

Simplification parallèle des groupes d'homologie persistante

Julien Tierny

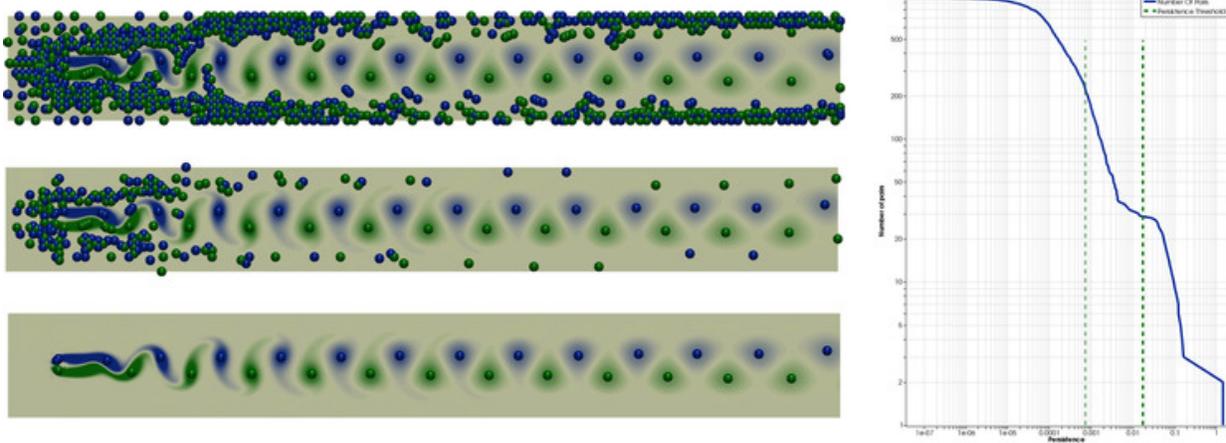


Fig. 1. Le sujet en une image – L'homologie persistante est un outil théorique puissant, qui permet en pratique d'introduire une mesure de bruit sur les structures topologiques, comme les singularités (sphères bleues et vertes) dans cette simulation 2D d'écoulement (le champ représenté du bleu vers le vert est la composante orthogonale du rotationnel du champ de vitesse derrière un obstacle, situé sur la gauche, exemple dit de l'allée de tourbillons de von Kármán [50]). Cette mesure de bruit permet de visualiser et de mesurer des structures topologiques à plusieurs échelles d'importance (du haut vers le bas), pour au final, n'extraire que les structures les plus proéminentes (en bas: centre des tourbillons). Une fois qu'un niveau de persistance satisfaisant a été sélectionné par l'utilisateur (typiquement, interactivement au vu de la distribution des singularités selon leur persistance, courbe à droite), il est nécessaire de **re-construire** une version simplifiée des données, n'admettant strictement que les singularités sélectionnées, avant d'effectuer toute autre analyse (segmentation, extraction de structures, etc.). Les algorithmes existants pour cette étape de re-construction sont intrinsèquement séquentiels et leur lenteur constitue aujourd'hui un goulot d'étranglement important en analyse topologique de données (jusqu'à 80% du temps de calcul). Dans ce stage, nous souhaitons répondre à ce problème en concevant une approche parallèle pour la simplification de groupes d'homologie persistante.

1 CONTEXTE

L'analyse topologique de données (TDA) [9, 12, 30, 35, 36] est une discipline à cheval entre informatique et mathématiques appliquées qui propose d'analyser des données complexes au vu de leur structure, de leur topologie [29]. Elle connaît un essor important depuis quelques années, dû principalement à ses succès récents en *data science* [30, 36] et en *machine learning* [8].

Parmi les différents outils d'analyse développés en TDA (comme le graphe de Reeb [19, 31, 41], le complexe de Morse-Smale [16, 22, 28], etc...), l'homologie persistante [12, 13] est un outil fondamental qui propose de mesurer l'importance des structures topologiques (composantes connexes, cycles, cavités, etc.) selon leur *durée de vie* (leur plage de valeur dans les données). Cette théorie permet d'introduire une mesure de bruit sur les structures topologiques, appelée *persistance*, dont la stabilité a été démontrée d'un point de vue théorique [11], et qui permet en pratique de distinguer avec efficacité et précision les structures d'intérêt du bruit (voir Fig. 1). L'efficacité pratique de l'homologie persistante a été documentée dans de nombreuses applications, comme en imagerie médicale [3, 7], en biologie cellulaire [23], en mécanique des fluides [10, 18, 25, 37], en physique des matériaux [15, 24, 27], en combustion [5, 6, 21, 26], en chimie moléculaire [4, 17], en astrophysique [33, 34], en traitement de surfaces [38, 40, 42, 43] ou encore en monitoring de simulations numériques haute-performance [32].

2 PROBLÈME SCIENTIFIQUE

Les données considérées sont typiquement représentées sous la forme d'une fonction scalaire linéaire par morceaux $f: \mathcal{M} \rightarrow \mathbb{R}$, associant une valeur réelle à chaque sommet d'une triangulation \mathcal{M} , qui représente un objet géométrique 2D ou 3D. En pratique, f représente dans les applications des niveaux de concentrations [15], des potentiels [17], des intensités [3], des températures [6], etc. Les sous-ensembles de niveau $f_{-\infty}^{-1}(i)$ sont définis comme la pré-image de l'intervalle ouvert $]-\infty, i[$ sur \mathcal{M} . Simplement, il s'agit de l'ensemble des points de l'objet au dessous d'une certaine valeur i . Quand i augmente, $f_{-\infty}^{-1}(i)$ change de topologie en un nombre fini de configurations: ses nombres de Betti [44] (nombres de composantes connexes, de cycles indépendants, de cavités, etc...) changent sur des points singuliers, appelés points critiques (sphères bleues et vertes, Fig. 1). Chaque structure topologique de $f_{-\infty}^{-1}(i)$ est donc créée sur un premier point critique à une valeur i , puis détruite sur un second point critique à une valeur $j > i$. Le diagramme de persistance $\mathcal{D}(f)$ [11, 13] (Fig. 2) est une représentation graphique de ce processus, où chaque classe d'homologie persistante (chaque structure topologique) est représentée par une barre verticale pour laquelle la coordonnée en abscisse correspond à la valeur i et les extrémités en ordonnées correspondent à i et j . La *persistance* de la classe est donnée par $|j - i|$. Dans ce diagramme, le bruit topologique apparaît donc sous la forme de petites barres, proche de la diagonale ($|j - i| \rightarrow 0$), voir [35].

Comme illustré Fig. 1, pour analyser les données en entrée à plusieurs échelles d'importance, il est nécessaire de pré-simplifier les données [39], c'est à dire de calculer une fonction $g: \mathcal{M} \rightarrow \mathbb{R}$, telle que (i) $\mathcal{D}(g) = \mathcal{C} \subset \mathcal{D}(f)$ et que (ii) $\|f - g\|_{\infty}$ soit aussi petite que

• Julien Tierny is with Sorbonne Université, CNRS, LIP6 UMR 7606, France. E-mails: julien.tierny@lip6.fr.

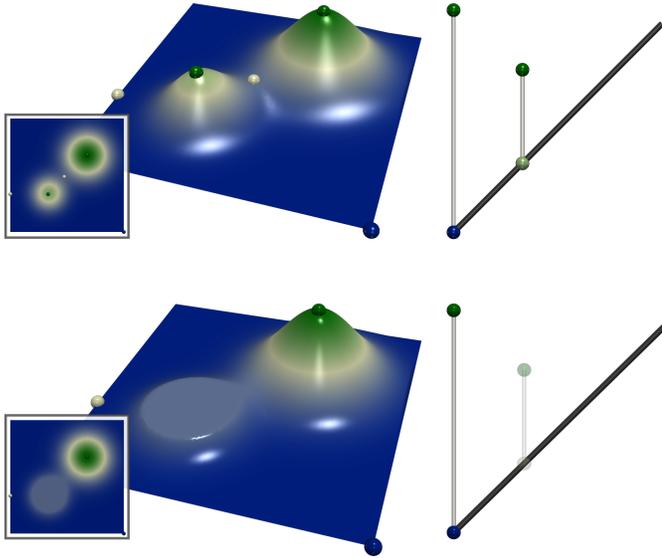


Fig. 2. Pré-simplification topologique d'une fonction d'élévation $f: \mathcal{M} \rightarrow \mathbb{R}$ (haut, représentée en 2D et avec un terrain 3D) en une fonction $g: \mathcal{M} \rightarrow \mathbb{R}$ (bas), admettant un diagramme de persistance $\mathcal{D}(g)$ simplifié.

possible, où \mathcal{C} correspond à l'ensemble des points critiques de f dont la persistance est supérieure à un seuil fixé par l'utilisateur, comme illustré Fig. 2. La contrainte (i) peut être vue comme un objectif de régularisation topologique, tandis que la contrainte (ii) peut être vue comme une contrainte d'attache aux données initiales. Cette étape de pré-simplification est souvent appelée étape de *re-construction*.

Il existe plusieurs algorithmes pour la re-construction de fonctions sous contraintes topologiques [1, 2, 14, 40, 42]. Cependant, tous ces algorithmes sont intrinsèquement séquentiels et leur lenteur constitue un réel goulot d'étranglement en pratique: jusqu'à 80% du temps de calcul sur des cas concrets.

L'objectif de ce stage est donc de définir un nouvel algorithme pour la pré-simplification topologique en **parallèle**, afin de rendre les temps de calcul de cette étape acceptables en pratique. Pour cela, il est nécessaire de revoir en profondeur l'intégralité de l'approche, puisque tous les algorithmes existants sont intrinsèquement séquentiels. Parmi les pistes possibles, nous envisagerons d'étudier et de nous inspirer d'une approche récente utilisant un parallélisme par tâche pour le calcul de diagrammes de persistance [20].

3 PERSPECTIVES

Ce stage est proposé dans l'optique d'une poursuite en thèse de doctorat sur le thème de l'analyse topologique de données haute performance. Ce sujet a été sélectionné comme thème possible pour une thèse CIFRE (co-encadrée par Julien Tierny et Julien Jomier, démarrage prévu fin 2018) dans le cadre d'un partenariat entre l'UPMC et Kitware [48], une société majeure dans le logiciel open-source (CMake [46], CDash [45], VTK [51], ITK [47], ParaView [49]) et la *data science*.

De manière plus générale, ce stage et sa possible poursuite en thèse apporteront un bagage de compétences scientifiques et techniques pointues et recherchées dans le domaine de la *data science* et de l'analyse et de la visualisation interactive de données scientifiques (TDA, TTK [39], ParaView [49]). Il constitue donc une expérience fortement valorisable pour accéder à des fonctions R&D sur ces thèmes, dans le monde académique comme industriel (Kitware, EDF, Total, CEA, etc.).

4 ORGANISATION DU STAGE

Le stage pourra se dérouler selon les étapes suivantes:

1. Etudier la bibliographie existante sur:
 - l'analyse topologique de données [12, 35];

- la pré-simplification de données [1, 2, 14, 40, 42];
- le parallélisme par tâche pour l'analyse topologique de données [20].

2. Imaginer et mettre en oeuvre un algorithme parallèle de pré-simplification de données;
3. Valider l'approche d'un point de vue expérimental sur une variété de jeux de données pratiques provenant de divers contextes applicatifs.

Les programmes d'expérimentation seront écrits en C++, sous la forme de modules pour la plate-forme open-source d'analyse topologique de données "Topology ToolKit" (TTK) [39] (intégrée à ParaView [49]).

Le stage peut durer de 16 à 24 semaines, selon les disponibilités du stagiaire. Il s'agit d'un stage rémunéré (rémunération académique standard, environ 500 euros par mois).

5 PROFIL

Nous recherchons un(e) étudiant(e) très motivé(e)! Curiosité, ouverture d'esprit, créativité, et ténacité sont les aptitudes de caractère que nous recherchons. Ce stage s'adresse aux étudiants en dernière année de master en informatique ou mathématiques appliquées (et domaines connexes) ou aux étudiants en dernière année d'école d'ingénieurs. Le stagiaire devra être à l'aise avec la programmation en C++, ou motivé pour le devenir. Un intérêt pour la 3D, la géométrie, la topologie et plus généralement pour les mathématiques et l'informatique est requis.

6 LIEU

Ce stage aura lieu au sein du département Calcul Scientifique du laboratoire d'informatique (LIP6) de Sorbonne Université, en plein coeur de Paris (arrêt Jussieu, lignes 7 et 10). Il sera encadré par Julien Tierny, chercheur au CNRS, expert en analyse topologique de données pour la visualisation et l'analyse de données scientifiques (<http://lip6.fr/Julien.Tierny>).

7 CANDIDATURES

Nous invitons les candidat(e)s à nous faire parvenir leur lettre de candidature accompagné d'un CV mis à jour à Julien Tierny (julien.tierny@lip6.fr). Nous vous encourageons à nous contacter par email pour toute question ou pour discuter davantage du sujet.

REFERENCES

- [1] D. Attali, M. Glisse, S. Hornus, F. Lazarus, and D. Morozov. Persistence-sensitive simplification of functions on surfaces in linear time. In *TopoInVis Workshop*, 2009.
- [2] U. Bauer, C. Lange, and M. Wardetzky. Optimal topological simplification of discrete functions on surfaces. *Disc. Compu. Geom.*, 2012.
- [3] A. Bock, H. Doraiswamy, A. Summers, and C. Silva. Topoangler: Interactive topology-based extraction of fishes. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2017.
- [4] R. A. Boto, J. C. Garcia, J. Tierny, and J.-P. Piquemal. Interpretation of the reduced density gradient. *Molecular Physics*, 2016.
- [5] P. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day, and J. Bell. A topological framework for the interactive exploration of large scale turbulent combustion. In *Proc. of IEEE eScience*, 2009.
- [6] P. Bremer, G. Weber, J. Tierny, V. Pascucci, M. Day, and J. Bell. Interactive exploration and analysis of large scale simulations using topology-based data segmentation. *IEEE Transactions on Visualization and Computer Graphics*, 2011.
- [7] H. Carr, J. Snoeyink, and M. van de Panne. Simplifying flexible isosurfaces using local geometric measures. In *IEEE VIS*, 2004.
- [8] F. Chazal, L. Guibas, S. Oudot, and P. Skraba. Persistence-based clustering in Riemannian manifolds. *Journal of the ACM*, 2013.
- [9] F. Chazal and J. Tierny. Topological data analysis, online class. <http://lip6.fr/Julien.Tierny/topologicalDataAnalysisClass.html>.
- [10] F. Chen, H. Obermaier, H. Hagen, B. Hamann, J. Tierny, and V. Pascucci. Topology analysis of time-dependent multi-fluid data using the reeb graph. *Computer Aided Geometric Design*, 2013.

- [11] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. In *Symp. on Comp. Geom.*, 2005.
- [12] H. Edelsbrunner and J. Harer. *Computational Topology: An Introduction*. American Mathematical Society, 2009.
- [13] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Disc. Comput. Geom.*, 2002.
- [14] H. Edelsbrunner, D. Morozov, and V. Pascucci. Persistence-sensitive simplification of functions on 2-manifolds. In *SoCG*, 2006.
- [15] G. Favelier, C. Gueunet, and J. Tierny. Visualizing ensembles of viscous fingers. In *IEEE SciVis Contest*, 2016.
- [16] R. Forman. A user's guide to discrete Morse theory. *Adv. in Math.*, 1998.
- [17] D. Guenther, R. Alvarez-Boto, J. Contreras-Garcia, J.-P. Piquemal, and J. Tierny. Characterizing molecular interactions in chemical systems. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2014.
- [18] D. Guenther, J. Salmon, and J. Tierny. Mandatory critical points of 2D uncertain scalar fields. *Computer Graphics Forum (Proc. of EuroVis)*, 2014.
- [19] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Contour forests: Fast multi-threaded augmented contour trees. In *IEEE LDAV*, 2016.
- [20] C. Gueunet, P. Fortin, J. Jomier, and J. Tierny. Task-based Augmented Merge Trees with Fibonacci Heaps. In *IEEE LDAV*, 2017.
- [21] A. Gyulassy, P. Bremer, R. Grout, H. Kolla, J. Chen, and V. Pascucci. Stability of dissipation elements: A case study in combustion. *Computer Graphics Forum (Proc. of EuroVis)*, 2014.
- [22] A. Gyulassy, P. T. Bremer, B. Hamann, and V. Pascucci. A practical approach to morse-smale complex computation: Scalability and generality. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2008.
- [23] A. Gyulassy, D. Guenther, J. A. Levine, J. Tierny, and V. Pascucci. Conforming morse-smale complexes. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2014.
- [24] A. Gyulassy, A. Knoll, K. Lau, B. Wang, P. Bremer, M. Papka, L. A. Curtiss, and V. Pascucci. Interstitial and interlayer ion diffusion geometry extraction in graphitic nanosphere battery materials. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2015.
- [25] J. Kasten, J. Reininghaus, I. Hotz, and H. Hege. Two-dimensional time-dependent vortex regions based on the acceleration magnitude. *IEEE Transactions on Visualization and Computer Graphics*, 2011.
- [26] D. E. Laney, P. Bremer, A. Mascarenhas, P. Miller, and V. Pascucci. Understanding the structure of the turbulent mixing layer in hydrodynamic instabilities. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2006.
- [27] J. Lukaszcyk, G. Aldrich, M. Steptoe, G. Favelier, C. Gueunet, J. Tierny, R. Maciejewski, B. Hamann, and H. Leitte. Viscous fingering: A topological visual analytic approach. In *Physical Modeling for Virtual Manufacturing Systems and Processes*, 2017.
- [28] J. Milnor. *Morse Theory*. Princeton U. Press, 1963.
- [29] MyWhyU. A humorous look at the topology of curved space. <https://www.youtube.com/watch?v=p2ofJPh2yMw&list=PL09E9E697F585A58C>.
- [30] V. Pascucci, X. Tricoche, H. Hagen, and J. Tierny. *Topological Methods in Data Analysis and Visualization: Theory, Algorithms and Applications*. Springer, 2010.
- [31] G. Reeb. Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *Acad. des Sci.*, 1946.
- [32] E. Santos, J. Tierny, A. Khan, B. Grimm, L. Lins, J. Freire, V. Pascucci, C. Silva, S. Klasky, R. Barreto, and N. Podhorszki. Enabling advanced visualization tools in a web-based simulation monitoring system. In *Proc. of IEEE eScience*, 2009.
- [33] N. Shivashankar, P. Pranav, V. Natarajan, R. van de Weygaert, E. P. Bos, and S. Rieder. Felix: A topology based framework for visual exploration of cosmic filaments. *IEEE Transactions on Visualization and Computer Graphics*, 2016. <http://vgl.serc.iisc.ernet.in/felix/index.html>.
- [34] T. Sousbie. The persistent cosmic web and its filamentary structure: Theory and implementations. *Royal Astronomical Society*, 2011. <http://www2.iap.fr/users/sousbie/web/html/indexd41d.html>.
- [35] J. Tierny. Introduction to topological data analysis. <https://hal.archives-ouvertes.fr/ce1-01581941/file/manuscript.pdf>.
- [36] J. Tierny. *Contributions to Topological Data Analysis for Scientific Visualization*. Habilitation (HDR), Sorbonne University UPMC, 2016.
- [37] J. Tierny and H. Carr. Jacobi fiber surfaces for bivariate Reeb space computation. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2016.
- [38] J. Tierny, J. Daniels, L. G. Nonato, V. Pascucci, and C. Silva. Interactive quadrangulation with Reeb atlases and connectivity textures. *IEEE Transactions on Visualization and Computer Graphics*, 2012.
- [39] J. Tierny, G. Favelier, J. A. Levine, C. Gueunet, and M. Michaux. The Topology ToolKit. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2017. <https://topology-tool-kit.github.io/>.
- [40] J. Tierny, D. Guenther, and V. Pascucci. Optimal general simplification of scalar fields on surfaces. In *Topological and Statistical Methods for Complex Data*. Springer, 2014.
- [41] J. Tierny, A. Gyulassy, E. Simon, and V. Pascucci. Loop surgery for volumetric meshes: Reeb graphs reduced to contour trees. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2009.
- [42] J. Tierny and V. Pascucci. Generalized topological simplification of scalar fields on surfaces. *IEEE Transactions on Visualization and Computer Graphics (Proc. of IEEE VIS)*, 2012.
- [43] A. Vintescu, F. Dupont, G. Lavoué, P. Memari, and J. Tierny. Conformal factor persistence for fast hierarchical cone extraction. In *Eurographics (short papers)*, 2017.
- [44] Wikipedia. Betti numbers. https://en.wikipedia.org/wiki/Betti_number.
- [45] Wikipedia. CDash. <https://www.cdash.org/>.
- [46] Wikipedia. CMake. <https://en.wikipedia.org/wiki/CMake>.
- [47] Wikipedia. ITK. https://en.wikipedia.org/wiki/Insight_Segmentation_and_Registration_Toolkit.
- [48] Wikipedia. Kitware Inc. <https://en.wikipedia.org/wiki/Kitware>.
- [49] Wikipedia. ParaView. <https://en.wikipedia.org/wiki/ParaView>.
- [50] Wikipedia. von Kármán vortex street. https://en.wikipedia.org/wiki/Karman_vortex_street.
- [51] Wikipedia. VTK. <https://en.wikipedia.org/wiki/VTK>.