Extraction d'information homologique des objets discrets s'appuyant sur des graphes orientés

Aldo Gonzalez-Lorenzo[1,2], Alexandra Bac[1], Jean-Luc Mari[1] et Pedro Real[2]

[1]Aix-Marseille Université, CNRS, LSIS UMR 7296 (France)
[2]Université de Séville, Institut de Mathématiques IMUS (Espagne)

Résumé

Tout objet discret n-dimensionnel peut être transformé en complexe cubique, dont on peut étudier les groupes d'homologie pour mieux comprendre l'objet original. Une approche classique consiste à calculer la Forme Normale de Smith de matrices encodant le complexe cubique. Un certain nombre de travaux développent des méthodes permettant de réduire la taille de ces matrices.
Dans cet article, nous proposons une nouvelle approche, initialement fondée sur la théorie discrète de Morse, calculant l'information homologique (nombres de Betti et cycles représentatifs) sans utiliser la Forme Normale de Smith. Notre approche s'applique en dimension quelconque et peut être utilisée pour n'importe quel type de complexe cellulaire régulier.

n-dimensional discrete objects can be interpreted as cubical complexes which are suitable for the study of their homology groups in order to understand the original discrete object. The classic approach consists in computing the Normal Smith Form of some matrices that encode the cubical complex. Further approaches deal mainly with a pre-processing of these matrices in order to reduce their size.
In this paper we propose a new approach, initially based on Discrete Morse Theory, which computes some homological information (Betti numbers and representative cycles) without calculating the Normal Smith Form. It works on any dimension, and it can also be applied to any kind of regular cell complex.

Mots clé : objet discret, complexe cubique, homologie, théorie de l'homologie effective, théorie discrète de Morse.

1. Introduction

L'homologie s'est révélée comme un très bon outil pour la classification et la compréhension des objets discrets, notamment quand ils ont plus de trois dimensions. Elle fournit une classe de descripteurs qui synthétisent la structure basique de la forme considérée.

L'homologie formalise la notion de « trou » dans un objet. On peut classifier les « trous » par leur dimension : les 0-trous sont les composantes connexes ; les 1-trous, les tunnels et les 2-trous, les cavités. Le calcul de l'information homologique a pour but, entre autres, de déterminer le nombre de ces trous (nombre de Betti) et un représentant de chacun (cycle représentatif ou générateur d'homologie). Cela reste intuitif jusqu'à trois dimensions, mais en dimension quatre, on peut trouver des coefficients de torsion dans les groupes d'homologie, ce qui n'a pas une interprétation intuitive mais permet cependant une classification des objets. Le calcul de l'homologie nécessite le choix préalable d'un anneau de coefficients ; afin de nous placer dans le cas le plus général et de pouvoir obtenir ces coefficients de torsion, nous avons choisi l'anneau des entiers.

Ce travail s'appuie sur deux approches antérieures : la théorie de l'homologie effective [Ser92] et la théorie discrète de Morse [For02]. La première explicite le lien entre l'objet original et ses groupes d'homologie (mais s'avère d'une complexité élevée, puisque basée sur des matrices typiquement géantes de taille $\mathcal{O}(n^2)$ où n est le nombre de cellules du complexe). La deuxième approche est en revanche très simple et de complexité beaucoup plus faible, mais ne permet pas toujours de calculer l'homologie exacte.

Dans cet article, nous proposons une approche qui combine ces deux théories, en en conservant les avantages tout en dépassant leurs limitations respectives. Nous construisons un objet qui généralise le *champ de vecteurs gradient discret* (outil principal de la théorie discrète de Morse) en autorisant les cycles, ce qui permet de calculer l'homologie exacte (et pas seulement une approximation). Tenant compte de ces cycles, on obtient la réduction (outil principal de la théorie de l'homologie effective) reliant le complexe cubique et ses groupes d'homologie. Cette réduction donne les générateurs de l'homologie.

Cette méthode s'applique en dimension quelconque et à n'importe quel type de complexe cellulaire régulier.

Pour plus de détails sur l'état de l'art et le contexte tant de la théorie des complexes cellulaires que de l'homologie, le lecteur pourra se reporter à notre article [GLBMR].

2. Préliminaires

Dans cette section nous introduisons les concepts nécessaires.

Un *objet discret* de dimension n est un ensemble de cubes n-dimensionnels dont le centre a des coordonnées entières.

Les complexes cubiques sont présentés de manière détaillée dans [KMM04]. Un *complexe cubique* est un type de complexe cellulaire dont les cellules sont des points (dimension 0), des arêtes (dimension 1), des carrés (dimension 2), des cubes (dimension 3), etc. Le *bord* d'une cellule de dimension q (q-cellule) est l'ensemble de $(q-1)$-cellules de sa frontière.
Un complexe cubique peut être défini par son *diagramme de Hasse*. C'est un graphe orienté où les sommets sont les cellules du complexe, et les arcs vont d'une cellule vers les cellules de son bord.

La façon la plus simple de définir un complexe cubique à partir d'un objet discret consiste à remplacer chaque cube de l'objet par une n-cellule avec ses sous-cellules de dimension inférieure. Une construction duale consiste à remplacer chaque cube par un point (0-cellule) et ajouter des cellules de dimension supérieure selon la relation d'adjacence entre les cubes. Cette dernière construction utilise moins de cellules, ce qui rend les méthodes de calcul de l'homologie plus performantes. La figure 1 montre ces deux constructions à partir d'un même objet discret.

Un *complexe de chaînes* (C_*, d_*) est une séquence de groupes $C_0, C_1, \ldots$ (appelés *groupes de chaînes*) et d'homomorphismes $d_1 : C_1 \to C_0, d_2 : C_2 \to C_1, \ldots$ (appelés *opérateurs différentiels* ou *de bord*) telle que $d_{q-1}d_q = 0, \forall q > 0$. On peut construire une complexe de chaînes à partir d'un complexe cubique : les groupes de chaînes sont les groupes libres engendrés par les cellules de chaque dimension ; l'opérateur de bord sur une cellule donne une combinaison linéaire des cellules de son bord. La formule exacte est explicitée dans [KMM04].

Le q-ème groupe d'homologie du complexe de chaînes (C_*, d_*) est le groupe quotient $H(C)_q = \ker(d_q)/\mathrm{im}(d_{q+1})$. Ce groupe est isomorphe à

$$\mathbb{Z}^{\beta_q} \times \mathbb{Z}/\lambda_1\mathbb{Z} \times \mathbb{Z}/\lambda_2\mathbb{Z} \times \ldots \times \mathbb{Z}/\lambda_t\mathbb{Z},$$

où chaque λ_i divise λ_{i+1}. β_q est le q-ème *nombre de Betti* et $\lambda_1, \ldots, \lambda_t$ sont appelés *coefficients de torsion*. Rappelons que si le complexe cubique est de dimension 3, il n'y a pas de *coefficients de torsion*.

La théorie de l'homologie effective introduit la notion de *réduction* : étant donné deux complexes de chaînes (C_*, d_*) et (C'_*, d'_*), une réduction est un triplet d'homomorphismes (h_*, f_*, g_*) tel que :

— $h_q : C_q \to C_{q+1}$ pour chaque $q \geq 0$
— $f_q : C_q \to C'_q$ est un morphisme de chaînes ($fd = d'f$)
— $g_q : C'_q \to C_q$ est aussi un morphisme de chaînes ($gd' = dg$)
— $gf = 1 - dh - hd$
— $fg = 1_{C'}$
— $hh = hf = hg = 0$

Une réduction est une sorte de simplification d'un complexe de chaînes (le deuxième complexe contient normalement des groupes dont les bases sont plus petites) préservant l'information homologique et fournissant un lien entre le complexe initial et le complexe simplifié. Quand le deuxième complexe est suffisamment simple pour ne plus contenir que les groupes d'homologie, il fournit alors les générateurs d'homologie du complexe initial.

Dans la suite de cette partie, nous présentons brièvement les notions de base de la théorie discrète de Morse. Pour une introduction plus rigoureuse, le lecteur pourra consulter [For02].
Un *champ de vecteurs discret* (*discrete vector field, DVF*) sur un complexe cubique est un couplage sur son diagramme de Hasse, c'est-à-dire un sous-graphe de ce dernier dont aucune des arêtes ne partage un même sommet. Étant donné un champ de vecteurs discret $\mathcal{V}$, le *graphe de Morse* est construit à partir du diagramme de Hasse, en renversant les flèches du sous-graphe $\mathcal{V}$. Les flèches issues de $\mathcal{V}$ sont alors appelées *intégrales*, les autres, *flèches différentielles*. Un *V-chemin* est un chemin sur le graphe de Morse qui alterne des flèches intégrales et différentielles. Ainsi, un champ de vecteurs discret qui ne contient pas de V-chemins fermés est un *champ de vecteurs gradient discret* (*discrete gradient vector field, DGVF*). Une *cellule critique* est une cellule non couplée dans le champ de vecteurs gradient discret.
L'un des résultats fondamentaux de la théorie discrète de Morse est que le nombre de q-cellules critiques est supérieur ou égal au q-ème nombre de Betti. Quand les nombres de Betti coïncident avec le nombre de cellules critiques, on dit que le champ de vecteurs gradient discret est *parfait*. Malheureusement, pour certains complexes cellulaires, un tel DGVF n'existe pas ; on appelle alors *optimal* un DGVF contenant un nombre minimal de cellules critiques (qui peut donc être supérieur au nombre de Betti).

A partir d'un champ de vecteurs gradient discret, on peut calculer une réduction associée. On définit tout d'abord l'opérateur linéaire suivant :

$$V(\sigma) = \begin{cases} \langle d(\tau), \sigma \rangle \cdot \tau, & (\sigma, \tau) \text{ est dans le couplage} \\ 0, & \text{sinon} \end{cases}$$

où $\langle d(\tau), \sigma \rangle$ est le coefficient de la cellule σ dans la chaîne $d(\tau)$. On pose alors :

$$\begin{array}{lll} h(\sigma) & = \sum_{k \geq 0} V(1 - dV)^k(\sigma) & = V(\sigma) + h(1 - dV)(\sigma) \\ f(\sigma) & = (1 - dh - hd)(\sigma) & = f(1 - dV)(\sigma) \\ g(\sigma) & = \sigma & \end{array} \tag{1}$$

3. Notre approche

Notre objectif est de dépasser les limitations de la théorie discrète de Morse en obtenant l'information homologique

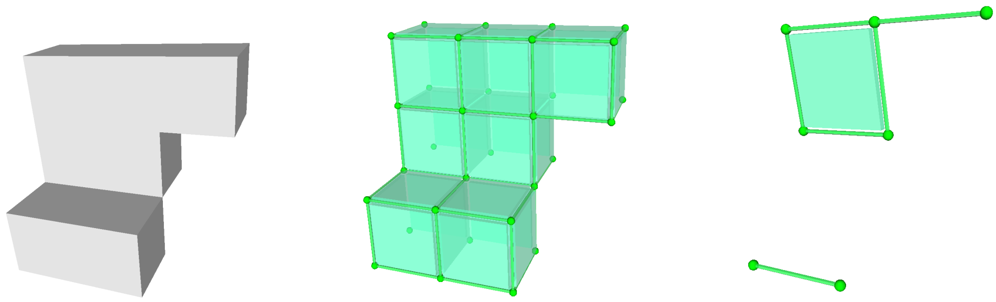

Figure 1: *Un objet discret de dimension 3 et ses deux complexes cubiques associés. Notons que dans le deuxième il y a deux composantes connexes.*

exacte à partir d'un champ de vecteurs gradient discret ; c'est une approche purement combinatoire (par opposition aux approches algébriques) qui, de plus, encode une réduction sous forme de graphe. Cependant, il a été prouvé que déterminer un DGVF optimal dans le cas général est un problème NP [LLT03] et que pour certains complexes, il n'existe pas de DGVF parfait (par exemple, la maison de Bing et le « dunce hat », voir [AFTV12]). Pour résoudre ce problème, notre approche repose sur l'enrichissement des DGVF par des cycles (soigneusement choisis et contrôlés). Nous appellerons ce type de champ de vecteurs discret un *champ de vecteurs discret homologique* (*homological discrete vector field, HDVF*).

Sans entrer dans les détails, notre approche est la suivante : étant donné un objet discret, nous construisons son complexe cubique associé et établissons un premier champ de vecteurs gradient discret quelconque. Puis, notre algorithme est basé sur la correction itérative de ce graphe. Chaque correction élimine deux cellules critiques en renversant un V-chemin entre elles, ce qui peut créer des V-chemins fermés. On peut à tout moment calculer la réduction associée au champ de vecteurs discret homologique en utilisant les formules récursives (1).
Obtenir l'image par h ou f d'une cellule σ par cette formule récursive revient à faire un parcours du graphe de Morse à partir de σ. Puisque notre HDVF peut contenir des cycles, il est nécessaire d'arrêter le parcours au bon moment. Nous introduisons la notion de *sommet de confluence* (sommet où deux cycles orientés se rejoignent) pour résoudre ce problème. Notre parcours s'arrête donc aux sommets de confluence, produisant un système d'équations linéaires (illustré dans l'exemple suivant), que nous résolvons pour obtenir la valeur de h ou f en σ.

À titre d'exemple, considérons le complexe de la figure 2. On peut noter qu'il y a trois V-chemins (en orange) entre les deux cellules critiques de dimension 1 et 2.

Lorsqu'on calcule fd sur la 2-cellule critique, on retrouve la 1-cellule critique ; s'il n'y avait qu'un V-chemin entre elles, on pourrait donc les annuler (comme en théorie discrète de Morse), c'est-à-dire, renverser ce V-chemin afin qu'elles ne soient plus critiques. Cependant, dans cet exemple, il y a trois V-chemins entre elles. Il est donc nécessaire d'en choisir un (ici, nous prenons le plus court), ce qui crée deux cycles orientés.

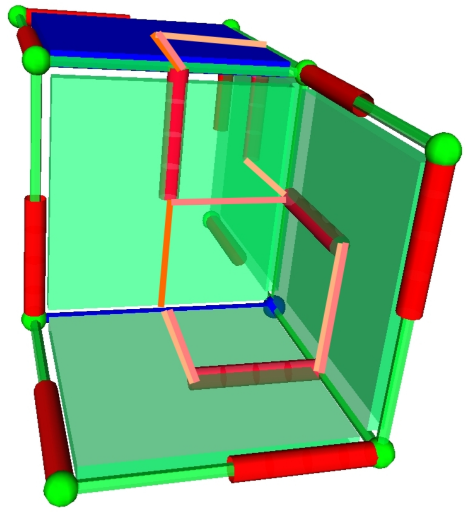

Figure 2: *Champ de vecteurs gradient discret (en rouge) sur un complexe cubique. Il y a trois cellules critiques en bleu, une de chaque dimension.*

Si l'on souhaite, à partir de ce graphe, calculer l'opérateur h (ou f) on doit d'abord déterminer sa valeur sur le sommet de confluence (nommé c, marqué en violet dans la figure 3). On obtient alors une équation du type :

$$h(c) = 2 \cdot h(c) + x \Rightarrow h(c) = -x \tag{2}$$

où x est la chaîne obtenue durant le parcours (non explicitée ici). Par simple résolution de cette équation, on obtient alors la valeur de $h(c)$, qu'on utilise ensuite pour le calcul de h en toute cellule σ (le parcours depuis σ est arrêté au sommet de confluence c et la valeur de $h(c)$ est alors substituée dans la formule récursive).

Dans le cas général on obtient une équation pour chaque sommet de confluence, produisant ainsi un système d'équations linéaires. Une fois ce processus de correction terminé, le nombre de cellules critiques dans chaque dimension coïncide avec les nombres de Betti s'il n'y a pas de coefficients

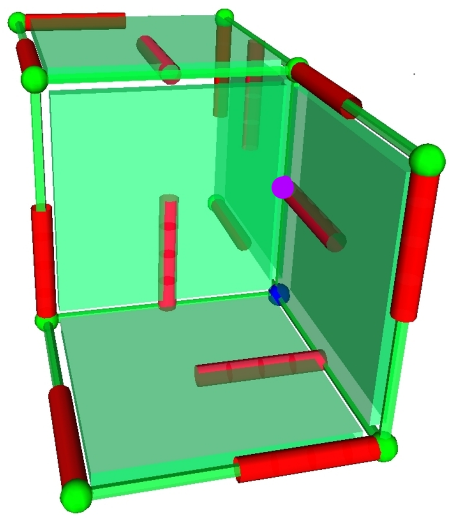

Figure 3: *Le champ de vecteurs discret (homologique) après le renversement du V-chemin. Le sommet de confluence est marqué par un point violet.*

de torsion. Le cas où il y a des coefficients de torsion requiert des explications plus précises, mais n'apporte pas de difficultés supplémentaires. Enfin, on peut obtenir les générateurs d'homologie en appliquant $f = 1 - dh - hd$ sur les cellules critiques.

Les figures 4 et 5 montrent deux exemples classiques où nous calculons un champ de vecteurs discret homologique. Le complexe (simplicial) pour le « dunce hat » a été obtenu de [Hac]. Notre algorithme a été implémenté en C++ avec les librairies DGtal [DGt] et Eigen [Eig].

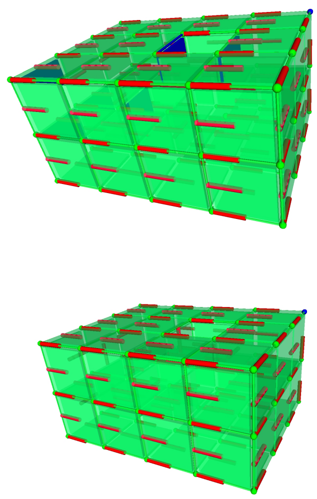

Figure 4: *Haut : un DGVF sur la maison de Bing. Bas : le HDVF après la correction du DGVF précédent. Il y a une seule cellule critique (en bleu).*

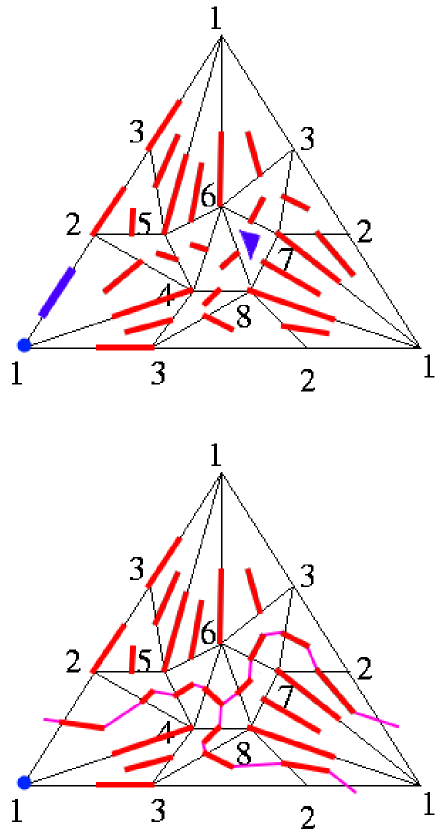

Figure 5: *Haut : DGVF sur le « dunce hat » avec trois cellules critiques. Bas : le HDVF après la correction du DGVF précédent. L'unique cellule critique est* 1*. Les deux cycles orientés du graphe de Morse sont affichés en violet.*

4. Conclusion

Nous présentons une généralisation du champ de vecteurs gradient discret pour laquelle nous introduisons un algorithme permettant d'obtenir les nombres de Betti ainsi que des générateurs d'homologie. Notre résultat n'est pas seulement théorique, mais une méthode efficace de calcul de l'information homologique. Dans les deux exemples bien connus de la maison de Bing et du « dunce hat », notre algorithme aboutit à un nombre correct de cellules critiques dans chaque dimension (correspondant aux nombres de Betti).

Références

[AFTV12] AYALA R., FERNÁNDEZ-TERNERO D., VILCHES J. A. : Perfect discrete Morse functions on 2-complexes. *Pattern Recogn. Lett.. Vol. 33*, Num. 11 (août 2012), 1495–1500.

[DGt] DGtal : Digital geometry tools and algorithms library. `http://dgtal.org`.

[Eig] Eigen. `http://eigen.tuxfamily.org`.

[For02] FORMAN R. : A user's guide to discrete Morse theory. *Séminaire Lotharingien de Combinatoire. Vol. 48* (2002), B48c.

[GLBMR] GONZALEZ-LORENZO A., BAC A., MARI J.-L., REAL P. : Computing homological information based on directed graphs within discrete objects.

[Hac] HACHIMORI M. : Simplicial complex library. `http://infoshako.sk.tsukuba.ac.jp/~hachi/math/library/index\_eng.html`.

[KMM04] KACZYNSKI T., MISCHAIKOW K., MROZEK M. : *Computational Homology*, vol. 157. Springer, 2004, ch. 2, 7, pp. 255–258.

[LLT03] LEWINER T., LOPES H., TAVARES G. : Toward optimality in discrete Morse theory. *Experimental Mathematics. Vol. 12*, Num. 3 (2003), 271–285.

[Ser92] SERGERAERT F. : Effective homology, a survey. `http://www-fourier.ujf-grenoble.fr/~sergerar/Papers/Survey.pdf`, 1992.