Visualizing Sparse Filtrations*

Nicholas J. Cavanna, Mahmoodreza Jahanseir, and Donald R. Sheehy

Department of Computer Science, University of Connecticut
Storrs, CT, USA
nicholas.j.cavanna@uconn.edu, {reza,donald}@engr.uconn.edu

Abstract

Over the last few years, there have been several approaches to building sparser complexes that still give good approximations to the persistent homology [5, 4, 3, 2, 1]. In this video, we have illustrated a geometric perspective on sparse filtrations that leads to simpler proofs, more general theorems, and a more visual explanation. We hope that as these techniques become easier to understand, they will also become easier to use.

1998 ACM Subject Classification F.2.2 Geometrical problems and computations, G.2.1 Combinatorial algorithms

Keywords and phrases Topological Data Analysis, Simplicial Complexes, Persistent Homology

Digital Object Identifier 10.4230/LIPIcs.SOCG.2015.23

1 The Shape of Data

Topological data analysis is concerned with finding the underlying shape of a data set. Often, a union of balls called offsets is used to approximate the shape and fill in the space between points. The topology of the balls can be represented as a simplicial complex called a nerve. Every subset of balls with a common intersection contributes one simplex to the nerve: an edge for each pairwise intersection, a triangle for each 3-way intersection, etc. (see Fig. 1).

Instead of looking at just one radius, we can look at the offsets at all radii from zero to infinity. A growing space like this is called a filtration. The nerves give a corresponding simplicial filtration. Persistent homology is a way to study the changes in topology over the course of a filtration. The output is called a persistent barcode and marks the components, holes, and voids as well as their lifespans. The Nerve Theorem and its persistent variant guarantee that the barcode for the offsets is the same as that of the nerve filtration.

Nerve complexes get very big very fast, even when restricting to subsets of constant size. A common variant that doesn't assume Euclidean metrics is the Rips complex and it suffers similar difficulties.

At larger scales, fewer points are needed to give a good approximation. The sparser subsample of our point set at scale α is obtained by calculating an ε-net, i.e. a subset where each pair is at least ε apart and the ε-radius balls centered on the points of the net cover the input. However, removing points cause the nerve of balls to no longer be a filtration, because a filtration is, by definition a monotonely growing space. For a simplicial complex, this means that simplices appear, but never disappear. We solve this problem by viewing the offset filtration as a nerve of objects one dimension higher as illustrated below.

If we visualize the scale parameter as another dimension, a growing ball traces out a cone (below, left). This cone is modified in two ways. First, we assign a maximum radius to each

* Partially supported by the National Science Foundation under grant number CCF-1464379.

31st International Symposium on Computational Geometry (SoCG'15).
Editors: Lars Arge and János Pach; pp. 23–25
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

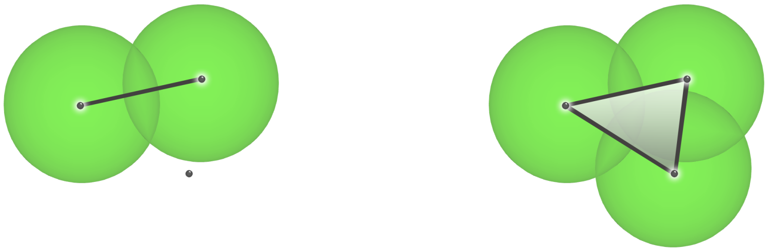


Figure 1 The nerve has an edge for each pairwise intersection, a triangle for each 3-way intersection (right), etc.

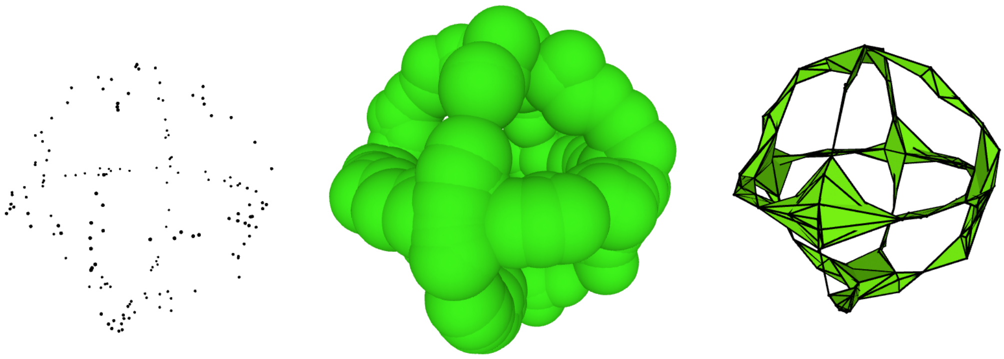
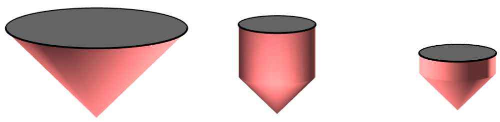
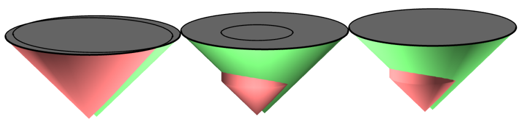


Figure 2 A point set sampled on a sphere, its offsets, and its (sparsified) nerve complex.

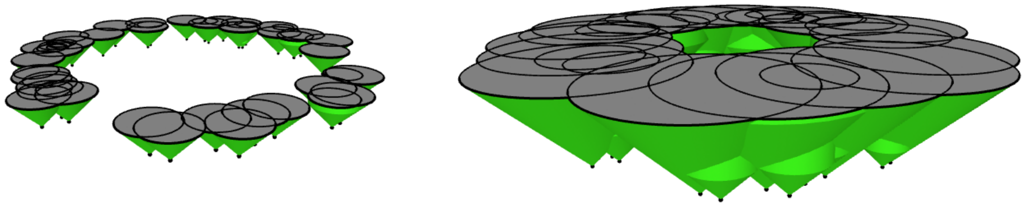
point (middle). Next, we truncate the cone at the height when the point will be removed from the filtration (right). Truncating the cone simulates the removal of the corresponding ball. The cone has not been removed, it just no longer intersects the time slices above the "removal" time.



The maximum radius and height are chosen so that the top of the cone is sure to be covered at the time it is removed.



These cones form a new filtration one dimension higher. Their nerve is the desired sparse filtration.



In this example, one can imagine flattening all the cones onto one level set, and since all the cones are stacked on each other, there won't be any loss of homological information. The sublevel sets of the cones are homotopy equivalent to the level sets, which implies they have the same homology. By the persistent nerve lemma we know that the nerve has the same persistent homology as the sublevel sets, thus we can calculate the persistent homology of our sparsified offsets by computing the persistent homology of the sparse nerve filtration.

The sparsification algorithm can cut down the asymptotic size of filtrations from polynomial to linear, while still achieving a close approximation to the persistence diagram. The video aims to provide a simple, geometric explanation for the topological guarantees of such sparse filtrations. It avoids the complexity of zig-zag inclusions maps from previous work by considering time as an extra spatial dimension. The construction and its analysis easily generalizes to Rips and other related complexes, and although the example input was two-dimensional, the construction works in any number of dimensions.

2 Production

In order to create the video, we used Processing to create the visualizations, iMovie to piece together the soundtrack and the visualizations, and Javaplex through Matlab to calculate the barcodes. We thank Julia Sheehy for supplying the narration.

References

1. Magnus Bakke Botnan and Gard Spreemann. Approximating persistent homology in Euclidean space through collapses. *Applicable Algebra in Engineering, Communication and Computing*, pages 1–29, 2015.
2. Mickaël Buchet, Frédéric Chazal, Steve Y. Oudot, and Donald Sheehy. Efficient and robust persistent homology for measures. In *ACM-SIAM Symposium on Discrete Algorithms*, 2015.
3. Tamal K. Dey, Fengtao Fan, and Yusu Wang. Computing topological persistence for simplicial maps. In *Proceedings of the 30th Annual Symposium on Computational Geometry*, pages 345–354, 2014.
4. Michael Kerber and R. Sharathkumar. Approximate Čech complexes in low and high dimensions. In *24th International Symposium on Algorithms and Computation (ISAAC 2013)*, volume LNCS 8283, pages 666–676, 2013.
5. Donald R. Sheehy. Linear-size approximations to the Vietoris-Rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.