

Réduction de dimension sous contraintes topologiques

Dimensionality reduction under topological constraints

Mattéo Clémot, Julie Digne, Julien Tierny

English Abstract—This paper presents a novel topology-aware dimensionality reduction approach aiming at accurately visualizing the cyclic patterns present in high dimensional data. To that end, we build on the *Topological Autoencoders* (TopoAE) [7] formulation. First, we provide a novel theoretical analysis of its associated loss and show that a zero loss indeed induces identical persistence pairs (in high and low dimensions) for the 0-dimensional persistent homology (PH^0) of the Rips filtration. We also provide a counter example showing that this property no longer holds for a naive extension of TopoAE to PH^d for $d \geq 1$. Based on this observation, we introduce a novel generalization of TopoAE to 1-dimensional persistent homology (PH^1), called TopoAE++, for the accurate generation of cycle-aware planar embeddings, addressing the above failure case. This generalization is based on the notion of *cascade distortion*, a new penalty term favoring an isometric embedding of the 2-chains filling persistent 1-cycles, hence resulting in more faithful geometrical reconstructions of the 1-cycles in the plane. We further introduce a novel, fast algorithm for the exact computation of PH for Rips filtrations in the plane, yielding improved runtimes over previously documented topology-aware methods. Our method also achieves a better balance between the topological accuracy, as measured by the Wasserstein distance, and the visual preservation of the cycles in low dimensions. The full paper is available at <https://arxiv.org/abs/2502.20215>.



1 INTRODUCTION

La réduction de dimension est un procédé courant, motivé par la sélection de dimensions significatives dans les données (pour éliminer les dimensions peu informatives et ainsi accélérer les post-traitements) ou par la visualisation (en réduisant la dimensionnalité à celle d'un espace visuel, c'est-à-dire en 2D ou 3D).

Alors que la réduction de dimension est souvent contrainte par des critères géométriques (par exemple la minimisation de distorsion), d'autres critères, notamment topologiques, peuvent s'avérer plus pertinents dans certaines applications : pour des applications de classification non-supervisée, il peut être opportun de préserver autant que possible le nombre de composantes connexes ; dans d'autres applications, il peut être pertinent de préserver les cycles pour caractériser et visualiser d'éventuels phénomènes périodiques. Pour traiter ces cas de figure, il faut alors concevoir des méthodes de réduction de dimension capables d'intégrer des critères de nature topologique.

L'analyse topologique des données (TDA) [4] fournit des outils pour détecter et représenter les éventuelles caractéristiques topologiques apparaissant dans un jeu de données, de façon robuste. En particulier, l'utilisation de l'homologie persistante (PH) de

la filtration de Rips est bien établie pour les nuages de points. De façon intuitive, l'homologie persistante de dimension 0 (PH^0) capture la présence de *composantes connexes* dans le nuage de points, tandis que celle de dimension 1 (PH^1) capture la présence de *cycles*. L'homologie persistante est souvent résumée de façon concise sous la forme de *diagrammes de persistance*. La distance de Wasserstein entre ces diagrammes permet de mesurer à quel point deux jeux de données présentent des caractéristiques topologiques similaires.

Récemment, des travaux [2], [3], [7] ont cherché à incorporer des contraintes sur la préservation de l'homologie persistante au sein de méthodes de réduction de dimension. En particulier, *Topological Autoencoders* [7] optimise un auto-encodeur – où la sortie de la couche centrale donne la projection en basse dimension – avec un terme de régularisation topologique ajouté au terme habituel de reconstruction. Ce terme de régularisation contraint les arêtes *critiques* en haute et basse dimension (c'est-à-dire qui créent ou détruisent une caractéristique topologique) de façon à ce qu'elle aient une longueur la plus proche possible en haute et basse dimension. Cependant, l'article original n'étudie expérimentalement la méthode que pour PH^0 , c'est-à-dire avec pour objectif de préserver les composantes connexes, et pas nécessairement les cycles. De plus, il est nécessaire de calculer à chaque itération l'homologie persistante de la représentation en basse dimension, ce qui peut être coûteux.

Dans cet article, nous cherchons à adapter de façon efficace l'approche de *Topological Autoencoders* à PH^1 , afin de préserver les cycles d'un jeu de données.

- *Mattéo Clémot*: Université Claude Bernard Lyon 1
E-mail: matteo.clemot@univ-lyon1.fr.
- *Julie Digne*: CNRS, Université Claude Bernard Lyon 1
E-mail: julie.digne@cnrs.fr.
- *Julien Tierny*: CNRS, Sorbonne Université
E-mail: julien.tierny@sorbonne-universite.fr.

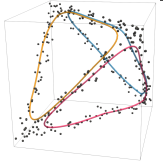
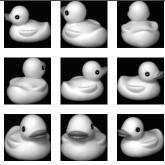
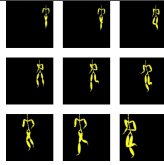


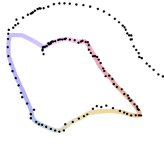
	(a)	(b)	(c)
Entrée (\mathbb{R}^h)	 $n = 300$ $h = 3$	 $n = 72$ $h = 4096$	 $n = 138$ $h = 62$
Projection (\mathbb{R}^2)			

Fig. 1. Projection dans \mathbb{R}^2 de trois jeux de données avec notre approche. n indique le nombre de points et h la dimension de l'espace de l'entrée. (a) Points échantillonnés autour des arêtes d'un tétraèdre régulier dans \mathbb{R}^3 . Des générateurs pour chacune des trois paires significativement persistantes sont représentés en haute et basse dimension en jaune, rouge et bleu. (b) Première classe du jeu d'image COIL-20 [8] qui consiste en un canard en plastique vu depuis 72 angles différents. (c) Capture de mouvement (mesures d'angles dans un squelette au cours du temps) d'un humain en train de courir [1]. Le caractère périodique du mouvement crée un cycle dans le jeu de données.

2 CONTRIBUTIONS

2.1 Analyse théorique

Nous commençons par étudier le terme de régularisation topologique proposé dans *Topological Autoencoders* [7], en montrant que sa formulation pour PH^0 donne une borne supérieure pour la distance de Wasserstein entre les diagrammes de persistance pour PH^0 en haute et basse dimension. En revanche, cela n'est plus vrai pour PH^k où $k \geq 1$. Nous exhibons un contre-exemple où les diagrammes de persistance pour PH^1 sont distincts en basse et haute dimension mais où ce terme est nul. En pratique, cela peut se traduire par la non-convergence du diagramme de persistance de la représentation vers celui de l'entrée malgré la convergence du terme de régularisation.

2.2 Distorsion de cascade

Pour remédier à ce problème, nous introduisons un terme de régularisation alternatif, appelé *distorsion de cascade*, qui coïncide avec celui de *Topological Autoencoders* pour PH^0 mais est plus restrictif pour PH^k où $k \geq 1$. Pour $k = 1$, ce terme consiste à contraindre non seulement les arêtes critiques, mais également les arêtes de 2-chaînes "remplissant" les 1-cycles, dites *arêtes de cascade*, qui jouent un rôle important lors du calcul de l'homologie persistante [10].

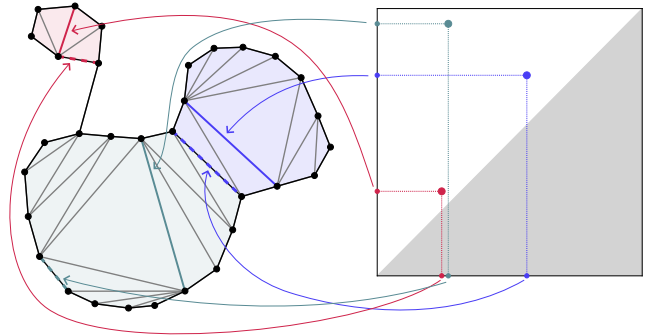


Fig. 2. Nous introduisons un algorithme rapide pour calculer l'homologie persistante de la filtration de Rips (et le diagramme de persistance associé, à droite) de nuages de points dans \mathbb{R}^2 en exploitant ses liens avec le *graphe de voisinage relatif* (à gauche, en noir) et les *triangulations minmax* (à gauche, en gris).

Nous proposons une étude expérimentale (Fig. 1) de notre approche sur des exemples synthétiques, acquis ou simulé présentant des cycles. Qualitativement, elle permet de projeter fidèlement ces jeux de données, dans le sens où la projection de ces cycles ne s'auto-intersecte pas. De plus, la *forme* globale des cycles, notamment leur élongation, est préservée. Plus quantitativement, la méthode permet d'obtenir une faible distance de Wasserstein entre les diagrammes de persistance de l'entrée et de sa projection.

2.3 Algorithme rapide

Pendant l'optimisation de l'auto-encodeur, le calcul de l'homologie persistante de la projection est effectuée à chaque étape de la descente de gradient, ce qui en fait une étape critique en termes de performances. Pour l'accélérer, nous introduisons un nouvel algorithme de calcul d'homologie persistante de la filtration de Rips d'un nuage de points dans \mathbb{R}^2 . Cet algorithme est uniquement géométrique et n'implique pas de réduction de matrice, à l'inverse des algorithmes de l'état de l'art. Il repose notamment sur le *graphe de voisinage relatif* [9] qui est connu pour contenir les arêtes faisant apparaître des cycles [6]. Nous montrons qu'il est possible d'utiliser les *triangulations minmax* [5] pour trouver les arêtes fermant ces cycles (Fig. 2), permettant de terminer le calcul des paires persistantes par dualité. Cela donne un algorithme plus rapide et économe en mémoire que l'état de l'art pour des nuages de points dans \mathbb{R}^2 , et permet l'accélération de l'optimisation de l'auto-encodeur lorsque la projection se fait vers \mathbb{R}^2 .

2.4 Implémentation

Le code de cet article est disponible à des fins de répliquabilité sur le dépôt <https://github.com/MClemot/TopologicalAutoencodersPlusPlus> et est en cours d'intégration dans la bibliothèque TTK (<https://topology-tool-kit.github.io/>).

REFERENCES

- [1] Carnegie Mellon University – CMU Graphics Lab – Motion capture library. <http://mocap.cs.cmu.edu/>.
- [2] M. Carriere, F. Chazal, M. Glisse, Y. Ike, H. Kannan, and Y. Umeda. Optimizing persistent homology based functions. In *ICML*, 2021.
- [3] H. Doraiswamy, J. Tierny, P. J. Silva, L. G. Nonato, and C. Silva. TopoMap: A 0-dimensional homology preserving projection of high-dimensional data. *IEEE TVCG*, 2020.
- [4] H. Edelsbrunner and J. L. Harer. *Computational topology: an introduction*. AMS, 2010.
- [5] H. Edelsbrunner and T. S. Tan. A quadratic time algorithm for the minmax length triangulation. *SIAM Journal on Computing*, 1993.
- [6] M. A. Koyama, F. Memoli, V. Robins, and K. Turner. Faster computation of degree-1 persistent homology using the reduced Vietoris–Rips filtration. *arXiv preprint*, 2023.
- [7] M. Moor, M. Horn, B. Rieck, and K. Borgwardt. Topological Autoencoders. In *International conference on machine learning*. PMLR, 2020.
- [8] S. A. Nene, S. K. Nayar, and H. Murase. Columbia Object Image Library (COIL-20), 1996.
- [9] K. J. Supowit. The relative neighborhood graph, with an application to minimum spanning trees. *JACM*, 1983.
- [10] A. Zomorodian. Computational topology. In *Algorithms and theory of computation handbook: special topics and techniques*. 2 edition, 2010.