A Novel Algorithm for Topological Persistence, with Application to Neuroscience

Greg Henselman **Robert Ghrist**
Departments of Mathematics and Electrical & Systems Engineering
University of Pennsylvania ; Philadelphia, PA
grh@seas.upenn.edu ; ghrist@math.upenn.edu

Abstract

A recent advance in computational homology gives an order-of-magnitude improvement in applications to neural coding analysis.

Key words: Neural coding, Geometric network, Correlation matrix, Computational homology, Discrete Morse Theory, Cellular matroid

1 Overview

The study of betti curves – topological shape descriptors -- of matrix order complexes has yielded recent insights into the structure of real symmetric matrices, with emerging applications in neural coding analysis [1, 2]. Unlike spectral quantifiers of structure and randomness, such descriptors are invariant under monotone nonlinear transformation, hence robust to order-preserving measurement noise. The principle challenge to computability of betti curves in this setting is exponential growth of combinatorial data relative to the size of the input matrix. To address this, we present a new reduction algorithm capable of lowering memory requirements by a factor of one to three orders of magnitude, enabling first-of-its kind analysis of large-scale neural connectivity data generated by fMRI, DTI, etc. As an aid to computation, we present a sparse topological LU factorization technique that produces, in addition, refined topological invariants, termed *barcode generators.* To our knowledge this is the first implemented library with this capability. Barcode generators are fundamental in topological data analysis [3]. The availability of finer invariants at reduced memory cost and greater scalability will enrich existing applications. Key to both algorithms are concepts from the theory of cellular matroids [4] and discrete Morse theory [5].

1 Background

A (combinatorial) simplicial complex X is a family of subsets (or simplices) of a ground set V, closed under set inclusion. The dimension of a simplex of cardinality $(k+1)$ is defined to be k, the set of all k-dimensional simplices in X is denoted $X^{(k)}$, and the set of all p-dimensional simplices with $p \leq k$ is denoted X^k. The k^{th} betti number of X with coefficients in field F, denoted $\beta_k(X;F)$, is the rank of the k^{th} homology – an algebraic invariant describing the number of F-linearly independent k-dimensional holes in X.

Let $K_m = (V,E)$ be the complete graph on m vertices, and $w\colon E \to \{1,\ldots,|E|\}$ be a bijective weight function. If $E_i = \{e \in E : w(e) \leq i\}$ is the set of all edges "of size smaller than i," one may define X_i to be the set of all cliques in (V,E_i). The function defined $c(i) = \beta_k(X_i)$ is the k^{th} betti curve of X with respect to w.

Say that a weight function w is *geometric* (or has *geometric structure*) if $w(\{i,j\}) = \#\{\{i' < j'\} : d(i',j') < d(i,j)\}$ for some metric d on V; given iid variables $(\phi_{i,j})$ for $0 < i < j \leq m$, we say the random variable w defined by $w(i,j) = \#\{\{i' < j'\} : \phi_{i',j'} < \phi_{i,j}\}$ is itself *iid*.

Curto et al. [2] show that geometric and iid weight functions induce markedly distinct distributions on betti curves. Further, Giusti et al. argue [1] that the distributions of betti curves for hippocampal place cells in rats are explained by the observation that neurons tuned to features that lie in a continuous coding space, such as orientation-tuned neurons [6] or hippocampal place cells [7], have correlations that decrease with distance. These observations suggest new methods of analysis and hypothesis testing vis-à-vis random and geometric structure in neural correlation matrices. To perform the same analysis on large matrices, however, users must address the issue of exponential growth in the size of the input complex, see Figure 1 [left].

The contributions presented are: first, an algorithm to reduce the number of simplices stored in memory, and second, a sparse factorization technique for homology computation, see Figure 1 for sample results.

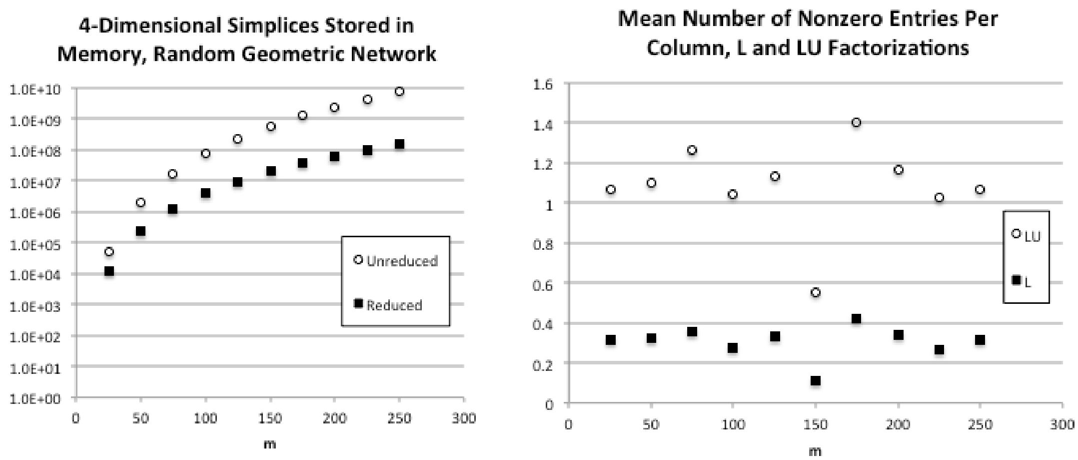

Figure 2: Sample results. Point clouds of cardinality m = 25, …, 250 were sampled from a uniform distribution on the unit cube of dimension 20. The induced filtrations on each simplex were coarsened to 200 levels, after which the reduction algorithm was applied. *Left* The number of 4-dimensional faces stored per sample, before and after reduction, log scale. *Right* The mean number of nonzero entries per column of the LU factorization in dimension 3, and of the L factorization in dimension 4 (a U component in the top dimension is not necessary to obtain generators, and is not computed). Note that the number of columns exceeds the number of nonzero entries stored in some cases, an artifact of the data structures employed.

Work in progress applies these contributions to experimental data from neural correlations in the context of learning.

References

[1] C. Giusti, E. Pastalkova, C. Curto, & V. Itskov, arXiv:1502.06172, 2015.
[2] C. Curto, C. Giusti , & V. Itskov, arXiv:1502.06173, 2015.
[3] R. Ghrist, *Elementary Applied Topology*, 2014.
[4] A. Bjorner *et al.*, *Oriented Matroids*, 1993.
[5] R. Foreman, *Advances in Mathematics* 134(1), p. 90*ff.*, 1998.
[6] D. Hubel, T. Wiesel, *Journal of Physiology* 148, p. 574, 1959.
[7] J. O’Keefe, J. Dostrovsky, *Brain Research* 34, p. 171, 1971.