EXPLORING PERSISTENT LOCAL HOMOLOGY IN TOPOLOGICAL DATA ANALYSIS

Brittany Terese Fasy[*]

Montana State University
Computer Science Department
Bozeman, MT

Bei Wang[†]

University of Utah
Scientific Computing and Imaging Institute
Salt Lake City, UT

ABSTRACT

Topological data analysis (TDA) has rapidly grown in popularity in recent years. One of the emerging tools is *persistent local homology*, which can be used to extract local structure from a dataset. In this paper, we provide a survey that explores this new tool, emphasizing its use in data analysis.

Index Terms— persistent homology, computational topology, stratified learning, data analysis

1. INTRODUCTION

Topological data analysis (TDA) bridges research from computational geometry and topology, algebraic topology, machine learning, and statistics in studying data-centric problems. Recently, the field has produced a collection of innovative techniques in studying the *shape of data*. Two fundamental tasks in the field of TDA are *reconstruction* (how to assemble discrete samples into global structures) and *inference* (how to infer structure).

In this paper, we present persistent local homology (PLH) as a recent development in TDA that is a combination of the concepts of persistent homology and local homology. Roughly speaking, the homology of a topological space measures its topological features such as connected components, tunnels, and voids. Local homology studies the homology within a local neighborhood relative to its boundary. PLH turns the algebraic concept of local homology into a multi-scale notion by constructing extended series of homology groups [1, 2]. It is, in our opinion, *the* tool within TDA that studies the local structure of data. We explore PLH from theoretical, algorithmic and application perspectives, with an emphasis on its application in data analysis. Furthermore, this paper unifies results developed by various researchers over the past few years.

Section 2 gives a brief theoretical and algorithmic exposition of PLH. Section 3 surveys applications of PLH in several key areas, including road network analysis, clustering and stratification learning. Section 4 concludes by contemplating possible future directions.

2. PERSISTENT LOCAL HOMOLOGY

We give theoretical background for defining PLH, namely, homology, persistent homology and local homology. We cover the relevant notions from their smooth/continuous setting to the corresponding discrete/simplicial setting, where the former is simple for theory and the latter is appropriate for algorithms. For more details, see [3, 4] for a gentle introduction to TDA and [5] for algorithmic foundations.

Homology and persistent homology. Homology deals with topological features of a space. Given a topological space $\mathbb{X}$, the zero-, one- and two-dimensional homology groups, denoted as $\mathsf{H}_0(\mathbb{X})$, $\mathsf{H}_1(\mathbb{X})$ and $\mathsf{H}_2(\mathbb{X})$ respectively, correspond to components, tunnels and voids of $\mathbb{X}$. Formally, the construction of homology groups begins with a chain complex $\mathsf{C}(\mathbb{X})$ that encodes information about $\mathbb{X}$, which is a sequence of abelian groups $\mathsf{C}_0(\mathbb{X}), \mathsf{C}_1(\mathbb{X}), \ldots$ connected by homomorphisms known as the boundary operators $\partial_k : \mathsf{C}_k(\mathbb{X}) \to \mathsf{C}_{k-1}(\mathbb{X})$. The k-th *homology* group $\mathsf{H}_k(\mathbb{X}) = \ker(\partial_k)/\mathrm{im}(\partial_{k+1})$. We work primarily with *relative homology* groups $\mathsf{H}_k(\mathbb{X}, \mathbb{A})$ (for $\mathbb{A} \subseteq \mathbb{X}$) which is defined by the same formula, but using boundary maps on the quotient spaces $\mathsf{C}_k(\mathbb{X})/\mathsf{C}_k(\mathbb{A}) \to \mathsf{C}_{k-1}(\mathbb{X})/\mathsf{C}_{k-1}(\mathbb{A})$.

A more nuanced way to describe the shape of $\mathbb{X}$ is using persistent homology, which is a multi-scale notion

[*]Partially supported by NSF and NIH award 1557716.

[†]This research is supported by NSF-IIS-1513616.

of homology. Its most common form applies to a sequence of topological spaces connected by inclusions, called a *filtration*. That is, we may consider the finite sequence $\emptyset = \mathbb{X}_0 \subseteq \mathbb{X}_1 \subseteq \cdots \subseteq \mathbb{X}_n = \mathbb{X}$. Applying homology to this sequence, the homology groups are connected from left to right by homomorphisms induced by inclusion $\mathbb{X}_i \hookrightarrow \mathbb{X}_j$ (for $i \leq j$), denoted as $f_k^{i,j} : \mathsf{H}_k(\mathbb{X}_i) \rightarrow \mathsf{H}_k(\mathbb{X}_j)$. The rank of the homology groups changes as the index increases. When the rank increases (equivalently, the map $f_k^{i-1,i}$ is not surjective), we call this a *birth* event at $\mathbb{X}_i$, and when the rank decreases (the map $f_k^{j-1,j}$ is not injective), we call it a *death* event at $\mathbb{X}_j$. Persistent homology pairs the birth events and the death events as a multi-set of points in the plane, called the *persistence diagram*; see [5].

Local homology. Homology studies the structure of the entire topological space; however, we are often interested in studying the local structures in the data as well. A standard tool to use for studying local structure is local homology. The k-th *local homology* group of $\mathbb{X}$ at a point $x_0 \in \mathbb{X}$ is the relative homology $\mathsf{H}_k(\mathbb{X}, \mathbb{X} - x_0)$. Alternatively, it can be defined as the limit of the homology of $\mathbb{X}$ relative to everything except a shrinking neighborhood around x, $\lim_{r\rightarrow 0} \mathsf{H}_k(\mathbb{X}, \mathbb{X}\backslash U_r)$, where U_r is a neighborhood of x_0 with radius r.

Persistent local homology. In PLH, we adapt a multiscale notion of local homology based on persistence. For a fixed distance α, we consider the "thickened" version of $\mathbb{X}$, that is, we let $\mathbb{X}_\alpha$ denote the subset of $\mathbb{R}^d$ that is at most distance α from $\mathbb{X}$. We fix a neighborhood $V \subset \mathbb{R}^d$ of x in the ambient space $\mathbb{R}^d$, and let $U_\alpha = V \cap \mathbb{X}_\alpha$. We then compute the homology of U_α relative to its boundary ∂U_α. We construct an *α-filtration* by allowing α to range from zero to the diameter of V. For all $\alpha < \alpha'$, the inclusion $U_\alpha \subset U_{\alpha'}$ induces a homomorphism on homology:

$$\mathsf{H}(U_\alpha, \partial U_\alpha) \rightarrow \mathsf{H}(U_{\alpha'}, \partial U_{\alpha'}).$$

In other words, we are interested in the persistent (local) homology defined by the relative homology groups $\mathsf{H}(U_\alpha, \partial U_\alpha)$, where α is the parameter that defines the filtration. We denote the corresponding persistence diagram by $\mathrm{Dgm}(\mathbb{X}; V)$. In addition, we could consider another notion of PLH, referred to as an r-filtration, constructed by fixing the thickening parameter α and varying the radius r of V, see [6] for details.

Distances between persistence diagrams. Given two persistence diagrams, D_1 and D_2, we may want to compare the diagrams. And, in fact, a well-defined distance between persistence diagrams is the *bottleneck distance*:

$$W_\infty(D_1, D_2) := \inf_{f\colon D_1 \rightarrow D_2} \sup_{x \in D_1} ||x - f(x)||, \quad (1)$$

where $f\colon D_1 \rightarrow D_2$ is bijection between the diagrams D_1 and D_2. This distance between diagrams is stable both for general filtrations [7, 8] as well as for local homology filtrations [9, 10].

Persistent local homology computation. Computing PLH from potentially noisy point cloud samples is more challenging than computing persistent homology in general (see [2]), due to the combinatorial complexity of computing ∂U_α, the boundary of local neighborhoods. Delaunay complexes and their variants have typically been employed to guarantee theoretical correctness [1, 11]; although, their construction does not scale well with dimension. On the other hand, Vietoris-Rips complexes have been used more commonly in TDA due to their algorithmic simplicity and robust computation in practice (e.g. [12, 13]). It has been shown that α- and r-filtrations of PLH could be approximated by constructing families of Vietoris-Rips complexes [6]. Some progress has been made towards computing PLH using even more compact combinatorial structures such as Graph Induced Complexes [14]. In other contexts, when the objects are convex shapes (e.g., straight road segments as studied in [9]), a family of Čech filtrations can provide quick computations, provided that the radius of the neighborhoods is small enough.

3. APPLICATIONS OF PLH

The original application of PLH is in stratification learning and clustering – partitioning an object of interest into pieces of uniform dimension. The role of PLH has expanded to include applications to road network analysis and beyond. We briefly explore these applications next.

3.1. Stratification Learning and Clustering

A classic problem in learning focuses on inferring the structure of data from point cloud samples. In manifold learning, we assume the samples are drawn from a manifold; more generally, in stratification learning, we assume the points are sampled from a stratified space (i.e., a mixture of possibly intersecting manifolds). Previous work in pure mathematics has focused on the study of stratified spaces under smooth and continuous settings without computational considerations of noisy and discrete data [15, 16]. Statistical approaches that rely on inferences of mixture models or local dimension estimation require either strict geometric assumptions (e.g. linearity) or may not handle complex singularities [17, 18] (as illustrated in Fig 1). Recently, topological approaches have led to new theoretical and algorithmic results by addressing these issues. The main objective is to cluster the point cloud samples, according to the structure of their underlying manifold pieces and how the pieces interact with one another, at multiple scales. In particular, the local structure of a sufficiently sampled stratified space could be studied based on PLH [1]; and point cloud data (PCD) could be clustered by the manifold pieces they belong to based on how the PLH of nearby sampled points map into one another [11].

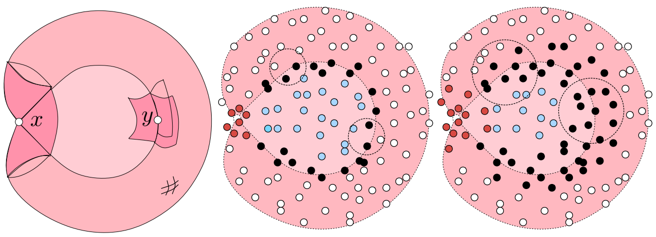


Fig. 1. Left: A stratified space is constructed by attaching a disk to the tunnel of a pinched torus. Structures surrounding complex singularities are highlighted at points x and y. It is difficult to cluster the sampled points based on local dimension estimation alone, especially surrounding these complex singularities. Middle and right: sampled points are clustered at two scales.

Stratification learning in the clustering setting relies on ingredients of PLH and intersection homology [19, 20, 21]. A crucial component is to study a persistent version of local homology intersection map. Intuitively, two nearby points belong to the same cluster, if they "look the same locally" and their local neighborhoods are "glued together in a nice way". Assume we are given a stratified space $\mathbb{X}$ embedded in $\mathbb{R}^d$, define $B_r(x)$ to be a ball of radius r in $\mathbb{R}^d$ centered at x that corresponds to a neighborhood of $x \in \mathbb{X}$. For a fixed radius r, and for every pair of points $p, q \in \mathbb{R}^d$ whose neighborhoods intersect, we define the following relative homology map $\phi^{\mathbb{X}}(p, q, r)$:

$$\begin{aligned} &\mathsf{H}(\mathbb{X} \cap B_r(p), \mathbb{X} \cap \partial B_r(p)) \to \\ &\quad \mathsf{H}(\mathbb{X} \cap B_r(p) \cap B_r(q), \mathbb{X} \cap \partial(B_r(p) \cap B_r(q))). \end{aligned} \tag{2}$$

For example, consider the space $\mathbb{X}$ in $\mathbb{R}^2$ in Fig. 2. For each pair (p, q), let $f = \phi^{\mathbb{X}}(p, q, r)$ and $g = \phi^{\mathbb{X}}(q, p, r)$. Then the points p and q are considered to have the same local structure if f and g are both isomorphisms, that is, their local structures map into each other bijectively via their intersection; equivalently, if $\ker f = \operatorname{cok} f = 0$ and if $\ker g = \operatorname{cok} g = 0$. For a fixed r, the local homology classes (in this case, "holes" or "tunnels" within the local neighborhoods) are labeled in their corresponding locations. In (a), $\mathbb{X}$ contains four pieces of one-dimensional manifolds (colored in red, pink, purple and blue). In (b), $\mathbb{X}$ itself is a one-dimensional manifold. Both p and q belong to different clusters in (a) and the same cluster in (b). However, the notion of "local" be-

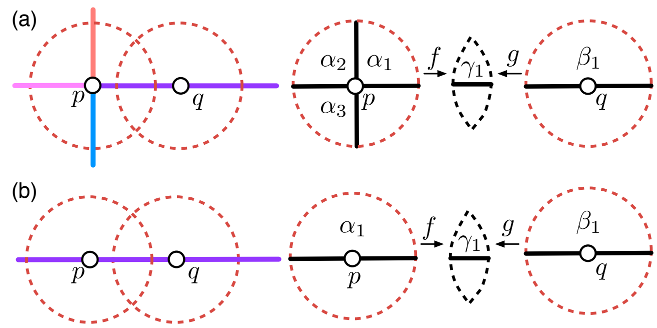


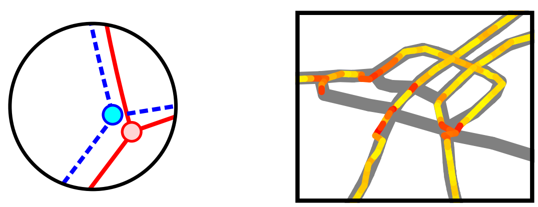
Fig. 2. (a) p and q do not have the same local structure at radius r since $\ker f \neq 0$, that is, extra local structure exists in the neighborhood of p; (b) p and q have the same local structure at radius r since both f and g are isomorphisms, that is, local structures of p and q map to the local structure of the intersection bijectively.

comes unclear in the context of the uncertainty induced from sampling, therefore a "persistence" version of the above local homology intersection map is constructed to define "same local structure" across multiple scales.

3.2. Road Network Comparison

Road networks are changing every day, due to evolving city architecture caused by new road construction, temporary (or permanent) road closures, etc. Understanding *where* and *by how much* the road network has changed is a new problem in transportation research. Traditional algorithms for comparing road networks have been heuristic in nature, failing to provide theoretical guarantees. The first non-heuristic distance measure between road networks was presented in [9]. This algorithm, explained next, uses PLH.

Suppose $\mathbb{X}$ and $\mathbb{Y}$ are two graphs embedded in D, a compact subspace of $\mathbb{R}^d$, and let $\varepsilon > 0$ be a *locality parameter* given a priori. The graphs $\mathbb{X}$ and $\mathbb{Y}$ represent the two road networks that we wish to compare. For each $x \in D$, we look at $V_x^\varepsilon = B_\varepsilon(x)$, the ball centered for $\mathbb{X}$ at x of radius ε. The *local signature* at x (for locality parameter ε) is the persistence diagram $S_\mathbb{X}(x, \varepsilon) := \mathrm{Dgm}(\mathbb{X}; V_x^\varepsilon)$. Likewise, we define the local signature for $\mathbb{Y}$ as $S_\mathbb{Y}(x, \varepsilon) := \mathrm{Dgm}(\mathbb{Y}; V_x^\varepsilon)$. A local signature at $x \in D$ is a descriptor (in this case, a persistence diagram) that is used to describe the local structure as witnessed by x. The parameter ε is used to determine how far x "can see".



(a) Local Structures (b) Signature restricted to $\mathbb{X}$

Fig. 3. PLH compares structures present in local neighborhoods; see (a). Choosing neighborhoods centered on each point of $\mathbb{X}$, we can visualize the PLH distance between graphs embedded in the same domain; see (b).

Instead of comparing the graphs $\mathbb{X}$ and $\mathbb{Y}$ directly, we compare the local signatures $S_\mathbb{X}(x, \varepsilon)$ and $S_\mathbb{Y}(x, \varepsilon)$. We define the *local distance signature* $\psi_\varepsilon \colon D \to \mathbb{R}$ by $\psi_\varepsilon(x) = W_\infty(S_\mathbb{X}(x, \varepsilon), S_\mathbb{Y}(x, \varepsilon))$, where $W_\infty(\cdot, \cdot)$ is the Bottleneck distance as defined in the previous section; see Figure 3, which illustrates the distance between two graphs generated from GPS trajectories in Athens.[1]

While the local signatures are very insightful, sometimes it is necessary to quantify the distance between two graphs with a single number. To this end, we integrate the local distance over D as well as over a range of values for ε in order to obtain a distance between our graphs $\mathbb{X}$ and $\mathbb{Y}$:

Definition 3.1. The *PLH distance metric* is:

$$d^{LH}(\mathbb{X}, \mathbb{Y}) = \int_0^{r_1} \omega(r) \int_D \eta(x) \psi_\varepsilon(x) \, dx \, d\varepsilon,$$

where $\eta \colon D \to \mathbb{R}$ and $\omega \colon [0, r_1] \to \mathbb{R}$ are non-negative weight functions that integrate to unity.

We note here that as long as both weight functions are everywhere positive, d^{LH} is a metric; see [9]. The PLH Metric is a powerful tool in road network analysis, as it provides an actual metric between road networks (represented as graphs embedded in the plane), as well as provides a visualization of the differences between the networks. In [22], this approach is compared against other distances between embedded networks.

[1] Data is available at www.mapconstruction.org.

3.3. Other Applications

PLH can be used in (local) dimension estimation [10, 23]. Its variant, persistent local cohomology, could be used to analyze branching structures in high-dimensional point cloud data for scientific visualization [24]. Stratification learning based on PLH could also be formulated in the context of graph reconstruction from noisy point cloud, to classify points as either vertices or edges, and to determine the degree of the vertices [25, 26].

Thinking of PLH as a local feature, it can be used as an input to a machine learning algorithm. Bendich et al. [10] describe a method, called multi-scale local shape analysis (MLSA), for extracting local geometric and topological structures in data sets. In particular, they use PLH as features in a machine learning algorithm to classify LIDAR data sets, demonstrating that MLSA outperforms PLH and PCA alone, when used to train an SVM classifier.

4. DISCUSSION

In this paper, we defined persistent local homology (PLH), illustrating the power of this tool in several applications, including stratified learning and clustering, road network analysis, and machine learning. We have discussed single parameter filtrations, but we acknowledge that the study of PLH can also benefit from advances in multi-parameter persistence, allowing us to vary multiple parameters (e.g., both α and r) simultaneously. Moreover, using PLH in new application domains will present new and interesting challenges. For example, PLH can be applied in scientific visualization and medical image analysis, where one studies structures such as vascular networks.

5. REFERENCES

[1] P. Bendich, D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov, “Inferring local homology from sampled stratified spaces,” *FOCS*, pp. 536 – 546, 2007.

[2] H. Edelsbrunner, D. Letscher, and A. J. Zomorodian, “Topological persistence and simplification,” *DCG*, vol. 28, pp. 511–533, 2002.

[3] G. Carlsson, “Topology and data,” *Bull. AMS*, vol. 46, no. 2, pp. 255–308, 2009.

[4] R. Ghrist, “Barcodes: The persistent topology of data,” *Bull. AMS*, vol. 45, pp. 61–75, 2008.

[5] H. Edelsbrunner and J. Harer, *Computational Topology: An Introduction*, AMS, 2010.

[6] P. Skraba and B. Wang, “Approximating local homology from samples,” *SODA*, pp. 174–192, 2014.

[7] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, “Stability of persistence diagrams,” *SoCG*, pp. 263–271, 2005.

[8] F. Chazal, D. Cohen-Steiner, M. Glisse, L.J. Guibas, and S.Y. Oudot, “Proximity of persistence modules and their diagrams,” *SoCG*, pp. 237–246, 2009.

[9] M. Ahmed, B. T. Fasy, and C. Wenk, “Local persistent homology based distance between maps,” *ACM SIGSPATIAL GIS*, 2014.

[10] P. Bendich, E. Gasparovic, J. Harer, R. Izmailov, and L. Ness, “Multi-scale local shape analysis for feature selection in machine learning applications,” *IJCNN*, 2015.

[11] P. Bendich, S. Mukherjee, and B. Wang, “Local homology transfer and stratification learning,” *SODA*, pp. 1355–1370, 2012.

[12] D. Attali, A. Lieutier, and D. Salinas, “Efficient data structure for representing and simplifying simplicial complexes in high dimensions,” *SoCG*, pp. 501–509, 2011.

[13] D. R. Sheehy, “Linear-size approximations to the Vietoris-Rips filtration,” *DCG*, vol. 49, no. 4, pp. 778–796, 2013.

[14] T. K. Dey, F. Fan, and Y. Wang, “Graph induced complex on point data,” *SoCG*, pp. 107–116, 2013.

[15] M. Goresky and R. MacPherson, *Stratified Morse Theory*, Springer-Verlag, 1988.

[16] S. Weinberger, “The topological classification of stratified spaces,” *Chicago Lect. Math.*, 1994.

[17] R. Vidal, Y. Ma, and S. Sastry, “Generalized principal component analysis,” *TPAMI*, vol. 27, pp. 1945 – 1959, 2005.

[18] G. Haro, G. Randall, and G. Sapiro, “Stratification learning: Detecting mixed density and dimensionality in high dimensional point clouds,” *NIPS*, vol. 17, 2005.

[19] M. Goresky and R. MacPherson, “Intersection homology i,” *Topology*, vol. 19, pp. 135–162, 1982.

[20] P. Bendich, *Analyzing Stratified Spaces Using Persistent Versions of Intersection and Local Homology*, Ph.D. thesis, Duke University, 2008.

[21] P. Bendich and J. Harer, “Persistent intersection homology,” *FoCM*, vol. 11, pp. 305–336, 2011.

[22] M. Ahmed, B. T. Fasy, and C. Wenk, “New techniques in road network comparison,” *Grace Hopper Celebr. Women Comput.*, 2014.

[23] T. K. Dey, F. Fan, and Y. Wang, “Dimension detection with local homology,” *CCCG*, 2014.

[24] B. Wang, B. Summa, V. Pascucci, and M. Vejdemo-Johansson, “Branching and circular features in high dimensional data,” *TVCG*, vol. 17, no. 12, pp. 1902–1911, 2011.

[25] A. Chernov and V. Kurlin, “Reconstructing persistent graph structures from noisy images,” *Image-a*, vol. 3, no. 5, pp. 19–22, 2013.

[26] M. Aanjaneya, F. Chazal, D. Chen, M. Glisse, L. Guibas, and D. Morozov, “Metric graph reconstruction from noisy data,” *IJCGA*, vol. 22, no. 04, pp. 305–325, 2012.