Programme de groupe de travail

Homologie persistante en analyse topologique de données : théorie et applications

Vadim Lebovici

Exposés basés sur le livre **Persistence Theory: From Quiver Representations to Data Analysis** de Steve Oudot

Septembre 2020-Janvier 2021

Résumé

L'objectif de ce groupe de travail sera de découvrir la théorie et les applications de l'homologie persistante, un outil phare de l'analyse topologique de données, destiné à extraire de l'information topologique d'un nuage fini de points.

D'un point de vue théorique, on présentera les modules de persistance, leur *code-barres*, i.e les résultats de décomposition des modules de persistance (ce sera l'occasion de découvrir la théorie des représentations de carquois) et les résultats de stabilité qui y sont liés. Un module de persistance est obtenu lorsque l'on considère les groupes d'homologie d'une famille paramétrée croissante d'espaces topologiques.

D'un point de vue pratique, on abordera d'abord les aspects algorithmiques liés à la construction et aux calculs des codes-barres, puis on étudiera l'apport de l'homologie persistante dans trois domaines : l'inférence topologique (comprendre la structure d'un objet topologique inconnu à partir la connaissance d'échantillons finis de points sur cet objet), le partitionnement (clustering), et les signatures des espaces métriques (comparaison quantitative entre jeux de données). Au cours du groupe de travail, des applications de ces outils dans le domaine scientifique, médical ou de l'entreprise seront présentées.

Si le temps le permet, on étudiera des résultats et des questions qui dépassent le cadre du livre, comme par exemple une généralisation de l'homologie persistante dont le besoin se fait sentir en pratique : l'homologie multi-persistante.

Modalités pratiques

— **Contact :** vadim.lebovici@ens.fr

— **Accès au livre :** Le livre est disponible en deux exemplaires à la bibliothèque et en pdf sur le site internet de Steve Oudot. Si vous empruntez un exemplaire à la bibliothèque, merci de le rendre après votre exposé pour que les autres puissent l'utiliser.

— **Préparation d'un exposé :** Je vous demande de venir me voir au moins une fois dans la semaine précédant l'exposé, et pas la veille de préférence, en ayant travaillé avant, pour parler de ce que vous traiterez et poser des questions si besoin. Si vous le souhaitez, vous pouvez faire l'exposé en entier devant moi.

— **Contenu des exposés :** Pour chaque chapitre, les • ci-dessous indiquent ce qui doit être traité dans l'exposé et les $\star$ indiquent ce qui peut être traité selon vos goûts personnels et le temps disponible. Il faut compter 1h15 par exposé pour laisser le temps pour des questions.

— **Support de l'exposé :** Vous pouvez utiliser le tableau ou préparer des slides. Il y aura très souvent des figures à montrer : vous pouvez combinez un exposé au tableau avec des images projetées lorsque c'est nécessaire. N'hésitez pas à prendre les figures dans le pdf du livre en ligne pour les projeter. **Pour bénéficier d'un projecteur, il faudra me prévenir avant pour que je le réserve.**

— **Critères d'évaluation :** Vous serez évalués sur trois critères : la compréhension et la présentation rigoureuse du contenu de votre exposé, le choix du contenu que vous mettez en valeur dans votre présentation, et enfin la qualité pédagogique de l'exposé.

Sur le contenu présenté

— Le livre donne une vue d'ensemble de l'analyse topologique de données, il y a donc peu de preuves à l'intérieur. Dans vos exposés, il faudra suivre cette ligne éditoriale : si certaines preuves vont nous intéresser, il faudra avant tout s'assurer que son auditoire comprenne la philosophie globale de ce que vous présentez, via l'explicitation des conséquences d'un théorème, des exemples concrets pour les algorithmes, des schémas, etc.

— Par ailleurs, l'analyse topologique de données étant à l'interface entre les mathématiques et l'informatique, il faudra bien mettre en avant les aspects applicatifs (complexité de calculs, optimisations algorithmiques possibles) puisqu'ils nourissent et motivent les résultats théoriques présentés.

— Si vous voulez traiter un chapitre, un article ou un résultat qui n'est pas prévu dans les lignes qui suivent, n'hésitez pas à me le demander, le programme est négociable !

Programme

0 Introduction (11 septembre, Vadim)

I Théorie

1 L'homologie dans une coquille de noix (18 septembre, David)

[Sec. 2.1 de Allen Hatcher, *Algebraic Topology*, Cambridge University Press, 2002]

- Δ-complexes
- Homologie simpliciale
- Homologie singulière
- Invariance par homotopie

⋆ Votre théorème préféré sur l'homologie

2 Partie algébrique de la persistance (29 septembre, Azélie)

[Chap. 1]

- Représentations de carquois, théorème de décompositions (Sec. 1)
- Modules de persistance et décomposition en intervalles (Sec. 2)
- Diagramme de persistance et codes-barres (Sec. 3)
- Extension aux modules indécomposables (Sec. 4)

⋆ Compléments et mise en perspective (Sec. 5)

3 Constructions topologiques et calculs (3 novembre, Vadim)

[Chap. 2]

- Filtrations topologiques (Sec. 1.1)
- Zigzags et filtration par les niveaux d'une fonction (Sec. 1.3)
- Calculs d'homologie persistante (Sec. 2.1)

- ⋆ Filtrations relative et persistance étendue (Sec. 1.2)
- ⋆ Noyaux, images et conoyaux (Sec. 1.4)
- ⋆ Optimisations possibles (Sec. 2.2)

4 Résultat de stabilité (10 novembre, Lydia)

[Chap. 3]

- Métriques (distance de Bottleneck et d'entrelacement) (Sec. 1)
- Preuve de la stabilité (inégalité sous-optimale) (Sec. 2.1)
- Fin de la preuve du théorème d'isométrie (Sec. 3)
- ⋆ Preuve de la stabilité (inégalité optimale) (Sec. 2.2)
- ⋆ Compléments (Sec. 4)

II Applications

5 Inférence topologique 1 (17 novembre, Alexis)

[Chap. 4]

- Inférence avec des fonctions distances (Sec. 1)
- Des décalages aux filtrations (Sec. 2)
- Des filtrations aux filtrations simpliciales (Sec. 3)

6 Inférence topologique 2 (20 novembre, David)

[Chap. 5]

- Enoncé du problème (début du Chap. 5)
- Des prédicats géométriques plus simples (Sec. 1)
- Filtrations de tailles linéaires (Sec. 2)
- ⋆ Prise en compte des points aberrants (Sec. 6)

7 Inférence topologique 3 (24 novembre, Lydia)

[Chap. 5]

- Prise en compte de la dimension intrinsèque des données (Sec. 3)
- Comparaison des différentes méthodes (Sec. 4)
- Applications aux images naturelles (Sec. 5)

8 Partitionnement (1 décembre, Azélie)

[Chap. 6]

- Contributions de la persistance (Sec. 1)
- L'algorithme ToMATo (Sec. 2)
- Garanties théoriques (Sec. 3)
- ⋆ Résultats expérimentaux (Sec. 4)
- ⋆ Structure de plus haute dimension (Sec. 5)

9 Signature pour les espaces métriques (8 décembre, Alexis)

[Chap. 7]

- Filtrations simpliciales pour les espaces métriques quelconques (Sec. 1)
- Stabilité pour les espaces métriques finis (Sec. 2)
- Stabilité pour les espaces métriques totalement bornés (Sec. 3, sauf 3.4)
- Calculs de signatures (Sec. 5.1)
- ⋆ Signatures pour les espaces métriques équippés d'une fonction (Sec. 4)

III Approfondissement

10 J. Perea, J. Harer - *Sliding Windows and Persistence: An Application of Topological Methods to Signal Analysis*, (15 décembre, Azélie)

- Définitions et motivation (Sec. 2)
- Théorème d'approximation (Sec. 4)
- Etude de l'influence de la taille du fenêtrage (Sec. 5.2)
- Application de l'homologie persistante (Sec. 6.1 & 6.2)
- Quantification de la périodicité (Sec. 7)

11 G. Carlsson, A. Zomorodian - *The Theory of Multidimensional Persistence*, (5 janvier, David)

- Introduction (Sec. 1)
- Correspondance modules multipersistants / modules n-gradués sur l'anneau des polynômes à n variables (Sec. 3) [1]
- Classification des classes d'isomorphismes (Sec. 4)

⋆ Invariant de rang (Sec. 6)

12 P. Bubenik - *Statistical Topological Data Analysis using Persistence Landscapes*, (12 janvier, Alexis)

13 V. Divol - *Minimax adaptive estimation in manifold inference*, (19 janvier, Lydia)

- Introduction (Sec. 1)
- Préliminaires (Sec. 2)
- Estimation minimax de variétés avec des enveloppes t-convexes (Sec. 3)
- Sélection des enveloppes t-convexes (Sec. 4)

⋆ Considérations numériques (Sec. 5)

1. Au besoin, regarder le cas de la dimension 1 d'abord : Thm. 2.1, Sec. 3.1, Sec. 3.2 de *Computing Persistent Homology*.