Topological Data Analysis Through Directed Filtrations

Gregory Bell (gcbell@uncg.edu) and Thanos Gentimis (agentim@ncsu.edu)
Univerisity of North Carolina at Greensboro and North Carolina State University

THE UNIVERSITY of NORTH CAROLINA GREENSBORO
Department of Mathematics & Statistics

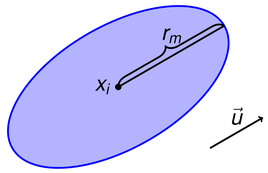
NC STATE UNIVERSITY

Abstract

Given a finite point set X, one constructs the Čech complex $\check{C}_r(X)$ at scale r by taking the nerve of the subset covered by r-balls centered at the points $x_i \in X$.
Given a direction vector $\vec{u}$ we construct covers by ellipsoids with major axis parallel to $\vec{u}$ and increasing radii. We then compute the homology of the complexes obtained as the nerves of those covers. This additional structure defines a generalized persistence module, yields a correspondence between rotations and interleavings of persistence modules and suggests an approach to detect directed homological features.
When X has a certain directed feature (say a cylindrical hole or void) we can use maximum persistence to find the direction of this feature.

Definitions

Given:
- $X = \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^2$
- $R = \{0 < r_1 < \ldots < r_K\}$
- $0 < c < 1$; and
- $\vec{u} \in S^1$,



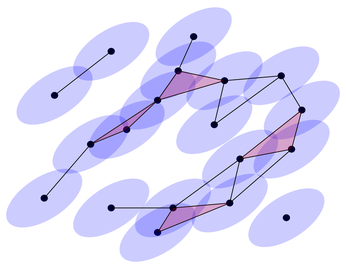
we define $E_{\vec{u}}(x_i, r_m)$ to be the ellipse centered at x_i, with major axis parallel to $\vec{u}$, with length of semi-major axis r_m and semi-minor axis $c \cdot r_m$.
More generally, we can consider $X \subset \mathbb{R}^n$, $\mathbf{c} = \{c_2, c_3, \ldots, c_{n-1}\} \in (0,1)^{n-1}$, and R as above. Then, given $\vec{u} \in S^{n-1}$, we find an orthonormal basis for R^n of the form $\{\vec{u}, \vec{u}_2, \vec{u}_3, \ldots, \vec{u}_n\}$. Define $E_{\vec{u}}(x_i, r_m)$ to be the ellipse centered at x_i, with semi-axes of length r_m parallel to $\vec{u}$ and $c_j \cdot r_m$, in the direction of $\vec{u}_j$.

A Skewed Čech Complex

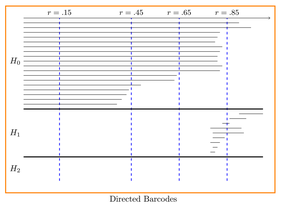
We fix a direction $\vec{u}$ and a number $r_i \in R$. Let $K_{r_i} = K_{\vec{u}}(X, r_i)$ denote the nerve of the union:

$$\bigcup_{x \in X} E_{\vec{u}}(x, r_i).$$

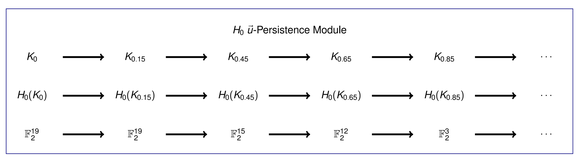
Then, we compute $H_*(K_{r_i})$ for each $r_i \in R$. This gives rise to a sequence of persistence vector spaces that are connected by inclusion maps. From this we can build directed barcodes (or persistence diagrams) and form persistence modules.



$r = .85$



Directed Barcodes

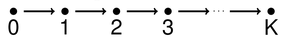


Generalized Persistence Modules

Definitions

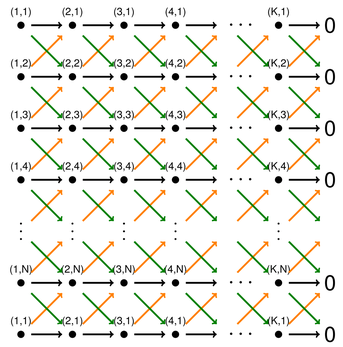
A Generalized persistence module in **D** over **P** is a functor $\mathbf{P} \to \mathbf{D}$, where
- **P** is any preordered set and
- **D** is any target category.

We wish to consider the case where the target category is $\mathbf{D} = \mathbf{Vect}_{\mathbb{F}_2}$.
For the preordered category **P** we consider two choices:
- **K**, the category of positive integers at most K.



The category **K**.

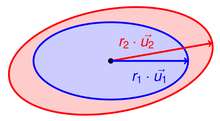
- $\mathbf{C_{NK}}$ a cylindrical category described on the left. Observe that the top and bottom rows of the diagram are identified.

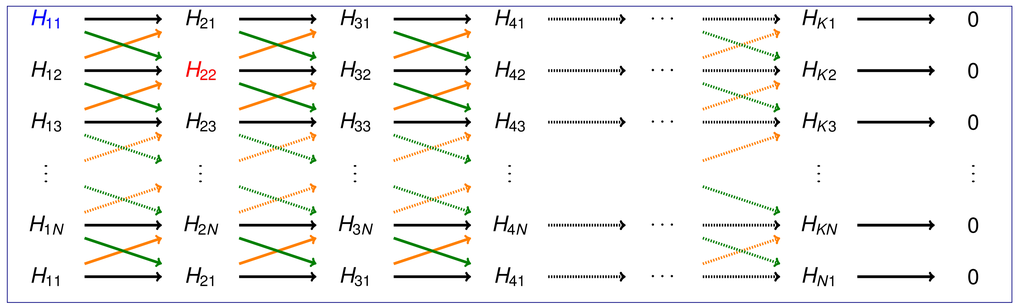


The category $\mathbf{C_{NK}}$.

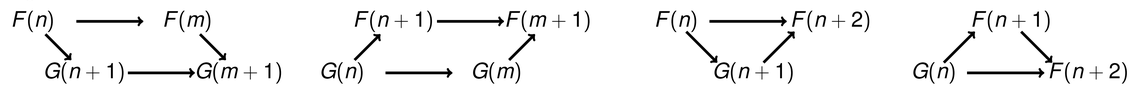
Examples

By careful choice of $\vec{u}_i$ and r_i, we have $E_{\vec{u}_1}(x, r_1) \subset E_{\vec{u}_2}(x, r_2)$, giving inclusion maps at the level of complexes, which induce linear maps on homology.





In this setting, a rotation and inclusion as described above corresponds to a translation on $\mathbf{C_{NK}}$, i.e., a function $\Gamma : \mathbf{C_{NK}} \to \mathbf{C_{NK}}$ that is monotone and which satisfies $x \leq \Gamma(x)$, for all x.
Alternatively, with $\Gamma(n) = n + 1$ on **K**, we get a Γ-interleaving as follows. With $F : \mathbf{K} \to \mathbf{Vect}_{\mathbb{F}_2}$ equal to the persistence with direction $\vec{u_1}$ and $G : \mathbf{K} \to \mathbf{Vect}_{\mathbb{F}_2}$ equal to the persistence with direction $\vec{u_2}$ we have, with $n < m$, that the following diagrams commute:



Thus, the diagrams corresponding to these "close" angles are interleaved.

Further Study

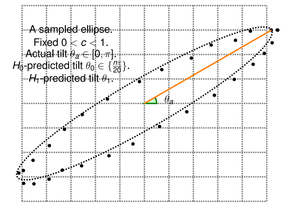
We would like to continue to pursue the following ideas:
- Can we find a method to discover an optimal c for a given dataset with a feature?
- In the case of $\mathbb{R}^2$ find an algorithm yielding the direction of a feature based on maximizing barcode length and entropy.
- What is the H_0- and H_1-detected angle for datasets with isomorphic features with different directions?
- Explore a local notion of directed features.
- Generalize our experiments to higher dimensional data.

Thanks

Work on this project began when the first author was a New Directions Professor at the IMA and the second author was visiting the IMA's Annual Thematic Program on Applications of Algebraic Topology. Both authors would like to express their thanks to the IMA for its support and hospitality.

IMA Institute for Mathematics and its Applications

Experimental Evidence



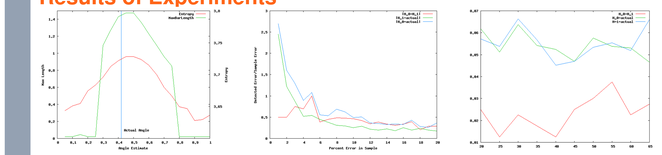
Experiments

We sampled N points on an ellipse with tilt $\theta_a \in [0, \pi]$ and fixed $c = 0.5$ and uniform sampling error ϵ.
- We computed directed filtrations with $c = 0.5$ and angles $\{\frac{\pi}{20}, \frac{2\pi}{20}, \ldots, \frac{19\pi}{20}\}$.
- For each angle, we compute the Shannon entropy of the H_0-barcodes:

$$-\sum_{j \in J} p_j \cdot \log(p_j)$$

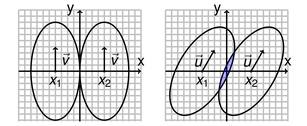
where the barcodes are $[a_j, b_j]$, $\ell_j = b_j - a_j$, $L = \sum \ell_j$ and $p_j = \frac{\ell_j}{L}$.
- For each angle, we compute the length of the longest H_1-barcode.

Results of Experiments



- Above left, we compare the entropy estimate θ_0 with the Max Length estimate θ_1 with the actual value (in blue).
- In the center we compute the difference in the estimate over the error in the sampling for increasing sampling errors.
- On the right we observe that increasing the number of sample points doesn't affect the estimates significantly.

Regular Directions



Direction $\vec{v}$ is irregular. Direction $\vec{u}$ is regular.

Regular Directions

For fixed X, R, c as above, call a direction $\vec{u}$ regular if every pair $E_{\vec{u}}(x_i, r_m)$, $E_{\vec{u}}(x_j, r_m)$ are disjoint or meet in a set with positive measure.

Theorem: Regular Directions are Common

For fixed X, R, c, irregular directions have measure zero.

Theorem: Direction Stability

Fix X, R, c as above, and let $\vec{u}$ be a regular direction. Then, there is some $\delta > 0$ so that $\|\vec{u} - \vec{v}\| < \delta$ implies the directed complexes associated to $\vec{u}$ and $\vec{v}$ are identical, i.e., $\mathcal{K}(\vec{u}) = \mathcal{K}(\vec{v})$, where $\mathcal{K}(\vec{u}) = \{K_{r_1}(\vec{u}) \subseteq K_{r_2}(\vec{u}) \subseteq \cdots \subseteq K_{r_n}(\vec{u})\}$ and $\mathcal{K}(\vec{v}) = \{K_{r_1}(\vec{v}) \subseteq K_{r_2}(\vec{v}) \subseteq \cdots \subseteq K_{r_n}(\vec{v})\}$

References

1. Peter Bubenik, Vin de Silva, and Jonathan Scott, *Metrics for generalized persistences modules*, arXiv:1312.3829 (2103).
2. Dan Burghelea and Tamal K. Dey, *Topological persistence for circle-valued maps*, Discrete Comput. Geom., 50(1) (2013) 69–98.
3. Gunnar Carlsson, *Topology and data*, Bull. Amer. Math. Soc. 46 (2009), no. 2 255–308.
4. Michael Lesnick, *The theory of the interleaving distance on multidimensional persistence modules*, arXiv:1106.5305 (2011).
5. Primoz Skraba and Mikael Vejdemo-Johansson, *Persistence modules: Algebra and algorithms*, arXiv:1302.2015 (2013).

Greg Bell UNCG Department of Mathematics and Statistics gcbell@uncg.edu
Thanos Gentimis NC State University Department of Electrical and Computer Engineering agentim@ncsu.edu