(1, 1, 1), f'_2(1, 1, 1), f_3(1, 1, 1))$ and $(f_1(1, 1, 1), f_2(1, 1, 1), f'_3(1, 1, 1))$ are also Markoff triples. \square

Keeping all the notations above, then we have the following two corollaries:

Corollary 4.2. $f_i(1, 1, 1)$ is a Markoff number for each $i = 1, 2, 3$.

Corollary 4.3. There is a bijection between the set of all clusters in \mathcal{A} and the set of all Markoff triples given by

$$(f_1, f_2, f_3) \longmapsto (f_1(1, 1, 1), f_2(1, 1, 1), f_3(1, 1, 1)).$$

References

- [ABCJP10] I. Assem, T. Brüstle, G. Charbonneau-Jodoin, and P.-G. Plamondon. Gentle algebras arising from surface triangulations. *Journal of Algebra and Number Theory*, 4(201-229), 2010.
- [BSJ12] D. Boulet-St-Jacques. Les algèbres amassées: Définitions de base et résultats. *Cahiers de Mathématique de l'Université de Sherbrooke*, 2(135-150), 2012.
- [Dup10] G. Dupont. Mutations de carquois. *Cahiers de Mathématique de l'Université de Sherbrooke*, 2(99-117), 2010.
- [Fro13] G. Frobenius. Über die Markoffschen Zahlen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, (458-487), 1913.
- [FZ02] S. Fomin and A. Zelevinsky. Cluster algebras I. Foundations. *J. Amer. Math. Soc.*, 15(2)(497-529), 2002.
- [LF09] D. Labardini-Fragoso. Quivers with potentials associated to triangulated surfaces. *Proc. London Math. Soc.*, 98(797-839), 2009.
- [Mar79] A. A. Markoff. Sur les formes quadratiques binaires indéfinies. *Math. Ann.*, 15(381-496), 1879.
- [Mar80] A. A. Markoff. Sur les formes quadratiques binaires indéfinies (second memoire). *Math. Ann.*, 17(379-399), 1880.

- [SFT08] M. S. S. Fomin and D. Thurston. Cluster algebras and triangulated surfaces. part I: Cluster complexes. *Acta Mathematica*, 201(83-146), 2008.

XUEYUAN PENG

DEPARTMENT OF MATHEMATICS, BISHOP'S UNIVERSITY

Courriel: `xuedoris0106@hotmail.com`

JIE ZHANG

DEPARTMENT OF MATHEMATICAL SCIENCES, TSINGHUA UNIVERSITY

Courriel: `jie.zhang@math.tsinghua.edu.cn`

PAVAGES ADDITIFS

JEAN-FRANÇOIS MARCEAU

RÉSUMÉ Le présent article en est un d'introduction aux pavages additifs, faisant également des liens avec les pavages multiplicatifs. Nous ajoutons également deux nouveaux théorèmes pour les pavages additifs (voir théorème 3.7 et théorème 3.11) et une conjecture sur les pavages multiplicatifs infinis. Nous souhaitons avec cet article ramener les recherches sur les pavages au goût du jour et montrer de nouvelles pistes pour tenter de découvrir une application aux pavages additifs.

1 Introduction

Au cours des dernières années, les avancées en matière d'algèbres amassées ont permis de révéler plusieurs liens avec divers domaines des mathématiques. Par exemple, un lien évident entre les frises de nombres générées à partir d'un carquois de type A_n et les pavages (multiplicatifs) de nombres développés par Conway et Coxeter [CC73a, CC73b] au cours des années 70. Bien que les pavages (multiplicatifs) de Conway et Coxeter soient bien connus, il existe certaines variantes de ceux-ci qui sont moins connues, par exemple les pavages additifs développés par G.C. Shephard [She73] ou les 2-frises étudiées par Morier-Genoud, Ovsienko et Tabachnikov [SMGT11, MG11]. Ce qui suit est une introduction aux pavages additifs de Shephard comportant deux nouveaux théorèmes (2 et 5) et une conjecture.

2 Pavages multiplicatifs

Un *pavage multiplicatif* est un assemblage de nombres arrangés en lignes décalées comme les briques d'un mur de telle sorte que la ligne du haut et la ligne du bas sont composées uniquement de 1 et que chaque losange

$$\begin{array}{ccc} & a & \\ b & & c \\ & d & \end{array}$$

satisfait à l'équation suivante :

$$bc - ad = 1.$$

Cette recherche a été rendue possible grâce à une bourse de recherche de premier cycle du CRSNG et grâce à David Smith de l'Université Bishop's qui m'a guidé tout au long de mes recherches.

Cette équation est appelée *règle unimodulaire*. Voici un exemple de pavage multiplicatif :

$$\begin{array}{cccccccccccc}
 \dots & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & \dots \\
 & \dots & 1 & & 2 & & 2 & & 2 & & 2 & & 1 & & 5 & & 1 & & \dots & & \dots \\
 \dots & 4 & & 1 & & 3 & & 3 & & 3 & & 1 & & 4 & & 4 & & 1 & & \dots & \dots \\
 & \dots & 3 & & 1 & & 4 & & 4 & & 1 & & 3 & & 3 & & 3 & & \dots & & \dots \\
 \dots & 2 & & 2 & & 1 & & 5 & & 1 & & 2 & & 2 & & 2 & & 2 & & 2 & \dots \\
 & \dots & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & 1 & & \dots
 \end{array}$$

Définition 2.1. Étant donné un pavage à n lignes, une *tranche* est un ensemble de n éléments adjacents, où chaque élément provient d’une ligne différente (en gras dans l’exemple suivant). Un *couple* est une paire d’éléments adjacents sur une tranche. Une *diagonale* est une tranche à travers laquelle il est possible de tracer une ligne droite (en italique dans l’exemple suivant).

$$\begin{array}{cccccccccccc}
 \dots & 1 & & 1 & & 1 & & \mathbf{1} & & 1 & & \mathit{1} & & 1 & & 1 & & 1 & & 1 & \dots \\
 & \dots & 1 & & 2 & & 2 & & \mathbf{2} & & 2 & & \mathit{1} & & 5 & & 1 & & \dots & & \dots \\
 \dots & 4 & & 1 & & 3 & & 3 & & \mathbf{3} & & 1 & & \mathit{4} & & 4 & & 1 & & \dots & \dots \\
 & \dots & 3 & & 1 & & 4 & & 4 & & 1 & & 3 & & \mathit{3} & & 3 & & \dots & & \dots \\
 \dots & 2 & & 2 & & 1 & & \mathbf{5} & & 1 & & 2 & & 2 & & 2 & & 2 & & 2 & \dots \\
 & \dots & 1 & & 1 & & 1 & & \mathbf{1} & & 1 & & 1 & & 1 & & \mathit{1} & & 1 & & \dots
 \end{array}$$

Dans ce document, nous utiliserons les notations suivantes afin de spécifier la position d’un nombre appartenant à un pavage :

$$\begin{array}{cccccccccccc}
 \dots & 1 & & & & 1 & & & & 1 & & & & & 1 & & \dots & \text{ligne 1} \\
 & \dots & m_{1,3} & & & m_{2,4} & & & & m_{3,5} & & & & & m_{4,6} & & \dots & \text{ligne 2} \\
 & & \dots & m_{1,4} & & & m_{2,5} & & & & m_{3,6} & & & & \dots & & \text{ligne 3} \\
 & & & \dots & m_{1,5} & & & m_{2,6} & & & \dots & & & & & & \text{ligne 4} \\
 & & & & \dots & m_{1,6} & & \dots & & & & & & & & & \text{ligne 5} \\
 & & & & & & & & & & & & & & & & \vdots \\
 & & & & & & & \dots & & & & & & & & & \vdots
 \end{array}$$

Remarque . Dans un pavage multiplicatif à n lignes, la ligne 1 et la ligne n sont composées uniquement de 1.

2.1 Propriétés

Les huit propriétés des pavages multiplicatifs qui suivent serviront de points de comparaison avec les propriétés des pavages additifs dont nous discuterons dans la section 3. Lors de nos recherches nous avons remarqué que, pour chaque propriété du cas additif, il existe une propriété analogue pour le cas multiplicatif. Il est à noter que toutes les propriétés qui suivent proviennent des travaux de Conway et Coxeter [CC73a, CC73b], à l’exception de la propriété 3 qui a été tirée de l’article de Fraser Martineau et Lavertu [ML10].

satisfait à l'équation suivante :

$$(b + c) - (a + d) = 1. \quad (17)$$

Exemple 3.1.

$$\begin{array}{cccccccc} \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ & \dots & 0 & 1 & 2 & 3 & 0 & 1 & \dots \\ \dots & 2 & 0 & 2 & 4 & 2 & 0 & 2 & \dots \\ & \dots & 1 & 0 & 3 & 2 & 1 & 0 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \end{array}$$

Exemple 3.2.

$$\begin{array}{cccccccc} \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ & \dots & 0 & 3.5 & 1.5 & 1 & 0 & 3.5 & \dots \\ \dots & 0 & 2.5 & 4 & 1.5 & 0 & 2.5 & 4 & \dots \\ & \dots & 1.5 & 2 & 3 & -0.5 & 1.5 & 2 & \dots \\ \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots \end{array}$$

Dans la suite du document, nous utiliserons les notations suivantes pour situer les éléments dans un pavage additif :

$$\begin{array}{cccccccc} \dots & 0 & 0 & 0 & 0 & 0 & \dots & \text{ligne 0} \\ & \dots & a_1 & a_2 & a_3 & a_4 & \dots & \text{ligne 1} \\ & & \dots & a_{12} & a_{23} & a_{34} & \dots & \text{ligne 2} \\ & & & \dots & a_{123} & a_{234} & \dots & \text{ligne 3} \\ & & & & \dots & a_{1234} & \dots & \text{ligne 4} \\ & & & & & \dots & & \vdots \\ & & & & & \dots & & \vdots \end{array}$$

Note : Dans un pavage additif à $n+1$ lignes, la ligne 0 et la ligne n sont composées uniquement de 0. Dans le contexte des pavages additifs, la ligne k sera aussi appelée la k^e ligne.

3.1 Propriétés du pavage additif

Cette section est l'essence du document, nous y présenterons les principaux théorèmes connus à propos des pavages additifs et nous fournirons des preuves à ceux-ci.

Le théorème suivant (tiré de [She73]) est l'analogie de la propriété 6 des pavages multiplicatifs.

Théorème 3.3. *Nous avons*

$$a_{r\dots s} = \left(\sum_{i=r}^s a_i \right) - t_{s-r},$$

où

$$t_n = \frac{n(n+1)}{2},$$

c'est-à-dire que t_n est le n^e nombre triangulaire.

Démonstration. Nous procéderons par récurrence.

Pour $s - r = 1$:

$$a_{rs} = a_r + a_s - 1 = \left(\sum_{i=r}^s a_i \right) - t_1.$$

Pour $s - r = 2$:

$$\begin{aligned} a_{rps} &= a_{rp} + a_{ps} - a_p - 1 \\ &= (a_r + a_p - 1) + (a_p + a_s - 1) - a_p - 1 \\ &= a_r + a_p + a_s - 3 \\ &= \left(\sum_{i=r}^s a_i \right) - t_2. \end{aligned}$$

Pour $s - r \geq 3$: en appliquant l'équation (17) au losange

$$\begin{array}{ccc} & a_{r+1\dots s-1} & \\ a_{r\dots s-1} & & a_{r+1\dots s} \\ & a_{r\dots s} & \end{array}$$

nous obtenons :

$$\begin{aligned} a_{r\dots s} &= a_{r\dots s-1} + a_{r+1\dots s} - a_{r+1\dots s-1} - 1 \\ &= \left(\sum_{i=r}^{s-1} a_i + \sum_{i=r+1}^s a_i - \sum_{i=r+1}^{s-1} a_i \right) - 2t_{s-r-1} + t_{s-r-2} - 1 \\ &= \left(\sum_{i=r}^s a_i \right) - \frac{2(s-r-1)(s-r)}{2} + \frac{(s-r-2)(s-r-1)}{2} - 1 \\ &= \left(\sum_{i=r}^s a_i \right) - \frac{(s^2 - 2sr + s + r^2 - r)}{2} \\ &= \left(\sum_{i=r}^s a_i \right) - \frac{(s-r)(s-r+1)}{2} \\ &= \left(\sum_{i=r}^s a_i \right) - t_{s-r}. \end{aligned}$$

□

Ce théorème entraîne trois corollaires qui nous permettent de caractériser les pavages additifs et qui se révéleront très importants pour démontrer certains théorèmes. Le premier corollaire est une condition suffisante pour que le pavage soit composé uniquement d'entiers.

Corollaire 3.4. *Le pavage est uniquement composé d'entiers si et seulement si les éléments de la première ligne sont tous des entiers.*

Le second corollaire est une condition suffisante pour que la n^e ligne soit composée uniquement de 0.

Corollaire 3.5. *La n^e ligne est une ligne de 0 si et seulement si la somme de n'importe quelle suite de n nombres consécutifs sur la première ligne est égale à t_{n-1} .*

Le dernier corollaire est une condition suffisante pour que le pavage additif soit composé uniquement de nombres positifs.

Corollaire 3.6. *Si pour tous $r, s \in \mathbb{Z}$ avec $s \leq n$, nous avons*

$$\sum_{i=r}^{r+s} a_i \geq t_s$$

alors le pavage additif est composé uniquement de nombres positifs.

Le théorème précédent nous a montré une relation permettant de calculer un élément quelconque du pavage à partir des éléments sur la première ligne. Maintenant, il serait intéressant de voir s'il existe une relation permettant de calculer les éléments de la première ligne à partir d'éléments quelconques. En fait, si un couple d'éléments est connu, il existe une telle relation. Celle-ci est présentée dans le théorème suivant qui se veut l'analogue de la propriété 7 des pavages multiplicatifs.

Théorème 3.7. *Pour tous r, s , avec $s > r$, nous avons :*

$$(a) \quad a_s = a_{r\dots s} - a_{r\dots s-1} + (s - r),$$

$$(b) \quad a_r = a_{r\dots s} - a_{r+1\dots s} + (s - r).$$

Démonstration.

(a) En vertu du théorème 3.3, nous avons :

$$a_{r\dots s} - a_{r\dots s-1} + (s - r) = \left(\sum_{i=r}^s a_i \right) - \left(\sum_{i=r}^{s-1} a_i \right) - t_{s-r} + t_{s-r-1} + (s - r) = a_s$$

(b) La preuve est similaire et laissée au lecteur.

□

Le théorème suivant, tiré de [She73], est l'analogue de la propriété 2 des pavages multiplicatifs.

Théorème 3.8. *Tout pavage additif à $n + 1$ lignes admet comme isométrie une translation de n positions vers la droite (ou la gauche).*

Démonstration. En vertu du théorème 3.3, nous savons que :

$$\left(\sum_{i=k}^{k+n-1} a_i \right) - t_{n-1} = 0 = \left(\sum_{i=k+1}^{k+n} a_i \right) - t_{n-1}.$$

En simplifiant, on obtient

$$a_k = a_{k+n}.$$

Ensuite,

$$\begin{aligned} a_{r\dots s} &= \left(\sum_{i=r}^s a_i \right) - t_{s-r} \\ &= \left(\sum_{i=r+n}^{s+n} a_i \right) - t_{s-r} \\ &= a_{r+n\dots s+n}. \end{aligned}$$

□

Le théorème suivant, également tiré de [She73], est l'analogie de la propriété 3 des pavages multiplicatifs.

Théorème 3.9. *Si un pavage additif à $n + 1$ lignes admet une tranche de 0, alors le pavage est composé uniquement d'entiers positifs.*

Démonstration. En utilisant l'équation $a_s = a_{r\dots s} - a_{r\dots s-1} + (s - r)$ (ou $a_r = a_{r\dots s} - a_{r+1\dots s} + (s - r)$) du théorème 3.7 à tous les couples de zéro qui composent la tranche, nous pouvons calculer n nombres consécutifs sur la première ligne. De cette manière, nous découvrons que la première ligne est en fait une permutation des nombres de 1 à n . Comme toutes les permutations de 1 à n aux conditions des trois corollaires du théorème 3.3, le pavage est composé uniquement d'entiers positifs. □

La propriété 8 des pavages multiplicatifs nous permet de générer un pavage à partir d'un autre. Il se trouve que nous avons découvert une technique similaire pour le faire avec les pavages additifs. En voici un exemple :

Exemple 3.10. Considérons le pavage original suivant :

$$\begin{array}{cccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}$$

En insérant un 5 dans la première rangée nous obtenons :

$$\begin{array}{cccccccc}
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 & 2 & 2 & 5 & 2 & 2 & 2 & 2 \\
 3 & 3 & 6 & 6 & 3 & 3 & 3 & 3 \\
 & 3 & 6 & 6 & 6 & 3 & 3 & 3 \\
 2 & 5 & 5 & 5 & 5 & 2 & 2 & 5 \\
 & 3 & 3 & 3 & 3 & 3 & 0 & 3 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{array}$$

Toutefois, il est évident qu'il est impossible d'obtenir le pavage original de cette façon, puisqu'il contient 5 lignes et qu'il n'y a aucun 4 sur la première ligne.

Voici le théorème :

Théorème 3.11. *Il est possible d'obtenir un pavage à $n+1$ lignes à partir d'un pavage à n lignes en insérant le nombre $n-1$ n'importe où dans la première ligne du pavage. De plus, si le pavage à n lignes est formé d'entiers positifs, alors il en est de même pour le pavage à $n+1$ lignes.*

Démonstration. Pour démontrer la première partie de l'énoncé, il suffit de vérifier que les hypothèses du corollaire 2 sont satisfaites pour le nouvel arrangement à $n+1$ lignes. Supposons un pavage additif à n lignes. Nous savons que :

$$\left(\sum_{i=k}^{k+n-2} a_i \right) = t_{n-2}$$

pour tout k . Alors :

$$\left(\sum_{i=k}^{k+n-2} a_i \right) + n - 1 = t_{n-2} + n - 1 = t_{n-1}.$$

de sorte que les hypothèses du corollaire 2 sont vérifiées pour le nouvel arrangement. On obtient donc un pavage à $n+1$ lignes.

Pour ce qui est de la seconde partie, il suffit de montrer que les hypothèses des corollaires 1 et 3 sont satisfaites pour le nouveau pavage à $n+1$ lignes. Pour montrer que les hypothèses du corollaire 3 sont satisfaites, il suffit de le montrer pour les sous-sommes qui incluent le nombre $n-1$ nouvellement ajouté à la première ligne. En vertu du théorème 3.3, nous savons que :

$$\left(\sum_{i=r}^{r+s} a_i \right) \geq t_s,$$

d'où

$$\left(\sum_{i=r}^{r+s} a_i \right) + (n-1) \geq t_s + (n-1) \geq t_s + (s+1) = t_{s+1}$$

car s est au maximum égal à $n-2$. Les hypothèses du corollaire 3 sont donc vérifiées pour le pavage à $n+1$ lignes. Finalement, puisque $n-1$ est un entier, l'hypothèse du corollaire 1 est vérifiée. \square

Note : Cette méthode fonctionne sur tous les pavages, mais elle ne permet pas de tous les obtenir.

Le théorème qui suit provient de [She73].

Théorème 3.12. *Supposons un parallélogramme ayant pour sommets :*

$$(a_{j\dots k}, a_{j\dots s}, a_{r\dots s}, a_{r\dots k})$$

où $r < j \leq k < s$. Alors :

$$a_{j\dots k} + a_{r\dots s} = a_{j\dots s} + a_{r\dots k} - (j - r)(s - k),$$

où $(j - r)(s - k)$ représente l'aire du parallélogramme si nous définissons l'aire des losanges comme étant égale à 1.

Démonstration.

$$\begin{aligned} & a_{r\dots k} + a_{j\dots s} - (j - r)(s - k) \\ &= \left(\left(\sum_{i=r}^k a_i \right) - t_{k-r} \right) + \left(\left(\sum_{i=j}^s a_i \right) - t_{s-j} \right) - (j - r)(s - k) \\ &= \left(\sum_{i=r}^k a_i \right) + \left(\sum_{i=j}^s a_i \right) - \frac{(k-r)(k-r+1)}{2} - \frac{(s-j)(s-j+1)}{2} \\ &\quad - (j - r)(s - k) \\ &= \left(\sum_{i=r}^s a_i \right) + \left(\sum_{i=j}^k a_i \right) - \frac{(k^2 - 2kr + k - r + r^2)}{2} - \frac{(s^2 - 2sj + s - j + j^2)}{2} \\ &\quad - \frac{2(js - jk - rs + rk)}{2} \\ &= \left(\sum_{i=r}^s a_i \right) + \left(\sum_{i=j}^k a_i \right) + \frac{(-k^2 - k + r - r^2 - s^2 - s + j - j^2 + 2jk + 2rs)}{2} \\ &= \left(\sum_{i=r}^s a_i \right) + \left(\sum_{i=j}^k a_i \right) - \frac{(k^2 - 2jk + k - j + j^2)}{2} - \frac{(s^2 - 2sr + s - r + r^2)}{2} \\ &= \left(\sum_{i=r}^s a_i \right) + \left(\sum_{i=j}^k a_i \right) - \frac{(k-j)(k-j+1)}{2} - \frac{(s-r)(s-r+1)}{2} \\ &= \left(\sum_{i=r}^s a_i \right) + \left(\sum_{i=j}^k a_i \right) - t_{k-j} - t_{s-r} \\ &= a_{j\dots k} + a_{r\dots s} \end{aligned}$$

□

4 Pavages constants

Un *pavage constant* est un pavage (additif ou multiplicatif) pour lequel sur chaque ligne tous les éléments sont égaux.

Exemple 4.1 (Exemple additif).

$$\begin{array}{cccccc}
 & 0 & 0 & 0 & 0 & 0 \\
 & & 2 & 2 & 2 & 2 \\
 \dots & 3 & 3 & 3 & 3 & 3 & \dots \\
 & & 3 & 3 & 3 & 3 \\
 & 2 & 2 & 2 & 2 & 2 \\
 & & 0 & 0 & 0 & 0
 \end{array}$$

Exemple 4.2 (Exemple multiplicatif).

$$\begin{array}{cccccc}
 & 1 & 1 & 1 & 1 & 1 \\
 & & x & x & x & x \\
 \dots & y & y & y & y & y & \dots \\
 & & x & x & x & x \\
 & 1 & 1 & 1 & 1 & 1
 \end{array}$$

où $x = \sqrt{3}$ et $y = 2$.

Une interprétation géométrique peut être associée à chacun des deux types de pavages.

4.1 Le cas multiplicatif :

Pour le cas multiplicatif, il a été démontré dans [CC73a, CC73b] que la valeur des éléments sur une ligne d'un pavage multiplicatif constant à n lignes représente la longueur des diagonales d'un polygone régulier à $n + 1$ côtés de longueur 1. L'exemple précédent illustre les deux longueurs possibles pour les diagonales d'un hexagone régulier avec des côtés de longueur 1 (voir Figure 8).

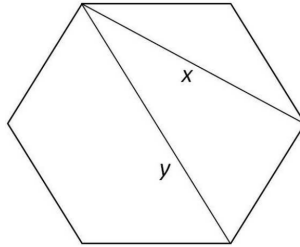


FIGURE 8 – Hexagone régulier avec des côtés de longueur 1.

4.2 Le cas additif :

Il est spécifié dans [She73] que tout pavage additif constant est directement relié à une parabole. Par exemple, le pavage

$$\begin{array}{cccccc}
 & 0 & 0 & 0 & 0 & 0 \\
 & & 2 & 2 & 2 & 2 \\
 \dots & 3 & 3 & 3 & 3 & 3 & \dots \\
 & & 3 & 3 & 3 & 3 \\
 & 2 & 2 & 2 & 2 & 2 \\
 & & 0 & 0 & 0 & 0
 \end{array}$$

et que nous adjoignons cette ligne à une ligne de 1 nous obtenons :

| | | | | | | | | | | | | | | | | |
|-----|-----|----|----|---|----|---|----|----|----|---|----|---|----|-----|-----|-----|
| ... | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | ... | | |
| | ... | 4 | | 1 | | 3 | | 3 | | 3 | | 1 | | 4 | ... | |
| ... | 15 | | 3 | | 2 | | 8 | | 8 | | 2 | | 3 | | 15 | ... |
| | ... | 11 | | 5 | | 5 | | 21 | | 5 | | 5 | | 11 | ... | |
| ... | 8 | | 18 | | 12 | | 13 | | 13 | | 12 | | 18 | | 8 | ... |
| | | | 13 | | 43 | | 31 | | 8 | | 31 | | 43 | | 13 | |
| | | | ⋮ | | ⋮ | | ⋮ | | ⋮ | | ⋮ | | ⋮ | | ⋮ | |

Ce genre de pavages pourrait faire l'objet de recherches plus approfondies pour de futurs projets.

Références

- [CC73a] J. CONWAY et H. COXETER : Triangulated polygons and frieze patterns. *Mathematical Gazette*, 57(400):87–94, 1973.
- [CC73b] J. CONWAY et H. COXETER. : Triangulated polygons and frieze patterns. *Mathematical Gazette*, 57(400):175–193, 1973.
- [MG11] S. MORIER-GENOUD : Arithmetic of 2-friezes. arXiv :1109.0917v1, 2011.
- [ML10] J.-S. F. MARTINEAU et D. LAVERTU. : Frise et triangulation de polygones. *CaMUS*, pages 39–59, 2010.
- [She73] G. SHEPHARD. : Additive frieze patterns and multiplication table. *Mathematical Gazette*, 60(413):178–184, 1973.
- [SMGT11] V. O. S. MORIER-GENOUD et S. TABACHNIKOV. : 2-frieze patterns and the cluster structure of the space of polygons. arXiv :1008.3359v5, 2011.

JEAN-FRANÇOIS MARCEAU
 DÉPARTEMENT DE MATHÉMATIQUES, BISHOP'S UNIVERSITY
Courriel: jmarceau10@ubishops.ca

GROUPES DE FRISE DES RÉPÉTITIONS DES CARQUOIS DE TYPE DYNKIN

DAVID BOULET-ST-JACQUES, MYRIAM CHABOT ET GUILLAUME
DOUVILLE

RÉSUMÉ La répétition d'un carquois Q consiste en un carquois infini, fait de copies de Q indexées par les entiers et reliées entre elles par des flèches supplémentaires. Dans cet article, on classe les répétitions des carquois de type Dynkin selon leur groupe de symétrie.

1 Introduction

Les **frises** sont des bandes sur lesquelles se répète de manière périodique un motif. Plusieurs mathématiciens, par exemple Bravais, ont étudié les frises en tant qu'objets mathématiques [RSA08]. En effet, leur étude a permis, grâce à la théorie des groupes des isométries du plan, de classer les frises en sept classes, chacune identifiée par l'action d'un groupe. Un tel groupe s'appelle **groupe de frise**. Il est notamment possible de créer des frises à partir d'un carquois Q fini et acyclique, et ce, en construisant la répétition $\mathbb{Z}Q$ de ce carquois. On obtient une bande infinie et périodique que l'on peut classer dans l'un des sept groupes de frise existants. En particulier, dans la théorie des algèbres amassées introduites au début du 21^e siècle par Fomin-Zelevinsky ([FZ02], [FZ03]), on s'intéresse particulièrement aux carquois de type Dynkin, en lien avec les algèbres amassées de type fini. Christine Riedtmann a d'ailleurs présenté une classification des automorphismes des répétitions de carquois de type Dynkin [Rie80]. Cet article propose une classification des répétitions associées aux carquois de type Dynkin parmi les sept groupes de frise. Pour ce faire, nous introduirons d'abord des notions élémentaires liées aux répétitions de carquois de type Dynkin, pour ensuite présenter les sept groupes de frise. Finalement, nous procéderons à la classification des répétitions de carquois de type Dynkin.

2 Éléments de base

On rappelle qu'un **carquois** est un quadruplet $Q = (Q_0, Q_1, s, b)$ où Q_0 , est l'ensemble des sommets, Q_1 est l'ensemble des flèches, et s et b sont des fonctions

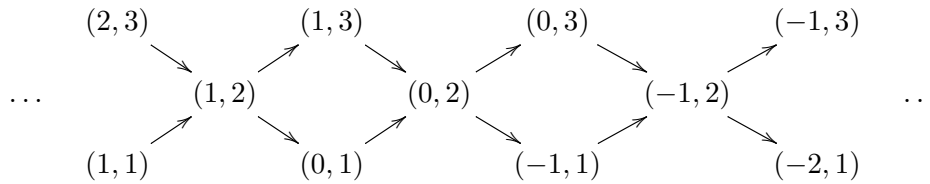
Les auteurs souhaitent remercier le LaCIM, l'ISM et le CRSNG pour leur financement respectif. De plus, les auteurs souhaitent remercier Ibrahim Assem pour la supervision de cet article.

À partir d'un carquois Q , nous pouvons construire sa répétition $\mathbb{Z}Q$.

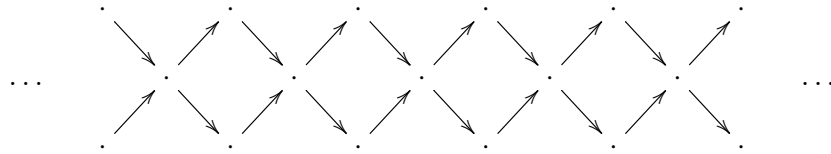
Définition 2.2. Soit Q un carquois fini et acyclique. La **répétition** de Q est le carquois infini $\mathbb{Z}Q$ tel que

$$\begin{aligned}
 (\mathbb{Z}Q)_0 &= \mathbb{Z} \times Q_0 = \{(n, i) \mid n \in \mathbb{Z}, i \in Q_0\} \\
 (\mathbb{Z}Q)_1 &= \{(n, \alpha) : (n, i) \rightarrow (n, j) \mid n \in \mathbb{Z}, \alpha : i \rightarrow j\} \cup \\
 &\quad \{(n, \alpha') : (n, j) \rightarrow (n-1, i) \mid n \in \mathbb{Z}\}
 \end{aligned}$$

Exemple 2.3. Considérons le carquois \vec{A}_3 . Alors la répétition $\mathbb{Z}\vec{A}_3$ est :



Ainsi, nous obtenons le carquois suivant :



3 Classification des frises

3.1 Définitions initiales

Le but de cette section est de présenter le théorème de classification des groupes de frise, ce qui mènera à une classification des répétitions $\mathbb{Z}Q$ parmi ces groupes. Pour arriver à ce théorème, on propose de définir la notion de frise, puis de trouver l'ensemble des isométries laissant une frise invariante, appelé groupe des isométries de la frise, en passant par le groupe des isométries du plan.

Définition 3.1. Soit $E \subseteq \mathbb{R}^2$, une **isométrie** est une fonction $f : E \rightarrow E$ qui préserve les distances, c'est-à-dire que pour tous points $(x_1, y_1), (x_2, y_2)$ appartenant à E , on a :

$$d((x_1, y_1), (x_2, y_2)) = d(f(x_1, y_1), f(x_2, y_2)).$$

Exemple 3.2. Soit $L \in \mathbb{R} \setminus \{0\}$. La translation horizontale t_L est une isométrie définie comme suit

$$\begin{aligned}
 t_L : \mathbb{R} \times [-1, 1] &\longrightarrow \mathbb{R} \times [-1, 1] \\
 (x, y) &\longmapsto (x + L, y)
 \end{aligned}$$

Définition 3.3. Une **frise** \mathcal{F} est un ensemble de points inscrits dans un sous-ensemble de hauteur fixe et de largeur infinie $\mathbb{R} \times [-1, 1]$ du plan \mathbb{R}^2 tel qu'il existe une longueur minimale non nulle L avec la propriété que l'isométrie t_L applique \mathcal{F} sur elle-même.

On rappelle que la **distance euclidienne** entre deux points (x_1, y_1) et (x_2, y_2) du plan \mathbb{R}^2 est

$$d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

Définition 3.4. Une **isométrie de frise** ou **symétrie de frise** S est une isométrie qui envoie une frise sur elle-même.

$$\begin{aligned} S : \quad \mathcal{F} &\longrightarrow \mathcal{F} \\ (x, y) &\longmapsto S(x, y) \end{aligned}$$

3.2 Isométries du plan et limitations pour la frise

Nous en venons à étudier un groupe d'isométries plus général, celui des isométries du plan. Nous restreindrons ce groupe à celui des isométries de frise en utilisant la hauteur finie d'une frise.

Théorème de classification des isométries du plan. *Toute isométrie du plan \mathbb{R}^2 est l'une des suivantes : Identité, translation, rotation, réflexion ou symétrie glissée. De plus, l'ensemble des isométries du plan forme un groupe pour la composition usuelle.* \square

On réfère le lecteur à [And95] pour une preuve de cette classification.

L'ensemble des isométries de frise possibles est le groupe des isométries du plan \mathbb{R}^2 auquel on retire toute isométrie qui n'envoie pas l'ensemble $\mathbb{R} \times [-1, 1]$ sur lui-même, car si l'on n'a pas cette condition, il est impossible d'envoyer la frise sur elle-même. Voyons maintenant les symétries de frise possibles.

Lemme 3.5. *Soit \mathcal{F} une frise de période L . Alors, les uniques générateurs du groupe des isométries de \mathcal{F} sont :*

1. La translation t_L de longueur L le long de l'axe $y = 0$, dans le sens positif;
2. La réflexion horizontale r_h par rapport à la droite $y = 0$;
3. La réflexion verticale $r_v|_{x=\frac{kL}{2}}$ d'axe $x = \frac{kL}{2}$, $k \in \mathbb{Z}$;
4. La rotation $r_\pi = r_h r_v$ d'angle π autour des intersections entre l'axe $y = 0$ et les droites $x = \frac{kL}{2}$, $k \in \mathbb{Z}$;
5. La symétrie glissée $s_{L/2} = t_{L/2} r_h$.

Démonstration. Commençons par les translations. Toute translation ayant une composante verticale est éliminée d'emblée, car l'image d'une telle translation ne sera pas dans $\mathbb{R} \times [-1, 1]$. De plus, la translation t_a est une isométrie de frise si et seulement si $a = kL$, puisque sinon il existe une composition t_{kL} avec $k \in \mathbb{Z}$ telle que $t_{kL}t_a = t_{a'}$ avec $0 < a' < L$, ce qui contredit la minimalité de L .

Puis, toute réflexion qui n'est pas d'axe $y = 0$ ou parallèle à l'axe $x = 0$ ne pourra avoir $\mathbb{R} \times [-1, 1]$ pour image, et les réflexions verticales sont limitées aux axes $x = \frac{kL}{2}$, $k \in \mathbb{Z}$, car la composition de $r_v|_{x=a}$ et de $r_v|_{x=b}$ donne $t_{2(b-a)}$ qui doit être égale à t_{kL} pour ne pas contredire la minimalité de L . Cela donne :

$$2(b - a) = kL$$

ce qui est le cas si et seulement si

$$b - a = \frac{kL}{2}$$

On ne conserve donc que ces deux types de réflexions, r_h et r_v , respectivement.

Ensuite, seule la rotation d'angle π peut avoir $\mathbb{R} \times [-1, 1]$ pour image.

Finalement, la symétrie glissée aura pour image $\mathbb{R} \times [-1, 1]$ si et seulement si sa composante de réflexion est r_h . De plus, comme la période L est minimale, la composante de translation doit non seulement être parallèle à $y = 0$, mais doit être aussi de longueur $\frac{L}{2}$, car sinon on obtient une contradiction à la minimalité de L en utilisant le fait que $s_a^{-1} = r_h t_{\frac{-a}{2}}$. \square

Graphiquement, chaque isométrie peut être représentée comme l'indique la figure 10. Nous avons appliqué chaque isométrie au motif initial I . Ainsi, le triangle a est le résultat de la translation t_L ; le triangle b , celui de la rotation r_h ; le triangle c , celui de la réflexion verticale r_v ; le triangle d , celui de la rotation r_π ; le triangle e , celui de la symétrie glissée s_g . On remarque que les lettres servent uniquement à identifier les générateurs : elles ne font pas partie de la frise.

Pour alléger la présentation, la symétrie glissée $s_{L/2}$ sera désormais notée s_g et r_v représentera l'ensemble des $r_v|_{x=\frac{kL}{2}}$.

3.3 Théorème de classification des groupes de frise

Notation 3.6. Soient a_1, \dots, a_n , où $n \in \mathbb{N}$, des éléments d'un groupe G . Alors $\langle a_1, \dots, a_n \rangle$ note le sous-groupe de G engendré par ces éléments.

Lemme 3.7. *Soit \mathcal{F} une frise. Alors, on a que $\langle r_h, r_v \rangle = \langle r_h, r_\pi \rangle = \langle r_v, r_\pi \rangle$.*

Démonstration. Cela suit du fait que $r_h r_v = r_\pi = r_v r_h$. \square

Nous arrivons au théorème principal de cette section. Nous avons vu que les isométries du plan forment un groupe pour la composition usuelle. Nous verrons maintenant quels sous-groupes d'isométries peuvent laisser la frise invariante. On réfère par exemple à [RSA08] pour une preuve alternative de ce théorème connu.

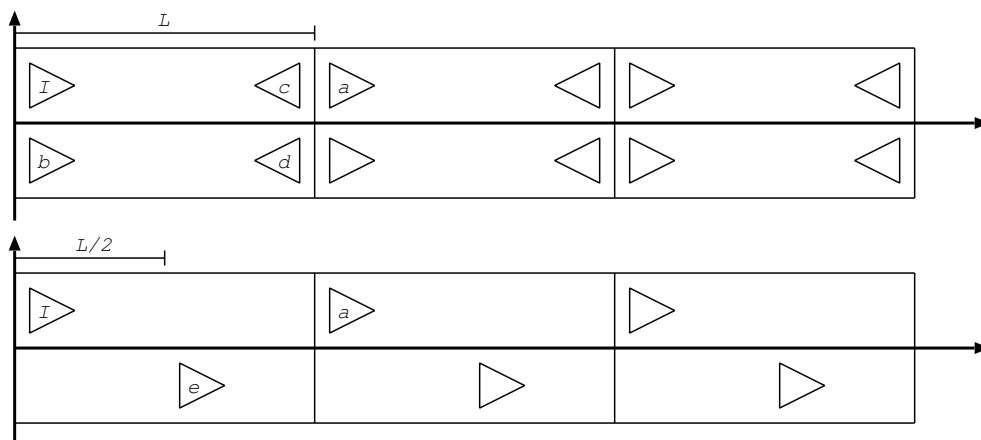


FIGURE 10 – Générateurs du groupe des isométries de frise

Théorème de classification des groupes de frise. Toute frise \mathcal{F} a pour groupe de frise l'un des sept suivants :

1. $\langle t_L \rangle$
2. $\langle t_L, r_v \rangle$
3. $\langle t_L, r_h \rangle$
4. $\langle t_L, r_\pi \rangle$
5. $\langle s_g \rangle$
6. $\langle s_g, r_\pi \rangle$
7. $\langle t_L, r_h, r_v \rangle$

Démonstration. En vertu du théorème 3.5, nous savons que les seuls générateurs possibles sont t_L, r_v, r_h, r_π et s_g . Puisque, en vertu de la définition, un groupe de frise contient t_L , il reste seulement $2^4 = 16$ ensembles de générateurs possibles. Comme on a que $s_g^2 = t_L$, si s_g est présent, on retire t_L pour obtenir la minimalité de l'ensemble des générateurs. De plus, en vertu du lemme 3.7, on retire 6 ensembles de générateurs qui ne sont pas uniques. Il nous reste :

1. $\langle t_L \rangle$
2. $\langle t_L, r_v \rangle$
3. $\langle t_L, r_h \rangle$
4. $\langle t_L, r_\pi \rangle$
5. $\langle s_g \rangle$

- 6. $\langle s_g, r_\pi \rangle$
- 7. $\langle t_L, r_h, r_v \rangle$
- 8. $\langle s_g, r_h \rangle$
- 9. $\langle s_g, r_v \rangle$
- 10. $\langle s_g, r_h, r_v \rangle$

Or, si r_h et s_g sont dans le même ensemble de générateurs, leur composition y est aussi, ce qui donne :

$$\begin{aligned} s_g \circ r_h &= (t_{L/2} r_h) \circ r_h \\ &= t_{L/2} \circ (r_h^2) \\ &= t_{L/2} \end{aligned}$$

ce qui est une contradiction à la minimalité de L . On élimine donc les cas 8 et 10 de la liste. Finalement, on a vu au théorème 3.3 qu'on peut appliquer r_v pour tout axe $x = \frac{nL}{2}$, ce qui implique que r_v et $t_{L/2} r_v$ engendrent la même frise (à un changement d'origine près), et donc les cas 6 et 9 engendrent le même groupe :

$$\begin{aligned} \langle s_g, r_\pi \rangle &= \langle s_g, s_g r_\pi \rangle \\ &= \langle s_g, t_{L/2} r_h \circ r_h r_v \rangle \\ &= \langle s_g, r_v \rangle \end{aligned}$$

Il subsiste donc 7 groupes de frise potentiels. Les figures 11 à 17 [RSA08] montrent que ces groupes existent. En effet, chacune des figures suivantes présente des frises laissées inchangées par les isométries appartenant à l'un de ces sept groupes de frise présentés précédemment.

□

4 Théorème de classification des répétitions de carquois Dynkin

Telles que définies en 2.2, les flèches de $\mathbb{Z}Q$, pour Q fini et acyclique, n'ont pas de longueur. Or, la classification selon les groupes de frise nécessite une notion

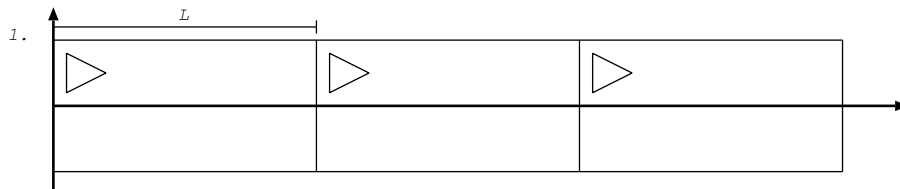
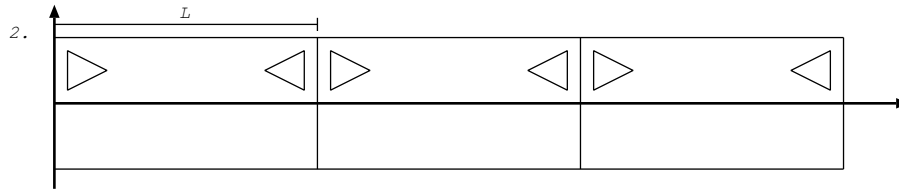
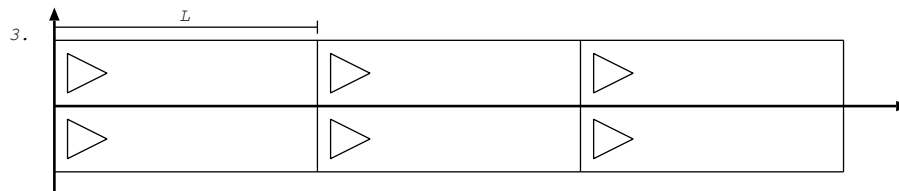
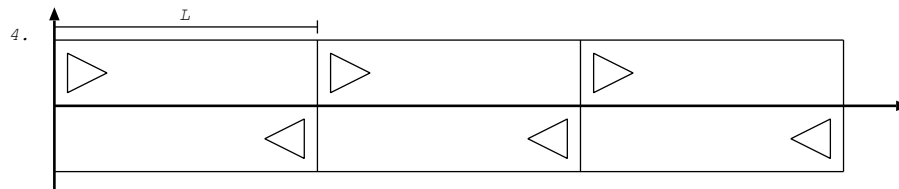
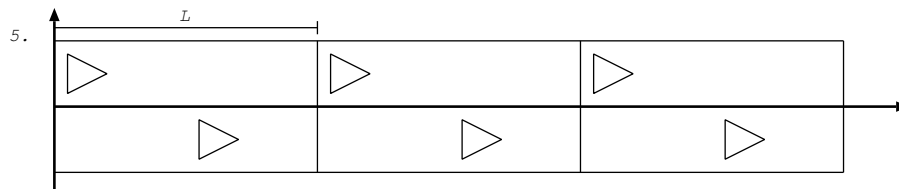
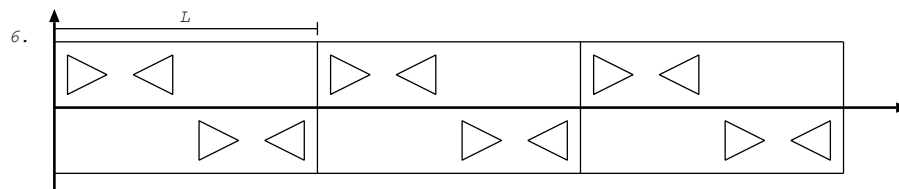


FIGURE 11 – Groupe de frise $\langle t_L \rangle$

FIGURE 12 – Groupe de frise $\langle t_L, r_v \rangle$ FIGURE 13 – Groupe de frise $\langle t_L, r_h \rangle$ FIGURE 14 – Groupe de frise $\langle t_L, r_\pi \rangle$ FIGURE 15 – Groupe de frise $\langle s_g \rangle$ FIGURE 16 – Groupe de frise $\langle s_g, r_\pi \rangle$

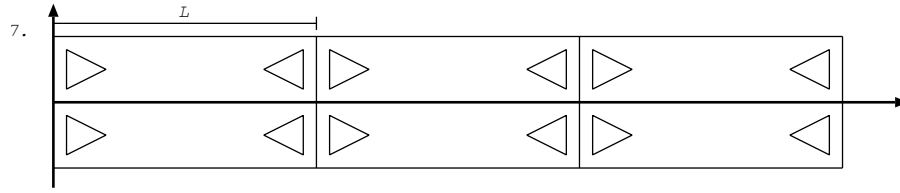


FIGURE 17 – Groupe de frise $\langle t_L, r_h, r_v \rangle$

de distance. Afin de pallier à ce problème, on définit la notion de distance de façon à maximiser le nombre de générateurs pour tout groupe de frise. Soit ainsi $\lambda, \mu > 0$ deux constantes réelles.

1. La répétition doit être inscrite dans l'ensemble $\mathbb{R} \times [-1, 1]$.
2. i) Pour $\vec{\mathbb{A}}_n$, toutes les flèches appartenant à $(\mathbb{Z}\vec{\mathbb{A}}_n)_1$ sont de longueur λ .
 ii) Pour $\vec{\mathbb{D}}_m$, les flèches appartenant à l'ensemble

$$A = \{(n, \alpha) : (n, m - 2) \rightarrow (n, m) \mid n \in \mathbb{Z}, \alpha : m - 2 \rightarrow m\} \cup \{(n, \alpha') : (n, m) \rightarrow (n - 1, m - 2) \mid n \in \mathbb{Z}\}$$

sont de longueur μ et parallèles à l'axe des abscisses ($y = 0$). Les flèches appartenant à $(\mathbb{Z}\vec{\mathbb{D}}_m)_1 \setminus A$ sont de longueur λ .

- iii) Pour $\vec{\mathbb{E}}_6, \vec{\mathbb{E}}_7$ et $\vec{\mathbb{E}}_8$, les flèches appartenant à

$$B = \{(n, \alpha) : (n, 3) \rightarrow (n, 4) \mid n \in \mathbb{Z}, \alpha : 3 \rightarrow 4\} \cup \{(n, \alpha') : (n, 4) \rightarrow (n - 1, 3) \mid n \in \mathbb{Z}\}$$

sont de longueur μ et parallèles à l'axe des abscisses ($y = 0$). Les flèches appartenant à $(\mathbb{Z}\vec{\mathbb{E}}_p)_1 \setminus A$, $p \in \{6, 7, 8\}$ sont de longueur λ .

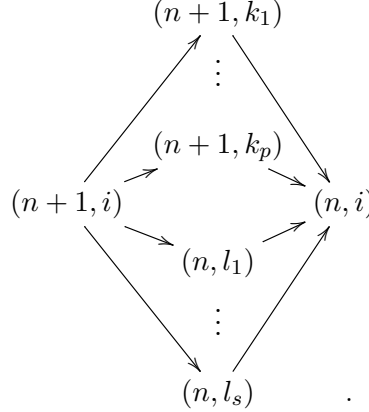
Avant d'énoncer la classification des carquois de type Dynkin selon les groupes de frise, nous présentons une propriété importante sur l'orientation initiale d'un carquois Q pour la répétition $\mathbb{Z}Q$.

Proposition 4.1. [ASS06] *Soient Q, Q' deux carquois finis et acycliques ayant le même arbre pour graphe sous-jacent. Alors $\mathbb{Z}Q$ est isomorphe à $\mathbb{Z}Q'$. \square*

Cette proposition permet de ne considérer que le carquois principalement orienté dans la classification des frises. Le prochain lemme permet de réduire les types d'isométries possibles dans les répétitions $\mathbb{Z}Q$ pour en venir, par la suite, à la classification des répétitions $\mathbb{Z}Q$, où Q est un carquois de type Dynkin, selon leur groupe de frise.

Lemme 4.2. *Soit Q un carquois fini et acyclique. Alors le groupe de frise de la répétition $\mathbb{Z}Q$ ne peut pas contenir l'un des générateurs r_v ou r_π .*

Démonstration. Soit Q un carquois fini et acyclique. Une maille de $\mathbb{Z}Q$ est représentée de cette manière :



Soit $\alpha \in (\mathbb{Z}Q)_1$, une flèche allant du sommet $(n+1, i)$ au sommet $(n+1, k_q)$, où $q \in \{1, 2, \dots, p\}$. Soient le sommet (n, l_r) , où $r \in \{1, 2, \dots, s\}$, l'image de la rotation appliquée au sommet $(n+1, k_q)$, et $\delta \in (\mathbb{Z}Q)_1$, une flèche allant du sommet $(n+1, i)$ au sommet (n, l_r) . Dans le cas où le sommet (n, l_r) n'existe pas, il est trivial que la rotation r_π n'est pas un générateur de la répétition $\mathbb{Z}Q$.

Pour qu'il y ait une réflexion verticale r_v passant par le sommet $(n+1, k_q)$, il doit y avoir une flèche β allant du sommet (n, i) au sommet $(n+1, k_q)$. De plus, en vertu de la définition d'une répétition $\mathbb{Z}Q$, il y a une flèche β' allant du sommet $(n+1, k_q)$ au sommet (n, i) , une contradiction, car α et β' forment un 2-cycle et Q est acyclique.

Pour qu'il y ait une rotation r_π ayant pour centre de rotation le centre de la maille, il doit y avoir une flèche λ allant du sommet (n, i) au sommet (n, l_r) . En vertu de la définition d'une répétition, puisqu'il y a une flèche γ allant du sommet $(n+1, i)$ au sommet (n, l_r) , il y a une flèche γ' allant du sommet (n, l_r) au sommet (n, i) , une contradiction, car λ et γ' forment un 2-cycle dans un carquois Q acyclique. \square

Remarque 4.3. Toute répétition est nécessairement laissée invariante par la translation qui envoie la n^e copie du carquois vers la $n-1^e$ copie. Ainsi, pour une répétition $\mathbb{Z}Q$, s'il n'existe aucune autre isométrie laissant la frise invariante, elle aura le groupe $\langle t_L \rangle$ pour groupe de frise.

Théorème 4.4. *Soit $n \in \mathbb{Z}_{>0}$ et $m \in \mathbb{Z}$, $m > 4$. Alors, on classe les répétitions de carquois Dynkin ainsi : $\mathbb{Z}\mathbb{D}_m$, $\mathbb{Z}\mathbb{E}_7$ et $\mathbb{Z}\mathbb{E}_8$ ont pour groupe de frise $\langle t_L \rangle$, $\mathbb{Z}\mathbb{A}_{2n+1}$ et $\mathbb{Z}\mathbb{D}_4$ ont pour groupe de frise $\langle t_L, r_h \rangle$, et finalement $\mathbb{Z}\mathbb{A}_{2n}$ et $\mathbb{Z}\mathbb{E}_6$ admettent $\langle s_g \rangle$ pour groupe de frise.*

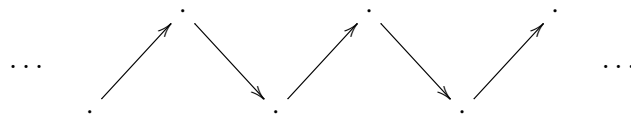
Démonstration. En vertu du lemme 4.2, il est impossible d'avoir une isométrie de type r_v ou r_π . Ainsi, toutes les répétitions ne peuvent être classifiées que dans les groupes $\langle t_L \rangle$, $\langle s_g \rangle$ ou $\langle t_L, r_H \rangle$. Considérons chacun des types de répétitions :

CAS 1 : La répétition $\mathbb{Z}\mathbb{A}_n$

Soit \mathcal{F} une frise formée par la répétition $\mathbb{Z}\mathbb{A}_n$ dans $\mathbb{R} \times [-1, 1]$. On subdivise alors la frise sur la hauteur de telle sorte que l'on obtienne $n - 1$ bandes de même hauteur $h = \frac{2}{n-1}$ et on subdivise la frise sur la longueur de telle sorte que l'on obtient une infinité de colonnes de longueur h . On obtient alors une infinité de cases carrées dans la frise. Soit $a_{1,0}$ la case contenant la flèche allant du sommet 1 au sommet 2 de la copie 0 et soit $a_{i,j}$ la case à l'intersection de la i^{e} ligne et de la j^{e} colonne. Alors, on peut traduire en termes de cases les isométries précédemment présentées : Pour tous i, j avec $i \in \{1, 2, \dots, n - 1\}$ et $j \in \mathbb{Z}$, la frise sera laissée invariante par s_g si et seulement si $a_{i,j} = a_{n-i,j}$ et elle sera invariante sous r_h si et seulement si $a_{i,j} = a_{n-i,j+1}$.

La répétition $\mathbb{Z}\mathbb{A}_{2n}$

Pour $n = 1$, la répétition $\mathbb{Z}\mathbb{A}_2$ est



Il est alors trivial de voir que $\mathbb{Z}\mathbb{A}_2$ est dans le groupe de frise $\langle s_g \rangle$.

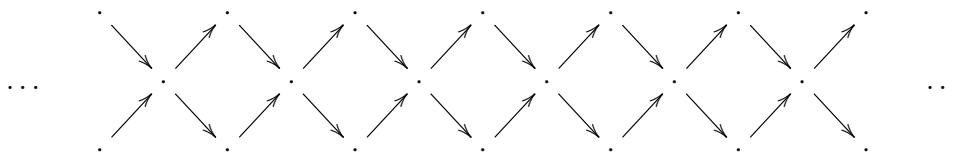
Supposons que la classification soit correcte pour le rang $n - 1$, c'est-à-dire que $\mathbb{Z}\mathbb{A}_{2n-2}$ soit dans le groupe de frise $\langle s_g \rangle$. Considérons la répétition $\mathbb{Z}\mathbb{A}_{2n}$. En vertu de l'hypothèse de récurrence, on a que

$$a_{i,j} = a_{2n-i,j},$$

pour tous i, j tels que $i \in \{2, \dots, 2n - 2\}$ et $j \in \mathbb{Z}$, puisqu'il s'agit de la répétition $\mathbb{Z}\mathbb{A}_{2n-2}$. Ainsi, si $a_{1,j} = a_{2n-1,j}$ pour tout $j \in \mathbb{Z}$, on aura que la répétition $\mathbb{Z}\mathbb{A}_{2n}$ se classe dans le groupe de frise $\langle s_g \rangle$. Or, puisque le carquois dont est issu la répétition est d'orientation principale, on a que $a_{1,j} = a_{2n-1,j+2n-2}$ pour tous $j \in \mathbb{Z}$. En vertu de la périodicité, $a_{2n-1,j+2n-2} = a_{2n-1,j+2n-2k}$. En particulier, en prenant $k = n$ on obtient que $a_{1,j} = a_{2n-1,j}$, ce qui termine la récurrence.

La répétition $\mathbb{Z}\mathbb{A}_{2n+1}$

Pour $n = 1$, la répétition $\mathbb{Z}\mathbb{A}_3$ est



Cette frise est laissée invariante par r_h , la répétition a donc pour groupe de frise $\langle t_L, r_h \rangle$.

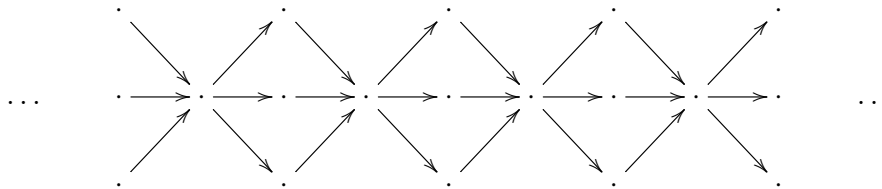
Supposons la classification correcte pour le rang $n-1$, c'est-à-dire que $\mathbb{Z}\mathbb{A}_{2n-1}$ soit dans le groupe de frise $\langle t_L, r_h \rangle$. Considérons la répétition $\mathbb{Z}\mathbb{A}_{2n+1}$. En vertu de l'hypothèse de récurrence, on a que

$$a_{i,j} = a_{2n-i+1,j+1}$$

pour tous i, j tels que $i \in \{2, \dots, 2n-1\}$, car il s'agit de la répétition $\mathbb{Z}\mathbb{A}_{2n-1}$. Ainsi, si $a_{1,j} = a_{2n,j+1}$ pour tout $j \in \mathbb{Z}$, on aura que la symétrie glissée laisse la frise invariante, donc la répétition $\mathbb{Z}\mathbb{A}_{2n+1}$ se classe dans le groupe de frise $\langle t_L, r_h \rangle$, car la translation est toujours présente en vertu de la périodicité. Or, puisque le carquois dont est issu la répétition est d'orientation principale, on a que $a_{1,j} = a_{2n,j+2n-1}$ pour tous $j \in \mathbb{Z}$. En vertu de la périodicité, $a_{2n,j+2n-1} = a_{2n,j+2n-1-2k}$. En particulier, en prenant $k = n-1$ on obtient que $a_{1,j} = a_{2n,j+1}$, ce qui termine la récurrence.

CAS 2 : La répétition $\mathbb{Z}\mathbb{D}_4$

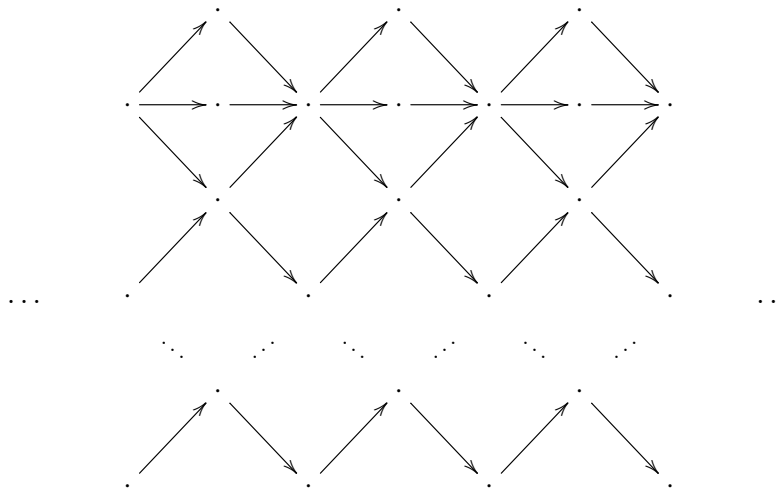
La répétition $\mathbb{Z}\mathbb{D}_4$ est :



Les seules isométries précédemment définies qui laissent la frise invariante sont r_h et t_L . Ainsi, la répétition $\mathbb{Z}\mathbb{D}_4$ a pour groupe de frise $\langle t_L, r_h \rangle$.

CAS 3 : La répétition $\mathbb{Z}\mathbb{D}_m$

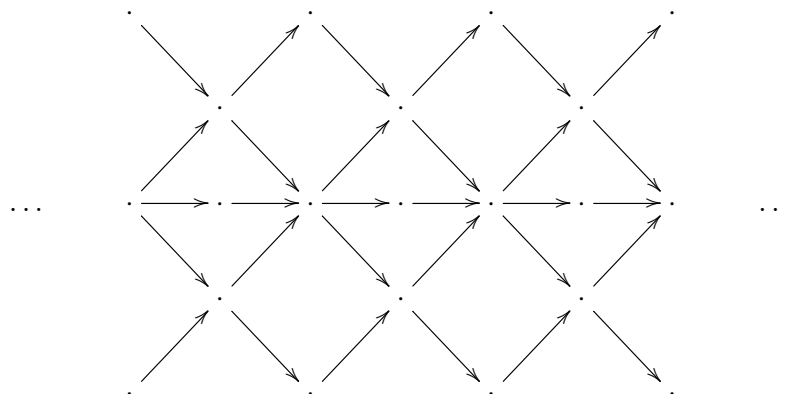
La répétition $\mathbb{Z}\mathbb{D}_m$, où $m > 4$, est :



Comme la ligne de flèches horizontales n'est présente que dans le haut de la répétition, ni la réflexion horizontale ni la symétrie glissée ne laissent celle-ci inchangée. La répétition a donc pour groupe de frise le groupe $\langle t_L \rangle$.

CAS 4 : La répétition $\mathbb{Z}\mathbb{E}_6$

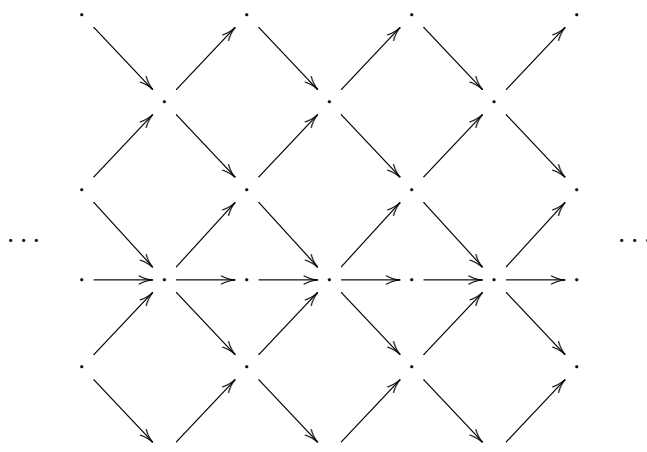
En vertu de la proposition 4.1, la répétition $\mathbb{Z}\mathbb{E}_6$ est :



Les seules isométries sont r_h et t_L . Ainsi, $\mathbb{Z}\mathbb{E}_6$ a pour groupe de frise $\langle t_L, r_h \rangle$.

CAS 5 : La répétition $\mathbb{Z}\mathbb{E}_7$

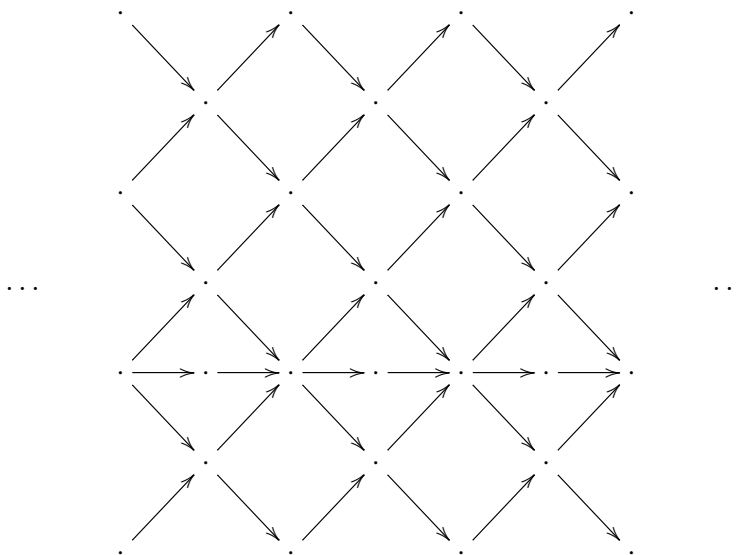
En vertu de la proposition 4.1, la répétition $\mathbb{Z}\mathbb{E}_7$ est :



La seule isométrie est t_L . Ainsi, $\mathbb{Z}\mathbb{E}_7$ a pour groupe de frise $\langle t_L \rangle$.

CAS 6 : La répétition $\mathbb{Z}\mathbb{E}_8$

En vertu de la proposition 4.1, la répétition $\mathbb{Z}\mathbb{E}_8$ est :



La seule isométrie applicable est t_L . Ainsi, $\mathbb{Z}\mathbb{E}_8$ a pour groupe de frise $\langle t_L \rangle$.

□

Références

- [And95] D. ANDREW : Isometries of the plane. *The American Mathematical Monthly*, 102(7):628–631, 1995.
- [ASS06] I. ASSEM, A. SKOWRONSKI et D. SIMSON : *Elements of representation theory of associative algebras, vol I : Techniques of representation theory*, volume 65 de *London mathematical society Student Texts*. Cambridge University Press, Cambridge, UK, 2006.
- [FZ02] S. FOMIN et A. ZELEVINSKY : Cluster algebras I : Foundations. *J. Amer. Math. Soc.*, (15(2)):497–529, 2002.
- [FZ03] S. FOMIN et A. ZELEVINSKY : Cluster algebras II : Finite type classification. *Invent. Math.*, (154):63–121, 2003.
- [Rie80] C. RIEDTMANN : Algebren, darstellungsköcher, ueberlagerungen and zurück. *Comment. Math. Helvetici*, (55):199–224, 1980.
- [RSA08] C. ROUSSEAU et Y. SAINT-AUBIN : Mathématiques et technologies. *Springer Undergraduate Texts in Mathematics and Technology*, pages 45–84, 2008.

DAVID BOULET-ST-JACQUES

DÉPARTEMENT DE MATHÉMATIQUES, UNIVERSITÉ DU QUÉBEC À MONTRÉAL

Courriel: boulet-st-jacques.david@courrier.uqam.ca

MYRIAM CHABOT

DÉPARTEMENT DE MATHÉMATIQUES, UNIVERSITÉ DE SHERBROOKE

Courriel: myriam.chabot@usherbrooke.ca

GUILLAUME DOUVILLE

DÉPARTEMENT DE MATHÉMATIQUES, UNIVERSITÉ DE SHERBROOKE

Courriel: guillaume.douville@usherbrooke.ca

RÉSEAUX DE NEURONES

ADAM SALVAIL-BÉRARD

RÉSUMÉ Le but de cet article est de donner une introduction simple, mais formelle, du procédé de construction des réseaux de neurones artificiels. En plus de la définition d'un réseau de neurones, sont données les versions scalaire, vectorielle et par lot des formules utilisées dans l'apprentissage des paramètres du réseau de neurones. Les algorithmes de propagation des données et de rétropropagation du gradient de l'erreur d'apprentissage sont abordés. Finalement, dans le contexte d'optimisation des réseaux de neurones, une méthode exacte de calcul du produit de la matrice hessienne et d'un vecteur est montrée.

1 Introduction

Les réseaux de neurones sont des constructions abstraites simulant l'activité d'un réseau de neurones biologique simplifié. Ils sont utilisés en apprentissage automatique pour construire un modèle à partir de données existantes dans le but d'effectuer des prédictions sur de nouvelles données soit à l'aide de régression, dans le cas continu, ou de classification, dans le cas discret. Dans [SB12] se trouve une introduction plus complète à l'apprentissage automatique et à son cousin, le forage de données.

L'apprentissage automatique¹ concerne l'ajustement de paramètres d'une « boîte » qui prend en entrée des données et fournit en sortie un résultat. C'est la métaphore classique décrivant une fonction. En statistiques, la boîte est souvent simple et peut utiliser l'analyse mathématique pour caractériser les paramètres optimaux. Par exemple, la régression classique utilise une « boîte » constituée d'une combinaison linéaire de fonctions élémentaires pertinentes et l'ajustement consiste simplement à optimiser les coefficients de la combinaison linéaire. Le contexte moderne du forage de données a motivé le développement de « boîtes » plus complexes, impossibles à optimiser analytiquement. Les réseaux de neurones sont un exemple de tels modèles complexes qui ont fait leurs preuves dans plusieurs domaines comme la détection de fraudes ou la reconnaissance de caractères manuscrits.

Ce texte s'intéresse donc aux réseaux de neurones, à ouvrir la boîte métaphorique qui est associée à cette classe de modèles. Un réseau de neurones

J'aimerais remercier Hugo Larochelle et Jean-Pierre Dussault pour leurs judicieux conseils et leurs relectures de ce documents. Merci également à Guillaume Douville et Vincent Morissette-Thomas pour leur relecture et leurs conseils.

¹De l'anglais *machine learning*.

est un graphe orienté de nœuds, appelés neurones, souvent disposés en couches. Ces nœuds reçoivent une information de leurs prédécesseurs (les neurones de la couche précédente) et combinent cette information selon une certaine pondération. Chaque nœud possède également un seuil d'activation. Autour de ce seuil, la valeur de sortie est beaucoup plus sensible aux changements de la valeur d'entrée.

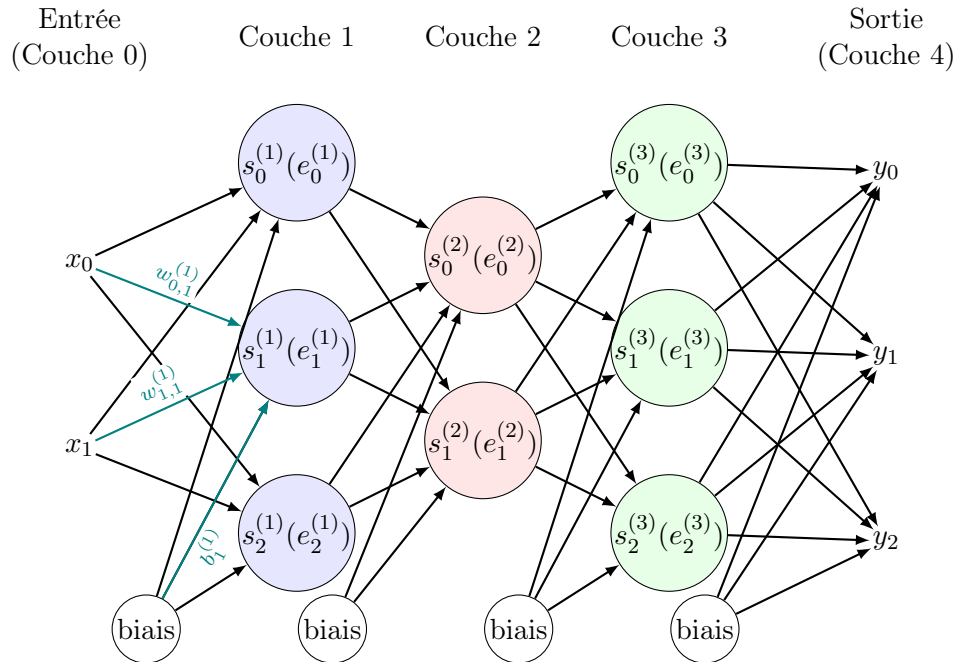


FIGURE 18 – Illustration d'un réseau de neurones *feed-forward* à 5 couches dont 3 couches cachées, une couche d'entrée et une couche de sortie. Ce type de réseau est caractérisé par le fait qu'un neurone ne tire ses entrées que de la couche précédente. À titre d'exemple, les entrées du deuxième nœud de la seconde couche ont été étiquetées en bleu.

Le comportement du réseau de neurones dépend donc de la pondération des apports des nœuds d'une couche à la suivante et des seuils d'activation de chaque nœud. Si ces paramètres sont fixés arbitrairement, les résultats fournis seront médiocres. Il convient d'utiliser un jeu de données nommé ensemble d'entraînement pour optimiser les paramètres afin de minimiser l'erreur de prédiction du réseau. Pour cet ensemble d'entraînement, la prédiction attendue est connue et l'optimisation consiste à minimiser la somme des carrés des différences entre les sorties calculées et les sorties attendues. Contrairement aux modèles classiques en statistique, il est impossible d'obtenir une solution analytique à ces problèmes d'optimisation. Il faut donc optimiser numériquement la fonction d'erreur.

Cet article présente une description rigoureuse de ces réseaux de neurones.

De plus, une méthode d'évaluation des dérivées premières et secondes (dans une direction donnée) d'une fonction d'erreur servant à identifier les paramètres optimaux est proposée. Cette information sur les dérivées est très utile pour les algorithmes d'optimisation qui s'en servent pour trouver rapidement une solution (localement) optimale au problème de minimisation de l'erreur.

2 Définitions

Un réseau de neurones agit sur des données. Ces données sont définies formellement comme étant un ensemble de couples (\mathbf{x}, \mathbf{c}) avec $\mathbf{x} \in \mathbb{R}^n$ et $\mathbf{c} \in \mathbb{R}^m$, $n, m \in \mathbb{N}^*$. En outre, un réseau de neurones peut être défini comme étant une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ qui prend en entrée les informations du vecteur \mathbf{x} et qui effectue une prédiction $\mathbf{y} \in \mathbb{R}^m$ du vecteur \mathbf{c} . Le but de la construction d'un réseau de neurones est de trouver la fonction f qui donne la meilleure approximation \mathbf{y} de la cible \mathbf{c} .

Comme son nom l'indique, un réseau de neurones est une collection de nœuds, nommés neurones, disposés en $r + 1$ couches. La couche 0 représente l'entrée et possède n neurones, soit la dimension du vecteur d'entrée \mathbf{x} . De même, la r -ième couche, la couche de sortie, possède m neurones, soit la dimension du vecteur de sortie \mathbf{y} . Les couches intermédiaires, nommées couches cachées, possèdent un nombre de neurones arbitraires, souvent déterminé par la complexité du problème à résoudre.

Afin de propager les valeurs d'entrée, chaque neurone de la couche i , $0 \leq i \leq r$, possède une connexion avec tous les neurones de la couche $i + 1$ ². Chacune de ces connexions est pondérée par un poids $w_{j,k}^{(i)}$ défini comme étant le poids liant la sortie $s_j^{(i-1)}$ du j -ième neurone de la couche $i - 1$ à l'entrée $e_k^{(i)}$ du k -ième neurone de la couche i . Les poids liant la couche $i - 1$ et la couche i forment une matrice notée $\mathbf{w}^{(i)}$. La Figure 18 illustre ces notations à l'aide d'une représentation graphique d'un réseau de neurones.

Les neurones d'un réseau de neurones représentent les unités de calcul du modèle. Chaque neurone possède une entrée (un scalaire), une fonction d'activation et une sortie. Par exemple, le k -ième neurone de la couche i reçoit une valeur d'entrée

$$e_k^{(i)} = \sum_j w_{j,k}^{(i)} s_j^{(i-1)} + b_k^{(i)} \quad (18)$$

et retourne la valeur $s_k^{(i)} = a_k^{(i)}(e_k^{(i)})$, soit la fonction d'activation appliquée à l'entrée $e_k^{(i)}$ du neurone.

La valeur $b_k^{(i)}$ est un biais ajouté à l'entrée du k -ième neurone de la couche i servant à déterminer la sensibilité d'un neurone à son entrée.

²Ceci est la représentation classique d'un réseau de neurones. D'autres topologies, c'est-à-dire dispositions des neurones, sont possibles. Dans cet article, seul ce type de réseau de neurones est discuté.

Finalement, il y a la fonction d'erreur

$$E = \frac{1}{2} \|\mathbf{y} - \mathbf{c}\|_2^2 = \frac{1}{2} \sum_{j=1}^m (y_j - c_j)^2, \quad (19)$$

qui dénote l'erreur du réseau de neurones dans son approximation de la sortie \mathbf{y} par rapport à la cible \mathbf{c} . C'est cette fonction qui détermine la qualité du modèle.

3 Réseau de neurones *feed-forward*

Un réseau de neurones typique est le réseau *feed-forward*. Ce réseau propage l'entrée du réseau vers les couches suivantes sans jamais revenir en arrière. Ce type de réseau est utilisé pour le reste de cet article.

En plus du type de réseau de neurones, il faut également choisir une fonction d'erreur et une fonction d'activation pour les neurones. Ces choix sont souvent guidés par le type de données traitées. Plus de détails sont disponibles dans l'ouvrage de Bishop [Bis06] au chapitre 5.

Dans cet article, la fonction d'activation utilisée est une fonction logistique définie par $a(x) = \frac{1}{1+\exp(-x)}$, sauf pour la couche de sortie qui garde la fonction identité comme fonction d'activation. De cette façon, la valeur de sortie

$$s_j^{(i)} = a_j^{(i)}(e_j^{(i)}) = \begin{cases} \frac{1}{1+\exp(-e_j^{(i)})} & \text{pour } i = 0, \dots, r-1 \\ e_j^{(i)} & \text{pour } i = r. \end{cases} \quad (20)$$

Pour la fonction d'erreur, un choix simple est d'utiliser la formule (19), soit la moitié du carré de la distance euclidienne entre la sortie du réseau et la cible.

Avec ces définitions, tous les outils nécessaires à l'établissement d'une prédiction sont présents. Pour effectuer la propagation avec un couple (\mathbf{x}, \mathbf{c}) , les sorties de la couche d'entrée sont assignées aux valeurs du vecteur \mathbf{x} , c'est-à-dire que $s_j^{(0)} = x_j$. Les valeurs sont ensuite propagées dans le réseau de neurones grâce aux formules (18) et (20) pour donner les sorties $y_j = s_j^{(r)}$ qui sont comparées aux cibles c_j par la fonction d'erreur E en (19). Il faut maintenant minimiser la moyenne des erreurs données par la fonction E sur l'ensemble des données fournies en entrée :

$$E_{\text{moyenne}} = \frac{1}{N} \sum_{t=1}^N E_t$$

où N est le nombre de couples donnés en entraînement au réseau de neurones et E_t représente la t -ième erreur d'apprentissage. Pour y arriver, il suffit d'ajuster les paramètres $w_{j,k}^{(i)}$ et $b_j^{(i)}$.

4 Construction du réseau de neurones

Bien entendu, au départ, un réseau de neurones avec des poids initialisés aléatoirement donne de pauvres résultats. C'est pourquoi il y a une étape d'appren-

tissage, c'est-à-dire d'ajustement des poids du réseau afin que ceux-ci puissent donner de meilleures prédictions.

Une façon de voir le problème est d'essayer de minimiser l'erreur de prédiction du modèle. Cette formulation transforme le problème d'apprentissage en un problème d'optimisation pour lequel plusieurs algorithmes existent. L'optimisation de la fonction d'erreur s'effectue par l'ajustement des paramètres de la fonction, c'est-à-dire les poids et les biais du réseau de neurones.

Soit la fonction $\dim(i)$ qui donne le nombre de neurones de la couche i . Soit

$$\boldsymbol{\theta} = \begin{bmatrix} w_{j,k}^{(i)} & b_j^{(i)} \end{bmatrix} \in \mathbb{R}^p \quad (21)$$

où $0 < i \leq r$, $0 \leq j < \dim(i-1)$ et $0 \leq k < \dim(i)$ qui représentent le vecteur rassemblant chacun des paramètres du réseau de neurones. Le problème d'optimisation associé est de trouver les paramètres du réseau de neurones permettant de minimiser l'erreur $E(\boldsymbol{\theta})$. Formellement, il faut trouver

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{t=1}^N E_t(\boldsymbol{\theta}). \quad (22)$$

Néanmoins, l'application d'algorithmes d'optimisation nécessite souvent d'avoir plus d'information sur la fonction d'erreur $E(\boldsymbol{\theta})$, comme ses dérivées de premier et second ordres par rapport aux paramètres $\boldsymbol{\theta}$ de la fonction d'erreur.

Trouver les dérivées de la fonction d'erreur est un problème complexe. Puisque la fonction d'erreur, avec toutes les couches du réseau, peut facilement devenir difficile à appréhender, il devient impossible de déduire une forme analytique de la solution au problème de minimisation de l'erreur moyenne. Il faut donc utiliser des méthodes itératives. La méthode la plus élégante pour l'obtention de la dérivée de premier ordre est la rétropropagation de l'erreur de prédiction. Pour la dérivée de second ordre, une méthode proposée par Pearlmutter [Pea94] est décrite. Cette méthode donne le résultat exact du produit de la matrice hessienne et d'un vecteur arbitraire.

4.1 Rétropropagation

L'idée de la rétropropagation est de faire circuler l'information sur la dérivée de la fonction d'erreur à partir de la couche de sortie, où l'erreur de prédiction est connue (la différence entre \mathbf{y} et \mathbf{c}), jusqu'à la couche d'entrée. Pour y arriver, il faut arriver à exprimer la dérivée de la fonction d'erreur d'un nœud en fonction de l'information donnée par les couches suivantes. De cette façon, les valeurs attendues de la couche de sortie pourront être rétropropagées vers l'entrée.

Tout d'abord, la dérivée, à la sortie, de la fonction d'erreur (19) est donnée par

$$\frac{\partial E}{\partial y_i} = y_i - c_i. \quad (23)$$

À partir de cette équation, il est possible de trouver la dérivée de la fonction d'erreur à l'entrée de la couche de sortie. La dérivée reste la même puisque la

fonction d'activation de la dernière couche est la fonction identité, c'est-à-dire

$$\frac{\partial E}{\partial e_j^{(r)}} = \frac{\partial E}{\partial s_j^{(r)}} = \frac{\partial E}{\partial y_j}.$$

À partir de ce moment, il faudra se servir de la formule de dérivée en chaîne pour trouver, de façon récursive, les autres dérivées. En d'autres mots, il faut pouvoir exprimer chaque dérivée de la fonction d'erreur en terme de dérivées venant plus loin dans le réseau de neurones.

Pour ce faire, il faut voir que la fonction d'erreur est affectée par la sortie d'un neurone de la i -ième couche à travers l'ensemble des neurones de la $(i + 1)$ -ième couche. Donc,

$$\frac{\partial E}{\partial s_j^{(i)}} = \sum_k w_{j,k}^{(i+1)} \frac{\partial E}{\partial e_k^{(i+1)}}. \quad (24)$$

De plus, la fonction d'erreur est affectée par l'entrée d'un neurone de la i -ième couche seulement au travers de la sortie de ce même neurone. Il en résulte que

$$\frac{\partial E}{\partial e_j^{(i)}} = \frac{\partial E}{\partial s_j^{(i)}} \frac{\partial s_j^{(i)}}{\partial e_j^{(i)}}.$$

Sachant que la dérivée de la fonction d'activation logistique est donnée par $\frac{\partial a}{\partial x} = a(1 - a)$, la dérivée de la fonction d'erreur par l'entrée d'un neurone est donc

$$\frac{\partial E}{\partial e_j^{(i)}} = \frac{\partial E}{\partial s_j^{(i)}} s_j^{(i)} (1 - s_j^{(i)}). \quad (25)$$

La formule (25) permet de rétropropager la dérivée de l'erreur jusqu'à la première couche. À l'aide de ces valeurs, il est possible d'arriver à la dérivée de la fonction d'erreur par rapport à chacun des poids du réseau de neurones, en appliquant la règle de dérivée en chaîne. Donc,

$$\frac{\partial E}{\partial w_{j,k}^{(i)}} = \frac{\partial E}{\partial e_k^{(i)}} \frac{\partial e_k^{(i)}}{\partial w_{j,k}^{(i)}} = \frac{\partial E}{\partial e_k^{(i)}} s_j^{(i-1)}. \quad (26)$$

Finalement, comme les biais sont également des paramètres impliqués dans l'apprentissage d'un réseau de neurones, il faut aussi pouvoir leur trouver une valeur optimale. Il est donc nécessaire de trouver la dérivée de l'erreur par rapport aux biais. De la même façon qu'avec les poids, un biais n'affecte l'erreur de prédiction qu'au travers de l'entrée du nœud auquel il est rattaché. Donc, grâce à la dérivée en chaîne et la formule (18),

$$\frac{\partial E}{\partial b_j^{(i)}} = \frac{\partial E}{\partial e_j^{(i)}} \frac{\partial e_j^{(i)}}{\partial b_j^{(i)}} = \frac{\partial E}{\partial e_j^{(i)}}. \quad (27)$$

4.2 Calcul de la dérivée seconde

Pour pouvoir minimiser la fonction d'erreur de manière efficace, plusieurs algorithmes d'optimisation nécessitent des informations sur la dérivée seconde qui représente le comportement de la fonction autour d'un point. Dans le cas de la fonction d'erreur, une fonction avec un domaine dans \mathbb{R}^p , sa dérivée seconde est en fait la matrice hessienne

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 E}{\partial \theta_1^2} & \frac{\partial^2 E}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 E}{\partial \theta_1 \partial \theta_p} \\ \frac{\partial^2 E}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 E}{\partial \theta_2^2} & \cdots & \frac{\partial^2 E}{\partial \theta_2 \partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial \theta_p \partial \theta_1} & \frac{\partial^2 E}{\partial \theta_p \partial \theta_2} & \cdots & \frac{\partial^2 E}{\partial \theta_p^2} \end{bmatrix}.$$

Malheureusement, cette matrice, pour de grandes dimensions, devient difficile à stocker et à calculer, même pour un ordinateur. Puisque la plupart des algorithmes utilisent plutôt le produit de la matrice hessienne avec un vecteur arbitraire \mathbf{v} (résultant en un vecteur de dimension p), il est plus facile de calculer directement ce vecteur plutôt que la matrice hessienne entière.

En se basant sur le principe de la propagation de valeurs exposé plus haut, il est possible de trouver la valeur de $\mathbf{H}\mathbf{v}$ en apportant quelques changements aux formules déjà données. Cette technique, sous la forme présentée dans cet article, est attribuée à Pearlmutter [Pea94].

Le produit hessienne-vecteur peut être défini par

$$\mathbf{H}\mathbf{v} = \left. \frac{\partial}{\partial q} \nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta} + q\mathbf{v}) \right|_{q=0}.$$

De cette définition, il est possible d'extraire un nouvel opérateur

$$\mathcal{R}\{f(\boldsymbol{\theta})\} = \left. \frac{\partial}{\partial q} f(\boldsymbol{\theta} + q\mathbf{v}) \right|_{q=0}$$

où \mathbf{v} représente le vecteur par lequel la matrice hessienne est multipliée et qui est de même dimension que le vecteur $\boldsymbol{\theta}$. Pour fins de concordance, les éléments $v_{j,k}^{(i)}$ et $v_j^{(i)}$ du vecteur \mathbf{v} sont respectivement à la même position que les éléments $w_{j,k}^{(i)}$ et $b_j^{(i)}$ du vecteur $\boldsymbol{\theta}$.

Par définition, $\mathcal{R}\{w_{j,k}^{(i)}\} = v_{j,k}^{(i)}$ et $\mathcal{R}\{b_k^{(i)}\} = v_k^{(i)}$ puisque les $w_{j,k}^{(i)}$ et les $b_k^{(i)}$ sont les paramètres à optimiser. En appliquant les règles usuelles de dérivée et de dérivée en chaîne, il est possible d'obtenir des formules de propagation pour les calculs de $\mathcal{R}\{\cdot\}$. Pour l'entrée dans un neurone, la formule devient, avec la

règle de la dérivée d'un produit,

$$\begin{aligned}
\mathcal{R} \left\{ e_k^{(i)} \right\} &= \mathcal{R} \left\{ \sum_j w_{j,k}^{(i)} s_j^{(i-1)} + b_k^{(i)} \right\} \\
&= \sum_j w_{j,k}^{(i)} \mathcal{R} \left\{ s_j^{(i-1)} \right\} + \mathcal{R} \left\{ w_{j,k}^{(i)} \right\} s_j^{(i-1)} + \mathcal{R} \left\{ b_k^{(i)} \right\} \\
&= \sum_j w_{j,k}^{(i)} \mathcal{R} \left\{ s_j^{(i-1)} \right\} + v_{j,k}^{(i)} s_j^{(i-1)} + v_k^{(i)}.
\end{aligned} \tag{28}$$

Pour la sortie d'un neurone, la formule devient

$$\mathcal{R} \left\{ s_j^{(i)} \right\} = s_j^{(i)} (1 - s_j^{(i)}) \mathcal{R} \left\{ e_j^{(i)} \right\}. \tag{29}$$

En initialisant $\mathcal{R} \left\{ s_j^{(0)} \right\}$ au vecteur nul et à partir des formules (28) et (29), il est possible de faire la propagation des valeurs données en entrée pour le calcul de la dérivée seconde.

À l'image de la rétropropagation, il faut également rétropropager la dérivée de la fonction d'erreur sur laquelle est appliqué l'opérateur $\mathcal{R} \{ \cdot \}$. Il faut donc appliquer l'opérateur $\mathcal{R} \{ \cdot \}$ à la formule de la dérivée de la sortie :

$$\mathcal{R} \left\{ \frac{\partial E}{\partial y_j} \right\} = \mathcal{R} \left\{ y_j - c_j \right\} = \mathcal{R} \left\{ y_j \right\} = \mathcal{R} \left\{ s_j^{(r)} \right\}. \tag{30}$$

De façon identique à la rétropropagation,

$$\mathcal{R} \left\{ \frac{\partial E}{\partial y_j} \right\} = \mathcal{R} \left\{ \frac{\partial E}{\partial s_j^{(r)}} \right\} = \mathcal{R} \left\{ \frac{\partial E}{\partial e_j^{(r)}} \right\}.$$

Les formules récursives de la rétropropagation deviennent

$$\begin{aligned}
\mathcal{R} \left\{ \frac{\partial E}{\partial e_j^{(i)}} \right\} &= \mathcal{R} \left\{ \frac{\partial E}{\partial s_j^{(i)}} \frac{\partial s_j^{(i)}}{\partial e_j^{(i)}} \right\} \\
&= \frac{\partial E}{\partial s_j^{(i)}} \mathcal{R} \left\{ \frac{\partial s_j^{(i)}}{\partial e_j^{(i)}} \right\} + \mathcal{R} \left\{ \frac{\partial E}{\partial s_j^{(i)}} \right\} \frac{\partial s_j^{(i)}}{\partial e_j^{(i)}} \\
&= \frac{\partial E}{\partial s_j^{(i)}} \frac{\partial^2 s_j^{(i)}}{\partial e_j^{(i)2}} \mathcal{R} \left\{ e_j^{(i)} \right\} + \mathcal{R} \left\{ \frac{\partial E}{\partial s_j^{(i)}} \right\} \frac{\partial s_j^{(i)}}{\partial e_j^{(i)}}
\end{aligned} \tag{31}$$

$$\text{où } \frac{\partial s_j^{(i)}}{\partial e_j^{(i)}} = s_j^{(i)} (1 - s_j^{(i)}) \text{ et } \frac{\partial^2 s_j^{(i)}}{\partial e_j^{(i)2}} = s_j^{(i)} (1 - s_j^{(i)}) (1 - 2s_j^{(i)}).$$

De même, pour la sortie d'un neurone,

$$\begin{aligned}
\mathcal{R} \left\{ \frac{\partial E}{\partial s_j^{(i)}} \right\} &= \mathcal{R} \left\{ \sum_j w_{j,k}^{(i+1)} \frac{\partial E}{\partial e_k^{(i+1)}} \right\} \\
&= \sum_j \left(w_{j,k}^{(i+1)} \mathcal{R} \left\{ \frac{\partial E}{\partial e_k^{(i+1)}} \right\} + \mathcal{R} \left\{ w_{j,k}^{(i+1)} \right\} \frac{\partial E}{\partial e_k^{(i+1)}} \right) \\
&= \sum_j \left(w_{j,k}^{(i+1)} \mathcal{R} \left\{ \frac{\partial E}{\partial e_k^{(i+1)}} \right\} + v_{j,k}^{(i+1)} \frac{\partial E}{\partial e_k^{(i+1)}} \right). \tag{32}
\end{aligned}$$

Finalement, pour les poids

$$\begin{aligned}
\mathcal{R} \left\{ \frac{\partial E}{\partial w_{j,k}^{(i)}} \right\} &= \mathcal{R} \left\{ s_j^{(i-1)} \frac{\partial E}{\partial e_k^{(i)}} \right\} \\
&= s_j^{(i-1)} \mathcal{R} \left\{ \frac{\partial E}{\partial e_k^{(i)}} \right\} + \mathcal{R} \left\{ s_j^{(i-1)} \right\} \frac{\partial E}{\partial e_k^{(i)}} \tag{33}
\end{aligned}$$

et pour les biais

$$\mathcal{R} \left\{ \frac{\partial E}{\partial b_k^{(i)}} \right\} = \mathcal{R} \left\{ \frac{\partial E}{\partial e_k^{(i)}} \right\}. \tag{34}$$

Après tous ces calculs, il est possible de reconstituer le vecteur du produit hessienne-vecteur

$$\begin{aligned}
\mathbf{H}\mathbf{v} &= \mathcal{R} \{ \nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta}) \} \\
&= \left[\mathcal{R} \left\{ \frac{\partial E}{\partial w_{0,0}^{(0)}} \right\} \dots \mathcal{R} \left\{ \frac{\partial E}{\partial w_{\dim(r-1)-1, m-1}^{(r)}} \right\} \mathcal{R} \left\{ b_0^{(0)} \right\} \dots \mathcal{R} \left\{ b_{m-1}^{(r)} \right\} \right]^\top. \tag{35}
\end{aligned}$$

5 Version vectorielle

Alors que la version scalaire du réseau de neurones est utile à la compréhension, une version vectorielle est plus concise. La plupart des implémentations de cet algorithme d'apprentissage utilisent d'ailleurs la version vectorielle pour utiliser la puissance de calcul des bibliothèques d'algèbre linéaire.

La notation vectorielle se base sur la notation scalaire à partir de laquelle les indices se rattachant à un neurone particulier ont été omis. Par exemple, l'entrée d'un neurone $\mathbf{e}^{(i)} = [e_0^{(i)} \ e_1^{(i)} \ \dots \ e_{\dim(i)-1}^{(i)}]^\top$.

L'utilisation de vecteurs apporte quelques changements dont il faut tenir compte. D'abord, l'ensemble des dérivées par rapport à l'erreur est remplacé par le gradient de cette même fonction. Dans cet article, le gradient d'une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est défini comme étant

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_0} \quad \frac{\partial f}{\partial x_1} \quad \cdots \quad \frac{\partial f}{\partial x_{n-1}} \right]^\top.$$

Une exception notable à cette règle (tiré du livre de Rojas [Roj96]) est la fonction d'activation qui est une fonction de \mathbb{R}^n dans \mathbb{R}^n , donc qui nécessite de calculer la matrice jacobienne de la fonction, c'est-à-dire que

$$\mathbf{J}_{\mathbf{e}^{(i)}}(\mathbf{a}^{(i)}) = \begin{bmatrix} \frac{\partial a_0^{(i)}}{\partial e_0^{(i)}} & \frac{\partial a_0^{(i)}}{\partial e_1^{(i)}} & \cdots & \frac{\partial a_0^{(i)}}{\partial e_{\dim(i)-1}^{(i)}} \\ \frac{\partial a_1^{(i)}}{\partial e_0^{(i)}} & \frac{\partial a_1^{(i)}}{\partial e_1^{(i)}} & \cdots & \frac{\partial a_1^{(i)}}{\partial e_{\dim(i)-1}^{(i)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial a_{\dim(i)-1}^{(i)}}{\partial e_0^{(i)}} & \frac{\partial a_{\dim(i)-1}^{(i)}}{\partial e_1^{(i)}} & \cdots & \frac{\partial a_{\dim(i)-1}^{(i)}}{\partial e_{\dim(i)-1}^{(i)}} \end{bmatrix}.$$

Mais comme la dérivée partielle de la fonction d'activation du j -ième neurone ne dépend que de l'entrée du même neurone, tous les éléments non diagonaux de la matrice $\mathbf{J}_{\mathbf{e}^{(i)}}(\mathbf{a}^{(i)})$ sont nuls. Donc, dans le cas où la fonction d'activation est la fonction logistique,

$$\mathbf{J}_{\mathbf{s}^{(i)}} = \mathbf{J}_{\mathbf{a}_{\text{logistique}}^{(i)}} = \begin{bmatrix} s_1^{(i)}(1 - s_1^{(i)}) & 0 & \cdots & 0 \\ 0 & s_2^{(i)}(1 - s_2^{(i)}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_n^{(i)}(1 - s_n^{(i)}) \end{bmatrix}. \quad (36)$$

Sachant ceci, il est plus efficace d'implémenter cette matrice comme un vecteur colonne qui est multiplié terme-à-terme à un autre. Par exemple, la formule (39) multiplie un gradient et une matrice jacobienne. Pour implémenter cette multiplication de manière efficace, il est possible de multiplier le gradient et le vecteur contenant les éléments diagonaux de la matrice jacobienne terme à terme. Il en résulte un nouveau vecteur de la même dimension que le gradient, ce qui correspond à la dimension du produit du gradient et de la matrice jacobienne. Cette manipulation évite de mettre en mémoire tous les zéros qui ne sont pas situés sur la diagonale de la matrice jacobienne, ce qui accélère également les calculs.

Pour profiter de ces avantages, l'opérateur $\text{diag}(\mathbf{A})$ est utilisé pour représenter le vecteur colonne composé des éléments de la diagonale de la matrice \mathbf{A} . De plus, l'opérateur \odot est utilisé pour signifier un produit terme à terme de deux matrices de mêmes dimensions.

5.1 Propagation

Pour propager les entrées dans le réseau de neurones, il convient de convertir les formules vues précédemment. Dans ce contexte, une entrée du réseau de neurones

est représentée par le vecteur $\mathbf{x} = \mathbf{s}^{(0)}$. Ensuite, pour calculer les entrées des neurones des couches suivantes, il faut transformer la formule (18) en

$$\mathbf{e}^{(i)} = \mathbf{w}^{(i)} \mathbf{s}^{(i-1)} + \mathbf{b}^{(i)}.$$

De même pour la sortie d'un neurone, la formule (20) devient

$$\mathbf{s}^{(i)} = \mathbf{a}^{(i)}(\mathbf{e}^{(i)}) = \left[\frac{1}{1 + \exp(-e_1^{(i)})} \quad \frac{1}{1 + \exp(-e_2^{(i)})} \quad \cdots \quad \frac{1}{1 + \exp(-e_n^{(i)})} \right]^\top.$$

La fonction d'erreur (19) se simplifie pour donner

$$E(\boldsymbol{\theta}) = \frac{1}{2} \|\mathbf{y} - \mathbf{c}\|_2^2 \quad (37)$$

dont le gradient est tout simplement

$$\nabla_{\mathbf{s}^{(r)}} E = \nabla_{\mathbf{y}} E = \mathbf{y} - \mathbf{c}. \quad (38)$$

Ces fonctions permettent de faire l'ensemble des calculs en faisant abstraction des neurones et en ne manipulant que les vecteurs ou matrices se rattachant aux couches.

5.2 Rétropropagation

Pour retrouver la version vectorielle de la rétropropagation, il faut refaire les mêmes opérations que pour la version scalaire à partir de la dernière couche. L'erreur par rapport à la sortie est déjà connue (38) et la fonction d'activation de la dernière couche est une fonction linéaire ($\mathbf{s}^{(r)} = \mathbf{e}^{(r)}$). Puisque la dérivée de cette fonction linéaire donne 1, l'entrée de la dernière couche est égale à sa sortie, c'est-à-dire que $\nabla_{\mathbf{e}^{(r)}} E = \nabla_{\mathbf{s}^{(r)}} E = \mathbf{y} - \mathbf{c}$.

À partir de là, il est possible de retrouver la formulation des gradients des autres couches de façon récursive. Le gradient de la sortie s'obtient, à l'image de la formule (24), par

$$\nabla_{\mathbf{s}^{(i)}} E = (\mathbf{w}^{(i+1)})^\top \nabla_{\mathbf{e}^{(i+1)}} E.$$

Quant à lui, le gradient de l'entrée d'un neurone suit le principe de la formule (25) :

$$\nabla_{\mathbf{e}^{(i)}} E = \mathbf{J}_{\mathbf{e}^{(i)}}(\mathbf{s}^{(i)}) \nabla_{\mathbf{s}^{(i)}} E = \text{diag} \left(\mathbf{J}_{\mathbf{e}^{(i)}}(\mathbf{s}^{(i)}) \right) \odot \nabla_{\mathbf{s}^{(i)}} E \quad (39)$$

où $\mathbf{J}_{\mathbf{e}^{(i)}}(\mathbf{s}^{(i)})$ est définie par la formule (36).

Ne reste qu'à utiliser ces deux vecteurs pour pouvoir calculer la valeur des gradients de l'erreur par rapport aux poids et aux biais, chacun décrit par les formules (26) et (27). Ils s'écrivent

$$\nabla_{\mathbf{w}^{(i)}} E = \nabla_{\mathbf{e}^{(i)}} E \nabla_{\mathbf{w}^{(i)}} \mathbf{e}^{(i)} = \nabla_{\mathbf{e}^{(i)}} E \mathbf{s}^{(i-1)\top}$$

et

$$\nabla_{\mathbf{b}^{(i)}} E = \nabla_{\mathbf{e}^{(i)}} E \nabla_{\mathbf{b}^{(i)}} \mathbf{e}^{(i)} = \nabla_{\mathbf{e}^{(i)}} E.$$

5.3 Calcul de la matrice hessienne dans une direction \mathbf{v}

Au lieu de parler de la dérivée seconde, il est maintenant question de trouver la hessienne de la fonction d'erreur. C'est celle-ci qui permet de donner une idée du comportement de la fonction d'erreur au point d'évaluation.

Par contre, la plupart des algorithmes d'optimisation s'intéressent plutôt à la hessienne dans une direction particulière v et non à toute la matrice hessienne. En d'autres termes, au lieu de calculer la hessienne \mathbf{H} , il est plus efficace de calculer directement le produit hessienne-vecteur $\mathbf{H}\mathbf{v}$ puisque la matrice \mathbf{H} n'a pas à être gardée en mémoire.

À l'image de la version scalaire où le produit hessienne-vecteur de la fonction était composé des dérivées secondes de la fonction d'erreur par rapport aux poids et aux biais, la version vectorielle manipule directement les matrices de poids et de biais.

De la même façon que la version vectorielle de la propagation des données, la version vectorielle du calcul du produit hessienne-vecteur peut s'écrire

$$\begin{aligned}\mathcal{R}\{\mathbf{e}^{(i)}\} &= \mathcal{R}\{\mathbf{w}^{(i)}\mathbf{s}^{(i-1)} + \mathbf{b}^{(i)}\} \\ &= \mathbf{w}^{(i)}\mathcal{R}\{\mathbf{s}^{(i-1)}\} + \mathcal{R}\{\mathbf{w}^{(i)}\}\mathbf{s}^{(i-1)} + \mathcal{R}\{\mathbf{b}^{(i)}\} \\ &= \mathbf{w}^{(i)}\mathcal{R}\{\mathbf{s}^{(i-1)}\} + \mathbf{v}_w^{(i)}\mathbf{s}^{(i-1)} + \mathbf{v}_b^{(i)}\end{aligned}$$

et

$$\mathcal{R}\{\mathbf{s}^{(i)}\} = \mathbf{J}_{\mathbf{e}^{(i)}}(\mathbf{s}^{(i)}) \mathcal{R}\{\mathbf{e}^{(i)}\}.$$

Encore une fois, l'algorithme est initialisé avec $\mathcal{R}\{\mathbf{s}^{(0)}\} = \mathbf{0}_{n \times 1}$.

La méthode reste la même pour la rétropropagation. Par contre, il faut savoir ce que représente la dérivée seconde de la sortie d'un neurone par rapport à son entrée, c'est-à-dire de trouver $\mathcal{R}\{\mathbf{J}_{\mathbf{e}^{(i)}}(\mathbf{s}^{(i)})\}$. Or, comme la fonction d'activation n'a pas une image à une seule dimension, la matrice résultante serait en fait un tenseur (une matrice à plus de deux indices). Pour simplifier les choses, il est possible de se concentrer sur la deuxième forme de la formule (39) qui fait intervenir un vecteur multiplié à un autre vecteur de même dimension de façon terme-à-terme.

Dans cette notation, le j -ième élément du vecteur de la dérivée seconde est donné par

$$\text{diag}\left(\mathcal{R}\left\{\mathbf{J}_{\mathbf{e}^{(i)}}(\mathbf{s}^{(i)})\right\}\right)_j = s_j^{(i)}(1 - s_j^{(i)})(1 - 2s_j^{(i)})\mathcal{R}\left\{e_j^{(i)}\right\}.$$

Avec cette information, les formules (30), (31) et (32) deviennent respectivement

$$\begin{aligned}\mathcal{R}\{\nabla_{\mathbf{y}}E\} &= \mathcal{R}\{\mathbf{y} - \mathbf{c}\} \\ &= \mathcal{R}\{\mathbf{y}\}\end{aligned}\tag{40}$$

$$\begin{aligned}
\mathcal{R}\{\nabla_{\mathbf{e}^{(i)}}\} &= \mathcal{R}\left\{\mathbf{J}_{\mathbf{e}^{(i)}}\left(\mathbf{s}^{(i)}\right) \nabla_{\mathbf{s}^{(i)}} E\right\} \\
&= \mathcal{R}\left\{\mathbf{J}_{\mathbf{e}^{(i)}}\left(\mathbf{s}^{(i)}\right)\right\} \nabla_{\mathbf{s}^{(i)}} E + \mathbf{J}_{\mathbf{e}^{(i)}}\left(\mathbf{s}^{(i)}\right) \mathcal{R}\{\nabla_{\mathbf{s}^{(i)}} E\} \\
&= \text{diag}\left(\mathcal{R}\left\{\mathbf{J}_{\mathbf{e}^{(i)}}\left(\mathbf{s}^{(i)}\right)\right\}\right) \odot \nabla_{\mathbf{s}^{(i)}} E + \mathbf{J}_{\mathbf{e}^{(i)}}\left(\mathbf{s}^{(i)}\right) \mathcal{R}\{\nabla_{\mathbf{s}^{(i)}} E\}
\end{aligned}$$

$$\begin{aligned}
\mathcal{R}\{\nabla_{\mathbf{s}^{(i)}} E\} &= \mathcal{R}\left\{\left(\mathbf{w}^{(i+1)}\right)^\top \nabla_{\mathbf{e}^{(i+1)}} E\right\} \\
&= \left(\mathbf{w}^{(i+1)}\right)^\top \mathcal{R}\{\nabla_{\mathbf{e}^{(i+1)}} E\} + \mathcal{R}\left\{\left(\mathbf{w}^{(i+1)}\right)^\top\right\} \nabla_{\mathbf{e}^{(i+1)}} E \\
&= \left(\mathbf{w}^{(i+1)}\right)^\top \mathcal{R}\{\nabla_{\mathbf{e}^{(i+1)}} E\} + \left(\mathbf{v}_w^{(i+1)}\right)^\top \nabla_{\mathbf{e}^{(i+1)}} E
\end{aligned}$$

Encore une fois, à partir de ces résultats il est possible de reconstituer le vecteur du produit $\mathbf{H}\mathbf{v}$ à partir de la concaténation des éléments des matrices, comme dans l'équation (35) :

$$\begin{aligned}
\mathcal{R}\{\nabla_{\mathbf{w}^{(i)}} E\} &= \mathcal{R}\left\{\nabla_{\mathbf{e}^{(i)}} E \left(\mathbf{s}^{(i-1)}\right)^\top\right\} \\
&= \nabla_{\mathbf{e}^{(i)}} E \mathcal{R}\left\{\left(\mathbf{s}^{(i-1)}\right)^\top\right\} + \mathcal{R}\{\nabla_{\mathbf{e}^{(i)}} E\} \left(\mathbf{s}^{(i-1)}\right)^\top
\end{aligned}$$

et des vecteurs

$$\mathcal{R}\{\nabla_{\mathbf{b}^{(i)}} E\} = \mathcal{R}\{\nabla_{\mathbf{e}^{(i)}} E\}$$

qui sont dérivés des formules (33) et (34).

6 Apprentissage d'un réseau de neurones

Comme il a été mentionné plus tôt, le but de la phase d'apprentissage est de trouver le vecteur de paramètres $\boldsymbol{\theta}$ défini en (21) qui minimise la fonction d'erreur comme illustré à la formule (22).

Cette section porte sur quelques façons d'y arriver, chacune présentant leurs avantages et leurs inconvénients.

6.1 Descente de gradient

Une première méthode consiste à appliquer un algorithme de descente de gradient sur l'ensemble des couples (\mathbf{x}, \mathbf{c}) . Cette façon de procéder consiste à trouver le gradient de l'erreur de chacun des couples et d'en faire la moyenne. Comme le gradient est un vecteur pointant vers la direction de plus forte croissance de la fonction d'erreur, déplacer les paramètres dans la direction opposée au gradient fait diminuer l'erreur.

Dans cette optique, les paramètres seront mis-à-jour selon l'équation $\boldsymbol{\theta} := \boldsymbol{\theta} - \alpha (\nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta})) = \alpha \left(\frac{1}{N} \sum_{t=1}^N \nabla E_t(\boldsymbol{\theta})\right)$ où α est appelé le taux d'apprentissage

et contrôle la vitesse d'apprentissage. Un taux trop élevé pourrait nuire à l'algorithme qui sauterait par-dessus le déplacement idéal.

Selon cette méthode, l'algorithme d'apprentissage ressemble à l'algorithme 1. Dans cet algorithme, le critère d'arrêt peut être en lien avec un seuil d'erreur à atteindre ou un nombre limite d'itérations.

Données : Une base de données \mathbf{B} où \mathbf{B}_t est le t -ième couple (\mathbf{x}, \mathbf{c}) et dont la taille est N .

Entrées : Une fonction $CritereArret()$ qui renvoie vrai quand l'apprentissage est terminé.
Un réseau de neurones muni d'une collection de paramètres θ .
Un hyper-paramètre α donnant la vitesse d'apprentissage de l'algorithme.

tant que $CritereArret(\theta)$ **faire**

pour chaque *couple* (\mathbf{x}, \mathbf{c}) **de** \mathbf{B} **faire**

$\mathbf{y} \leftarrow Propagation(\mathbf{x})$; /* La prédiction du réseau */

$\mathbf{g}_t \leftarrow Retropropagation(\mathbf{y} - \mathbf{c})$; /* Le t -ième gradient */

$\mathbf{g} \leftarrow \frac{1}{N} \sum_{t=1}^N \mathbf{g}_t$; /* Le gradient moyen */

$\theta \leftarrow \theta - \alpha \mathbf{g}$; /* Mise à jour des paramètres */

$E \leftarrow \frac{1}{N} \sum_{t=1}^N \|y_t - c_t\|_2^2$; /* Erreur moyenne de prédiction */

Sorties : Un réseau de neurones ayant des paramètres θ optimisés.

Algorithme 1: Algorithme de la descente de gradient pour l'apprentissage d'un réseau de neurones.

6.2 Descente de gradient par lots

Une autre méthode, l'apprentissage par lots³ tente d'améliorer la vitesse d'apprentissage en utilisant un certain nombre de couples (\mathbf{x}, \mathbf{c}) par lot plutôt que l'ensemble des éléments de la base de données.

Puisque les calculs ne se font pas sur l'ensemble des données, mais seulement sur un échantillon, l'estimation de la moyenne des gradients reste sans biais alors que la variance de l'estimation de la moyenne augmente. Or, il peut être plus efficace de procéder ainsi pour avancer plus rapidement dans l'espace solution. L'algorithme 2 présente le fonctionnement de cette méthode.

Une façon d'accélérer les choses, surtout dans l'optique de l'utilisation de bibliothèques d'algèbre linéaire optimisées, est de passer à la propagation tous les vecteurs d'entrée d'un lot à la fois, sous forme de matrice. De cette façon, les calculs peuvent tous se faire qu'en une seule propagation et rétropropagation.

Dans ce cas, un lot est un couple (\mathbf{X}, \mathbf{C}) où \mathbf{X} et \mathbf{C} sont des matrices de dimensions $n \times l$ composées de l vecteurs \mathbf{x} et \mathbf{c} respectivement.

À partir de ceci, chaque vecteur colonne des formules de la section 5 devient une matrice à l colonnes pour accommoder l'entrée et la cible de dimension l .

³De l'anglais *mini-batch training*.

Données : Une base de donnée \mathbf{B} où \mathbf{B}_t est le t -ième couple (\mathbf{x}, \mathbf{c}) et dont la taille est N .

Entrées : Une fonction $CritereArret()$ qui renvoie vrai quand l'apprentissage est terminé.
 Un réseau de neurones muni d'une collection de paramètres θ .
 Un hyper-paramètre α donnant la vitesse d'apprentissage de l'algorithme.
 Le nombre de données l à mettre dans chaque lots.

tant que $CritereArret(E)$ **faire**

$\mathbf{B}_{lot} \leftarrow l$ couples (\mathbf{x}, \mathbf{c}) tirés au hasard et sans remise de \mathbf{B} ;

pour chaque *couple* (\mathbf{x}, \mathbf{c}) de \mathbf{B}_{lot} **faire**

$\mathbf{y} \leftarrow Propagation(\mathbf{x})$; /* La prédiction du réseau */

$\mathbf{g}_t \leftarrow Retropropagation(\mathbf{y} - \mathbf{c})$; /* Le t -ième gradient */

$\mathbf{g} \leftarrow \frac{1}{N} \sum_{t=1}^N \mathbf{g}_t$; /* Le gradient moyen */

$\theta \leftarrow \theta - \alpha \mathbf{g}$; /* Mise à jour des paramètres */

$E \leftarrow \frac{1}{N} \sum_{t=1}^N \|y_t - c_t\|_2^2$; /* Erreur moyenne de prédiction */

Sorties : Un réseau de neurones ayant des paramètres θ optimisés.

Algorithme 2: Algorithme de la descente de gradient pour l'apprentissage d'un réseau de neurones par lots.

Les formules vectorielles restent valides. Seulement les dimensions des entrées/sorties et de leurs gradients changent. Toutefois, une différence est à noter au niveau de la fonction d'erreur (37) qui, pour compenser la gestion de plusieurs vecteurs à la fois, devient, selon l'article de Møller [Mø192],

$$E = \frac{1}{2l} \|\mathbf{Y} - \mathbf{C}\|_F^2$$

où \mathbf{Y} est la matrice des sorties prédites. En fait, ce calcul revient à trouver l'erreur moyenne commise sur chacun des couples de vecteurs (\mathbf{x}, \mathbf{c}) . La dérivée de cette fonction d'erreur est $\frac{\partial E}{\partial \mathbf{Y}} = \frac{1}{l}(\mathbf{Y} - \mathbf{C})$.

Pour le calcul du produit hessienne-vecteur, il faut également normaliser par le facteur $\frac{1}{l}$ la formule (40) qui devient

$$\mathcal{R}\{\nabla_{\mathbf{y}} E\} = \frac{1}{l} \mathcal{R}\{\mathbf{y}\}.$$

Grâce à ces deux petits changements, il est possible d'utiliser l'apprentissage par lots sans autres modifications aux formules présentées à la section 5.

6.3 Descente de gradient stochastique

Il est également possible de pousser plus loin l'idée des lots en utilisant des lots de taille 1. Pour ce faire, la méthode consiste à prendre chaque couple (\mathbf{x}, \mathbf{c}) individuellement et à appliquer les formules de la section 5. Cette façon de faire

est appelée l'apprentissage en ligne⁴ lorsqu'elle se fait dans l'ordre des données ou apprentissage par descente de gradient stochastique lorsque le couple en entrée est choisi au hasard.

6.4 Méthodes d'optimisation de second ordre

Une autre possibilité est d'utiliser des algorithmes d'optimisation se servant d'informations sur la courbure de la fonction d'erreur, c'est-à-dire sur la matrice hessienne de E . Pour aider ces algorithmes, une méthode semblable à la rétro-propagation a été exposée dans cet article pour trouver le produit de la matrice hessienne et d'un vecteur. Ce produit est souvent utilisé dans ces techniques pour observer la courbure de la fonction d'erreur dans une direction donnée, ce qui ne nécessite pas le calcul de la hessienne en entier.

Bien entendu, il est possible de combiner ces méthodes avec la philosophie des lots afin d'accélérer l'apprentissage.

Références

- [Bis06] C. M. BISHOP : *Pattern recognition and machine learning*. Springer New York, 2006.
- [Mø192] M. MØLLER : Supervised learning on large redundant training sets. *In Neural Networks for Signal Processing [1992] II., Proceedings of the 1992 IEEE-SP Workshop*, pages 79–89. IEEE, 1992.
- [Pea94] B. A. PEARLMUTTER : Fast exact multiplication by the hessian. *Neural Computation*, 6(1):147–160, 1994.
- [Roj96] R. ROJAS : *Neural networks : a systematic introduction*. Springer, 1996.
- [SB12] A. SALVAIL-BÉRARD : Les arbres de décision hybrides. *Cahier de Mathématique de l'Université de Sherbrooke*, 2:34–58, 2012.

ADAM SALVAIL-BÉRARD
DÉPARTEMENT D'INFORMATIQUE, UNIVERSITÉ DE SHERBROOKE
Courriel: Adam.Salvail-Berard@USherbrooke.ca

⁴De l'anglais *online training*.

EXEMPLES D'EXTENSIONS GALOISIENNES DE DEGRÉ 24 SUR \mathbb{Q}

NICOLAS BUREAU

RÉSUMÉ Le théorème fondamental de la théorie de Galois permet de donner la liste des sous-corps d'une extension algébrique de \mathbb{Q} en utilisant la liste des sous-groupes de son groupe de Galois. L'objet de cet article est de décrire les sous-corps d'une classe d'extensions de degré 24 sur le corps \mathbb{Q} des rationnels.

1 Introduction

Cet article tient pour acquis que la lectrice et le lecteur soient familiers avec la théorie des groupes ainsi qu'avec la théorie des extensions de corps.

La théorie de Galois est très utile dans plusieurs sphères des mathématiques telles que l'arithmétique ainsi que l'algèbre. Le théorème fondamental de cette théorie est inévitable lorsque vient le moment de faire l'éventail des sous-corps d'une extension donnée de \mathbb{Q} . En effet, elle permet de passer d'un problème *a priori* continu à un problème discret, celui des sous-groupes d'un groupe. Ce changement transforme la chasse aux sous-corps en une chasse aux sous-groupes.

Le but de cet article est de mettre la lumière sur un exemple d'application du théorème fondamental de la théorie de Galois [DF04]. Il s'agit d'étudier la classe d'extensions de la forme $\mathbb{Q}(\omega, \theta, \xi)/\mathbb{Q}$ où

$$\omega = \sqrt[6]{M}, \quad \zeta = \frac{1}{2}(1 + \theta), \quad \theta = \sqrt{-3}, \quad \xi = \sqrt{m},$$

avec M et m deux entiers copremiers > 1 sans facteur carré et ζ une sixième racine primitive de l'unité. Dès maintenant, nous pouvons remarquer que $\mathbb{Q}(\omega, \theta, \xi)$ est une extension de degré 24 sur \mathbb{Q} . En effet,

$$[\mathbb{Q}(\omega, \theta, \xi) : \mathbb{Q}] = [\mathbb{Q}(\omega, \theta, \xi) : \mathbb{Q}(\theta, \xi)] \times [\mathbb{Q}(\theta, \xi) : \mathbb{Q}(\xi)] \times [\mathbb{Q}(\xi) : \mathbb{Q}] = 6 \times 2 \times 2 = 24.$$

Nous verrons dans les pages qui suivent que le diagramme de Hasse des sous-corps de $\mathbb{Q}(\omega, \theta, \xi)$ a l'allure de la Figure 19.

L'auteur tient à remercier son directeur de maîtrise, monsieur Claude Levesque, de l'Université Laval, sans qui cet article n'aurait jamais vu le jour. Ses nombreux conseils ainsi que son support financier sont grandement appréciés. L'auteur tient également à remercier le département de mathématiques de l'Université Laval pour sa contribution monétaire.

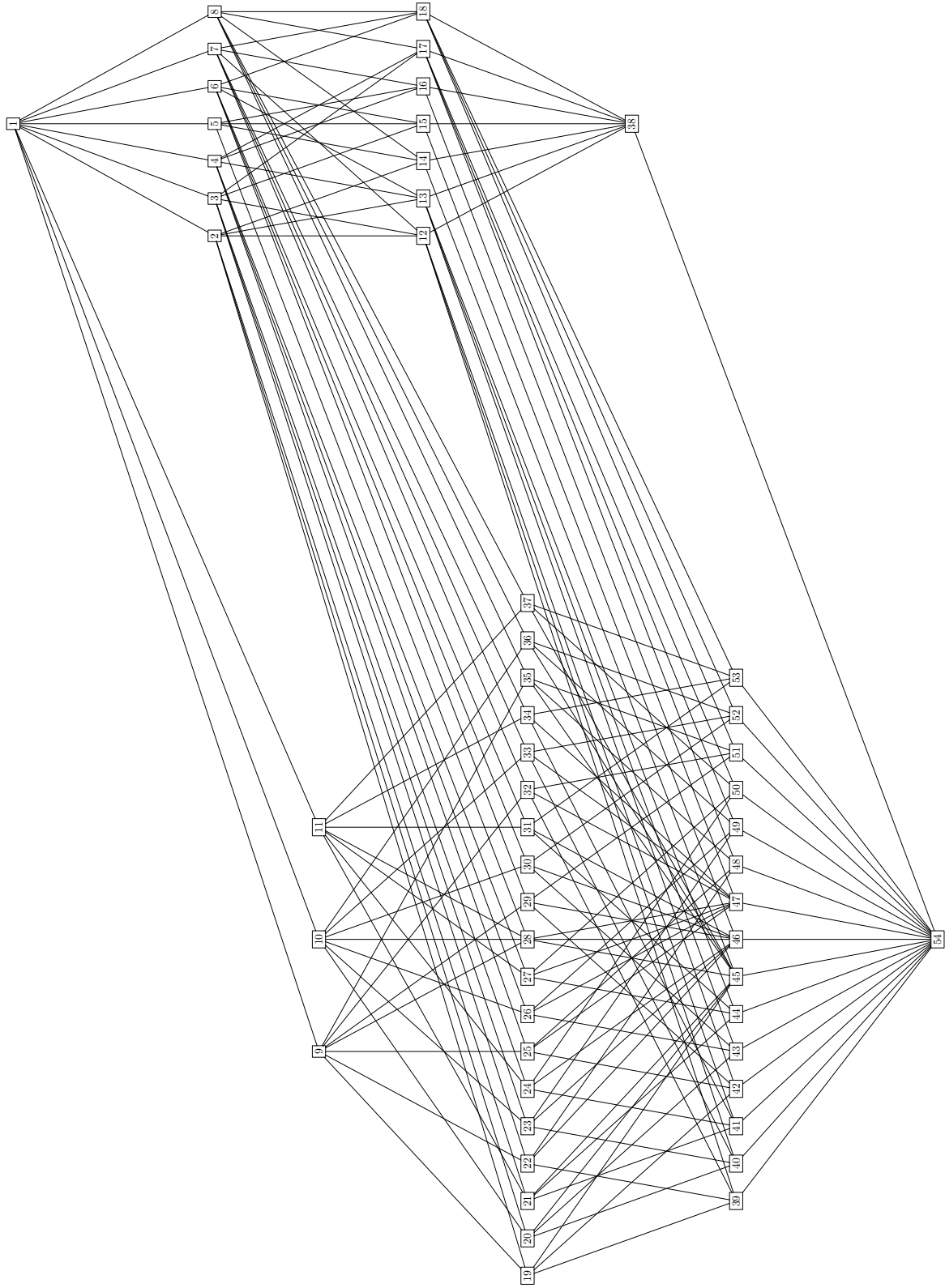


FIGURE 19 – Diagramme de Hasse des sous-groupes de G et des sous-corps de K (en inversant l'ordre)

2 Théorème fondamental

Avant d'énoncer le théorème fondamental de la théorie de Galois, quelques rappels seront utiles. Nous disons qu'un automorphisme σ d'un corps K fixe un élément $\alpha \in K$ si $\sigma(\alpha) = \alpha$. Si un tel automorphisme fixe tous les éléments d'un sous-corps F de K , dans ce cas nous disons simplement que σ fixe F . Inversement, étant donné un sous-groupe H du groupe $Aut(K)$ des automorphismes de K , le sous-corps de K fixé par tous les automorphismes de ce sous-groupe est dit le corps *laissé fixe* par H . Dans le cas où nous avons une extension de corps K/F , nous notons alors par $Aut(K/F)$ le groupe des automorphismes de K qui fixent F . Cette extension est dite *galoisienne* si son degré $[K : F]$ est fini et égal au cardinal de $Aut(K/F)$. De plus, en supposant l'extension K/F galoisienne, nous identifions $Aut(K/F)$ comme étant le *groupe de Galois* de K/F et nous le notons $Gal(K/F)$.

Cette petite mise en contexte nous amène à énoncer le résultat fondamental suivant.

Théorème 2.1. (Théorème fondamental de la théorie de Galois) *Soit K/F une extension galoisienne et soit $G = Gal(K/F)$. Alors il y a une bijection entre les ensembles suivants :*

$$\begin{array}{ccc} \{\text{sous-corps } E \text{ de } K \text{ contenant } F\} & \longleftrightarrow & \{\text{sous-groupes } H \text{ de } G\} \\ K & \longleftrightarrow & \mathbb{I} \\ | & & | \\ E & \longleftrightarrow & H \\ | & & | \\ F & \longleftrightarrow & G \end{array}$$

avec $\mathbb{I} = \{id\}$ où id est l'élément neutre de G . Cette bijection est donnée par les correspondances

$$\begin{aligned} E &\longmapsto \{\text{les éléments de } G \text{ qui laissent } E \text{ fixe}\}, \\ \{\text{les éléments du corps laissé fixe par } H\} &\longleftarrow H, \end{aligned}$$

qui sont inverses l'une de l'autre. De plus, nous avons les résultats suivants :

- (1) Si les corps E_1 et E_2 correspondent respectivement aux groupes H_1 et H_2 , alors $E_1 \subseteq E_2$ si et seulement si $H_2 \leq H_1$.
- (2) Le degré $[K : E]$ de l'extension K/E est égal à la cardinalité du groupe H associé à E . En d'autres termes, $[K : E] = |H|$. De plus, le degré de l'extension E/F est égal à l'indice du sous-groupe H dans G . En d'autres termes, $[E : F] = [G : H]$:

$$\begin{array}{ccc} K & & \\ | & \} & |H| \\ E & & \\ | & \} & |G : H| \\ F & & \end{array}$$

- (3) L'extension K/E est toujours galoisienne avec $H = Gal(K/E)$ comme groupe de Galois.
- (4) L'extension E/F est galoisienne si et seulement si H est un sous-groupe distingué de G . Si c'est le cas, alors le groupe de Galois $Gal(E/F)$ est isomorphe au quotient de groupes G/H .
- (5) Si les corps E_1 et E_2 correspondent respectivement aux groupes H_1 et H_2 , alors l'intersection $E_1 \cap E_2$ correspond au groupe $\langle H_1, H_2 \rangle$ engendré par H_1 et H_2 . De plus, le corps composé $E_1 E_2$ est associé à l'intersection $H_1 \cap H_2$.

3 En quête des sous-corps de $\mathbb{Q}(\omega, \theta, \xi) / \mathbb{Q}$

Par l'intermédiaire du théorème fondamental de la théorie de Galois, nous trouverons la liste des sous-corps de l'extension galoisienne K/\mathbb{Q} où $K = \mathbb{Q}(\omega, \theta, \xi)$. En premier lieu, sachant que le degré de l'extension est 24, nous définirons les automorphismes de K qui forment le groupe de Galois de K/\mathbb{Q} . Finalement, nous expliciterons tous les sous-groupes de $G = Gal(K/\mathbb{Q})$ ainsi que tous les sous-corps laissés fixes par ces derniers.

3.1 Automorphismes du groupe de Galois

Tel que susmentionné, l'extension galoisienne est de degré 24. Nous sommes donc à la recherche d'un groupe d'automorphismes ayant ce même nombre d'éléments. La Table 2 explicite les trois générateurs de $Gal(K/\mathbb{Q})$ ainsi que leurs effets sur certains éléments clefs du corps K . Remarquons que $\omega\zeta^2 = \sigma(\omega\zeta) = \sigma(\omega)\sigma(\zeta) = \omega\zeta \cdot \sigma(\zeta)$ de sorte que $\sigma(\zeta) = \zeta$. De plus $\tau(\zeta) = \tau(\frac{1}{2}(1 + \theta)) = \frac{1}{2}(1 - \theta) = -[\frac{1}{2}(1 + \theta)]^2 = -\zeta^2$.

| | | | | | | | | | |
|----------|---------------|------------------|-----------------|----------------|------------------|------------------|-----------|--------|------------|
| | ω | $\omega\zeta$ | $\omega\zeta^2$ | $-\omega$ | $-\omega\zeta$ | $-\omega\zeta^2$ | θ | ξ | ζ |
| σ | $\omega\zeta$ | $\omega\zeta^2$ | $-\omega$ | $-\omega\zeta$ | $-\omega\zeta^2$ | ω | θ | ξ | ζ |
| τ | ω | $-\omega\zeta^2$ | $-\omega\zeta$ | $-\omega$ | $\omega\zeta^2$ | $\omega\zeta$ | $-\theta$ | ξ | $-\zeta^2$ |
| ρ | ω | $\omega\zeta$ | $\omega\zeta^2$ | $-\omega$ | $-\omega\zeta$ | $-\omega\zeta^2$ | θ | $-\xi$ | ζ |

TABLE 2 – Table des générateurs σ , τ et ρ de G et leurs effets

Montrons que l'ordre du premier automorphisme σ est 6 et que $\tau\sigma^5 = \sigma\tau$. En effet,

$$\begin{aligned} \sigma^6(\omega) &= \sigma^5(\omega\zeta) = \sigma^4(\omega\zeta^2) = \sigma^3(-\omega) = -\sigma^2(\omega\zeta) = -\sigma(\omega\zeta^2) = \omega, \\ \sigma^6(\theta) &= \theta \quad \text{et} \quad \sigma^6(\xi) = \xi, \\ \tau\sigma^5(\omega) &= \tau\sigma^4(\omega\zeta) = \tau\sigma^3(\omega\zeta^2) = \tau\sigma^2(-\omega) = -\tau\sigma(\omega\zeta) = \omega\zeta = \sigma\tau(\omega), \\ \tau\sigma^5(\theta) &= \tau(\theta) = -\theta = \sigma(-\theta) = \sigma\tau(\theta) \quad \text{et} \quad \tau\sigma^5(\xi) = \xi = \sigma\tau(\xi). \end{aligned}$$

Les deux derniers automorphismes τ et ρ sont en fait d'ordre 2. Nous obtenons donc que le groupe de Galois $G = Gal(K/F) = \langle \sigma, \tau, \rho \rangle$ est isomorphe à $D_{12} \times$

$\mathbb{Z}/2\mathbb{Z}$, où D_{12} est le groupe diédral d'ordre 12. La table de Cayley pour G est représentée dans la Table 3.

3.2 Sous-groupes du groupe de Galois

En vertu du théorème de Lagrange, nous savons que les cardinalités des sous-groupes de G sont des diviseurs de 24. La liste de tous les sous-groupes se trouve dans la colonne appropriée du tableau de la page 78. Selon Neubiiser [Neu67], il existe sept groupes de cardinalité 12, trois de cardinalité 8, sept de cardinalité 6, dix-neuf de cardinalité 4, un seul de cardinalité 3 et finalement quinze de cardinalité 2, pour un total de 54 sous-groupes, si nous comptons le groupe lui-même $G = D_{12} \times \mathbb{Z}/2\mathbb{Z}$ et le sous-groupe trivial \mathbb{I} .

3.3 Liste des sous-corps

Le cinquième résultat du théorème fondamental donne une méthode efficace pour trouver tous les sous-corps à partir des sous-groupes. L'idée est de commencer avec tous les groupes engendrés par un seul élément (il y en a 18, si nous ignorons le groupe trivial) et de déterminer le sous-corps associé à chacun. Pour ce faire, il faut à prime abord appliquer l'automorphisme générateur à un élément typique de K et nous vérifions quels termes sont fixés.

Par exemple, trouvons le sous-corps associé au groupe $\langle \tau\sigma^4 \rangle$. Un élément typique z de K s'écrit

$$\begin{aligned} z = & a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3 + a_4\omega^4 + a_5\omega^5 \\ & + b_0\theta + b_1\omega\theta + b_2\omega^2\theta + b_3\omega^3\theta + b_4\omega^4\theta + b_5\omega^5\theta \\ & + c_0\xi + c_1\omega\xi + c_2\omega^2\xi + c_3\omega^3\xi + c_4\omega^4\xi + c_5\omega^5\xi \\ & + d_0\theta\xi + d_1\omega\theta\xi + d_2\omega^2\theta\xi + d_3\omega^3\theta\xi + d_4\omega^4\theta\xi + d_5\omega^5\theta\xi. \end{aligned}$$

Par la suite, nous appliquons l'automorphisme $\tau\sigma^4$ sur z :

$$\begin{aligned} \tau\sigma^4(z) = & a_0 - \frac{a_1}{2}\omega(1-\theta) + \frac{a_2}{4}\omega^2(1-\theta)^2 - \frac{a_3}{8}\omega^3(1-\theta)^3 + \frac{a_4}{16}\omega^4(1-\theta)^4 \\ & - \frac{a_5}{32}\omega^5(1-\theta)^5 - b_0\theta + \frac{b_1}{2}\omega(1-\theta)\theta - \frac{b_2}{4}\omega^2(1-\theta)^2\theta \\ & + \frac{b_3}{8}\omega^3(1-\theta)^3\theta - \frac{b_4}{16}\omega^4(1-\theta)^4\theta + \frac{b_5}{32}\omega^5(1-\theta)^5\theta + c_0\theta \\ & - \frac{c_1}{2}\omega(1-\theta)\xi + \frac{c_2}{4}\omega^2(1-\theta)^2\xi - \frac{c_3}{8}\omega^3(1-\theta)^3\xi + \frac{c_4}{16}\omega^4(1-\theta)^4\xi \\ & - \frac{c_5}{32}\omega^5(1-\theta)^5\xi - d_0\theta\xi + \frac{d_1}{2}\omega(1-\theta)\theta\xi - \frac{d_2}{4}\omega^2(1-\theta)^2\theta\xi \\ & + \frac{d_3}{8}\omega^3(1-\theta)^3\theta\xi - \frac{d_4}{16}\omega^4(1-\theta)^4\theta\xi + \frac{d_5}{32}\omega^5(1-\theta)^5\theta\xi \\ = & a_0 - \frac{a_1}{2}\omega(1-\theta) - \frac{a_2}{2}\omega^2(1+\theta) + a_3\omega^3 - \frac{a_4}{2}\omega^4(1-\theta) - \frac{a_5}{2}\omega^5(1+\theta) \\ & - b_0\theta + \frac{b_1}{2}\omega(1-\theta)\theta + \frac{b_2}{2}\omega^2(1+\theta)\theta - b_3\omega^3\theta + \frac{b_4}{2}\omega^4(1-\theta)\theta \\ & + \frac{b_5}{2}\omega^5(1+\theta)\theta + c_0 - \frac{c_1}{2}\omega(1-\theta)\xi - \frac{c_2}{2}\omega^2(1+\theta)\xi + c_3\omega^3\xi \end{aligned}$$

$$\begin{aligned}
& -\frac{c_4}{2}\omega^4(1-\theta)\xi - \frac{c_5}{2}\omega^5(1+\theta)\xi - d_0\theta\xi + \frac{d_1}{2}\omega(1-\theta)\theta\xi \\
& + \frac{d_2}{2}\omega^2(1+\theta)\theta\xi - d_3\omega^3\theta\xi + \frac{d_4}{2}\omega^4(1-\theta)\theta\xi + \frac{d_5}{2}\omega^5(1+\theta)\theta\xi \\
= & a_0 + \left(-\frac{a_1}{2} + \frac{3b_1}{2}\right)\omega + \left(-\frac{a_2}{2} - \frac{3b_2}{2}\right)\omega^2 + a_3\omega^3 \\
& + \left(-\frac{a_4}{2} + \frac{3b_4}{2}\right)\omega^4 + \left(-\frac{a_5}{2} - \frac{3b_5}{2}\right)\omega^5 - b_0\theta + \left(\frac{a_1}{2} + \frac{b_1}{2}\right)\omega\theta \\
& + \left(-\frac{a_2}{2} + \frac{b_2}{2}\right)\omega^2\theta - b_3\omega^3\theta + \left(\frac{a_4}{2} + \frac{b_4}{2}\right)\omega^4\theta + \left(-\frac{a_5}{2} + \frac{b_5}{2}\right)\omega^5\theta \\
& + c_0\xi + \left(-\frac{c_1}{2} + \frac{3d_1}{2}\right)\omega\xi + \left(-\frac{c_2}{2} - \frac{3d_2}{2}\right)\omega^2\xi + c_3\omega^3\xi \\
& + \left(-\frac{c_4}{2} + \frac{3d_4}{2}\right)\omega^4\xi + \left(-\frac{c_5}{2} - \frac{3d_5}{2}\right)\omega^5\xi - d_0\theta\xi + \left(\frac{c_1}{2} + \frac{d_1}{2}\right)\omega\theta\xi \\
& + \left(-\frac{c_2}{2} + \frac{d_2}{2}\right)\omega^2\theta\xi - d_3\omega^3\theta\xi + \left(\frac{c_4}{2} + \frac{d_4}{2}\right)\omega^4\theta\xi \\
& + \left(-\frac{c_5}{2} + \frac{d_5}{2}\right)\omega^5\theta\xi.
\end{aligned}$$

Or, comme nous cherchons le corps laissé fixe par $\langle \tau\sigma^4 \rangle$, nous avons $\tau\sigma^4(z) = z$, ce qui entraîne que $b_0 = 0$, $b_1 = a_1$, $b_2 = -a_2$, $b_3 = 0$, $b_4 = a_4$, $b_5 = -a_5$, $d_0 = 0$, $d_1 = c_1$, $d_2 = -c_2$, $d_3 = 0$, $d_4 = c_4$ et $d_5 = -c_5$. Ainsi, z se réduit à

$$\begin{aligned}
z = & a_0 + a_1\omega(1+\theta) + a_2\omega^2(1-\theta) + a_3\omega^3 + a_4\omega^4(1+\theta) + a_5\omega^5(1-\theta) \\
& + c_0\xi + c_1\omega(1+\theta)\xi + c_2\omega^2(1-\theta)\xi + c_3\omega^3\xi + c_4\omega^4(1+\theta)\xi \\
& + c_5\omega^5(1-\theta)\xi \\
= & a_0 + a_1(\omega(1+\theta)) + a'_2(\omega(1+\theta))^2 + a'_3(\omega(1+\theta))^3 + a'_4(\omega(1+\theta))^4 \\
& + a'_5(\omega(1+\theta))^5 + c_0\xi + c_1(\omega(1+\theta))\xi + c'_2(\omega(1+\theta))^2\xi \\
& + c'_3(\omega(1+\theta))^3\xi + c'_4(\omega(1+\theta))^4\xi + c'_5(\omega(1+\theta))^5\xi.
\end{aligned}$$

Nous pouvons ainsi conclure que comme z est un élément typique de $\mathbb{Q}(\omega(1+\theta), \xi)$, alors ce dernier corps est le corps laissé fixe par $\langle \tau\sigma^4 \rangle$.

Après avoir trouvé tous les corps laissés fixes par les groupes cycliques, il ne reste qu'à tenir compte des sous-groupes engendrés par deux automorphismes ou plus. En vertu du cinquième résultat du théorème fondamental de la théorie de Galois, il suffit ensuite de prendre l'intersection des sous-corps qui correspondent aux groupes monogènes qui le composent.

Par exemple, le sous-groupe $\langle \tau \rangle$ correspond au sous-corps $\mathbb{Q}(\omega, \xi)$. En effet, $\tau(\omega) = \omega$, $\tau(\xi) = \xi$ et $|\langle \tau \rangle| = [K : \mathbb{Q}(\omega, \xi)] = 2$, alors le deuxième résultat du théorème nous assure que $\langle \tau \rangle$ ne fixe pas un corps plus grand que $\mathbb{Q}(\omega, \xi)$. De même, le sous-groupe $\langle \sigma \rangle$ est associé à $\mathbb{Q}(\theta, \xi)$, car $\sigma(\theta) = \theta$, $\sigma(\xi) = \xi$ et $\langle \sigma \rangle$ est d'ordre 6, ce qui correspond au degré de $K/\mathbb{Q}(\theta, \xi)$. Nous pouvons donc aisément déduire que le sous-groupe $\langle \tau, \sigma \rangle$ correspond à l'intersection $\mathbb{Q}(\omega, \xi) \cap \mathbb{Q}(\theta, \xi) = \mathbb{Q}(\xi)$.

Il existe des raccourcis pour déterminer les corps laissés fixes par les sous-groupes de G , mais cette technique requiert une bonne intuition et une certaine expérience. À partir d'un sous-groupe donné H , il s'agit de déduire à prime abord un bon candidat pour le corps E , puis de vérifier que les générateurs de H le fixent bel et bien, mais qu'ils ne fixent aucun corps plus grand en comparant l'ordre de H avec le degré de l'extension K/E .

Pour illustrer cette méthode, considérons le groupe $H_{11} = \langle \sigma^3, \tau\sigma, \rho \rangle$. Notre flair nous amène à choisir $E_{11} = \mathbb{Q}(\omega^2(1-\theta))$ comme corps laissé fixe. En premier, nous vérifions les effets des générateurs sur E_{11} :

$$\begin{aligned} \sigma^3(\omega^2(1-\theta)) &= (\sigma^3(\omega))^2 \cdot \sigma^3(1-\theta) = (-\omega)^2(1-\theta) = \omega^2(1-\theta), \\ \tau\sigma(\omega^2(1-\theta)) &= (\tau\sigma(\omega))^2 \cdot \tau\sigma(1-\theta) = (\tau(\omega\zeta))^2(1+\theta) \\ &= \frac{1}{4}\omega^2(1-\theta)^2(1+\theta) = \omega^2(1-\theta), \\ \rho(\omega^2(1-\theta)) &= \omega^2(1-\theta). \end{aligned}$$

Les trois générateurs de H_{11} fixent donc E_{11} . Finalement, comme l'ordre de H_{11} vaut 8 et que le degré de K/E_{11} donne également 8, alors le candidat est le bon.

À l'aide de la Table 4, nous pouvons illustrer les relations groupe-sous-groupe ou inversement corps-sous-corps. De plus, le diagramme de Hasse que nous avons exhibé nous donne une vision globale des sous-corps reproduisant le treillis des sous-groupes.

| n | Sous-groupes | Sous-corps | n | Sous-groupes | Sous-corps |
|-----|---|--|-----|---|--|
| 1 | $G = \langle \sigma, \tau, \rho \rangle$ | $F = \mathbb{Q}$ | 28 | $H_{28} = \langle \sigma^3, \rho \rangle$ | $E_{28} = \mathbb{Q}(\omega^2, \theta)$ |
| 2 | $H_2 = \langle \rho\sigma, \tau \rangle$ | $E_2 = \mathbb{Q}(\omega^3\xi)$ | 29 | $H_{29} = \langle \sigma^3, \rho\tau\sigma^2 \rangle$ | $E_{29} = \mathbb{Q}(\omega^2(1+\theta), \theta\xi)$ |
| 3 | $H_3 = \langle \sigma, \tau \rangle$ | $E_3 = \mathbb{Q}(\xi)$ | 30 | $H_{30} = \langle \sigma^3, \rho\tau \rangle$ | $E_{30} = \mathbb{Q}(\omega^2, \theta\xi)$ |
| 4 | $H_4 = \langle \sigma^2, \tau\sigma, \rho \rangle$ | $E_4 = \mathbb{Q}(\omega^3\theta)$ | 31 | $H_{31} = \langle \sigma^3, \rho\tau\sigma \rangle$ | $E_{31} = \mathbb{Q}(\omega^2(1-\theta), \theta\xi)$ |
| 5 | $H_5 = \langle \sigma, \rho \rangle$ | $E_5 = \mathbb{Q}(\theta)$ | 32 | $H_{32} = \langle \tau\sigma^2, \rho \rangle$ | $E_{32} = \mathbb{Q}(\omega(1-\theta))$ |
| 6 | $H_6 = \langle \sigma, \rho\tau \rangle$ | $E_6 = \mathbb{Q}(\theta\xi)$ | 33 | $H_{33} = \langle \tau, \rho \rangle$ | $E_{33} = \mathbb{Q}(\omega)$ |
| 7 | $H_7 = \langle \sigma^2, \tau, \rho \rangle$ | $E_7 = \mathbb{Q}(\omega^3)$ | 34 | $H_{34} = \langle \tau\sigma^4, \rho \rangle$ | $E_{34} = \mathbb{Q}(\omega(1+\theta))$ |
| 8 | $H_8 = \langle \rho\sigma, \rho\tau \rangle$ | $E_8 = \mathbb{Q}(\omega^3\theta\xi)$ | 35 | $H_{35} = \langle \rho\sigma^3, \rho\tau\sigma^2 \rangle$ | $E_{35} = \mathbb{Q}(\omega\theta(1-\theta)\xi)$ |
| 9 | $H_9 = \langle \sigma^3, \tau\sigma^2, \rho \rangle$ | $E_9 = \mathbb{Q}(\omega^2(1+\theta))$ | 36 | $H_{36} = \langle \rho\sigma^3, \rho\tau \rangle$ | $E_{36} = \mathbb{Q}(\omega\theta\xi)$ |
| 10 | $H_{10} = \langle \sigma^3, \tau, \rho \rangle$ | $E_{10} = \mathbb{Q}(\omega^2)$ | 37 | $H_{37} = \langle \rho\sigma^3, \tau\sigma \rangle$ | $E_{37} = \mathbb{Q}(\omega\theta(1+\theta)\xi)$ |
| 11 | $H_{11} = \langle \sigma^3, \tau\sigma, \rho \rangle$ | $E_{11} = \mathbb{Q}(\omega^2(1-\theta))$ | 38 | $H_{38} = \langle \sigma^2 \rangle$ | $E_{38} = \mathbb{Q}(\omega^3, \theta, \xi)$ |
| 12 | $H_{12} = \langle \sigma^2, \tau \rangle$ | $E_{12} = \mathbb{Q}(\omega^3, \xi)$ | 39 | $H_{39} = \langle \tau\sigma^2 \rangle$ | $E_{39} = \mathbb{Q}(\omega(1-\theta), \xi)$ |
| 13 | $H_{13} = \langle \sigma^2, \rho\tau\sigma \rangle$ | $E_{13} = \mathbb{Q}(\omega^3\xi, \theta\xi)$ | 40 | $H_{40} = \langle \tau \rangle$ | $E_{40} = \mathbb{Q}(\omega, \xi)$ |
| 14 | $H_{14} = \langle \rho\sigma \rangle$ | $E_{14} = \mathbb{Q}(\omega^3\xi, \theta)$ | 41 | $H_{41} = \langle \tau\sigma^4 \rangle$ | $E_{41} = \mathbb{Q}(\omega(1+\theta), \xi)$ |
| 15 | $H_{15} = \langle \sigma \rangle$ | $E_{15} = \mathbb{Q}(\theta, \xi)$ | 42 | $H_{42} = \langle \rho\tau\sigma^5 \rangle$ | $E_{42} = \mathbb{Q}(\omega\theta(1-\theta), \theta\xi)$ |
| 16 | $H_{16} = \langle \sigma^2, \rho \rangle$ | $E_{16} = \mathbb{Q}(\omega^3, \theta)$ | 43 | $H_{43} = \langle \rho\tau\sigma^3 \rangle$ | $E_{43} = \mathbb{Q}(\omega\theta, \theta\xi)$ |
| 17 | $H_{17} = \langle \sigma^2, \tau\sigma \rangle$ | $E_{17} = \mathbb{Q}(\omega^3\theta, \xi)$ | 44 | $H_{44} = \langle \rho\tau\sigma \rangle$ | $E_{44} = \mathbb{Q}(\omega\theta(1+\theta), \theta\xi)$ |
| 18 | $H_{18} = \langle \sigma^2, \rho\tau \rangle$ | $E_{18} = \mathbb{Q}(\omega^3, \theta\xi)$ | 45 | $H_{45} = \langle \rho\sigma^3 \rangle$ | $E_{45} = \mathbb{Q}(\omega\xi, \theta)$ |
| 19 | $H_{19} = \langle \rho\sigma^3, \tau\sigma^2 \rangle$ | $E_{19} = \mathbb{Q}(\omega(1-\theta)\xi)$ | 46 | $H_{46} = \langle \sigma^3 \rangle$ | $E_{46} = \mathbb{Q}(\omega^2, \theta, \xi)$ |
| 20 | $H_{20} = \langle \rho\sigma^3, \tau \rangle$ | $E_{20} = \mathbb{Q}(\omega\xi)$ | 47 | $H_{47} = \langle \rho \rangle$ | $E_{47} = \mathbb{Q}(\omega, \theta)$ |
| 21 | $H_{21} = \langle \rho\sigma^3, \tau\sigma^4 \rangle$ | $E_{21} = \mathbb{Q}(\omega(1+\theta)\xi)$ | 48 | $H_{48} = \langle \tau\sigma^5 \rangle$ | $E_{48} = \mathbb{Q}(\omega\theta(1-\theta), \xi)$ |
| 22 | $H_{22} = \langle \sigma^3, \tau\sigma^2 \rangle$ | $E_{22} = \mathbb{Q}(\omega^2(1+\theta), \xi)$ | 49 | $H_{49} = \langle \tau\sigma^3 \rangle$ | $E_{49} = \mathbb{Q}(\omega\theta, \xi)$ |
| 23 | $H_{23} = \langle \sigma^3, \tau \rangle$ | $E_{23} = \mathbb{Q}(\omega^2, \xi)$ | 50 | $H_{50} = \langle \tau\sigma \rangle$ | $E_{50} = \mathbb{Q}(\omega\theta(1+\theta), \xi)$ |
| 24 | $H_{24} = \langle \sigma^3, \tau\sigma \rangle$ | $E_{24} = \mathbb{Q}(\omega^2(1-\theta), \xi)$ | 51 | $H_{51} = \langle \rho\tau\sigma^2 \rangle$ | $E_{51} = \mathbb{Q}(\omega(1-\theta), \theta\xi)$ |
| 25 | $H_{25} = \langle \tau\sigma^5, \rho \rangle$ | $E_{25} = \mathbb{Q}(\omega\theta(1-\theta))$ | 52 | $H_{52} = \langle \rho\tau \rangle$ | $E_{52} = \mathbb{Q}(\omega, \theta\xi)$ |
| 26 | $H_{26} = \langle \tau\sigma^3, \rho \rangle$ | $E_{26} = \mathbb{Q}(\omega\theta)$ | 53 | $H_{53} = \langle \rho\tau\sigma^4 \rangle$ | $E_{53} = \mathbb{Q}(\omega(1+\theta), \theta\xi)$ |
| 27 | $H_{27} = \langle \tau\sigma, \rho \rangle$ | $E_{27} = \mathbb{Q}(\omega\theta(1+\theta))$ | 54 | \mathbb{I} | $K = \mathbb{Q}(\omega, \theta, \xi)$ |

TABLE 4 – Liste des sous-groupes et des sous-corps de $\mathbb{Q}(\omega, \theta, \xi)$

4 Conclusion

En fin de compte, l'utilisation du théorème fondamental de la théorie de Galois a été grandement utile dans l'énumération des sous-corps de l'extension algébrique $\mathbb{Q}(\omega, \theta, \xi)/\mathbb{Q}$. Sans la bijection avec les sous-groupes du groupe de Galois, la tâche aurait été impossible. Dans le domaine de la chasse aux groupes de Galois, il existe un document assez important et très intéressant. Il s'agit du mémoire de maîtrise du mathématicien Leonard H. Soicher, qui étudiait alors à l'Université Concordia [Soi81].

Références

- [DF04] D. DUMMIT et R. FOOTE : *Abstract Algebra*. Wiley, 2004.
- [Neu67] J. NEUBÜSER : Die untergruppenverbande der gruppen der ordnungen ≤ 100 mit ausnahme der ordnungen 64 und 96. Habilitationsschrift, Universität Kiel, Kiel, Germany, 1967.
- [Soi81] L. SOICHER : The computation of galois groups. Mémoire de D.E.A., Concordia University, Montreal, Canada, 1981.

NICOLAS BUREAU
DÉPARTEMENT DE MATHÉMATIQUES ET DE STATISTIQUE, UNIVERSITÉ LAVAL
Courriel: nicolas.bureau.2@ulaval.ca

CaMUS est une revue mathématique publiée par le Département de mathématiques de l’Université de Sherbrooke. Le but de ces cahiers est de permettre aux étudiants de présenter leurs travaux effectués dans le cadre d’activités comme les stages de recherche du premier cycle, les présentations au Club Mathématique et les cours d’initiation à la recherche. Les auteurs sont principalement des étudiants du premier cycle, notamment en mathématiques et au Baccalauréat en enseignement au secondaire avec profil en mathématiques. Cette revue est publiée à l’intention de tous ceux qui s’intéressent aux mathématiques, à l’Université de Sherbrooke et ailleurs. Ses objectifs sont de favoriser :

- l’apprentissage de la rédaction d’articles dès le premier cycle,
- l’amélioration de la communication écrite,
- le développement de la rigueur d’expression,
- la motivation des étudiants pour la recherche en général.

Le comité de rédaction de CaMUS est composé des étudiants :

- Krystelle Chouinard
- Guillaume Douville
- Adam Salvail-Bérard (rédacteur gérant)

et des professeurs :

- Ibrahim Assem
- Taoufik Bouezmarni
- Virginie Charette
- Tomasz Kaczynski (rédacteur en chef)

Information aux auteurs : Les articles doivent constituer des présentations originales mais on ne demande pas qu’ils contiennent des résultats originaux : pour la prépublication de ces derniers, la série de Rapports de recherche du Département de mathématiques est un meilleur médium. Les articles doivent être rédigés en \LaTeX et soumis suivant les instructions données à la page Web de CaMUS :

<http://camus.math.usherbrooke.ca>

Abonnement, achat d’exemplaires et autres renseignements : CaMUS est une publication électronique sans frais avec un nombre limité d’exemplaires imprimés à vendre. Pour des informations sur le prix courant, le mode de paiement ou pour nous contacter, rendez-vous à l’adresse Web indiquée ci-dessus.

CaMUS · Département de mathématiques · Faculté des sciences · Université de Sherbrooke · 2500, boulevard de l’Université · Sherbrooke (Québec), Canada J1K 2R1