Topological Techniques for Working With Data

Michelle H. Feng, mhfeng@caltech.edu

Outline

Algebraic Topology

Applied Topology

Persistent Homology

- Simplicial Complexes (and other complexes)
- Persistent Homology

Algebraic Topology

A Brief History

- How do you reduce a topological space down into invariants?
- Used to be called combinatorial topology (Betti numbers, Euler characteristics)
- Emmy Noether: we can work with algebraic objects like groups instead
- These days, much of the field have gone down the categorical rabbit hole

What is it good for?

Key insight: sometimes, we care less about the precise shape of an object than the way it's put together. Using this, we can:

- classify topological spaces (up to homeomorphism)
- understand what types of vector fields can exist on a manifold
- understand fixed points and their existence

Why should we care? It's useful to think about how data is put together!

Applied Topology

Some applications

- Impossibility theorems
- Data structure
- Knot theory
- Circuit topology

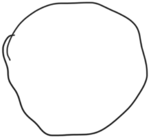
Data

- Connectedness
- “Shape”
- Patterns and cycles

Persistent Homology

Persistent Homology (PH)

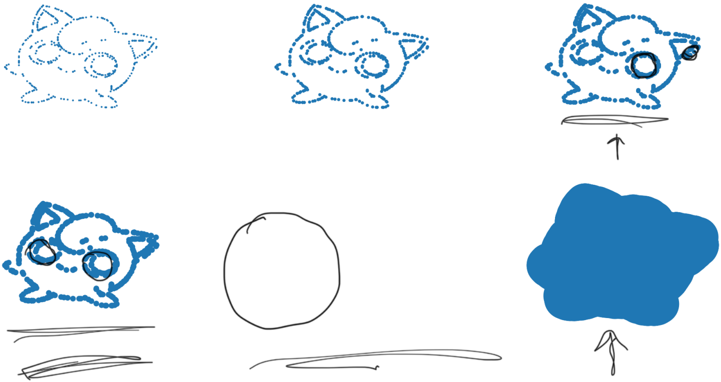
- One of the most popular tools in TDA
- Applies homology theory to point clouds (usually)
- Loosely, characterizes the shape of a data set at a variety of different scales



S_1

What is the shape of a point cloud?

What is the shape of a point cloud?



Why do we care about persistence?

Features that exist across a variety of scales are the "truest"

Basic workflow

Data set $\longrightarrow$ Topological object $\xrightarrow{\text{Homology computation}}$ Homology groups

Data set
- point cloud
- Graph/network
- Images

Topological object
- Simplicial complexes

Simplicial complexes

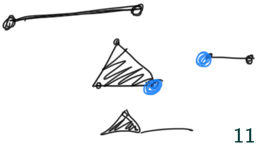
Definition

A *k-simplex* is a k-dimensional polytope which is the convex hull of its $k+1$ vertices.

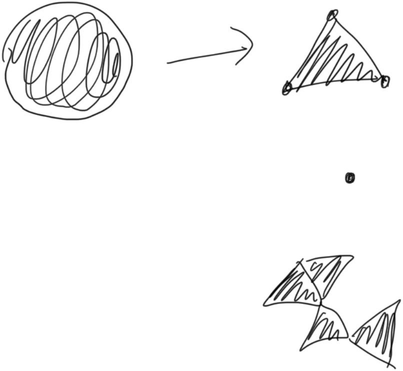
An *m-face* of a k-simplex is a subset of size $m+1$, and is an m-simplex itself.

A simplicial complex S is a set of simplices that satisfies the following conditions:

1. Every face of a simplex from S is also in S
2. The non-empty intersection of any two simplices $\sigma_1, \sigma_2 \in S$ is a face of both σ_1 and σ_2.



What do simplicial complexes look like?

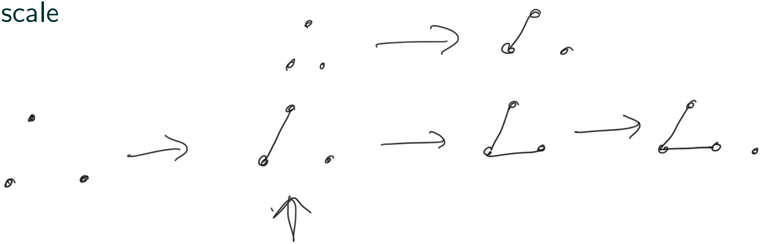


Filtered simplicial complexes

A filtered simplicial complex is a series

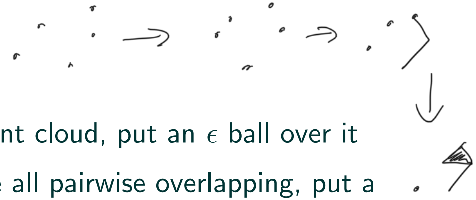
$$S_0 \subseteq S_1 \subseteq S_2 \subseteq \ldots \subseteq S_n$$

Intuitively, think of each S_i as being a glimpse of your object at a different scale

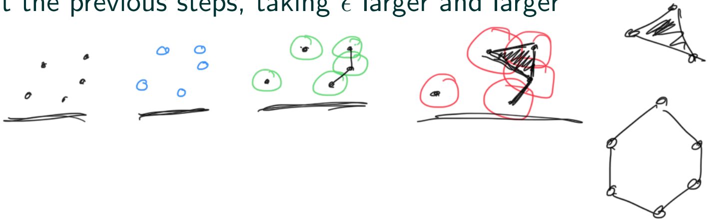


Vietoris–Rips

Constructed as follows:

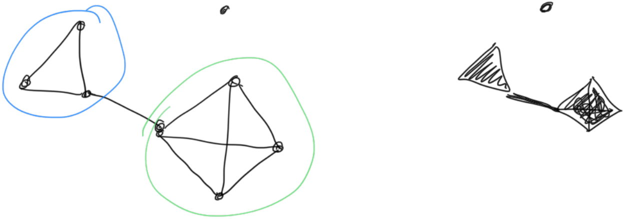


1. Fix ϵ
2. For every point in your point cloud, put an ϵ ball over it
3. Whenever $k+1$ ϵ balls are all pairwise overlapping, put a create a k-simplex
4. The resulting collection of simplices is a simplicial complex (call it S_ϵ)
5. Repeat the previous steps, taking ϵ larger and larger



Clique complex

Let G be a graph. The clique complex $X(G)$ of G is a simplicial complex formed by the sets of vertices in the cliques of G. Alternatively, $X(G)$ is a topological space in which each clique of k vertices is represented by a $(k-1)$-simplex.



Gotta go fast

- Alpha complexes: Delaunay triangulations for fun and profit
- Witness complexes: Sampling is your friend
- Flag complexes

Get creative!

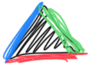
Homology theories

- Associate a sequence of algebraic objects to topological spaces
- Give us a way to rigorously define and categorize holes in a manifold
- Can compute homology on a variety of different topological objects
  - Simplicial
  - Cubical
  - Others

Homology

Let X be a simplicial complex. We construct a chain complex on X, which is a sequence of abelian groups $C_0, C_1, \ldots$ connected by homomorphisms $\partial_n : C_n \to C_{n-1}$, which are called boundary operators. We do this in such a way that

$$\partial_n \circ \partial_{n+1} = 0$$



Then the n-th homology group of X is

$$H_n(X) := \ker(\partial_n)/\mathrm{im}(\partial_{n+1})$$

$$C_2 \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0$$

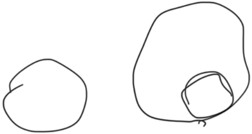
Key takeaway: **The n-th homology group of X is entirely determined by what n-simplices and $n+1$-simplices in X look like**

$$
\begin{tikzcd}
S_0 \arrow[r, "i"] \arrow[d] & S_1 \arrow[r] \arrow[d] & \cdots \arrow[r] & S_n \arrow[d] \\
H_i(S_0) \arrow[r] & H_i(S_1) \arrow[r] & \cdots \arrow[r] & H_i(S_n)
\end{tikzcd}
$$

What comes out of a PH computation?

Usually, a list of features, which correspond to elements of the homology groups, with some additional information:

For $x \in H_m(S_i)$, we get:



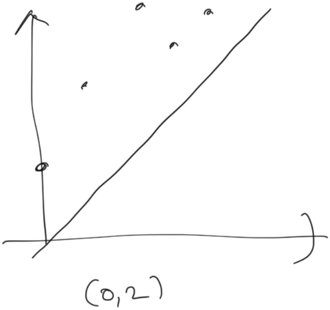
- Dimension m of the feature
- Non-unique generators of x
- A time b_x, called the "birth" of x, which is the lowest value such that $x \in H_m(S_{b_x})$
- A time d_x, called the "death" of x, which is the lowest value greater than b_x such that $x \notin H_m(S_{d_x})$ $\quad x \to x \to x \to x \to 0$
- Note that a feature can theoretically live forever, in which case we sometimes say it dies at ∞

Barcodes

- Represents each feature x with a bar stretching across $[b_x, d_x)$
- Usually use arrows or something similar to represent infinite persistence
- Handle different dimensions in a variety of ways (separate barcodes, colors, etc)

Persistence Diagrams (PDs)

- Plot each feature x at the point (b_x, d_x)
- Can compute distances between PDs (bottleneck, Wasserstein)



Persistence Images (PIs)

- Turns a PD into a vector
- Very roughly, perform a transformation which maps each feature to $(b_x, d_x - b_x)$, replace the point with a probability distribution, and discretize into an image
- The PI for each dimension is a vector
- PIs for all dimensions can be concatenated, resulting in a vector format for persistence

- Allen Hatcher. *Algebraic Topology*. Cambridge, UK: Cambridge University Press, 2002, p. 556
- H Edelsbrunner, D Kirkpatrick, and R Seidel. "On the shape of a set of points in the plane". In: *IEEE Transactions on Information Theory* 29.4 (July 1983), pp. 551–559. ISSN: 0018-9448
- Leopold Vietoris. "Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen". In: *Mathematische Annalen* 97.1 (Dec. 1927), pp. 454–472. ISSN: 00255831. DOI: 10.1007/BF01447877
- Afra Zomorodian and Gunnar Carlsson. "Computing persistent homology". In: *Proceedings of the Annual Symposium on Computational Geometry* 33.2 (Feb. 2004), pp. 347–356. ISSN: 1432-0444. DOI: 10.1145/997817.997870
- Herbert Edelsbrunner and John Harer. *Computational Topology: An Introduction*. Providence, RI, USA: American Mathematical Society, 2010. ISBN: 978-0-8218-4925-5
- Tomasz Kaczynski, Konstantin Michael Mischaikow, and Marian Mrozek. *Computational homology*. Vol. 157. 04. New York, New York, USA: Springer, 2004. ISBN: 9780387215972. DOI: https://doi.org/10.1007/b97315
- Nina Otter et al. "A roadmap for the computation of persistent homology". In: *EPJ Data Science* 6.1 (2017), p. 17. ISSN: 21931127. DOI: 10.1140/epjds/s13688-017-0109-5
- Afra Zomorodian. "Fast construction of the Vietoris-Rips complex". In: *Computers and Graphics* 34.3 (June 2010), pp. 263–271. ISSN: 00978493. DOI: 10.1016/j.cag.2010.03.007

- Michael Kerber and R. Sharathkumar. "Approximate Čech complex in low and high dimensions". In: *Algorithms and Computation. ISAAC 2013*. Ed. by Leizhen Cai, Siu-Wing Cheng, and Tak-Wah Lam. Vol. 8283. Lecture Notes in Computer Science. 2013. ISBN: 9783642450297. DOI: 10.1007/978-3-642-45030-3_62
- Michał Adamaszek and Henry Adams. "The Vietoris-Rips complexes of a circle". In: *Pacific Journal of Mathematics* 290.1 (July 2017), pp. 1–40. ISSN: 00308730. DOI: 10.2140/pjm.2017.290.1
- Dominique Attali, Herbert Edelsbrunner, and Yuriy Mileyko. "Weak witnesses for Delaunay triangulations of submanifolds". In: *Proceedings - SPM 2007: ACM Symposium on Solid and Physical Modeling*. New York, New York, USA: ACM Press, 2007, pp. 143–150. ISBN: 1595936661. DOI: 10.1145/1236246.1236267
- Robert Ghrist. "Barcodes: The persistent topology of data". In: *Bulletin of the American Mathematical Society*. Vol. 45. 1. Jan. 2008, pp. 61–75. DOI: 10.1090/S0273-0979-07-01191-3
- Elizabeth Munch et al. "Probabilistic Fréchet means for time varying persistence diagrams". In: *Electronic Journal of Statistics* 9.1 (July 2015), pp. 1173–1204. ISSN: 19357524. DOI: 10.1214/15-EJS1030
- Henry Adams et al. "Persistence images: A stable vector representation of persistent homology". In: *Journal of Machine Learning Research* 18.1 (Jan. 2017), pp. 1–35. ISSN: 15337928
- Peter Bubenik. "Statistical topological data analysis using persistence landscapes". In: *Journal of Machine Learning Research* 16.1 (Jan. 2015), pp. 77–102. ISSN: 15337928
- Peter Bubenik et al. "Persistent homology detects curvature". In: *Inverse Problems* 36.2 (2020). ISSN: 13616420. DOI: 10.1088/1361-6420/ab4ac0

References: Algebraic Topology and Developments in TDA iii

- Omer Bobrowski, Sayan Mukherjee, and Jonathan E. Taylor. "Topological consistency via kernel estimation". In: *Bernoulli* 23.1 (2017), pp. 288–328. ISSN: 13507265. DOI: 10.3150/15-BEJ744
- Roland Kwitt et al. "Statistical topological data analysis-A kernel perspective". In: *Advances in Neural Information Processing Systems 28*. Ed. by C Cortes et al. Curran Associates, Inc., 2015, pp. 3070–3078
- Xiaojin Zhu et al. "Stochastic multiresolution persistent homology kernel". In: *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. IJCAI'16. AAAI Press, 2016, pp. 2449–2455. ISBN: 978-1-57735-770-4
- Jan Reininghaus et al. "A stable multi-scale kernel for topological machine learning". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 07-12-June. 2015, pp. 4741–4748. ISBN: 9781467369640. DOI: 10.1109/CVPR.2015.7299106
- Andrew J. Blumberg and Michael A. Mandell. "Quantitative Homotopy Theory in Topological Data Analysis". In: *Foundations of Computational Mathematics* 13.6 (Dec. 2013), pp. 885–911. ISSN: 16153375. DOI: 10.1007/s10208-013-9177-5
- Facundo Mémoli and Ling Zhou. "Persistent Homotopy Groups of Metric Spaces". In: *arXiv:1912.12399* (Dec. 2019)

References: Applications of PH i

- Yu Wu et al. "Congestion barcodes: Exploring the topology of urban congestion using persistent homology". In: *2017 IEEE Conference on Intelligent Transportation Systems, Proceedings (ITSC)*. Yokohama: Institute of Electrical and Electronics Engineers Inc., 2017, pp. 1–6. ISBN: 9781538615256. DOI: 10.1109/ITSC.2017.8317777
- Moon Duchin, Tom Needham, and Thomas Weighill. *The (homological) persistence of gerrymandering*. 2020
- Michelle Feng and Mason A. Porter. "Persistent Homology of Geospatial Data: A Case Study with Voting". In: *arXiv:1902.05911* (2019)
- Michelle Feng and Mason A. Porter. "Spatial Applications of Topological Data Analysis: Cities, Snowflakes, Random Structures, and Spiders Spinning Under the Influence". In: *arXiv:2001.01872* (2020)
- Bernadette J Stolz, Heather A Harrington, and Mason A. Porter. "The Topological "Shape" of Brexit". In: *arXiv:1610.00752* (2016)
- Paul Samuel P. Ignacio and Isabel K. Darcy. "Tracing patterns and shapes in remittance and migration networks via persistent homology". In: *EPJ Data Science* 8 (Jan. 2019), p. 1. ISSN: 21931127. DOI: 10.1140/epjds/s13688-018-0179-z
- Barbara I. Mahler, Ulrike Tillmann, and Mason A. Porter. "Analysis of contagion maps on a class of networks that are spatially embedded in a torus". In: *arXiv:1812.09806* (2018)
- Dane Taylor et al. "Erratum: Topological data analysis of contagion maps for examining spreading processes on networks". In: *Nature Communications* 6 (2015), p. 8374. ISSN: 20411723. DOI: 10.1038/ncomms9374

References: Applications of PH ii

- Derek Lo and Briton Park. "Modeling the spread of the Zika virus using topological data analysis". In: *PLoS ONE* 13.2 (2018), e0192120. ISSN: 19326203. DOI: 10.1371/journal.pone.0192120
- Paul Bendich et al. "Persistent homology analysis of brain artery trees". In: *Annals of Applied Statistics* 10.1 (2016), pp. 198–218. ISSN: 19417330. DOI: 10.1214/15-AOAS886
- Lida Kanari et al. "Quantifying topological invariants of neuronal morphologies". In: *arXiv:1603.08432* (2016)
- Kevin Emmett, Benjamin Schweinhart, and Raul Rabadan. "Multiscale topology of chromatin folding". In: *EAI International Conference on Bio-inspired Information and Communications Technologies (BICT)*. BICT'15. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2015, pp. 177–180. ISBN: 978-1-63190-100-3. DOI: 10.4108/eai.3-12-2015.2262453
- Yuri Dabaghian, Vicky L. Brandt, and Loren M. Frank. "Reconceiving the hippocampal map as a topological template". In: *eLife* 3 (Aug. 2014). Ed. by Howard Eichenbaum, e03476. ISSN: 2050084X. DOI: 10.7554/eLife.03476
- Carina Curto. "What can topology tell us about the neural code?" In: *Bulletin of the American Mathematical Society* 54.1 (2017), pp. 63–78. ISSN: 02730979. DOI: 10.1090/bull/1554
- Louis David Lord et al. "Insights into brain architectures from the homological scaffolds of functional connectivity networks". In: *Frontiers in Systems Neuroscience* 10 (2016), p. 85. ISSN: 16625137. DOI: 10.3389/fnsys.2016.00085

References: Applications of PH iii

- Kelin Xia and Guo Wei Wei. "Persistent homology analysis of protein structure, flexibility, and folding". In: *International Journal for Numerical Methods in Biomedical Engineering* 30.8 (2014), pp. 814–844. ISSN: 20407947. DOI: 10.1002/cnm.2655
- Marcio Gameiro et al. "A topological measurement of protein compressibility". In: *Japan Journal of Industrial and Applied Mathematics* 32.1 (Mar. 2015), pp. 1–17. ISSN: 1868937X. DOI: 10.1007/s13160-014-0153-5
- Jaejun Yoo et al. "Topological persistence vineyard for dynamic functional brain connectivity during resting and gaming stages". In: *Journal of Neuroscience Methods* 267 (2016), pp. 1–13. ISSN: 1872678X. DOI: 10.1016/j.jneumeth.2016.04.001
- G. Petri et al. "Homological scaffolds of brain functional networks". In: *Journal of the Royal Society Interface* 11 (Dec. 2014), p. 20140873. ISSN: 17425662. DOI: 10.1098/rsif.2014.0873
- Helen M. Byrne et al. "Topology characterises tumour vasculature". In: *Mathematics Today* 5.5 (2019). ISSN: 1361-2042
- Henrik Ronellenfitsch et al. "Topological Phenotypes Constitute a New Dimension in the Phenotypic Space of Leaf Venation Networks". In: *PLoS Computational Biology* 11.12 (Dec. 2015). Ed. by Enrico Scarpella, e1004680. ISSN: 15537358. DOI: 10.1371/journal.pcbi.1004680
- Chad M. Topaz, Lori Ziegelmeier, and Tom Halverson. "Topological data analysis of biological aggregation models". In: *PLoS ONE* 10.5 (2015), e0126383. ISSN: 19326203. DOI: 10.1371/journal.pone.0126383
- Jason W. Rocks, Andrea J. Liu, and Eleni Katifori. "Revealing structure-function relationships in functional flow networks via persistent homology". In: *arXiv:1901.00822* (2019)

References: Applications of PH iv

- Gökhan Yalnlz and Nazmi Burak Budanur. "Inferring symbolic dynamics of chaotic flows from persistence". In: *Chaos* 30.3 (Mar. 2020), p. 033109. ISSN: 10897682. DOI: 10.1063/1.5122969
- Bernadette J. Stolz, Heather A. Harrington, and Mason A. Porter. "Persistent homology of time-dependent functional networks constructed from coupled time series". In: *Chaos* 27.4 (2017), p. 047410. ISSN: 10541500. DOI: 10.1063/1.4978997
- Gunnar Carlsson et al. "On the local behavior of spaces of natural images". In: *International Journal of Computer Vision* 76.1 (Jan. 2008), pp. 1–12. ISSN: 09205691. DOI: 10.1007/s11263-007-0056-x
- Mickaël Buchet, Yasuaki Hiraoka, and Ippei Obayashi. "Persistent homology and materials informatics". In: *Nanoinformatics*. Heidelberg, Germany: Springer-Verlag, Jan. 2018, pp. 75–95. ISBN: 9789811076176. DOI: 10.1007/978-981-10-7617-6_5
- Miroslav Kramá et al. "Persistence of force networks in compressed granular media". In: *Physical Review E* 87.4 (Apr. 2013), p. 042207. ISSN: 15393755. DOI: 10.1103/PhysRevE.87.042207
- Sarah Tymochko et al. "Using persistent homology to quantify a diurnal cycle in hurricanes". In: *Pattern Recognition Letters* 133 (2020), pp. 137–143. ISSN: 01678655. DOI: 10.1016/j.patrec.2020.02.022
- Leo Speidel et al. "Topological data analysis of continuum percolation with disks". In: *Physical Review E* 98.1 (2018), p. 012318. ISSN: 24700053. DOI: 10.1103/PhysRevE.98.012318

References: Other Topological Applications i

- Gurjeet Singh, Facundo Mémoli, and Gunnar Carlsson. "Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition". In: *Eurographics Symposium on Point-Based Graphics*. Ed. by M Botsch et al. The Eurographics Association, 2007, pp. 91–100. ISBN: 0011051000. DOI: 10.2312/SPBG/SPBG07/091-100
- Jesse S. F. Levitt, Mustafa Hajij, and Radmila Sazdanovic. "Big Data Approaches to Knot Theory: Understanding the Structure of the Jones Polynomial". In: *arXiv:1912.10086* (Dec. 2019)
- Devon P. Humphreys et al. "Fast estimation of recombination rates using topological data analysis". In: *Genetics* 211.4 (2019), pp. 1191–1204. ISSN: 19432631. DOI: 10.1534/genetics.118.301565
- Chad Giusti, Robert Ghrist, and Danielle S. Bassett. "Two's company, three (or more) is a simplex". In: *Journal of Computational Neuroscience* 41.1 (Aug. 2016), pp. 1–14. ISSN: 0929-5313. DOI: 10.1007/s10827-016-0608-6
- Iacopo Iacopini et al. "Simplicial models of social contagion". In: *Nature Communications* 10.1 (Dec. 2019), pp. 1–9. ISSN: 20411723. DOI: 10.1038/s41467-019-10431-6
- Leonhard Horstmeyer and Christian Kuehn. "Adaptive voter model on simplicial complexes". In: *Physical Review E* 101.2 (Feb. 2020), p. 022305. ISSN: 24700053. DOI: 10.1103/PhysRevE.101.022305
- David I. Spivak. *Simplicial Databases*. 2009