Algorithmes et structures de données en topologie algorithmique

Clément Maria [1]

Prix de thèse Gilles Kahn 2015

Clément Maria a soutenu sa thèse [2] en octobre 2014 à l'Université de Nice-Sophia Antipolis, sous la direction de Jean-Daniel Boissonnat, au sein de l'équipe Inria Geometrica. Il occupe actuellement un poste de Postdoctoral Research Fellow de l'Université du Queensland, en Australie. Il y étudie les aspects algorithmiques et topologiques de la théorie des nœuds et des variétés de dimension 3.

L'*analyse topologique des données* est un domaine scientifique récent et en pleine expansion. Son approche de la donnée est peu conventionnelle, en ce qu'il considère toute information statistique comme un nuage de points géométriques, dessinant une forme. L'enjeu est de décrire l'apparence de cette forme ainsi esquissée ; notamment des motifs comme son nombre de trous ou de cavités, ou encore son entortillement.

Malgré son jeune âge, l'analyse topologique des données a déjà été employée avec succès pour résoudre des problèmes appliqués, fournissant par exemple une nouvelle génération de signatures en intelligence articielle [4] ou une vision originale du comportement de systèmes dynamiques [7]. Ces champs applicatifs génèrent cependant

1. The University of Queensland, Australia.
2. Manuscrit consultable à l'adresse `https://hal.inria.fr/tel-01097764`

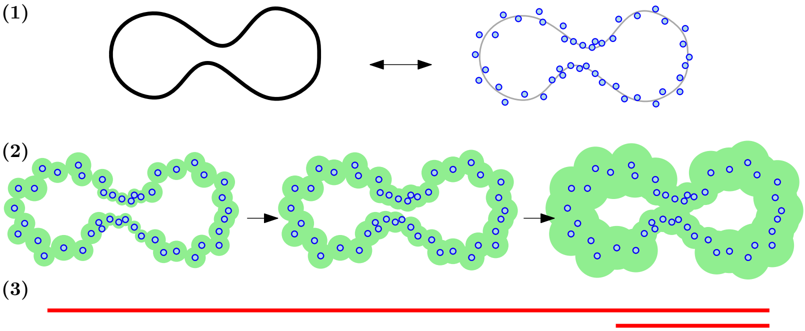

FIGURE 1. (1) Forme géométrique représentée par un ensemble de points l'échantillonnant. L'analyse topologique de ce nuage de points consiste à approcher la forme sous-jacente à diverses échelles — ici, en faisant grossir des disques (verts) centrés en chaque point — puis à étudier l'évolution de son apparence. Le diagramme (3) représente, par des barres (rouges), l'apparition, la survie puis la disparition de trous. La longue barre indique la *persistance* d'un trou, hérité de la forme d'origine.

des données complexes à analyser, du fait de leur dimensionalité (bien au-delà de nos trois dimensions habituelles) et de la présence de bruit.

L'objet de ma thèse est donc de recourir à des méthodes algorithmiques efficaces pour résoudre la question de l'analyse topologique des données dans sa généralité, y compris pour ces cas complexes. La beauté du problème est qu'il touche à une grande variété de domaines des mathématiques. En effet, reconstruire correctement une forme, à partir d'un échantillon, relève de la géométrie en grande dimension et de la topologie. Son analyse topologique fait appel à la topologie algébrique (avec la notion d'*homologie* qui formalise mathématiquement l'idée qualitative de trou et d'entortillement) et à la théorie des carquois (à travers la théorie de l'*homologie persistante* [8], qui permet l'approche multi-échelle de la Figure 1).

Les aspects algorithmiques du problème ne sont pas moins divers. Reconstruire une forme se fait par le biais de *complexes simpliciaux*, qui sont des structures combinatoires composées d'un ensemble de briques de base — telles des arêtes, des triangles et des tétrahèdres — collées entre elles par des faces communes. Ainsi, nous pouvons par exemple représenter, de manière approchée, une surface par un ensemble de triangles collés le long de leurs arêtes. Dans [3], nous introduisons une nouvelle structure de données pour construire et représenter, de façon optimale, ces complexes simpliciaux dans le cadre de l'analyse topologique des données.

Nous proposons dans [1] une structure de données pour calculer l'homologie persistante de complexes simpliciaux, en mêlant des techniques de calcul formel et des méthodes de compression de matrice. Nous raffinons également dans [2] l'analyse topologique, en capturant une notion d'entortillement (torsion homologique) persistant, grâce à une astuce arithmétique. Finalement, nous proposons un algorithme pour calculer l'homologie persistante *zigzag*, une généralisation significative de l'homologie persistante.

Ces méthodes permettent d'améliorer les temps de calcul de plusieurs ordres de grandeur, ainsi que de passer à l'échelle des problèmes d'analyse de données en sciences appliquées. Nous avons donc cherché à promouvoir et diffuser ces structures de données à travers des bibliothèques de calcul scientifique libres ; voir [5] pour une implémentation en `C++` et [6] pour une interface en `R` avec traitements statistiques.

Références

[1] Jean-Daniel Boissonnat, Tamal K. Dey, and Clément Maria. The compressed annotation matrix : An efficient data structure for computing persistent cohomology. *Algorithmica*, 73(3) :607–619, 2015.

[2] Jean-Daniel Boissonnat and Clément Maria. Computing persistent homology with various coefficient fields in a single pass. In *Algorithms - ESA 2014 - 22th Annual European Symposium, Wroclaw, Poland, September 8-10, 2014. Proceedings*, pages 185–196, 2014.

[3] Jean-Daniel Boissonnat and Clément Maria. The simplex tree : An efficient data structure for general simplicial complexes. *Algorithmica*, 70(3) :406–427, 2014.

[4] Frédéric Chazal, David Cohen-Steiner, Leonidas J. Guibas, Facundo Mémoli, and Steve Oudot. Gromov-hausdorff stable signatures for shapes using persistence. *Comput. Graph. Forum*, 28(5) :1393–1403, 2009.

[5] Brittany Terese Fasy, Jisu Kim, Fabrizio Lecci, and Clément Maria. Introduction to the R package TDA. *CoRR*, abs/1411.1830, 2014.

[6] Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The gudhi library : Simplicial complexes and persistent homology. In *Mathematical Software - ICMS 2014 - 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings*, pages 167–174, 2014.

[7] S. Martin, A. Thompson, E.A. Coutsias, and J. Watson. Topology of cyclo-octane energy landscape. *J Chem Phys*, 132(23) :234115, 2010.

[8] Scmuel Weinberger. WHAT IS... persistent homology ? *Notices of the American Mathematical Society*, 58(1), 2011.