Dimension Detection by Local Homology

Tamal K. Dey* Fengtao Fan† Yusu Wang‡

Abstract

Detecting the dimension of a hidden manifold from a point sample has become an important problem in the current data-driven era. Indeed, estimating the shape dimension is often the first step in studying the processes or phenomena associated to the data. Among the many dimension detection algorithms proposed in various fields, a few can provide theoretical guarantee on the correctness of the estimated dimension. However, the correctness usually requires certain regularity of the input: the input points are either uniformly randomly sampled in a statistical setting, or they form the so-called (ε, δ)-sample which can be neither too dense nor too sparse.

Here, we propose a purely topological technique to detect dimensions. Our algorithm is provably correct and works under a more relaxed sampling condition: we do not require uniformity, and we also allow Hausdorff noise. Our approach detects dimension by determining local homology. The computation of this topological structure is much less sensitive to the local distribution of points, which leads to the relaxation of the sampling conditions. Furthermore, by leveraging various developments in computational topology, we show that this local homology at a point z can be computed *exactly* for manifolds using Vietoris-Rips complexes whose vertices are confined within a local neighborhood of z. We implement our algorithm and demonstrate the accuracy and robustness of our method using both synthetic and real data sets.

*Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210, USA. Email: tamaldey@cse.ohio-state.edu

†Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210, USA. Email: fanf@cse.ohio-state.edu

‡Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210, USA. Email: yusu@cse.ohio-state.edu

1 Introduction

A fundamental problem in the current data-centric era is to estimate various qualitative structures from input data. Very often, the data is represented as a set of points sampled from a hidden domain. In particular, recent years have witnessed tremendous interest and progress in the field of *manifold learning*, where the hidden domain is assumed to be a manifold M embedded in an ambient Euclidean space $\mathbb{R}^d$. The intrinsic dimension of the manifold M is one of the simplest, yet still very important, quantities that one would like to infer from input data. Indeed, the dimension of M reflects the degree of freedom of the dynamic process that generates the data, and/or the number of variables necessary to describe the hidden domain. Hence, its estimation is crucial to our understanding of the processes or phenomena associated to the data.

In this paper, we present an algorithm to estimate the intrinsic dimension of a manifold M from a set of noisy point samples $P \subset \mathbb{R}^d$ *on and around* M. Our algorithm is based on the topological concept of local homology which was first investigated by Bendich et al. in the discrete setting [2]. We show that our estimation is provably correct under appropriate sampling conditions and choice of parameters.

Related work. The problem of dimension estimation has been studied in various fields including pattern recognition, artificial intelligence and machine learning; see e.g., surveys [5, 29]. If the domain of interest is linear, then the principal component analysis (PCA) [19] is perhaps the most popular method to estimate its dimension. However, PCA fails for non-linear domains and the curvature of the domain tends to cause PCA to overestimate the dimension. Fukunaga and Olsen pioneered the idea of using a local PCA applied to points within small neighborhoods for the non-linear case [15], and several variants have been developed along this direction [4, 22]. In particular, Little et al. developed a multi-scaled version of the local PCA idea [22] that can achieve certain guarantee for points possibly corrupted with Gaussian noise, but uniformly sampled from a hidden manifold. A different approach estimates the manifold dimension based on the growth rate of the volume (or some analog of it) of an intrinsic ball [6, 14, 17, 18, 25]. Both types of approaches above usually work in the statistical setting, where the input points are assumed to be sampled from some probabilistic distribution whose support is concentrated on the hidden manifold.

In the computational geometry community, Dey et al. [11] provided the first provably correct approach to estimate the dimension of a manifold M from a so-called (ε, δ)-sample of M, which enforces a regularity of the point samples by requiring that these points are both ε-dense and δ-sparse. Their approach requires constructing the Voronoi diagram for input points, the computational cost of which becomes prohibitive when the ambient dimension is high. Requiring the same (ε, δ)-sampling condition from input points, Giesen and Wagner [16] introduced the so-called adaptive neighborhood graph, and then locally fit (approximately) the best affine subspace under the L_∞ norm to each sample point p and its neighbors in this graph. The time complexity of their algorithm is exponential only in the intrinsic dimension and the detected dimension is correct for appropriate parameters. Cheng et al. improved this result by applying a local PCA to each sample point and its neighbors in the adaptive neighborhood graph [9]. They also showed that a small amount of Hausdorff noise (of the order ε^2 times the local feature size) and a sparse set of outliers can be tolerated in the input points. More recently, Cheng et al. [8] proposed an algorithm to estimate dimension by detecting the so-called slivers. This algorithm works in a statistical setting, and assumes that the input points are sampled from the hidden manifold using a Poisson process without noise.

In this paper we develop a dimension-detection method based on the topological concept of local homology. The idea of using local homology to understand spaces from sampled points was first proposed by Bendich et al. [2]. Specifically, they introduced multi-scale representations of local homology to infer on stratified spaces, and developed algorithms to compute these representations using the weighted Delaunay triangulation. This line of work was further developed in [3] where the so-called local homology transfer was proposed to cluster points from different strata. In a recent paper [27], Skraba and Wang proposed to approximate the multi-scale representations of local homology using families of Rips complexes. Rips com-

plexes are more suitable than the Delaunay triangulations for points sampled from *low* dimensional compact sets embedded in *high* dimensional space and have attracted much attention in topology inference [1, 7, 27].

Our results. Given a smooth m-dimensional manifold M embedded in $\mathbb{R}^d$, the local homology group $\mathsf{H}(\mathsf{M}, \mathsf{M}-z)$ at a point $z \in \mathsf{M}$ is isomorphic to the reduced homology group of a m-dimensional sphere, that is $\mathsf{H}(\mathsf{M}, \mathsf{M}-z) \cong \tilde{\mathsf{H}}(\mathbf{S}^m)$. Hence, given a set of noisy sample points P of M, we aim to detect the dimension of M by estimating $\mathsf{H}(\mathsf{M}, \mathsf{M}-z)$ from P. Specifically, we assume that P is an ε-sample[1] of M in the sense that the Hausdorff distance between P and M is at most ε. Our main result is that by inspecting two nested neighborhoods around a sample point $p \in P$ and considering certain relative homology groups computed from the Rips complexes induced by points within these neighborhoods, one can recover the local homology *exactly*; see Theorem 5.3. This in turn provides a provably correct dimension-detection algorithm for an ε-sample P of a hidden manifold M when ε is small enough.

Compared with previous provable results in [8, 9, 11, 14, 16, 22], our theoretical guarantee on the estimated dimension is obtained with a more relaxed sampling condition on P. Specifically, there is no uniformity requirement for the sample points P, which was required by all previous dimension-estimation algorithms with theoretical guarantees: either in the form of a uniform random sampling in the statistical setting [8, 14, 22] or the (ε, δ)-sampling in the deterministic setting [9, 11, 16]. We also allow larger amount of noise (ε vs. ε^2 as in [8]). Such a relaxation in the sampling condition is primarily made possible by considering the topological information, which is much less sensitive to the distribution of points compared to the approaches based on local fitting.

In Section 6, we provide preliminary experimental results of our algorithm on both synthetic and real data. For synthetic data our method detects the right dimension robustly. For real data some of which are laden with high noise and undersampling, not all points return the correct dimension. But, taking advantage of the fact that local homology is trivial in all but zero and intrinsic dimension of the manifold, we can eliminate most false positives and estimate the correct dimension from appropriately chosen points.

Finally, we remark that similar to the recent work in [27], our computation of local homology uses the Rips complex, which is much easier to construct than the ambient Delaunay triangulation as was originally required in [2]. Different from [27], we aim to compute $\mathsf{H}(\mathsf{M}, \mathsf{M}-z)$ *exactly* for the special case when M is a manifold, while the work in [27] *approximates* the multiscale representations of local homology (the persistence diagram of certain filtration) for more general compact sets. We also note that, unlike [27] our algorithm operates with Rips complexes that span vertices within a local neighborhood, thus saving computations. The goals from these two works are somewhat complementary and the two approaches address different technical issues.

2 Preliminaries and Notations

Manifold and sample. Let M be a compact smooth m-dimensional manifold without boundary embedded in an Euclidean space $\mathbb{R}^d$. The *reach* $\rho(\mathsf{M})$ is the minimum distance of any point in M to its medial axis. A finite point set $P \subset \mathbb{R}^d$ is an *ε-sample* of M if every point $z \in \mathsf{M}$ satisfies $d(z, P) \leq \varepsilon$ and every point $p \in P$ satisfies $d(p, \mathsf{M}) \leq \varepsilon$; in other words, the *Hausdorff distance* between P and M is at most ε.

Balls. An Euclidean closed ball with radius r and center z is denoted $B_r(z)$. The open ball with the same center and radius is denoted $\mathring{B}_r(z)$ and its complement $\mathbb{R}^d \setminus \mathring{B}_r(z)$ is denoted $B^r(z)$.

Homology. We denote the i-th dimensional homology group of a topological space X as $\mathsf{H}_i(X)$. We drop i and write $\mathsf{H}(X)$ when a statement holds for all dimensions. We mean by $\mathsf{H}(X)$ the singular homology if X

[1] Note that this definition of ε-sample allows points in P to be ε distance off the manifold M. Our ε-sampling condition is with respect to the *reach* of M while that used in [8, 9, 11, 16] is with respect to local feature size and thus adaptive.

is a manifold or a subset of $\mathbb{R}^d$, and simplicial homology if X is a simplicial complex. Both homologies are assumed to be defined with $\mathbb{Z}_2$ coefficients. We make similar assumptions to denote the relative homology groups $\mathsf{H}(X, A)$ for $A \subseteq X$. Notice that both $\mathsf{H}(X)$ and $\mathsf{H}(X, A)$ are vector spaces because they are defined with $\mathbb{Z}_2$ coefficients. The following two known results will be used several times in this paper.

Proposition 2.1 ([7]) *Let $\mathsf{H}(A) \to \mathsf{H}(B) \to \mathsf{H}(C) \to \mathsf{H}(D) \to \mathsf{H}(E) \to \mathsf{H}(F)$ be a sequence of homomorphisms. If $\operatorname{rank}(\mathsf{H}(A) \to \mathsf{H}(F)) = \operatorname{rank}(\mathsf{H}(C) \to \mathsf{H}(D)) = k$, then $\operatorname{rank}(\mathsf{H}(B) \to \mathsf{H}(E)) = k$.*

Proposition 2.2 (Steenrod-five lemma (Lemma 24.3 in [23])) *Suppose we have the commutative diagram of homology groups and homomorphisms:*

$$\begin{array}{ccccccccc}
\mathsf{H}_i(A) & \longrightarrow & \mathsf{H}_i(X) & \longrightarrow & \mathsf{H}_i(X, A) & \longrightarrow & \mathsf{H}_{i-1}(A) & \longrightarrow & \mathsf{H}_{i-1}(X) \\
\downarrow{\scriptstyle f_1} & & \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle f_3} & & \downarrow{\scriptstyle f_4} & & \downarrow{\scriptstyle f_5} \\
\mathsf{H}_i(B) & \longrightarrow & \mathsf{H}_i(Y) & \longrightarrow & \mathsf{H}_i(Y, B) & \longrightarrow & \mathsf{H}_{i-1}(B) & \longrightarrow & \mathsf{H}_{i-1}(Y)
\end{array}$$

where the horizontal sequences are exact. If f_1, f_2, f_4, and f_5 are isomorphisms, so is f_3.

Overview of approach. We are given an ε-sample $P = \{p_i\}_{i=1}^n$ of a compact smooth m-manifold M embedded in $\mathbb{R}^d$. However, the intrinsic dimension m of M is not known, and our goal is to estimate m from the point sample P. Note that for any point $z \in \mathsf{M}$, we have that $\mathsf{H}(\mathsf{M}, \mathsf{M} - z) \cong \tilde{\mathsf{H}}(\mathbf{S}^m)$ where $\tilde{\mathsf{H}}(\cdot)$ denotes the reduced homology. Thus $\operatorname{rank}(\mathsf{H}_i(\mathsf{M}, \mathsf{M} - z)) = 1$ if and only if $i = m$. Hence, if we can compute the rank of $\mathsf{H}_i(\mathsf{M}, \mathsf{M} - z)$ for every i, then we can recover the dimension of M. This is the approach we will follow. In Section 4, we first relate $\mathsf{H}(\mathsf{M}, \mathsf{M} - z)$ with the topology of the offset of the point set P. This requires us to inspect the deformation retraction from the offset to M carefully. The relation to the offset, in turns, allows us to provably recover the rank of $\mathsf{H}(\mathsf{M}, \mathsf{M} - z)$ using the so-called Vietoris Rips complex, which we detail in Section 5. One key ingredient here is to use only local neighborhoods of a sample point to obtain the estimate. First, in Section 3, we derive several technical results to prepare for the development of our approach in Section 4 and 5.

3 Local Homology of M and its Offsets

Local homology $\mathsf{H}(\mathsf{M}, \mathsf{M} - z)$. In this section, we develop a few results that we use later. First, we relate the target local homology groups $\mathsf{H}(\mathsf{M}, \mathsf{M} - z)$ to some other local homology which becomes useful later for connecting to the local homology of Rips complexes that are ultimately used in the algorithm. We start by quoting the following known result:

Proposition 3.1 ([10]) *Let $B_r(p)$ be a closed Euclidean ball so that it intersects the m-manifold M in more than one point. If $r < \rho(\mathsf{M})$, then $\mathsf{M} \cap B_r(p)$ is a closed topological m-ball.*

Proposition 3.2 *Let $D \subset \mathsf{M}$ be a closed topological m-ball from the m-manifold M, and $z \in \mathsf{M}$ a point contained in the interior $\mathring{D}$ of D. Then $\mathsf{H}(\mathsf{M}, \mathsf{M} - \mathring{D}) \xrightarrow{i_*} \mathsf{H}(\mathsf{M}, \mathsf{M} - z)$ is an isomorphism.*

Proof: Consider the following diagram where the two horizontal sequences are exact and all vertical maps are induced by inclusions:

$$\begin{array}{ccccccccc}
\mathsf{H}_i(\mathsf{M} - \mathring{D}) & \longrightarrow & \mathsf{H}_i(\mathsf{M}) & \longrightarrow & \mathsf{H}_i(\mathsf{M}, \mathsf{M} - \mathring{D}) & \longrightarrow & \mathsf{H}_{i-1}(\mathsf{M} - \mathring{D}) & \longrightarrow & \mathsf{H}_{i-1}(\mathsf{M}) \\
\downarrow{\scriptstyle i'_*} & & \downarrow{\scriptstyle \cong} & & \downarrow{\scriptstyle i_*} & & \downarrow{\scriptstyle i'_*} & & \downarrow{\scriptstyle \cong} \\
\mathsf{H}_i(\mathsf{M} - z) & \longrightarrow & \mathsf{H}_i(\mathsf{M}) & \longrightarrow & \mathsf{H}_i(\mathsf{M}, \mathsf{M} - z) & \longrightarrow & \mathsf{H}_{i-1}(\mathsf{M} - z) & \longrightarrow & \mathsf{H}_{i-1}(\mathsf{M})
\end{array}$$

As all vertical homomorphisms are induced by inclusions, the above diagram commutes, see Theorem 5.8 in Rotman [26]. Consider the inclusion $(\mathsf{M} - \mathring{D}) \overset{i'}{\hookrightarrow} (\mathsf{M} - z)$. Since D is a closed topological ball, $\mathsf{M} - z$ deformation retracts to $\mathsf{M} - \mathring{D}$. The inclusion i' is a homotopy inverse of the retraction $(\mathsf{M} - z) \to (\mathsf{M} - \mathring{D})$ and hence i'_* is an isomorphism. Since the first, second, fourth and fifth vertical homomorphisms in the above diagram are isomorphisms, i_* is also an isomorphism by Proposition 2.2. ■

We can extend Proposition 3.2 a little further. See Appendix A for the proof.

Proposition 3.3 *Let D_1 and D_2 be two closed topological balls containing z in the interior where $D_1 \subseteq D_2 \subseteq \mathsf{M}$. The inclusion-induced homomorphisms i'_* and i_* in the following sequence are isomorphisms:*

$$\mathsf{H}(\mathsf{M}, \mathsf{M} - \mathring{D}_2) \overset{i'_*}{\to} \mathsf{H}(\mathsf{M}, \mathsf{M} - \mathring{D}_1) \overset{i_*}{\to} (\mathsf{M}, \mathsf{M} - z).$$

Local homology of the offset. Later we wish to relate the local homology $\mathsf{H}(\mathsf{M}, \mathsf{M} - z)$ at a point z to the local homology of an α-offset of an ε-sample $P = \{p_i\}_{i=1}^n$, defined as

$$\mathbb{X}_\alpha = \cup_{i=1}^n B_\alpha(p_i), \text{ the union of balls centered at every } p_i \text{ with radius } \alpha.$$

For this, we will need a map to connect the two spaces, which is provided by the following projection map:

$$\pi_\alpha : \mathbb{X}_\alpha \to \mathsf{M} \text{ given by } x \mapsto \operatorname{argmin}_{z \in \mathsf{M}} d(x, z).$$

Choose $\alpha < \rho(\mathsf{M}) - \varepsilon$. Since P is an ε-sample, no point of $\mathbb{X}_\alpha$ is $\rho(\mathsf{M})$ or more away from M. This means that no point of the medial axis of M is included in $\mathbb{X}_\alpha$. Therefore, the map π is well defined. Furthermore, by the following result of [24], π is a deformation retraction for appropriate choices of parameters. In fact, under this projection map, the pre-image of a point has a nice structure (star-shaped).

Proposition 3.4 (pp.22, [24]) *If P is an ε-sample of M with reach $\rho = \rho(\mathsf{M})$ where $0 < \varepsilon < (3 - \sqrt{8})\rho$ and $\alpha \in (\frac{(\varepsilon+\rho) - \sqrt{\varepsilon^2+\rho^2-6\varepsilon\rho}}{2}, \frac{(\varepsilon+\rho) + \sqrt{\varepsilon^2+\rho^2-6\varepsilon\rho}}{2})$, then, for any $x \in \pi_\alpha^{-1}(z)$, the segment xz lies in $\pi_\alpha^{-1}(z)$.*

For convenience denote $\theta_1 = \frac{(\varepsilon+\rho) - \sqrt{\varepsilon^2+\rho^2-6\varepsilon\rho}}{2}$ and $\theta_2 = \frac{(\varepsilon+\rho) + \sqrt{\varepsilon^2+\rho^2-6\varepsilon\rho}}{2}$ and observe that $\varepsilon \leq \theta_1$ and $\theta_2 \leq \rho(\mathsf{M}) - \varepsilon$ for $\varepsilon, \rho > 0$. We have:

Proposition 3.5 *Let $0 < \varepsilon < (3 - \sqrt{8})\rho(\mathsf{M})$ and $\theta_1 \leq \alpha \leq \theta_2$. Let $\mathbb{A}_\alpha = \pi_\alpha^{-1}(\mathsf{N})$ where $\mathsf{N} \subseteq \mathsf{M}$ may be either an open or a closed subset. Then $\pi_\alpha : \mathbb{A}_\alpha \to \mathsf{N}$ is a retraction and N is a deformation retract of $\mathbb{A}_\alpha$.*

Proof: Notice that due to Proposition 3.4, $\pi_\alpha^{-1}(z)$ is star shaped meaning that every point $x \in \pi_\alpha^{-1}(z)$ has the segment xz lying in $\pi_\alpha^{-1}(z)$. It follows that $\mathsf{N} \subseteq \mathbb{A}_\alpha$ and there exists a straight line deformation retraction $F : \mathbb{A}_\alpha \times I \to \mathbb{A}_\alpha$ defined as $F(x, t) = (1 - t)x + t\pi(x)$. The proposition then follows. ■

Based on the above observation, the map $\pi_\alpha : (\mathbb{X}_\alpha, \mathbb{A}_\alpha) \to (\mathsf{M}, \mathsf{N})$ seen as a map on the pairs provides an isomorphism at the homology level.

Proposition 3.6 *Let $0 < \varepsilon < (3 - \sqrt{8})\rho$ and $\theta_1 \leq \alpha \leq \theta_2$. The homomorphism $\pi_{\alpha *} : \mathsf{H}(\mathbb{X}_\alpha, \mathbb{A}_\alpha) \to \mathsf{H}(\mathsf{M}, \mathsf{N})$ is an isomorphism.*

Proof: The map π_α provides the following commutative diagram (Theorem 5.8, Rotman [26]):

$$\begin{array}{ccccccccccc}
\longrightarrow & \mathsf{H}_i(\mathbb{A}_\alpha) & \longrightarrow & \mathsf{H}_i(\mathbb{X}_\alpha) & \longrightarrow & \mathsf{H}_i(\mathbb{X}_\alpha, \mathbb{A}_\alpha) & \longrightarrow & \mathsf{H}_{i-1}(\mathbb{A}_\alpha) & \longrightarrow & \mathsf{H}_{i-1}(\mathbb{X}_\alpha) & \longrightarrow \\
 & \downarrow{\scriptstyle \pi_{\alpha *}} & & \downarrow{\scriptstyle \pi_{\alpha *}} & & \downarrow{\scriptstyle \pi_{\alpha *}} & & \downarrow{\scriptstyle \pi_{\alpha *}} & & \downarrow{\scriptstyle \pi_{\alpha *}} & \\
\longrightarrow & \mathsf{H}_i(\mathsf{N}) & \longrightarrow & \mathsf{H}_i(\mathsf{M}) & \longrightarrow & \mathsf{H}_i(\mathsf{M}, \mathsf{N}) & \longrightarrow & \mathsf{H}_{i-1}(\mathsf{N}) & \longrightarrow & \mathsf{H}_{i-1}(\mathsf{M}) & \longrightarrow
\end{array}$$

The first, second, fourth, and fifth vertical maps are restrictions of $\pi_{\alpha *}$ and thus are all isomorphisms by Proposition 3.5. It follows from Proposition 2.2 that the third vertical map is an isomorphism as well. ■

Proposition 3.7 *Let $0 < \varepsilon < (3-\sqrt{8})\rho$, and $\theta_1 \leq \alpha < \alpha' \leq \theta_2$. Let $\mathsf{N} \subset \mathsf{N}'$ be two closed (or open) sets of M, and $\mathbb{A}_\alpha = \pi_\alpha^{-1}(\mathsf{N})$ and $\mathbb{A}_{\alpha'} = \pi_{\alpha'}^{-1}(\mathsf{N}')$. Denoting by $\mathrm{im}(\cdot)$ the image of a map, we have*

$$\mathrm{im}\left(\mathsf{H}(\mathbb{X}_\alpha, \mathbb{A}_\alpha) \to \mathsf{H}(\mathbb{X}_{\alpha'}, \mathbb{A}_{\alpha'})\right) \cong \mathrm{im}\left(\mathsf{H}(\mathsf{M}, \mathsf{N}) \to \mathsf{H}(\mathsf{M}, \mathsf{N}')\right).$$

Proof: The projection maps π_α and $\pi_{\alpha'}$ (both being maps of pairs) result in the following commutative diagram of pairs.

$$\begin{array}{ccc} (\mathbb{X}_\alpha, \mathbb{A}_\alpha) & \hookrightarrow & (\mathbb{X}_{\alpha'}, \mathbb{A}_{\alpha'}) \\ \downarrow{\scriptstyle \pi_\alpha} & & \downarrow{\scriptstyle \pi_{\alpha'}} \\ (\mathsf{M}, \mathsf{N}) & \hookrightarrow & (\mathsf{M}, \mathsf{N}') \end{array}$$

This diagram induces a commutative diagram at homology level, where $\pi_{\alpha *}$ and $\pi_{\alpha' *}$ are isomorphisms by Proposition 3.6. The claim now is immediate by the Persistence Equivalence Theorem [12], page 159. ■

4 Local Interleaving of Offsets

Let $p \in P$ be any sample point. We show how to obtain the local homology of the projected point $\pi(p)$ on M from pairs of p's local neighborhoods in $\mathbb{X}_\alpha$. The results from the previous section already allow us to relate the local homology of the projected point $\pi(p)$ with the local homology of some local neighborhoods in $\mathbb{X}_\alpha$ (which are the pre-image of some sets in M). We now use interleaving to relate them further to local neighborhoods that are intersection of $\mathbb{X}_\alpha$ with Euclidean balls. Since $\pi(p)$ plays an important role here, we use a special symbol $\bar{p} = \pi(p)$ for it. For convenience, we introduce notations (see Figure 1):

$$\mathbb{M}_{\alpha,\beta} = \pi_\alpha^{-1}(\mathring{B}_\beta(p) \cap \mathsf{M}),\ \mathbb{M}^{\alpha,\beta} = \mathbb{X}_\alpha - \mathbb{M}_{\alpha,\beta},\ \text{and } \mathbb{B}_{\alpha,\beta} = \mathring{B}_\beta(p) \cap \mathbb{X}_\alpha,\ \mathbb{B}^{\alpha,\beta} = \mathbb{X}_\alpha - \mathbb{B}_{\alpha,\beta}.$$

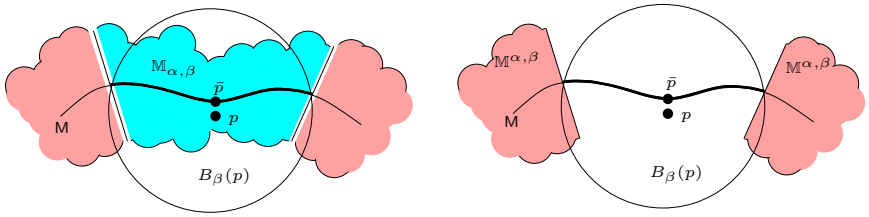


Figure 1: The spaces $\mathbb{M}_{\alpha,\beta}$ shown in cyan (left) and $\mathbb{M}^{\alpha,\beta}$ shown in pink (right).

The following simple observation follows from Propositions 3.2, 3.1, and 3.5.

Proposition 4.1 *Let $D_\beta = B_\beta(p) \cap \mathsf{M}$. For $0 < \varepsilon < (3-\sqrt{8})\rho$, $\varepsilon < \beta < \rho(\mathsf{M})$ and $\theta_1 \leq \alpha \leq \theta_2$, the maps $\pi_{\alpha *}$ and i_* are isomorphisms in the sequence: $\mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\beta}) \xrightarrow{\pi_{\alpha *}} \mathsf{H}(\mathsf{M}, \mathsf{M} - \mathring{D}_\beta) \xrightarrow{i_*} \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p})$.*

Now set $\delta = \alpha + 3\varepsilon$. Consider any $z \in \mathsf{M}$. Since any point $x \in \pi_\alpha^{-1}(z)$ resides within a ball $B_\alpha(p_i)$ for some $p_i \in P$, we have that

$$d(x,z) = d(x, \pi(x)) \leq d(x, \pi(p_i)) \leq d(x, p_i) + d(p_i, \pi(p_i)) \leq \alpha + \varepsilon = \delta - 2\varepsilon. \tag{1}$$

It follows that for any $\lambda \in (\varepsilon, \rho(\mathsf{M}) - \delta)$ we get the following inclusions(see Appendix B for details):

$$\mathbb{M}_{\alpha,\lambda} \subset \mathbb{B}_{\alpha,\lambda+\delta} \subset \mathbb{M}_{\alpha,\lambda+2\delta} \subset \mathbb{B}_{\alpha,\lambda+3\delta} \subset \mathbb{M}_{\alpha,\lambda+4\delta}.$$

Taking the complements, a new filtration in the reverse direction is generated:

$$\mathbb{M}^{\alpha,\lambda+4\delta} \subset \mathbb{B}^{\alpha,\lambda+3\delta} \subset \mathbb{M}^{\alpha,\lambda+2\delta} \subset \mathbb{B}^{\alpha,\lambda+\delta} \subset \mathbb{M}^{\alpha,\lambda}.$$

Considering each space as a topological pair, the nested sequence becomes

$$(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+4\delta}) \subset (\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+3\delta}) \subset (\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+2\delta}) \subset (\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+\delta}) \subset (\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda}) \tag{2}$$

Inclusion between topological pairs induces a homomorphism between their relative homology groups. Therefore, the following relative homology sequence holds.

$$\mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+4\delta}) \to \mathsf{H}(\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+3\delta}) \to \mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+2\delta}) \to \mathsf{H}(\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+\delta}) \to \mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda}) \tag{3}$$

Let $\epsilon \leq \alpha' \leq \rho(\mathsf{M}) - \epsilon$ and $\delta' = \alpha' + 3\varepsilon$. Similar to sequence (2), for any $\lambda' \in (\varepsilon, \rho(\mathbb{M}) - 4\delta')$ we have:

$$(\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'+4\delta'}) \subset (\mathbb{X}_{\alpha'}, \mathbb{B}^{\alpha',\lambda'+3\delta'}) \subset (\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'+2\delta'}) \subset (\mathbb{X}_{\alpha'}, \mathbb{B}^{\alpha',\lambda'+\delta'}) \subset (\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'}) \tag{4}$$

The stated range of λ, λ' is valid if $\alpha, \alpha' < \frac{\rho(\mathsf{M})-13\varepsilon}{4}$. We also need $\theta_1 \leq \alpha, \alpha'$. These two conditions are satisfied for $\varepsilon < \frac{\rho(\mathsf{M})}{22}$. Let $\theta_2' = \frac{\rho(\mathsf{M})-13\varepsilon}{4}$.

Proposition 4.2 *Let $0 < \varepsilon < \frac{\rho(\mathsf{M})}{22}$, and $\theta_1 \leq \alpha \leq \alpha' \leq \theta_2'$. Set $\delta = \alpha + 3\varepsilon$ and $\delta' = \alpha' + 3\varepsilon$. For $\varepsilon < \lambda' < \rho(\mathsf{M}) - 4\delta'$ and $\lambda \geq \lambda' + 2(\alpha' - \alpha)$, we have,*

$$\operatorname{im}\Big(\mathsf{H}(\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+3\delta}) \to \mathsf{H}(\mathbb{X}_{\alpha'}, \mathbb{B}^{\alpha',\lambda'+\delta'})\Big) \cong \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p}). \tag{5}$$

In particular, $\operatorname{im}\big(\mathsf{H}(\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+3\delta}) \to \mathsf{H}(\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+\delta})\big) \cong \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p})$.

Proof: Due to our choice of parameters, we have that $\lambda + 2\delta \geq \lambda' + 2\delta'$. From Eqn (2) and (4), we obtain the following sequence of homomorphisms induced by inclusions:

$$\begin{aligned}&\mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+4\delta}) \to \mathsf{H}(\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+3\delta}) \to \mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+2\delta}) \to \\ &\mathsf{H}(\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'+2\delta'}) \to \mathsf{H}(\mathbb{X}_{\alpha'}, \mathbb{B}^{\alpha',\lambda'+\delta'}) \to \mathsf{H}(\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'}).\end{aligned}$$

We first show

$$\operatorname{im}\Big(\mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+4\delta}) \to \mathsf{H}(\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'})\Big) \cong \operatorname{im}\Big(\mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+2\delta}) \to (\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'+2\delta'})\Big) \cong \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p}). \tag{6}$$

Consider the following commutative diagram where π_α and $\pi_{\alpha'}$ are seen as maps on pairs:

$$\begin{array}{ccc}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+4\delta}) & \hookrightarrow & (\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'}) \\ \downarrow \pi_\alpha & & \downarrow \pi_{\alpha'} \\ (\mathsf{M}, \mathsf{M} - \mathring{D}_{\lambda+4\delta}) & \hookrightarrow & (\mathsf{M}, \mathsf{M} - \mathring{D}_{\lambda'}))\end{array}$$

where $D_\beta = B_\beta(p) \cap \mathsf{M}$. By Proposition 3.3, we have

$$\operatorname{im}\Big(\mathsf{H}(\mathsf{M}, \mathsf{M} - \mathring{D}_{\lambda+4\delta}) \to \mathsf{H}(\mathsf{M}, \mathsf{M} - \mathring{D}_{\lambda'})\Big) \cong \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p}).$$

Hence, $\mathrm{im}\left(\mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+4\delta}) \to \mathsf{H}(\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'})\right) \cong \mathsf{H}(\mathsf{M}, \mathsf{M}-\bar{p})$ by Proposition 3.7. The same argument implies that $\mathrm{im}\left(\mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\lambda+2\delta}) \to (\mathbb{X}_{\alpha'}, \mathbb{M}^{\alpha',\lambda'+2\delta'})\right) \cong \mathsf{H}(\mathsf{M}, \mathsf{M}-\bar{p})$ which establishes the claim in (6). Eqn (5) then follows from Proposition 2.1. In particular, if $\alpha' = \alpha$, we have

$$\mathrm{im}\left(\mathsf{H}(\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+3\delta}) \to \mathsf{H}(\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\lambda+\delta})\right) \cong \mathsf{H}(\mathsf{M}, \mathsf{M}-\bar{p}).$$

■

Finally, we intersect each set with a sufficiently large ball $B_r(p)$ so that we only need to inspect within the neighborhood $B_r(p)$ of p. Specifically, denote $\mathbb{X}_{\alpha,r} = \mathbb{X}_\alpha \cap B_r(p)$ and $\mathbb{X}^\beta_{\alpha,r} = \mathbb{X}_{\alpha,r} \cap B^\beta(p)$. We obtain the next proposition by applying the Excision theorem (details in Appendix B).

Proposition 4.3 *Let all the parameters satisfy the same conditions as in Proposition 4.2. Then, for $r > \lambda + 5\delta$, we have:*

$$\mathrm{im}\left(\mathsf{H}(\mathbb{X}_{\alpha,r}, \mathbb{X}^{\lambda+3\delta}_{\alpha,r}) \to \mathsf{H}(\mathbb{X}_{\alpha',r}, \mathbb{X}^{\lambda'+\delta'}_{\alpha',r})\right) \cong \mathsf{H}(\mathsf{M}, \mathsf{M}-\bar{p}).$$

In particular, $\mathrm{im}\left(\mathsf{H}(\mathbb{X}_{\alpha,r}, \mathbb{X}^{\lambda+3\delta}_{\alpha,r}) \to \mathsf{H}(\mathbb{X}_{\alpha,r}, \mathbb{X}^{\lambda+\delta}_{\alpha,r})\right) \cong \mathsf{H}(\mathsf{M}, \mathsf{M}-\bar{p})$.

In fact, one can relax the parameters, and the image homology $\mathrm{im}\left(\mathsf{H}(\mathbb{X}_{\alpha,r}, \mathbb{X}^{\beta_2}_{\alpha,r}) \to \mathsf{H}(\mathbb{X}_{\alpha',r}, \mathbb{X}^{\beta_1}_{\alpha',r})\right)$ captures (that is, is isomorphic to) the local homology $\mathsf{H}(\mathsf{M}, \mathsf{M}-\bar{p})$ as long as $\beta_1 \geq \alpha' + 4\varepsilon$, $\beta_2 \geq \beta_1 + \alpha + \alpha' + 6\varepsilon$ and $r > \beta_2 + 2\alpha + 6\varepsilon$.

5 Interleaving Nerves and Rips complexes

We now relate the relative homology of pairs as in Proposition 4.3 to the relative homology of pairs in Rips complexes. Our algorithm works on these pairs of Rips complexes to derive the local homology at a point on M. As before, let $p \in P$ be a point from the sample.

Nerves of spaces. Consider the space $\mathbb{X}_{\alpha,r} = \mathbb{X}_\alpha \cap B_r(p)$. The connection of such spaces with simplicial complexes (Vietoris-Rips complex in particular) is made through the so-called nerve of a cover. In general, let $\mathcal{U}$ be a finite collection of sets. The *nerve $\mathcal{NU}$ of $\mathcal{U}$* is a simplicial complex whose simplices are given by all subsets of $\mathcal{U}$ whose members have a non-empty common intersection. That is,

$$\mathcal{NU} := \{\mathcal{A} \subseteq \mathcal{U} \mid \cap\mathcal{A} \neq \emptyset\}.$$

The set $\mathcal{U}$ forms a *good cover* of the union $\bigcup\mathcal{U}$ if the intersection of any subsets of $\mathcal{U}$ is either empty or contractible. The Nerve Lemma states that if $\mathcal{U}$ is a good cover of $\bigcup\mathcal{U}$, then $\mathcal{NU}$ is homotopic to $\bigcup\mathcal{U}$, denoted by $\mathcal{NU} \approx \bigcup\mathcal{U}$.

Now consider the set of sets $\mathcal{X}_{\alpha,r} = \{B_\alpha(p_i) \cap B_r(p) \mid p_i \in P\}$; note that $\mathbb{X}_{\alpha,r} = \bigcup\mathcal{X}_{\alpha,r}$. Since each set in $\mathcal{X}_{\alpha,r}$ is convex, $\mathcal{X}_{\alpha,r}$ forms a good cover of $\mathbb{X}_{\alpha,r}$ and thus $\mathcal{NX}_{\alpha,r} \approx \mathbb{X}_{\alpha,r}$ by the Nerve Lemma. Furthermore, it follows from Lemma A.5 of [27] that for $r > \beta + 2\alpha$, the set $\mathcal{X}^\beta_{\alpha,r} = \{B_\alpha(p_i) \cap B_r(p) \cap B^\beta(p)\}_{i\in[1,n]}$ also form a good cover of $\bigcup\mathcal{X}^\beta_{\alpha,r} (= \mathbb{X}^\beta_{\alpha,r})$; see Appendix C.1 for details. Thus, we have $\mathcal{NX}^\beta_{\alpha,r} \approx \mathbb{X}^\beta_{\alpha,r}$. We can now convert the relative homology between $\mathbb{X}_{\alpha,r}$ and $\mathbb{X}^\beta_{\alpha,r}$ to the homology of their nerves. In particular, we have the following result. The proof is in Appendix C.2, and it relies heavily on the proof of Lemma 3.4 of [7] which gives a crucial commutative result for the space and its nerve.

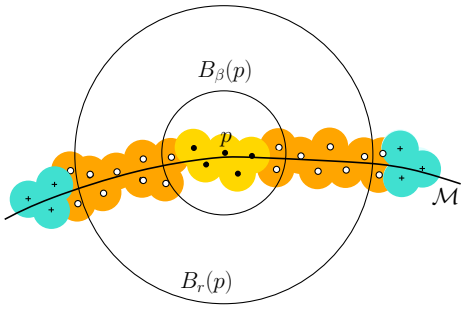
Lemma 5.1 *Let all the parameters satisfy the same conditions as in Proposition 4.2. Then, for $r > \lambda + 5\delta$:*

$$\mathrm{im}\left(\mathsf{H}(\mathcal{NX}_{\alpha,r}, \mathcal{NX}^{\lambda+3\delta}_{\alpha,r}) \to \mathsf{H}(\mathcal{NX}_{\alpha',r}, \mathcal{NX}^{\lambda'+\delta'}_{\alpha',r})\right) \cong \mathsf{H}(\mathsf{M}, \mathsf{M}-\bar{p}).$$

Relating nerves and Rips complexes. First, we recall that for $\alpha \geq 0$, the *Čech complex* $C^\alpha(Q)$ of a point set Q is the nerve of the cover $\{B_\alpha(q_i) : q_i \in Q\}$ of $\cup B_\alpha(q_i) = \mathbb{X}_\alpha$. The *Vietoris-Rips* (Rips in short) complex $\mathcal{R}^\alpha(Q)$ is the maximal complex induced by the edge set $\{(p_j, p_k) \mid d(p_j, p_k) \leq \alpha\}$. It is well known that for any point set Q, the following holds:

$$C^\alpha(Q) \subset \mathcal{R}^{2\alpha}(Q) \subset C^{2\alpha}(Q).$$

Define $P_{\alpha,r} = \{p_i \in P \mid B_\alpha(p_i) \cap B_r(p) \neq \emptyset\}$. Obviously, $P_{\alpha,r}$ forms the vertex set for the nerve $\mathcal{NX}_{\alpha,r}$. Similarly, let $P^\beta_{\alpha,r} = \{p_i \in P_{\alpha,r} \mid B_\alpha(p_i) \cap B^\beta(p) \neq \emptyset\}$ denote the vertex set of $\mathcal{NX}^\beta_{\alpha,r}$. See the figure on right for an example, where the union of solid and empty dots forms the set of points $P_{\alpha,r}$, while $P^\beta_{\alpha,r}$ consists the set of empty dots. Note that from the definition, it follows that $P^\beta_{\alpha,r} \subset P_{\alpha,r}$ and $P^\beta_{\alpha,r} \subset P^{\beta'}_{\alpha,r}$ for $\beta' < \beta$. Furthermore, as the offset $\mathbb{X}_\alpha$ grows, it is immediate that $P_{\alpha,r} \subset P_{\alpha',r}$ and $P^\beta_{\alpha,r} \subset P^\beta_{\alpha',r}$ for $\alpha < \alpha'$.



Each element in the good cover $\mathcal{X}_{\alpha,r}$ or $\mathcal{X}^\beta_{\alpha,r}$ is in the form of $B_\alpha(p_i) \cap B_r(p)$ or $B_\alpha(p_i) \cap B_r(p) \cap B^\beta(p)$. Since the Čech complex of a set is the nerve of the set of balls $B_\alpha(p_i)$, it follows easily that

$$\mathcal{NX}_{\alpha,r} \subset C^\alpha(P_{\alpha,r}) \subset \mathcal{R}^{2\alpha}(P_{\alpha,r}) \text{ and } \mathcal{NX}^\beta_{\alpha,r} \subset C^\alpha(P^\beta_{\alpha,r}) \subset \mathcal{R}^{2\alpha}(P^\beta_{\alpha,r}). \tag{7}$$

Claim 5.2 *(i)* $\mathcal{R}^{2\alpha}(P_{\alpha,r}) \subset \mathcal{NX}_{3\alpha,r}$*, and (ii)* $\mathcal{R}^{2\alpha}(P^\beta_{\alpha,r}) \subset \mathcal{NX}^\beta_{3\alpha,r}$.

Proof: To prove (i), consider an arbitrary simplex $\sigma = [p_0 p_1 \ldots p_\ell] \in \mathcal{R}^{2\alpha}(P_{\alpha,r})$. By definition of Rips complex, $d(p_i, p_j) \leq 2\alpha$ for $0 \leq i, j \leq \ell$. Then, for any point $x \in B_\alpha(p_0) \cap B_r(p)$, we have that $d(x, p_i) < d(x, p_0) + d(p_0, p_i) < 3\alpha$ implying $x \in \cap_{i=0}^\ell B_{3\alpha}(p_i)$ and $(\cap_{i=0}^\ell B_{3\alpha}(p_i)) \cap B_r(p) \neq \emptyset$. In other words, $\sigma \in \mathcal{NX}_{3\alpha,r}$, thus proving Claim (i). Claim (ii) can be shown by a similar argument. ■

Set $\eta_1 = \lambda + 9\alpha + 3\varepsilon$ and $\eta_2 \geq \eta_1 + 12\alpha + 6\varepsilon$ for any $\lambda > \varepsilon$. Combining Eqn (7) and Claim 5.2, we get three nested sequences

$$\mathcal{NX}_{\alpha,r} \subset \mathcal{R}^{2\alpha}(P_{\alpha,r}) \subset \mathcal{NX}_{3\alpha,r} \subset \mathcal{R}^{6\alpha}(P_{3\alpha,r}) \subset \mathcal{NX}_{9\alpha,r}$$
$$\mathcal{NX}^{\eta_1}_{\alpha,r} \subset \mathcal{R}^{2\alpha}(P^{\eta_1}_{\alpha,r}) \subset \mathcal{NX}^{\eta_1}_{3\alpha,r} \subset \mathcal{R}^{6\alpha}(P^{\eta_1}_{3\alpha,r}) \subset \mathcal{NX}^{\eta_1}_{9\alpha,r}$$
$$\mathcal{NX}^{\eta_2}_{\alpha,r} \subset \mathcal{R}^{2\alpha}(P^{\eta_2}_{\alpha,r}) \subset \mathcal{NX}^{\eta_2}_{3\alpha,r} \subset \mathcal{R}^{6\alpha}(P^{\eta_2}_{3\alpha,r}) \subset \mathcal{NX}^{\eta_2}_{9\alpha,r}$$

These give rise to the following sequence of pairs

$$(K_\alpha, K^{\eta_2}_\alpha) \hookrightarrow (R_\alpha, R^{\eta_2}_\alpha) \hookrightarrow (K_{3\alpha}, K^{\eta_2}_{3\alpha}) \hookrightarrow (K_{3\alpha}, K^{\eta_1}_{3\alpha}) \hookrightarrow (R_{3\alpha}, R^{\eta_1}_{3\alpha}) \hookrightarrow (K_{9\alpha}, K^{\eta_1}_{9\alpha})$$

where $K_\alpha = \mathcal{NX}_{\alpha,r}$, $K^\beta_\alpha = \mathcal{NX}^\beta_{\alpha,r}$, $R_\alpha = \mathcal{R}^{2\alpha}(P_{\alpha,r})$ and $R^\beta_\alpha = \mathcal{R}^{2\alpha}(P^\beta_{\alpha,r})$. From Proposition 4.3 and Lemma 5.1, it is immediate that $\mathrm{im}(i_{\alpha *}) \cong \mathrm{im}(i_{3\alpha *}) \cong \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p})$ where $i_{\alpha *}$ and $i_{3\alpha *}$ are induced from $i_\alpha : (K_\alpha, K^{\eta_2}_\alpha) \hookrightarrow (K_{9\alpha}, K^{\eta_1}_{9\alpha})$ and $i_{3\alpha} : (K_{3\alpha}, K^{\eta_2}_{3\alpha}) \hookrightarrow (K_{3\alpha}, K^{\eta_1}_{3\alpha})$. It follows from Proposition 2.1 that $\mathrm{im}(j_{\alpha *}) \cong \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p})$ where $j_{\alpha *}$ is induced from $j_\alpha : (R_\alpha, R^{\eta_2}_\alpha) \hookrightarrow (R_{3\alpha}, R^{\eta_1}_{3\alpha})$. To apply Proposition 4.3, we need the condition required by Eq. 6, which is $\eta_2 + \alpha + 3\varepsilon < \rho(\mathsf{M})$ here. This condition together with $\eta_2 \geq \eta_1 + 12\alpha + 6\varepsilon$ require that $\alpha < \frac{\rho(\mathsf{M}) - 13\varepsilon}{22}$. We also need $\theta_1 \leq \alpha$. Both conditions are satisfied when $0 < \varepsilon < \frac{\rho(\mathsf{M})}{58}$. Thus, we have our main result:

Theorem 5.3 *Let* $0 < \varepsilon < \frac{\rho(\mathsf{M})}{58}$ *and* $\theta_1 \leq \alpha \leq \frac{\rho(\mathsf{M}) - 13\varepsilon}{22}$. *Furthermore, let* η_1 *and* η_2 *be such that* $\varepsilon < \eta_1, \eta_2 < \rho(\mathsf{M})$, $\eta_1 \geq 9\alpha + 4\varepsilon$, *and* $\eta_2 \geq \eta_1 + 12\alpha + 6\varepsilon$. *The inclusion*

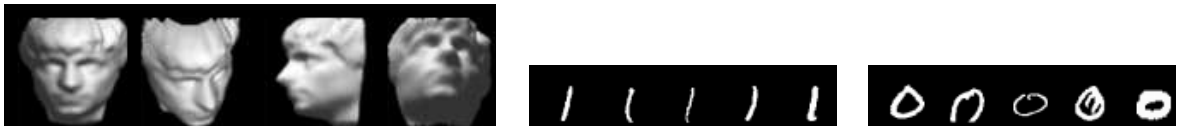
$$j_\alpha : (\mathcal{R}^{2\alpha}(P_{\alpha,r}), \mathcal{R}^{2\alpha}(P^{\eta_2}_{\alpha,r})) \hookrightarrow (\mathcal{R}^{6\alpha}(P_{3\alpha,r}), \mathcal{R}^{6\alpha}(P^{\eta_1}_{3\alpha,r}))$$

satisfies $\mathrm{im}(j_{\alpha *}) \cong \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p})$ *for any* $r \geq \eta_1 + \eta_2$.

Algorithm. Given a sample point $p = p_i$, our algorithm first constructs the necessary Rips complexes as specified in Theorem 5.3 for some parameters $\alpha < \eta_1 < \eta_2 < r$. For simplicity, rewrite $j_\alpha : (A_1, B_1) \hookrightarrow (A_2, B_2)$ where $B_1 \subset A_1 \subset A_2$ and $B_1 \subset B_2 \subset A_2$. After obtaining the necessary Rips complexes, one possible method for computing $\mathrm{im}(j_{\alpha *})$ would be to cone the subcomplexes B_1 and B_2 with a dummy vertex w to obtain an inclusion $\iota : A_1 \cup (w * B_1) \hookrightarrow A_2 \cup (w * B_2)$ where $w * B_j = B_j \cup \{w * \sigma | \sigma \in B_j\}$ is the cone on B_j $(j = 1, 2)$. It is easy to see that $\mathrm{im}(j_{\alpha *}) \cong \mathrm{im}(\iota_*)$. Then, the standard persistent homology algorithm can be applied. However, the cone operations may add many unnecessary simplices slowing down the computation. Instead, we order the simplices in A_2 properly to build a filtration so that the rank of $\mathrm{im}(j_{\alpha *})$ can be read off from the reduced boundary matrix built from the filtration. The details of this algorithm can be found in Appendix D.

6 Experimental results

We present some preliminary experimental results on several synthesized and real data. Recall that our method only needs points in the neighborhood of a base point. While the theoretical result guarantees the correct detection of dimension for correct choices of parameters, in practice, the choice of the base point plays an important role. If the points sample only a patch of a manifold, then the local homology of points near the boundary of that patch will be trivial, which results in plenty of base points with trivial local homology. Furthermore, noise and inadequate density make the dimension estimation difficult. To overcome these hurdles, we explore some practical strategies.



(a) **Head** (b) **D1** (c) **D0**

Figure 2: Image data : rotating head (**Head**), handwritten ones (**D1**) and zeros (**D0**).

For the synthesized data, which is uniform and dense, we take a sparse and uniform subsample from the input as a set of base points. At each base point, the local homology is estimated by our program. We discard the result in which the computed homology is trivial or does not coincide with $\tilde{\mathsf{H}}(\mathbf{S}^n)$ for any n, as these are obviously not correct. The remaining base points return the homology of an n-sphere, that is $\mathrm{rank}(\mathsf{H}_i) = 1$ *iff* $i = n$ for some n. These are called *valid base points*. These points are grouped according to which n-sphere homology they have, and we return the dimension n of the group with most members as the detected dimension.

	SAMPLE POINTS	AVG. NEIGHB.	NOT n-SPHERE	TRIVIAL	n-SPHERE		CORRECT RATIO
$\mathbf{S}^3$	4096	19	0/60	38/60	n=3	22/60	100%(22/22)
$\mathbf{S}^4$	4097	34	0/46	40/46	n=4	6/46	100%(6/6)
$\mathbf{S}^5$	32769	52	0/74	69/74	n=5	5/74	100%(5/5)
$\mathbf{S}^6$	262145	74	0/220	213/220	n=6	7/220	100%(7/7)
Shift	2240	37	0/67	15/67	n=2	52/67	100%(52/52)
M^3	2796	316	0/54	40/54	n=2 n=3	1/54 13/54	92.8%(13/14)

Table 1: Results for synthetic data

For the real data, which mostly comes from a small part of a manifold, we use a different strategy because these data are non-uniform and contain high noise and outliers. Three data **Head**, **D1**, and **D0** (some samples shown in Figure 2) are considered. We first identify some sample points called centers away from the boundary and undersampled regions using a graph based method described in the Appendix E. Then, we estimate the local homology at these points. Table 2 in the Appendix E provides the results on estimated dimensions.

Our synthetic data consists of points sampled from spherical caps of n-spheres $\mathbf{S}^n$ for $n = 3, 4, 5, 6$; a 3-manifold $M^3 \subset \mathbb{R}^{50}$ with boundary (computed from a parametric equation); and a 2D translation of a smaller image within a black image with resolution 60×84(**Shift**) (see [8]). The input for each $\mathbf{S}^n$ is a uniform 0.0125-sample of a spherical cap (thus is a manifold *with* boundary) with no noise. The **Shift** data is also noiseless. The sample points of M^3 is noisy with a 0.05 unit Hausdorff noise. The results on the synthetic data are summarized in Table 1. AVG. NEIGHB. column gives the average number of points in the local neighborhood of each base point used to estimate local homology. CORRECT RATIO column shows the ratio of correct dimension detection over all valid base points. Among all valid base points, our algorithm produces no false positives for all the $\mathbf{S}^n$ data sets. For the noisy sample of M^3, we have only one false positive out of 14 valid points. The high number of points that return trivial homology (5th column) is mainly due to points near the boundary of the manifold. For the **Shift** data, our method detects its dimension 2 with high confidence. The **Shift** was used and compared in [8].

	Shift	**Head**	**D1**	**D0**
Ours	2	3	4	3
SLIVER	3	4	3	2
MLE	4.27	4.31	11.47	14.86
MA	3.35	4.47	10.77	13.93
PN	3.62	3.98	6.22	8.86
LPCA	3	3	5	8.86
ISOMAP	2	3	5	$[3, 6]$

In the table on left, we show comparisons with other methods. Although **Shift** is uniform and noise free, only ISOMAP and ours get the correct dimension. The real data contains 698 images of a rotating head (**Head**, Fig. 2(a)), 6742 images of handwritten ones (**D1**, Fig. 2(b)) and 5923 images of handwritten zeros (**D0**, Fig. 2(c)) from MNIST database. These three data were also explored and compared in [8], where Cheng and Chiu [8] compared their dimension detection method via sliver (SLIVER) with other methods: the maximum likelihood estimation (MLE) [21], the manifold adaptive method (MA) [14], the packing number method (PN) [20], the local PCA (LPCA) [9], and the isomap method (ISOMAP) [28]. Since we test our method on the same data, we include the comparison results on these three data along with **Shift** data from [8] in the table where all rows except the first row are from [8]. Details and statistics of our experiments on real data are presented in the Appendix E.

7 Conclusions

In this paper, we present a topological method to estimate the dimension of a manifold from its point samples with a theoretical guarantee. The use of local topological structures helps to alleviate the dependency of our method on the regularity of point samples, and the use of persistent homology for a pair of homology groups (instead of a single homology group) helps to increase its robustness.

It will be interesting to investigate other data analysis problems where topological methods, especially those based on local topological information (yields to efficient computations), may be useful. Currently, we have conducted some preliminary experiments to demonstrate the performance of our algorithm. It will be interesting to conduct large-scale experiments under a broad range of practical scenarios, so as to better understand data in those contexts.

Acknowledgment:

We acknowledge the support of the NSF grants CCF 1318595, CCF 1319406, CCF 1116258.

References

[1] D. Attali, A. Lieutier, and D. Salinas. Vietoris-Rips complexes also provide topologically correct reconstructions of sampled shapes. In *Proc. 27th. Annu. Sympos. Comput. Geom.*, pages 491–500, 2011.

[2] P. Bendich, D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov. Inferring local homology from sampled stratified spaces. In *Proc. 48th Ann. IEEE Sympos. Foundat. Comp. Sci.*, pages 536–546, 2007.

[3] P. Bendich, B. Wang, and S. Mukherjee. Local homology transfer and stratification learning. In *Proc. 23rd Ann. ACM-SIAM Sympos. Discrete Alg.*, pages 1355–1370, 2012.

[4] J. Bruske and G. Sommer. Intrinsic dimensionality estimation with optimally topology preserving maps. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(5):572–575, 1998.

[5] F. Camastra. Data dimensionality estimation methods: a survey. *Pattern Recognition*, 36(12):2945–2954, 2003.

[6] F. Camastra and A. Vinciarelli. Estimating the intrinsic dimension of data with a fractal-based method. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(10):1404–1407, 2002.

[7] F. Chazal and S. Oudot. Towards persistence-based reconstruction in euclidean spaces. In *Proceedings of the twenty-fourth annual symposium on Computational geometry*, SCG '08, pages 232–241, 2008.

[8] S.-W. Cheng and M.-K. Chiu. Dimension detection via slivers. In *Proc. 20th Ann. ACM-SIAM Sympos. Discrete Alg.*, pages 1001–1010, 2009.

[9] S.-W. Cheng, Y. Wang, and Z. Wu. Provable dimension detection using principal component analysis. In *Proc. 21st Ann. Sympos. Comput. Geom.*, pages 208–217, 2005.

[10] T. K. Dey. *Curve and surface reconstruction: Algorithms with mathematical analysis*. Cambridge University Press, New York, 2006.

[11] T. K. Dey, J. Giesen, S. Goswami, and W. Zhao. Shape dimension and approximation from samples. *Discrete Comput. Geom.*, 29:419–434, 2003.

[12] H. Edelsbrunner and J. Harer. *Computational Topology*. American Mathematical Society, 2009.

[13] H. Edelsbrunner and J. Harer. *Computational Topology, An Introduction*. American Mathematical Society, 2010.

[14] A. M. Farahmand, C. Szepesvári, and J.-Y. Audibert. Manifold-adaptive dimension estimation. In *Proc. 24th Conf. Machine Learning (ICML)*., pages 265–272, 2007.

[15] K. Fukunaga and D. R. Olsen. An algorithm for finding intrinsic dimensionality of data. *IEEE Trans. Computers*, C-20(2):176–183, 1971.

[16] J. Giesen and U. Wagner. Shape dimension and intrinsic metric from samples of manifolds with high co-dimension. In *Proc. 19th Ann. Sympos. Comput. Geom.*, pages 329–337, 2003.

[17] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D*, 9:189–208, 1983.

[18] M. Hein and J.-Y. Audibert. Intrinsic dimensionality estimation of submanifolds in R^d. In *Proc. 22nd Conf. Machine Learning (ICML)*, pages 289–296, 2005.

[19] I. T. Jollife. *Principal Component Analysis*. Springer series in statistics. Springer, NY, 2002.

[20] B. Kégl. Intrinsic dimension estimation using packing numbers. In *Neural Infor. Proc. Sys. Foundation (NIPS)*, pages 681–688, 2002.

[21] E. levina and P. J. bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems 17*, pages 777–784, 2005.

[22] A. V. Little, M. Maggioni, and L. Rosasco. Multiscale geometric methods for estimating intrinsic dimension. In *Proc. SampTA*, 2011.

[23] J. R. Munkres. *Topology, a First Course*. Prentice Hall, New Jersey, 1975.

[24] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39:419–441, March 2008.

[25] K. W. Pettis, T. A. Bailey, A. K. Jain, and R. C. Dubes. An intrinsic dimensionality estimator from near neighbor information. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1:25–37, 1979.

[26] J. J. Rotman. *An Introduction to Algebraic Topology*. Springer, New York, 1988.

[27] P. Skraba and B. Wang. Approximating local homology from samples. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 174–192, 2014.

[28] J. B. Tenenbaum, V. Silva, and J. C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, 2000.

[29] P. J. Verveer and R. P. W. Duin. An evaluation of intrinsic dimensionality estimators. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(1):81–86, 1995.

A Proof for Proposition 3.3

We only need to show that i'_* is an isomorphism as Proposition 3.2 proves it for i_*. Since the inclusion induced homomorphisms $j_* : \mathsf{H}(\mathsf{M}, \mathsf{M} - \mathring{D}_2) \to \mathsf{H}(\mathsf{M}, \mathsf{M} - z)$ and $i_* : \mathsf{H}(\mathsf{M}, \mathsf{M} - \mathring{D}_1) \to \mathsf{H}(\mathsf{M}, \mathsf{M} - z)$ are isomorphisms by Proposition 3.2 and $j_* = i_* \circ i'_*$, we have that i'_* is an isomorphism as well.

B Missing Details in Section 4

Proof of Proposition 4.1. By Proposition 3.5, the map $\pi_{\alpha *}$ is an isomorphism. By Proposition 3.1, D_β is a closed topological ball as $\beta < \rho(\mathsf{M})$. Hence, $(\mathsf{M}, \mathsf{M} - \mathring{D}_\beta) \hookrightarrow (\mathsf{M}, \mathsf{M} - \bar{p})$ induces the isomorphism i_* at the homology level, see Proposition 3.2. The observation then follows.

Missing details for interleaving in section 4. From Eq. 1, it follows that for any $\lambda \in (\varepsilon, \rho(\mathsf{M}) - \delta)$:

$$\mathbb{M}_{\alpha,\lambda} = \pi_\alpha^{-1}(\mathring{B}_\lambda(p) \cap \mathsf{M}) \subset \pi_\alpha^{-1}(\mathring{B}_{\lambda+\varepsilon}(\bar{p}) \cap \mathsf{M}) \subset \mathring{B}_{\lambda+\delta-\varepsilon}(\bar{p}) \cap \mathbb{X}_\alpha \subset \mathring{B}_{\lambda+\delta}(p) \cap \mathbb{X}_\alpha = \mathbb{B}_{\alpha,\lambda+\delta}. \quad (8)$$

Now take a point $x \in B_\lambda(p) \cap \mathbb{X}_\alpha$. Then $d(\pi(x), p) \leq d(\pi(x), x) + d(x, p) \leq (\alpha + \varepsilon) + \lambda$. Therefore,

$$\mathbb{B}_{\alpha,\lambda} = B_\lambda(p) \cap \mathbb{X}_\alpha \subset \pi_\alpha^{-1}(B_{\lambda+\delta-2\varepsilon}(p) \cap \mathsf{M}) = \mathbb{M}_{\alpha,\lambda+\delta-2\varepsilon} \subset \mathbb{M}_{\alpha,\lambda+\delta}. \quad (9)$$

Eq. 8 and Eq. 9 provide the required nesting:

$$\mathbb{M}_{\alpha,\lambda} \subseteq \mathbb{B}_{\alpha,\lambda+\delta} \subseteq \mathbb{M}_{\alpha,\lambda+2\delta}.$$

Proof of Proposition 4.3. Recall that by definition $\mathbb{B}^{\alpha,r} = \mathbb{X}_\alpha - \mathring{B}_r(p)$. Then, for sufficient large $r > \beta + \alpha + 3\varepsilon$, the closure of $\operatorname{int} \mathbb{B}^{\alpha,r}$ is a subset of $\operatorname{int}(\mathbb{M}^{\alpha,\beta})$ or $\operatorname{int}(\mathbb{B}^{\alpha,\beta})$. By the excision theorem, it follows that

$$\mathsf{H}(\mathbb{X}_\alpha, \mathbb{M}^{\alpha,\beta}) \cong \mathsf{H}(\mathbb{X}_\alpha - \operatorname{int} \mathbb{B}^{\alpha,r}, \mathbb{M}^{\alpha,\beta} - \operatorname{int} \mathbb{B}^{\alpha,r})$$

and

$$\mathsf{H}(\mathbb{X}_\alpha, \mathbb{B}^{\alpha,\beta}) \cong \mathsf{H}(\mathbb{X}_\alpha - \operatorname{int} \mathbb{B}^{\alpha,r}, \mathbb{B}^{\alpha,\beta} - \operatorname{int} \mathbb{B}^{\alpha,r}) = \mathsf{H}(\mathbb{X}_{\alpha,r}, \mathbb{X}^\beta_{\alpha,r}),$$

where the isomorphisms are induced from canonical inclusions. The nested sequence of pairs involves only inclusion maps. If we repeat the arguments for Proposition 4.2 for sets intersecting the ball $B_r(z)$ and use Persistence Equivalence Theorem [12], we get the claim of this proposition. To make sure that r is large enough, we need that $r > \lambda + 4\delta + \alpha + 3\varepsilon$, as well as $r > \lambda' + 2\delta + \alpha' + 3\varepsilon$. We choose $r > \lambda + 5\delta$ to guarantee that.

C Missing Details in Section 5

C.1 Good Cover

Here we prove that the set of sets $\mathcal{X}^\beta_{\alpha,r} := \{B_\alpha(p_i) \cap B_r(p) \cap B^\beta(p) \mid p_i \in P\}$ is a good cover for $\bigcup \mathcal{X}^\beta_{\alpha,r} = \mathbb{X}^\beta_{\alpha,r}$.

For convenience, denote $F_j = B_\alpha(p_i) \cap B_r(p) \cap B^\beta(p)$. Note that since $r > \beta + 2\alpha$, we have that any ball $B_\alpha(p_i)$ may intersect the boundary $\partial B_r(p)$ of $B_r(p)$, or the boundary $\partial B^\beta(p)$ of $B_\beta(p)$, (or none of the two boundaries,) but not both. In other words, the set F_j can be of three types: (i) a complete ball $B_\alpha(p_i)$; (ii) a convex set which is the intersection between $B_\alpha(p_i)$ and $B_r(p)$, but not intersecting the boundary $\partial B_\beta(p)$; and (iii) a potentially non-convex set which is the difference $B_\alpha(p_i) - \mathring{B}_\beta(p)$, but not intersecting the boundary $\partial B_r(p)$.

Now consider any subset of $\mathcal{X}^{\beta}_{\alpha,r}$ with non-empty intersection: Since $B_\alpha(p_i)$ cannot intersect $\partial B_r(p)$ and $\partial B_\beta(p)$ simultaneously, such a subset either only consists of balls from type (i) and (ii), or from type (i) and (iii). Since type (i) and (ii) are both convex, their intersection must be contractible. If the subset consists of type (i) and (iii), then the result from Lemma A.5 of [27] shows that it is also contractible. Hence, the intersection of any subset of $\mathcal{X}^{\beta}_{\alpha,r}$ is contractible, and as such $\mathcal{X}^{\beta}_{\alpha,r}$ forms a good cover for $\mathbb{X}^{\beta}_{\alpha,r}$. By Nerve Lemma, this implies that $\mathcal{N}\mathcal{X}^{\beta}_{\alpha,r}$ is homotopic to $\mathbb{X}^{\beta}_{\alpha,r}$; that is, $\mathcal{N}\mathcal{X}^{\beta}_{\alpha,r} \approx \mathbb{X}^{\beta}_{\alpha,r}$.

C.2 Proof of Lemma 5.1

First, we quote the following result shown in [7], which states that the isomorphism induced by the homotopy equivalence between a nerve and its space commute with the canonical inclusions on the spaces at the homology level. To be consistent with the notations of [7], let $\mathcal{N}\mathcal{U}$ denote the nerve on a good cover $\mathcal{U}$.

Proposition C.1 (Lemma 3.4 in [7]) *Let $X \subset X'$ be two paracompact spaces, and Let $\mathcal{U} = \{U_i\}_{i\in J}$ and $\mathcal{U}' = \{U'_i\}_{i\in J}$ be two good open covers of X and X' respectively, based on a same finite parameter set J, such that $U_i \subset U'_i$ for all $i \in J$. Then, there exist homotopy equivalences $\mathcal{N}\mathcal{U} \to X$ and $\mathcal{N}\mathcal{U}' \to X'$ which commute with the canonical inclusions $X \hookrightarrow X'$ and $\mathcal{N}\mathcal{U} \hookrightarrow \mathcal{N}\mathcal{U}'$ at homology and homotopy levels.*

Extending the arguments in the proof of this lemma, we have the following relative homology version. We first give the proof of this result here, after which we explain how Lemma 5.1 follows from this result.

Proposition C.2 *Let Let $X \subset X' \subset X''$ be two paracompact spaces, and Let $\mathcal{U} = \{U_i\}_{i\in J}$, $\mathcal{U}' = \{U'_i\}_{i\in J}$ and $\mathcal{U}'' = \{U''_i\}_{i\in J}$ be three good open covers of X, X' and X'' respectively, based on a same finite parameter set J, such that $U_i \subset U'_i \subset U''_i$ for all $i \in J$. There exist commutative diagrams,*

$$\begin{array}{ccc} \mathsf{H}(X',X) & \longrightarrow & \mathsf{H}(X'',X) \\ \downarrow & & \downarrow \\ \mathsf{H}(\mathcal{N}\mathcal{U}',\mathcal{N}\mathcal{U}) & \longrightarrow & \mathsf{H}(\mathcal{N}\mathcal{U}'',\mathcal{N}\mathcal{U}) \end{array} \qquad \begin{array}{ccc} \mathsf{H}(X'',X) & \longrightarrow & \mathsf{H}(X'',X') \\ \downarrow & & \downarrow \\ \mathsf{H}(\mathcal{N}\mathcal{U}'',\mathcal{N}\mathcal{U}) & \longrightarrow & \mathsf{H}(\mathcal{N}\mathcal{U}'',\mathcal{N}\mathcal{U}') \end{array}$$

where horizontal maps are induced from canonical inclusions $(X',X) \hookrightarrow (X'',X)$, $(\mathcal{N}\mathcal{U}',\mathcal{N}\mathcal{U}) \hookrightarrow (\mathcal{N}\mathcal{U}'',\mathcal{N}\mathcal{U})$, $(X'',X) \hookrightarrow (X'',X')$, ($\mathcal{N}\mathcal{U}''$, and $\mathcal{N}\mathcal{U}) \hookrightarrow (\mathcal{N}\mathcal{U}'',\mathcal{N}\mathcal{U}')$; while vertical maps are isomorphisms.

Proof: From the good covers $\mathcal{U}$ of X, one can construct a topological space ΔX as in [7] such that the following diagram commutes

$$\begin{array}{ccccc} X & \hookrightarrow & X' & \hookrightarrow & X'' \\ \uparrow{\scriptstyle p} & & \uparrow{\scriptstyle p'} & & \uparrow{\scriptstyle p''} \\ \Delta X & \hookrightarrow & \Delta X' & \hookrightarrow & \Delta X'' \end{array}$$

where p and p' are restrictions of p'' to ΔX and $\Delta X'$ respectively, and p,p' and p'' are homotopy equivalences. Therefore, we have a map of pairs $p' : (\Delta X', \Delta X) \to (X', X)$. Considering the two long exact sequences of pairs $(\Delta X', \Delta X)$ and (X', X) and using the same arguments in Proposition 3.6, it follows that $p'_* : \mathsf{H}(\Delta X', \Delta X) \to \mathsf{H}(X', X)$ is an isomorphism. Similarly, p'' is also a map of pairs, and the induced homomorphisms $p''_* : \mathsf{H}(\Delta X'', \Delta X) \to \mathsf{H}(X'', X)$ is also an isomorphism. Given that both p' and p'' are maps of pairs, we have the following commutative diagram of pairs:

$$\begin{array}{ccc} (X',X) & \hookrightarrow & (X'',X) \\ \uparrow{\scriptstyle p'} & & \uparrow{\scriptstyle p''} \\ (\Delta X', \Delta X) & \hookrightarrow & (\Delta X'', \Delta X) \end{array} \quad .$$

It induces the following commutative diagram at homology level where vertical maps are isomorphisms, and horizontal maps are induced from canonical inclusions.

$$\begin{array}{ccc} \mathsf{H}(X', X) & \longrightarrow & \mathsf{H}(X'', X) \\ \uparrow{\scriptstyle p'_*} & & \uparrow{\scriptstyle p''_*} \\ \mathsf{H}(\Delta X', \Delta X) & \longrightarrow & \mathsf{H}(\Delta X'', \Delta X) \end{array} .$$

Next, let Γ be the first barycentric subdivision of $\mathcal{NU}$, and Γ' for $\mathcal{NU}'$ and Γ'' for $\mathcal{NU}''$, respectively. It is shown in [7] that the following diagram commutes

$$\begin{array}{ccccc} \Delta X & \hookrightarrow & \Delta X' & \hookrightarrow & \Delta X'' \\ \downarrow{\scriptstyle q} & & \downarrow{\scriptstyle q'} & & \downarrow{\scriptstyle q''} \\ \Gamma & \hookrightarrow & \Gamma' & \hookrightarrow & \Gamma'' \end{array}$$

where q and q' are the restrictions of q'' to ΔX and $\Delta X'$, respectively. Following the same arguments as above, one obtains the following commutative diagram at homology level with vertical isomorphisms and horizontal maps induced from canonical inclusions.

$$\begin{array}{ccc} \mathsf{H}(\Delta X', \Delta X) & \longrightarrow & \mathsf{H}(\Delta X'', \Delta X) \\ \downarrow{\scriptstyle q'_*} & & \downarrow{\scriptstyle q''_*} \\ \mathsf{H}(\Gamma', \Gamma) & \longrightarrow & \mathsf{H}(\Gamma'', \Gamma) \end{array}$$

It is known that simplicial approximation $g'' : \Gamma'' \to \mathcal{NU}''$ of the identity map $id : |\Gamma''| \to |\mathcal{NU}|$ commutes with canonical inclusions and induces an isomorphism between homology [23]. Therefore, we have the following commutative diagram

$$\begin{array}{ccccc} \Gamma & \hookrightarrow & \Gamma' & \hookrightarrow & \Gamma'' \\ \downarrow{\scriptstyle g} & & \downarrow{\scriptstyle g'} & & \downarrow{\scriptstyle g''} \\ \mathcal{NU} & \hookrightarrow & \mathcal{NU}' & \hookrightarrow & \mathcal{NU}'' \end{array}$$

where g and g' are the restriction of g'' to Γ and Γ' respectively. As before, there exists following commutative diagram at homology level with vertical isomorphisms and horizontal maps induced by canonical inclusions,

$$\begin{array}{ccc} \mathsf{H}(\Gamma', \Gamma) & \longrightarrow & \mathsf{H}(\Gamma'', \Gamma) \\ \downarrow{\scriptstyle g'_*} & & \downarrow{\scriptstyle g''_*} \\ \mathsf{H}(\mathcal{NU}', \mathcal{NU}) & \longrightarrow & \mathsf{H}(\mathcal{NU}'', \mathcal{NU}) \end{array}$$

Combining these three commutative diagrams at homology level, the first commutative diagram in the proposition follows immediately. A similar argument shows that the second commutative diagram in the proposition holds as well. ■

Using this proposition, one can easily obtain that

$$\mathrm{im}\left(\mathsf{H}(\mathcal{NX}_{\alpha,r}, \mathcal{NX}_{\alpha,r}^{\lambda+3\delta}) \to \mathsf{H}(\mathcal{NX}_{\alpha',r}, \mathcal{NX}_{\alpha',r}^{\lambda'+\delta'})\right) \cong \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p})$$

because

$$\mathrm{im}\left(\mathsf{H}(\mathbb{X}_{\alpha,r}, \mathbb{X}_{\alpha,r}^{\lambda+3\delta}) \to \mathsf{H}(\mathbb{X}_{\alpha',r}, \mathbb{X}_{\alpha',r}^{\lambda'+\delta'})\right) \cong \mathsf{H}(\mathsf{M}, \mathsf{M} - \bar{p}).$$

Indeed, for convenience, set $A_1 = \mathbb{X}_{\alpha,r}$, $B_1 = \mathbb{X}_{\alpha,r}^{\lambda+3\delta}$, $A_2 = \mathbb{X}_{\alpha',r}$ and $B_2 = \mathbb{X}_{\alpha',r}^{\lambda'+\delta'}$; and set $\mathcal{A}_1 = \mathcal{X}_{\alpha,r}$, $\mathcal{B}_1 = \mathcal{X}_{\alpha,r}^{\lambda+3\delta}$, $\mathcal{A}_2 = \mathcal{X}_{\alpha',r}$ and $\mathcal{B}_2 = \mathcal{X}_{\alpha',r}^{\lambda'+\delta'}$. We apply the above proposition twice, once to the three spaces $B_1 \subset A_1 \subset A_2$, and once to the three spaces $B_1 \subset B_2 \subset A_2$. This provides the following diagram, where the commutativity of each square follows from Proposition C.2.

$$\begin{array}{ccccc} \mathsf{H}(A_1, B_1) & \longrightarrow & \mathsf{H}(A_2, B_1) & \longrightarrow & \mathsf{H}(A_2, B_2) \\ \downarrow & & \downarrow & & \downarrow \\ \mathsf{H}(\mathcal{N}\mathcal{A}_1, \mathcal{N}\mathcal{B}_1) & \longrightarrow & \mathsf{H}(\mathcal{N}\mathcal{A}_2, \mathcal{N}\mathcal{B}_1) & \longrightarrow & \mathsf{H}(\mathcal{N}\mathcal{A}_2, \mathcal{N}\mathcal{B}_2) \end{array}$$

Since all vertical homomorphisms are isomorphisms, we have that

$$\operatorname{im}\left(\mathsf{H}(A_1, B_1) \to \mathsf{H}(A_2, B_2)\right) \cong \operatorname{im}\left(\mathsf{H}(\mathcal{N}\mathcal{A}_1, \mathcal{N}\mathcal{B}_1), \mathsf{H}(\mathcal{N}\mathcal{A}_2, \mathcal{N}\mathcal{B}_2)\right).$$

This finishes the proof of Lemma 5.1.

D The Algorithm to Compute $\mathrm{im}(j_{\alpha *})$

Recall that j_α is the inclusion of pairs $j_\alpha : (A_1, B_1) \hookrightarrow (A_2, B_2)$, where $B_1 \subset A_1 \subset A_2$ and $B_1 \subset B_2 \subset A_2$. To compute $\mathrm{im}(j_{\alpha *})$, we order the simplices of A_2 in a proper way to build a filtration such that the rank of $\mathrm{im}(j_{\alpha *})$ can be read off from the reduced boundary matrix built from the filtration. Precisely, the filtration adds the simplices of A_2 as follows. The simplices in $B_2 \setminus A_1$ appear first. Then the simplices in B_1, $(B_2 \setminus B_1) \cap A_1$, $A_1 \setminus B_2$ and $A_2 \setminus (A_1 \cup B_2)$ follow sequentially. This order is illustrated in Figure 3. For simplicity, let $\mathrm{R}(\mathsf{x}, \mathsf{y})$ denote submatrix occupying the rectangle region with x as its top left corner point and y as its bottom right corner point in Figure 3. It is known [13] that the rank of $\mathsf{H}(A_1, B_1)$ (or $\mathsf{H}(A_2, B_2)$)

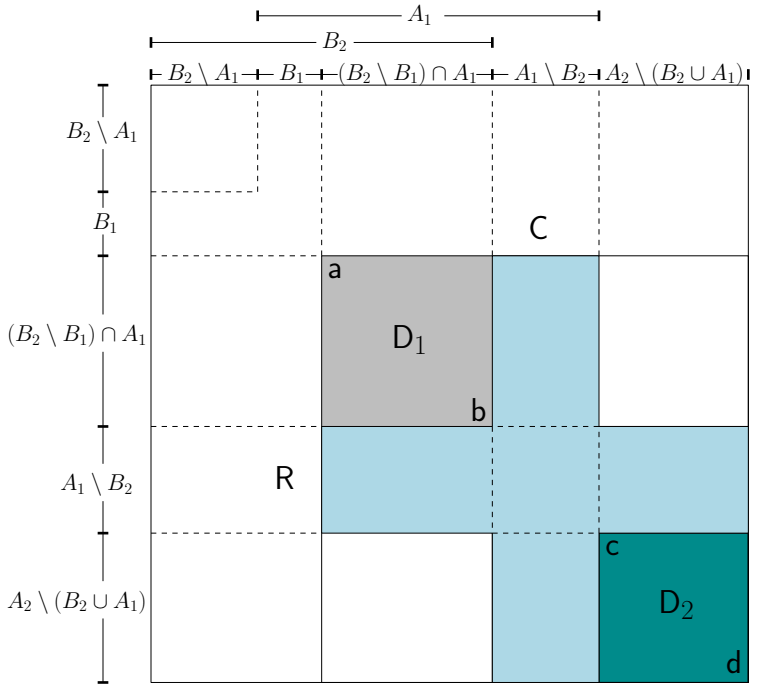


Figure 3: The order of the simplices in the filtration for A_2.

can be computed by reducing the submatrix $M_1 = \mathrm{R}(\mathsf{a}, \mathsf{c})$ (or $M_2 = \mathrm{R}(\mathsf{b}, \mathsf{d})$) in Figure 3. For our purpose, the submatrix $M = \mathrm{R}(\mathsf{a}, \mathsf{d})$ in Figure 3, which contains both M_1 and M_2, will be reduced in the same way as the classical persistent homology algorithm does [13]. Let $\widehat{M}$ denote the matrix reduced from M. It will be shown that the rank of $\mathrm{im}(j_{\alpha *})$ can be read off from $\widehat{M}$.

Recall that the $\mathrm{im}(j_{\alpha *})$ in dimension k contains the k-cycles of $\mathsf{H}_k(A_1, B_1)$ which are nontrivial in both $\mathsf{H}_k(A_1, B_1)$ and $\mathsf{H}_k(A_2, B_2)$. In particular, each k-simplex in the collection of simplices $A_1 \setminus B_2$ whose column in the reduced matrix $\widehat{M}$ is a zero column (i.e., a zero column corresponding to a k-simplex in the light blue column region C of Figure 3) represents a k-cycle in both $\mathsf{H}_k(A_1, B_1)$ and $\mathsf{H}_k(A_2, B_2)$. Let $\#Zero_k$ denote the number of such zero columns in C. If one such k-simplex is paired by a $(k+1)$-simplex in $A_2 \setminus B_2$ (i.e., the row in $\widehat{M}$ corresponding to this simplex which is in the light blue row region R of Figure 3 has a unique 1), its corresponding k-cycle is a k-boundary in $\mathsf{H}_k(A_2, B_2)$. Let $\#Bdry_k$ denote the number of such k-simplices. Since the k-cycles in $\mathsf{H}_k(A_1, B_1)$ corresponding to zero columns which appear before the columns in C contain only simplices from B_2, they all have trivial image in $\mathsf{H}_k(A_2, B_2)$. It is then immediate that the rank of $\mathrm{im}(j_{\alpha *})$ in dimension k equals $\#Zero_k - \#Bdry_k$, namely the number of zero columns in C which correspond to unpaired k-simplices. Once the matrix M is reduced, it is straightforward to compute $\#Zero_k - \#Bdry_k$. If there are n simplices in $A_2 \setminus B_2$, this algorithm runs in $O(n^3)$ time due to the reduction of M.

E Graph Based Central Points and Experimental Details on Real Data

A graph on sample points is built by connecting two points within certain distance. For every vertex v of each component of this graph, the shortest path tree with root v is computed and then the largest distance from v to leaves of this shortest path tree is recorded. The vertex whose distance to leaves of its shortest path tree is the minimum among those vertices in the component containing it, is considered to be the center of its component. Intuitively, these centers are away from the boundary and less likely to be outliers. We then discard the centers of components with few points. For remaining centers, we compute the local homology and report the intrinsic dimension of the manifold as that of the n-sphere whose homology is the same as the most common local homology of these centers. To accelerate the computation, if a component has a significantly large number of vertices, we generate a uniform sparse subsample from the points within some radius of its center and then compute local homology on the subsample points.

	n-SPHERE		EST. DIM	PERCENTAGE
Head	n=3	$53/53$	3	$100\%(53/53)$
D1	n=3	$4/37$	4	$83.7\%(31/37)$
	n=4	$31/37$		
	n=5	$2/37$		
D0	n=2	$2/9$	3	$77.7\%(7/9)$
	n=3	$7/9$		

Table 2: Estimated dimension for real data

We applied this strategy on **Head**, **D1** and **D0**. All of them have only one major component in the graph which connects two points within a distance that is several times the distance of the closest pair in the sample points. For **Head**, a subsample of around 138 points was taken from 505 points in the neighborhood of the center of the major component. We took a subsample of around 148 points from 943 points in the neighborhood of of **D1**'s center, and around 102 points from 3494 points in the neighborhood of **D0**'s center. Since the uniform subsamples were taken randomly, one will be biased to claim the result from one particular

subsample. Therefore, we repeated the local homology computation at the center with fixed parameters 100 times. Note that the points in the subsamples changed each time due to random sampling. Among these 100 computations, we only counted the valid ones which returned the local homology of $\tilde{\mathsf{H}}(\mathbf{S}^n)$ for some n. The distribution of valid computations is shown in Table 2. The n-SPHERE column shows the number of valid computations with the reduced homology of $\tilde{\mathsf{H}}(\mathbf{S}^n)$ for each n. The total number of valid computations is also included in this column. The EST. DIM column gives the estimated dimension. The PERCENTAGE column shows the percentage of computations with the estimated dimension in all valid computations. For the **Head** data, the detected dimension from our method matches the ground truth which is 3. Although the ground truth dimensions for **D1** and **D0** are unknown, ours along with SLIVER, PN, LPCA and ISOMAP report dimension in range $[3, 7]$ for **D1** and in range $[2, 9]$ for **D0**.