Seminar: Topological date analysis

- Responsible: Thomas Schick
- Subject: **Algebraic topology and date analysis**
- Area: Algebraic topology, with relation to applied mathematics
- Target audience: students of math, data science, computer science from semester 4
- Time: Tue 14:15-15:55, Sitzungszimmer (?)
- **Orga-meeting, Friday 7.3., 13:15 in Sitzungssaal**
- Kontakt: thomas.schick@math.uni-goettingen.de, Tel. 39-27799

Science today is marked by the collection of huge sets of data. The bottleneck is more and more the evaluation of this data.

In particular, one has to retrieve the decisive qualitative information in an efficient way (typically from noisy and high dimensional data which is presented in inappropriate coordinates).

Topology is (from the point of view of geometry in pure mathematics) the area which does precisely such a kind of job. In the last decades there has now also been a very active development to implement this in practical terms.

Example questions: given a scan of living tissue on the scale of cells: distinguish the different components (the membranes), detect connections (in particular their time evolution), but do this with noisy data and suppress irrelevant artefacts.

One suggestion to develop and apply algebraic topology to solve these problems will be the theme of the seminar, where we try to get all the way to some more or less real applications. The larger part, will be dedicated to develop the algebraic topology and geometry basics.

More precisely, our tool are homology groups (there is one for each integer n). Very roughly, these count n-dimensional holes in a topological space. E.g., an n-dimensinal sphere has precisely one n-dimensional hole, whereas a disk (of any dimension) has no hole at all (trivial homology).

To apply this to point sets (this is, what data measurement will produce), we construct a sequence of interesting topological spaces from this point set and apply homology to each of the spaces in the sequence. The *persistent homology* (our main tool) focuses on those homology features which persist for a long time in the sequence of spaces.

The seminar does start with a quick intro into the relevant aspects of topology, focusing then on the aspects which are relevant for topological data analysis. This way, it is suitable for the students which have taken a course in algebraic topology (but can't obtain credit if covering a talk introducing topological material they have already learned), but also for newcomers.

In later parts, will then see how the theory is used in practice, and learn about some fundamental theoretical features of persistent homology.

Examples for applications:

- given a set of points in R^3 wich represent (centers of) atoms of a large molecule, determine tunnels and cavities in the molecule
- determine the placement of sensors in sensor networks

The most significant topological basics consist of

- knowing and applying simplicial complexes
- know homology (of simplicial complexes), compute it and interpret the information
- understand homology as a functor
- knowing advanced computation tools for homology (from homological algebra): exact sequences

Specific for topological data analysis are

- skilled construction of simplicial complexes from data, e.g. Rips complex, Delauny-triangulation,. . . and comparison of these
- persisten homology and bar codes
- efficient algorithms to compute homology
- Flow complexes (and Morse theory)

Program

Nr	Thema	Quelle	Name	Termin
1	simplicial complexes	M, (EH, Z)		
2	(simplicial) homology: intro	M, H (EH, Z)		
3	(simpicial) homology II	EH, (M, Z)		
4	Simplicial complexes associated with point clouds	EH III.2-III.4		
5	persistent homology	EH VII.1, Z, 6.1		
6	Algebraic classification of persistent homology	ZC 3		
7	Algebraic classification of continuous persistence homology (advanced)	CB		
8	algorithms for (persistent) homology	EH IV.2, VII.1		
9	Fast algorithm for homology in $\mathbb{R}^3$	DH		
10	stability of persistent homology	EH VIII.2		
11	Künneth formulas in persistent homology (advanced!)	GP		
12	Application: pockets in proteins	IX.2, GJ1		
13	Application: coverage in sensor networks	SG		
14	A_∞-persistence (advanced)	BM, B		

1. Simplicial complexes ([13, Sections 1-3,14,16], also [8, III.1], [15, 2.3]

   Content: Definition of simplicial complexes and maps (abstract and geometric realization, triangulation, subdivision); formulation of simplicial approximation theorem and nerve theorem (probably at best with sketch of proof). Examples.

2. simplicial homology [13], [12, Theorem 2.44, p. 230] (also [8, IV.1,IV.2, III.2], [15, 4.2]

   Introduce algebraic chain complexes and their homology and then simplicial homology of a simplicial complex (briefly discuss role of coefficients). Introduce Betti numbers. Prove homotopy invariance and explain its significance. Examples

3. (simplicial) homology II (choice from [8, IV.3,IV.4], [13, 1,2]: introduce important slightly advanced properties/constructions of homology and computational tools. Relative homology, pair sequence, exact sequences, Mayer-Vietoris, more example computations.

4. simplicial complexes assiciated to point clouds [8, III.2-4].

   fundamental principle: understand a geometric object/data by choosing coverings and the combinatoris of intersection patterns.

Discuss classical constructions to assign to a discrete subset of $\mathbb{R}^n$ sensible thickenings.

Content of the presentation: coverings, nerve, nerve-theorem, Cech-complex, Rips-complex; spherical inversion, Voronoi-cells, Voronoi decomposition, Delaunay and α-complexes

5. persisten homology [8, VII.1], [15, 6.1 part]:

   Content: start with exmple of death and birth of connected components. State principle of older, define persistent homology and persistence-diagrams, give examples (in particular from [15]

6. Classification and main theorem of persistent homology [16, Section 3].

   Explain (with field coefficients) the classification of persistent homology/persistence modules (compare also [2, Theorem 1.2.4]. Discuss the problems with more general coefficients and in the context of multidimensional persistence [4].

7. Classification of continuous persistent cohomology [5]

   Persistent homology inspires delicate algebraic problems: how to classify the resulting modules? This comes up due to the variety of different contexts. For example, α-shapes and Rips complexes are indexed by the (positive) reals. Eminent algebraist have studied the problem. Discuss the result found in [5] and perhaps also in [1].

8. Algorithms for persistent homology [8, IV.2, VII.1], [15, 7.3], [13, 10-11]

   goal: describe efficiently usable algorithms to compute persistent homology.

   Cover: matrix reduction for homology and for persistent homology. Describe algorithmn, present example, analyse efficiently (potentially: discussion of sparse matrix techniques).

9. Fast algorithm for homology in $\mathbb{R}^3$ [8, V.4] and [6] .

   Describe algorithm to compute homology of sub-simplicial complexes of $\mathbb{R}^3$. Discuss efficiency. Briefly cover Alexander duality as the basic tool to reduce to degree 1. Desired extension: use for persistence as in [8, VII.2].

10. stability of persistence [8, VIII.2]

    Cover: bottleneck metric, stability in bottleneck metric (for tame functions); Wasserstein metric, stability for Wasserstein metric (reduced amount of detail. If time permits: cover aspects of computability, following [8, VIII.1]

11. Application: pockets in proteins [8, IX.2] [11]

    Content: Describe the problem, discuss the 1-dimensiona; “toy”-case from [8]; brief discussion of the Ansatz from [11]; including an introduction to the flow complex (which is a glimpse to Morse theory)

12. Application: Coverage in sensor networks [7].

    Explain the setup, main result, and prove of this result. The paper deals with the question how sensors have to be placed to cover a bounded domain if the topological type is unknown (assuming conditions on the regularity of the boundary).

13. A_∞-persistence (advanced). Follow [3] or [2, Chapter 2 and 3].

    Cover the additional structure one can give to (co)homology given by products and higher coherence structure products. Note that such structure can be used to distinguish non-homotopic spaces with the same homology. Give an idea of how persistence can be extend to A_∞-persistence.

Literatur

[1] Magnus Bakke Botnan and William Crawley-Boevey, *Decomposition of persistence modules*, Proc. Amer. Math. Soc. **148** (2020), no. 11, 4581–4596, DOI 10.1090/proc/14790. MR4143378

[2] Francisco Belchí, *A_∞-persistence*, PhD thesis, Universidad de Malaga, 2015.

[3] Francisco Belchí and Aniceto Murillo, *A_∞-persistence*, Appl. Algebra Engrg. Comm. Comput. **26** (2015), no. 1-2, 121–139, DOI 10.1007/s00200-014-0241-4. MR3320908

[4] Gunnar Carlsson and Afra Zomorodian, *The theory of multidimensional persistence*, Discrete Comput. Geom. **42** (2009), no. 1, 71–93, DOI 10.1007/s00454-009-9176-0. MR2506738

[5] William Crawley-Boevey, *Decomposition of pointwise finite-dimensional persistence modules*, J. Algebra Appl. **14** (2015), no. 5, 1550066, 8, DOI 10.1142/S0219498815500668. MR3323327

[6] Cecil Jose A. Delfinado and Herbert Edelsbrunner, *An incremental algorithm for Betti numbers of simplicial complexes on the* 3*-sphere*, Comput. Aided Geom. Design **12** (1995), no. 7, 771–784, DOI 10.1016/0167-8396(95)00016-Y. Grid generation, finite elements, and geometric design. MR1365107

[7] Vin de Silva and Robert Ghrist, *Coverage in sensor networks via persistent homology*, Algebr. Geom. Topol. **7** (2007), 339–358, DOI 10.2140/agt.2007.7.339. MR2308949

[8] Herbert Edelsbrunner and John L. Harer, *Computational topology*, American Mathematical Society, Providence, RI, 2010. An introduction. MR2572029

[9] U. Fugacci, S. Scaramuccia, F. Iuricich, and L. De Floriani, *Persistent homology: a step-by-step introduction for newcomers* (2016). STAG: Smart Tools and Apps in computer Graphics (2016) Giovanni Pintore and Filippo Stanco (Editors).

[10] Hitesh Gakhar and Jose A. Perea, *Künneth formulae in persistent homology*, 2019. arXiv:1910.05656.

[11] Joachim Giesen and Matthias John, *The flow complex: a data structure for geometric modeling*, Comput. Geom. **39** (2008), no. 3, 178–190, DOI 10.1016/j.comgeo.2007.01.002. MR2381026

[12] Allen Hatcher, *Algebraic topology*, Cambridge University Press, Cambridge, 2002. available at https://www.math.cornell. edu/ hatcher/AT/AT.pdf. MR1867354

[13] James R. Munkres, *Elements of algebraic topology*, Addison-Wesley Publishing Company, Menlo Park, CA, 1984. MR755006

[14] Patrick Schnider, *Introduction to Topological Data Analysis; Lecture Notes FS 2024*, 2024. Lecture notes, Informatics, ETH Zürich.

[15] Afra J. Zomorodian, *Topology for computing*, Cambridge Monographs on Applied and Computational Mathematics, vol. 16, Cambridge University Press, Cambridge, 2005. MR2111929

[16] Afra Zomorodian and Gunnar Carlsson, *Computing persistent homology*, Discrete Comput. Geom. **33** (2005), no. 2, 249–274, DOI 10.1007/s00454-004-1146-y. MR2121296