Seminar “Topological Data Analysis”

Olivier Haution

LMU, Sommersemester 2019

Introduction

This seminar will be an introduction to topological data analysis, a branch of computational topology. Good presentations of this field may be found in the surveys of Carlsson [Car09] and Ghrist [Ghr08]. Basics of topology, algebra and geometry will be expected. However no previous knowledge of algebraic topology will be required, and this seminar may constitute an introduction to algebraic topology. The last talks may consists in presentations of recent research papers. The program will be the following:

— Simplicial complexes
— Complexes associated with point clouds
— Homology
— Discrete Morse theory
— Persistence
— Applications

Remarks

- The participants are supposed to attend all talks.
- Each talk should last 60 to 75 minutes, and normally be conducted in English.
- It can be either a Beamer or board talk, or a mixture of both.
- Please practice your talk before giving it, at least in order to make sure that it fits into the specified time frame.
- I expect each participant to meet with me at least once, and at least a week before the talk (for instance this could happen in the seminar room, just after the previous talk).

- It is not necessary to hand out a written version of the talk, but this is possible if you find it helpful.
- An indicative plan of the main ideas covered in the talk is provided below. Deviation from this may be possible, but should be discussed with me beforehand. Of course alternative references may be used.

Talks

1 Simplicial complexes

[EH10, III.1] : Simplicial complexes, geometric realisation, barycentric subdivision, mesh lemma, simplicial approximation Theorem.

Additional reference: [Mun84, Chapter 1].

2 Cech and Rips complexes

[EH10, III.2]: Helly's Theorem, homotopy type, nerve Theorem, Čech complex, Vietoris-Rips complex.

Remark: A proof of Vietoris-Rips Lemma can be found in [dSG07, Theorem 2.5], but there is probably not enough time to explain it.

3 Delaunay and Alpha complexes

[EH10, III.3, III.4]: Inversion, stereographic projection, Voronoi diagram, Delaunay triangulation, alpha complex, filtration, collapses preserve the homotopy type.

Remark : If there is time, witness complexes may be mentioned [Car09, Definitions 2.7, 2.8].

4 Simplicial homology

[EH10, IV.1, IV.2, IV.3, IV.4]: Homology, reduced homology, morphisms, Betti numbers, Euler-Poincaré formula, relative homology, excision, exact sequence of a pair, Mayer-Vietoris sequence. Skip Brouwer's fixed point theorem and most of IV.2 (except the paragraph on the Euler-Poincaré formula).

Additional references: [Hat02, §2.1], [Mun84, Chapter 1]

5 Persistent homology

[EH10, VII.1 until p.152]: Persistent Betti numbers, persistent diagram.
[Car09, §2.3] (overview) and [ZC05, §3] (details): Interpretation in terms of $F[t]$-modules, classification theorem for graded modules over a PID, barcodes.

Remark: This talk may require a more algebraic background.

6 Algorithms computing homology

[EH10, IV.2] and [ZC05, §2.5,§4]: Reduction algorithm, Smith normal form, algorithm computing persistent homology.

7 PL-Morse functions, Reeb graphs

[EH10, VI.3,VI.4]: Lower star filtration, critical points, Euler characteristic, PL Morse inequalities, Reeb graph, loops in Reeb graphs (time permitting), algorithmic construction.

Remark: The usual Morse theory (see [EH10, VI.1,VI.2]) may be mentioned for comparison purposes, but the focus should be on the discrete version. A categorical approach to the Reeb graph is described in [dSMP16].

8 Stability

[EH10, VIII.2]: Bottleneck distance, vineyards, stability Theorem for tame functions.

Remark: A proof of the stability Theorem can be found in [CSEH05]. A detailed explanation of the proof is of course not expected, but some elements of that paper may be incorporated in the talk if there is time.

9 Mapper

[Car09, §3] (overview) and [SMC07] (details): A method to visualise data sets and its implementation.

10 Optional: Identifying a type of breast cancers

[NLC11]: One of the most famous application of topological data analysis.

11 Optional: Natural images

[Car09, §2.4], [Ghr08, §3], [CIdSZ08], [LPM03]: The analysis of a large collection of natural images reveal a bias towards vertical and horizontal directions; a Klein bottle appears in the data set.

12 Optional: An Impossibility Theorem for Clustering

[Kle02]: A proof that no clustering function satisfies simultaneously scale-invariance, richness and consistency.

13 Optional: Population activity in visual cortex

[SMI+08]: An illustration of "how computational topology can help tackle elementary questions about the representation of information in the nervous system".

14 Optional: Branching neuronal morphologies

[KDS+18]: Analysis of neuronal morphologies using persistence and barcodes.

References

[Car09] Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308, 2009.

[CIdSZ08] Gunnar Carlsson, Tigran Ishkhanov, Vin de Silva, and Afra Zomorodian. On the local behavior of spaces of natural images. *Int. J. Comput. Vis.*, 76(1):1–12, 2008.

[CSEH05] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *Proceedings of the Twenty-first Annual Symposium on Computational Geometry*, SCG '05, pages 263–271, New York, NY, USA, 2005. ACM.

[dSG07] Vin de Silva and Robert Ghrist. Coverage in sensor networks via persistent homology. *Algebr. Geom. Topol.*, 7:339–358, 2007.

[dSMP16] Vin de Silva, Elizabeth Munch, and Amit Patel. Categorified Reeb graphs. *Discrete Comput. Geom.*, 55(4):854–906, 2016.

[EH10] Herbert Edelsbrunner and John L. Harer. *Computational topology. An Introduction.* American Mathematical Society, Providence, RI, 2010.

[Ghr08] Robert Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N.S.)*, 45(1):61–75, 2008.

[Hat02] Allen Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

[KDS+18] Lida Kanari, Paweł Dłotko, Martina Scolamiero, Ran Levi, Julian Shillcock, Kathryn Hess, and Henry Markram. A topological representation of branching neuronal morphologies. *Neuroinformatics*, 16(1):3–13, Jan 2018.

[Kle02] Jon Kleinberg. An impossibility theorem for clustering. In *Proceedings of the 15th International Conference on Neural Information Processing Systems*, NIPS'02, pages 463–470, Cambridge, MA, USA, 2002. MIT Press.

[LPM03] Ann B. Lee, Kim S. Pedersen, and David Mumford. The nonlinear statistics of high-contrast patches in natural images. *Int. J. Comput. Vision*, 54(1-3):83–103, August 2003.

[Mun84] James R. Munkres. *Elements of algebraic topology*. Addison-Wesley Publishing Company, Menlo Park, CA, 1984.

[NLC11] Monica Nicolau, Arnold J. Levine, and Gunnar Carlsson. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. *Proceedings of the National Academy of Sciences*, 108(17):7265–7270, 2011.

[SMC07] Gurjeet Singh, Facundo Memoli, and Gunnar Carlsson. Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. In M. Botsch, R. Pajarola, B. Chen, and M. Zwicker, editors, *Eurographics Symposium on Point-Based Graphics*. The Eurographics Association, 2007.

[SMI+08] Gurjeet Singh, Facundo Memoli, Tigran Ishkhanov, Guillermo Sapiro, Gunnar Carlsson, and Dario L. Ringach. Topological analysis of population activity in visual cortex. *Journal of Vision*, 8(8):11–11, 06 2008.

[ZC05] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.