Advances in Applied Mathematics 102 (2019) 113–142

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

A higher-dimensional homologically persistent skeleton

Sara Kališnik [a,*], Vitaliy Kurlin [b,1], Davorin Lešnik [c,2]

[a] *Department of Mathematics and Computer Science, Wesleyan University, United States of America*
[b] *Materials Innovation Factory, University of Liverpool, United Kingdom of Great Britain and Northern Ireland*
[c] *Department of Mathematics, University of Ljubljana, Slovenia*

A R T I C L E I N F O

Article history:
Received 25 October 2017
Received in revised form 9 July 2018
Accepted 9 July 2018
Available online xxxx

MSC:
18G99
05E45

A B S T R A C T

Real data is often given as a point cloud, i.e. a finite set of points with pairwise distances between them. An important problem is to detect the topological shape of data – for example, to approximate a point cloud by a low-dimensional non-linear subspace such as an embedded graph or a simplicial complex. Classical clustering methods and principal component analysis work well when data points split into good clusters or lie near linear subspaces of a Euclidean space.

Methods from topological data analysis in general metric spaces detect more complicated patterns such as holes and voids that persist for a large interval in a 1-parameter family of shapes associated to a cloud. These features can be visualized in the form of a 1-dimensional homologically persistent skeleton, which optimally extends a minimum spanning tree of a point cloud to a graph with cycles. We generalize this skeleton to higher dimensions and prove its optimality among

* Corresponding author.
E-mail addresses: skalisnikver@wesleyan.edu (S. Kališnik), vkurlin@liverpool.ac.uk (V. Kurlin), davorin.lesnik@fmf.uni-lj.si (D. Lešnik).

[1] This author was supported by the Leverhulme Research Centre for Functional Materials Design (UK) and EPSRC grant "Application-driven Topological Data Analysis" EP/R018472/1.
[2] This author was partially supported by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF under Award No. FA9550-14-1-0096.

https://doi.org/10.1016/j.aam.2018.07.004

all complexes that preserve topological features of data at any scale.

1. Introduction

1.1. Motivations and data skeletonization problem

Real data is often unstructured and comes in the form of a non-empty finite metric space, called a *point cloud*. Such a point cloud can consist of points in 2D images or of high-dimensional vector descriptors of a molecule. A typical problem is to study interesting groups or clusters within a point cloud.

Real data rarely splits into well-separated clusters, though it often has an intrinsic low-dimensional structure. For example, the point cloud of mean-centered and normalized 3×3 patches in natural grayscale images has its 50% densest points distributed near a 2-dimensional Klein bottle in a 7-dimensional space [5]. This example motivates the following problem.

Data skeletonization problem. Given a point cloud $\mathcal{C}$ in a metric space M, find a complex $\mathcal{S} \subseteq M$ of a minimal weight to approximate $\mathcal{C}$ geometrically and topologically in a way that the inclusions of certain subcomplexes of $\mathcal{S}$ into *offsets* of $\mathcal{C}$ (unions of balls with a fixed radius α and centers at points of $\mathcal{C}$) induce homology isomorphisms up to a given dimension for all α.

The above problem is harder than describing the topological shape of a point cloud. Indeed, for a noisy random sample $\mathcal{C}$ of a circle, we aim not only to detect a circular shape $\mathcal{C}$, but also to approximate an unknown circle by a 1-dimensional graph $\mathcal{S}$ that should have exactly one cycle and be close to $\mathcal{C}$.

A homologically persistent skeleton (HoPeS) introduced by V. Kurlin [23] has solved the 1-dimensional case. Cycles of a $\mathtt{HoPeS}(\mathcal{C})$ are in a 1-1 correspondence with all 1D persistent homology classes of $\mathcal{C}$. The current paper extends the construction and optimality of HoPeS to higher dimensions. We denote the d-dimensional homologically persistent skeleton by $\mathtt{HoPeS}^{(d)}$.

1.2. Review of closely related past work

A metric graph reconstruction is related to the data skeletonization problem above. The output is an abstract metric graph or a higher-dimensional complex, which should be topologically similar to an input point cloud $\mathcal{C}$, but not embedded into the same space as $\mathcal{C}$, which makes the problem easier.

The classical Reeb graph is such an abstract graph defined for a function $f \colon \mathcal{Q} \to \mathbb{R}$, where $\mathcal{Q}$ is a simplicial complex built on the points of a given point cloud $\mathcal{C}$. For example,