

Approximating Local Homology from Samples

Primož Skraba*

Bei Wang†

Abstract

Recently, multi-scale notions of local homology (a variant of persistent homology) have been used to study the local structure of spaces around a given point from a point cloud sample. Current reconstruction guarantees rely on constructing embedded complexes which become difficult to construct in higher dimensions. We show that the persistence diagrams used for estimating local homology can be approximated using families of Vietoris-Rips complexes, whose simpler construction are robust in any dimension. To the best of our knowledge, our results, for the first time make applications based on local homology, such as stratification learning, feasible in high dimensions.

1 Introduction

Advances in scientific and computational experiments have improved our ability to gather large collections of data points in high-dimensional spaces. One aspect in topological data analysis is to infer the topological structure of a space given a point cloud sample. We often assume the space has manifold structure, however, more interesting cases arise when we relax our assumptions to include spaces that contains singularities and mixed dimensionality, such as stratified spaces for example.

Stratified spaces can be decomposed into manifold pieces that are glued together in some uniform way. An important tool in studying these spaces is the study of the neighborhoods surrounding singularities, where manifolds of different dimensionality intersect. We focus on sampling conditions for such neighborhoods, which allow us to begin examining how successful certain reconstruction techniques are with respect to the geometric properties of the underlying shape. Our goal is to infer sampling conditions for recovering local structures of stratified spaces, in particular, the *local homology groups*, from a possibly noisy sampled point set.

Stratification learning. In stratification learning (or mixed manifold learning), a point cloud is assumed to be sampled from a mixture of (possibly intersecting) manifolds. The objective is to recover the different

pieces, often treated as clusters, of the data associated with different manifolds of varying dimensions. Stratified spaces has been studied extensively in mathematics, see the seminal work in [24, 30]. Recently, topological data analysis, relying heavily on ingredients from computational topology [22, 33] and intersection homology [23, 3, 5] has gained momentum in stratification learning. In particular, the work in [4] focuses on studying the local structure of a sampled stratified spaces based on a multi-scale notion of local homology (see Section 2). More recent work [6] studies how point cloud data could be clustered by strata based on how the local homology of nearby sampled points map into one another.

Reconstruction and sampling. Reconstructing shapes from potential noisy point cloud samples has been studied in many different fields. Most often the work is heavily tied to a reconstruction criteria (e.g. homotopic, homeomorphic, etc.) and the assumptions on the underlying space (e.g. manifold). Combinatorial algorithms in geometry are generally derived from Delaunay triangulations [9] and alpha shapes [20], and provide correctness proofs associated with such reconstructions [18]. As the dimension increases, reconstruction efforts have been redirected towards alternative combinatorial structures such as tangential Delaunay complexes [7], witness complexes [16], Čech complexes and the closely related Vietoris-Rips complexes [10, 1, 2].

However, these existing techniques are primarily concerned with global reconstruction. Providing reconstruction guarantees for local structures is more challenging. To guarantee theoretical correctness in computing persistence local homology, both [4] and [6] use Delaunay complexes and their variants. However constructing Delaunay complexes in high dimensions is known to be difficult due to scaling and numerical issues with predicates. On the other hand, the Vietoris-Rips complex is one of the most commonly used combinatorial structure in topological data analysis due to its algorithmic simplicity and robust computation in practice. Methods for fast [31] and efficient [1, 32] construction and approximation [28] of Vietoris-Rips complexes are available, and there have been theoretical advances on their topology-preserving qualities, making it appealing for computations in high dimensions. The goal of this paper is to make persistent local homology com-

*Jozef Stefan Institute, Slovenia

†Scientific Computing and Imaging Institute, University of Utah.

putation more practical through approximations based upon Vietoris-Rips complexes.

Contributions. Our contributions focus on providing sampling conditions to recover the local structure of a space from a point cloud sample, based on previously introduced [4] multi-scale notions of local homology. Our main results are:

- We extend previously introduced algebraic constructions in the analysis of scalar fields over point cloud data [12] to two multi-scale notions of local homology.
- For both multi-scale notions of local homology, we approximate their persistence diagrams by constructing families of Vietoris-Rips complexes based on a set of sample points, formalized within Theorem 3.2 and 4.1. The simplicity and efficiency of building these complexes in any dimension, in combination with their small-size sparse approximation [28, 32] makes, for the first time, applications based on local homology such as stratification learning feasible in high dimensions.
- We show that relative persistent modules are interleaved if the respective absolute persistent modules are interleaved. We consider such a technical result (Theorem 3.1) of independent interest.
- Our results imply algorithms for computing the local homology either by a reduction to standard persistence or a known variant which we describe in Section 5.

2 Background

The background material focuses on the introduction of persistence modules [11], local homology and its multi-scale notions [4]. We assume a basic knowledge of homology and persistent homology, see [27, 26] for a readable background of the former, and [21] for a computational treatment of the latter.

Persistence Modules. We use the definition of persistence modules adapted from [11]. A *persistence module* $\mathcal{F} = \{F_\alpha\}_{\alpha \in \mathbb{R}}$ is a collection of vector spaces F_α (over any fields) together with a family $\{f_\alpha^\beta : F_\alpha \rightarrow F_\beta\}_{\alpha \leq \beta}$ of linear maps such that $\alpha \leq \beta \leq \gamma$ implies $f_\alpha^\gamma = f_\beta^\gamma \circ f_\alpha^\beta$, and $f_\alpha^\alpha = id_{F_\alpha}$. A persistence module is *tame* if it has finite number of critical values and all F_α are of finite rank. In this paper, we suppose all persistence modules we encounter are tame.

Two persistence modules $\{F\}_\alpha$ and $\{G\}_\alpha$ are (*strongly*) ε -*interleaved* if there exists two families of homomorphisms, $\mu_\alpha : F_\alpha \rightarrow G_{\alpha+\varepsilon}$ and $\nu_\alpha : G_\alpha \rightarrow F_{\alpha+\varepsilon}$, that make the following diagrams (Fig. 1) commute for all $\alpha \leq \beta \in \mathbb{R}$ [11]. The information contained in a persistence module can be encoded by a multi-set of points in the extended plane $\bar{\mathbb{R}}^2$ (where $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$),

called a *persistence diagram* [14]. If two tame persistence modules are ε -interleaved, the bottleneck distance between their persistence diagrams are upper bounded by ε ([14], Theorem 4.4). In this paper, we consider persistence modules of homology groups and relative homology groups over a field. Given a family of topological spaces $\{\mathbb{X}_\alpha\}_\alpha$ connected by inclusions $\mathbb{X}_\alpha \hookrightarrow \mathbb{X}_\beta$, the inclusions induce a sequence of homology groups connected by homomorphisms, $\mathbf{H}_k(\mathbb{X}_\alpha) \rightarrow \mathbf{H}_k(\mathbb{X}_\beta)$, where k is the homological dimension. We therefore obtain persistence modules of the form $\{\mathbf{H}_k(\mathbb{X}_\alpha)\}_\alpha$. Specifically, when the linear maps associated with two persistence modules $\{\mathbf{H}_k(\mathbb{X}_\alpha)\}_\alpha$ and $\{\mathbf{H}_k(\mathbb{Y}_\alpha)\}_\alpha$ are induced by inclusions at the space level $\mathbb{X}_\alpha \hookrightarrow \mathbb{Y}_{\alpha+\varepsilon}$ and $\mathbb{Y}_\alpha \hookrightarrow \mathbb{X}_{\alpha+\varepsilon}$, their k -th persistence modules are ε -interleaved [11]. For the rest of the paper, we sometimes abuse this notation by omitting the k -th homology functor unless necessary. We work with singular homology here but our results are applicable in the simplicial setting as well.

Local Homology. The *local homology groups* at a point $x \in \mathbb{X}$ are defined as the relative homology groups $\mathbf{H}(\mathbb{X}, \mathbb{X} - x)$ ([27], page 126). In this paper, we assume that the topological space \mathbb{X} is embedded in some Euclidean space \mathbb{R}^d ¹. Let $d_x : \mathbb{R}^d \rightarrow \mathbb{R}$ be the Euclidean distance function from a fixed $x \in \mathbb{X}$, $d_x(y) := d(x, y) = \|y - x\|$. Let $B_r = B_r(x) = d_x^{-1}[0, r]$ and $B^r = B^r(x) = d_x^{-1}[r, \infty)$ be the sublevel sets and superlevel sets of d_x . Taking a small enough r , the local homology groups in questions are in fact the *direct limit* of relative homology groups, $\lim_{r \rightarrow 0} \mathbf{H}(\mathbb{X}, \mathbb{X} \cap B^r)$, or alternatively $\lim_{r \rightarrow 0} \mathbf{H}(\mathbb{X} \cap B_r, \mathbb{X} \cap \partial B_r)$ [3], see Fig. 2. We adapt two multi-scale notions of this concept based on persistence (which are first introduced in [4]), referred to as the *r-filtration* and the α -*filtration*. The goal of this paper is to derive sampling conditions that are appropriate to compute the persistence diagrams with respect to these filtrations, therefore approximating the local homology at $x \in \mathbb{X}$.

For a fixed $\alpha \geq 0$, let \mathbb{X}_α be the “thickened” or “offset” version of \mathbb{X} , that is, the space of points in \mathbb{R}^d at Euclidean distance at most α from \mathbb{X} . Suppose L is a finite set of points sampled from \mathbb{X} ², where $L \subset \mathbb{X}$ and $L_\alpha = \cup_{x \in L} B_\alpha(x)$. In subsequent sections, we put further restrictions on L where we suppose L is an ε -*sample* of \mathbb{X} , that is,

$$\forall x \in \mathbb{X}, \quad d(x, L) := \inf_{y \in L} d(x, y) \leq \varepsilon.$$

The *r-filtration* (Fig. 3) is a sequence of relative

¹This assumption can be relaxed in several ways, but this setting is most common in our applications.

²Our results would hold with minor modifications in the setting of sampling with noise, where elements of L lie on or near \mathbb{X} .

$$\begin{array}{ccccccccccc}
F_{\alpha-\varepsilon} & \longrightarrow & F_{\beta+\varepsilon} & F_{\alpha+\varepsilon} & \longrightarrow & F_{\beta+\varepsilon} & F_{\alpha} & \longrightarrow & F_{\beta} & F_{\alpha} & \longrightarrow & F_{\beta} \\
& \searrow & & \nearrow & & \nearrow & \nearrow & & \searrow & \searrow & & \searrow \\
& & G_{\alpha} & \longrightarrow & G_{\beta} & G_{\alpha} & \longrightarrow & G_{\beta} & G_{\alpha-\varepsilon} & \longrightarrow & G_{\beta+\varepsilon} & G_{\alpha+\varepsilon} & \longrightarrow & G_{\beta+\varepsilon}
\end{array}$$

Figure 1: Strongly ε -interleaved persistence modules.

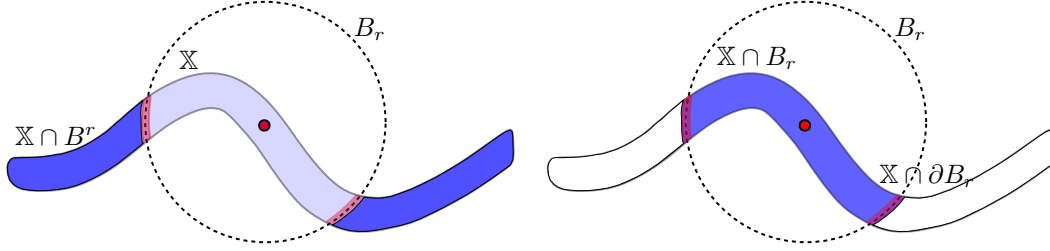


Figure 2: Local homology as the direct limit, $\lim_{r \rightarrow 0} \mathbf{H}(\mathbb{X}, \mathbb{X} \cap B^r)$ (left) or $\lim_{r \rightarrow 0} \mathbf{H}(\mathbb{X} \cap B_r, \mathbb{X} \cap \partial B_r)$ (right).

homology groups connected by linear maps induced by inclusion and excision, constructed by fixing a thickening parameter α and varying parameter r , for $r' > r$,

$$\cdots \rightarrow \mathbf{H}(\mathbb{X}_\alpha, \mathbb{X}_\alpha \cap B^{r'}) \rightarrow \mathbf{H}(\mathbb{X}_\alpha, \mathbb{X}_\alpha \cap B^r) \rightarrow \cdots$$

The same filtration could be built on a set of points L sampled from \mathbb{X} , that is,

$$\cdots \rightarrow \mathbf{H}(L_\alpha, L_\alpha \cap B^{r'}) \rightarrow \mathbf{H}(L_\alpha, L_\alpha \cap B^r) \rightarrow \cdots$$

Here, we fix the space at resolution α , and vary the scale r at which we analyze the local neighborhood, analog to changing the lens from the front of the camera.

The α -filtration³ (Fig. 4) is a sequence of relative homology groups connected by inclusion, constructed by fixing r and varying α , for $\alpha < \alpha'$,

$$\begin{aligned}
&\cdots \rightarrow \mathbf{H}(\mathbb{X}_\alpha \cap B_r, \mathbb{X}_\alpha \cap \partial B_r) \rightarrow \cdots \\
&\rightarrow \mathbf{H}(\mathbb{X}_{\alpha'} \cap B_r, \mathbb{X}_{\alpha'} \cap \partial B_r) \rightarrow \cdots
\end{aligned}$$

Its discrete counterpart built on a set of points L sampled from \mathbb{X} is,

$$\begin{aligned}
&\cdots \rightarrow \mathbf{H}(L_\alpha \cap B_r, L_\alpha \cap \partial B_r) \rightarrow \cdots \\
&\rightarrow \mathbf{H}(L_{\alpha'} \cap B_r, L_{\alpha'} \cap \partial B_r) \rightarrow \cdots
\end{aligned}$$

Here, we fix the size of the ball which defines the locality, i.e. the size r of the local neighborhood, and we vary the scale α at which we analyze the space.

³Technically, the r -filtration and the α -filtration are both persistence modules that arise from their corresponding filtrations, we refer to them as such for simplicity.

Čech and Vietoris-Rips Complexes. Suppose L is a finite point set in \mathbb{R}^d and $L_\alpha = \cup_{x \in L} B_\alpha(x)$. The nerve of L_α is the simplicial complex induced by all the non-empty intersections of subcollections of balls in L_α and is called the *Čech complex* of L , denoted as $\mathcal{C}_\alpha = \mathcal{C}_\alpha(L)$ (omitting L from the notation unless necessary). The *Vietoris-Rips complex* of L is denoted as \mathcal{R}_α , whose simplices correspond to non-empty subsets of L of diameter less than α . For Euclidean metric space, we have, $\forall \alpha > 0$, $\mathcal{C}_{\alpha/2} \subseteq \mathcal{R}_\alpha \subseteq \mathcal{C}_\alpha \subseteq \mathcal{R}_{2\alpha}$ ⁴. This implies that the persistence modules $\{\mathbf{H}(\mathcal{C}_\alpha)\}_\alpha$ and $\{\mathbf{H}(\mathcal{R}_\alpha)\}_\alpha$ are α -interleaved⁵.

3 Approximating the α -Filtration

In the α -filtration, since we will be computing relative persistent homology, there are certain requirements on the pairs, such that the maps of the relative filtration are well-defined. Two persistence modules, $\mathcal{A} = \{A_\alpha\}_{\alpha \in \mathbb{R}}$ and $\mathcal{F} = \{F_\alpha\}_{\alpha \in \mathbb{R}}$ are called *compatible* if for all $\alpha \leq \beta$, the following diagram commutes:

$$\begin{array}{ccc}
A_\alpha & \longrightarrow & F_\alpha \\
\downarrow & & \downarrow \\
A_\beta & \longrightarrow & F_\beta
\end{array}$$

⁴Jung's Theorem gives a tighter relation between Vietoris-Rips and Čech complexes. We use the slightly looser relation in our paper for simplicity.

⁵We emphasize that for our results the interleaving parameter does depend on the parameter of the filtration. In other words, for a fixed filtration scale parameter, we have a certain "constant" interleaving.

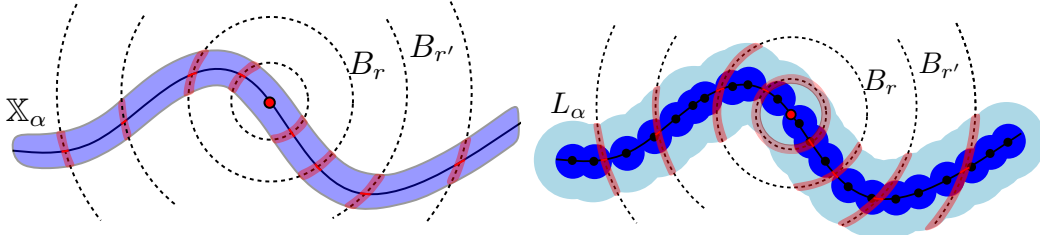


Figure 3: The r -filtration for space \mathbb{X} and its offsets (left), and the same filtration built on a set of points L , sampled from \mathbb{X} .

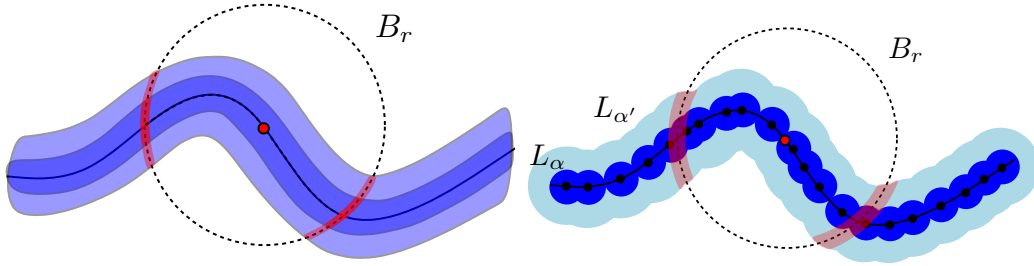


Figure 4: The α -filtration for space \mathbb{X} and its offset (left), and on the right, the same filtration built on a set of points L , sampled from \mathbb{X} .

This ensures that the relative persistence module is well-defined⁶. In our context, all the maps are induced by inclusions hence the above diagram commutes. We highlight steps involved to obtain our approximation results:

- First, we show that under certain conditions, the relative homology of a ball modulo its boundary is isomorphic to that of the entire space modulo the subspace outside the ball.
- Second, we prove that if we have two compatible persistence modules \mathcal{F} and \mathcal{A} which are respectively interleaved with \mathcal{G} and \mathcal{B} , the relative persistent homology $H(\mathcal{F}, \mathcal{A})$ is approximated by $H(\mathcal{G}, \mathcal{B})$. This result may be of independent interest.
- Last, we prove a series of interleavings to show that both filtrations in our case can be interleaved with a Vietoris-Rips construction on the samples.

We first show that the following two filtrations are equivalent (where $\alpha < \alpha'$):

$$(3.1) \quad 0 \rightarrow H(\mathbb{X}_\alpha \cap B_r, \mathbb{X}_\alpha \cap \partial B_r) \rightarrow H(\mathbb{X}_{\alpha'} \cap B_r, \mathbb{X}_{\alpha'} \cap \partial B_r) \rightarrow \dots \rightarrow H(B_r, \partial B_r),$$

$$(3.2) \quad 0 \rightarrow H(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r) \rightarrow H(\mathbb{X}_{\alpha'}, \mathbb{X}_{\alpha'} - \text{int } B_r) \rightarrow \dots \rightarrow H(\mathbb{R}^n, \mathbb{R}^n - \text{int } B_r).$$

Note that $\mathbb{X}_\alpha - \text{int } B_r = \mathbb{X}_\alpha - (\mathbb{X}_\alpha \cap \text{int } B_r)$. Unless otherwise specified, $\alpha, \alpha' \in [0, \infty)$. Graphically, these filtrations are shown in Fig. 5. As it turns out, it is easier to argue about the filtration in Fig. 5(right) than Fig. 5(left), as shown in the following lemma.

LEMMA 3.1. *Assuming that spaces \mathbb{X}_α and $\mathbb{X}_\alpha - \text{int } B_r$ form a good pair, then $H(\mathbb{X}_\alpha \cap B_r, \mathbb{X}_\alpha \cap \partial B_r) \cong H(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r)$.*

Proof Sketch. This follows from the Excision Theorem ([25], Theorem 15.1, page 82) and the Excision Extension Theorem ([25], Theorem 15.2, page 82). We excise the space $\mathbb{X}_\alpha - B_r$ from the pair $(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r)$, and obtain $H(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r) \cong H(\mathbb{X}_\alpha - (\mathbb{X}_\alpha - B_r), \mathbb{X}_\alpha -$

⁶This is slightly more general than the condition required when computing image, kernel and cokernel persistence of on pairs of filtrations in [15].

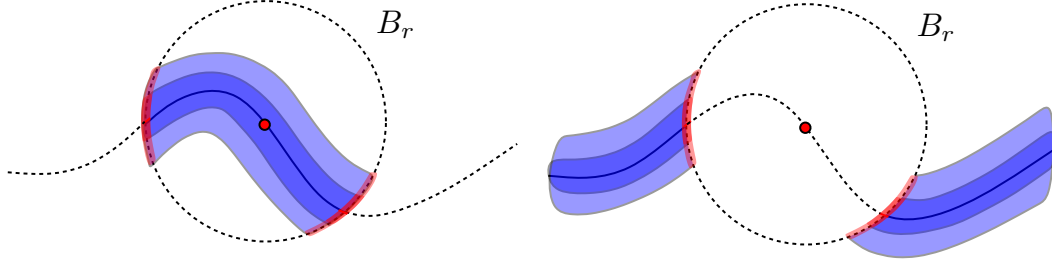


Figure 5: Left: the α -filtration with respect to the pair $(\mathbb{X}_\alpha \cap B_r, \mathbb{X}_\alpha \cap \partial B_r)$. Right: the filtration with respect to the pair $(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r)$.

$\text{int } B_r - (\mathbb{X}_\alpha - B_r) \cong \mathbb{H}(\mathbb{X}_\alpha \cap B_r, \mathbb{X}_\alpha \cap \partial B_r)$. Since the closure of $\mathbb{X}_\alpha - B_r$ needs not be contained in the interior of $\mathbb{X}_\alpha - \text{int } B_r$, there are some technical conditions which require some care. The key point is that since the spaces form a good pair, this is sufficient for the isomorphism to hold.

Recall that a pair of spaces (\mathbb{A}, \mathbb{B}) forms a *good pair* if \mathbb{B} is a nonempty closed subspace that is a deformation retract of some neighborhood in \mathbb{A} ([26], page 114). This condition is required for the above excision to be applicable. Without this condition, we cannot guarantee that the homology of a quotient space is isomorphic to the relative homology of the pair. See Appendix A for details. \square

We now show that we can approximate local homology at multi-scale via the α -filtration using sample points. We begin with sequence (3.2). Specifically, we first consider the persistence module corresponding to the whole space $\{\mathbb{X}_\alpha\}$, and then the persistence module corresponding to the subspace we quotient by, $\{\mathbb{X}_\alpha - \text{int } B_r\}$. The key is a technical result described in Theorem 3.1 which says that if we can interleave persistence modules independently, we can interleave their corresponding quotient persistence modules. We consider this result to be of independent interest.

THEOREM 3.1. *Consider two pairs of compatible persistence modules. Let $\mathcal{A} = \{A_\alpha\}_{\alpha \in \mathbb{R}}$ be compatible with $\mathcal{F} = \{F_\alpha\}_{\alpha \in \mathbb{R}}$ and $\mathcal{B} = \{B_\alpha\}_{\alpha \in \mathbb{R}}$ be compatible with $\mathcal{G} = \{G_\alpha\}_{\alpha \in \mathbb{R}}$. If the modules \mathcal{A} and \mathcal{B} are ε_1 -interleaved and \mathcal{F} and \mathcal{G} are ε_2 -interleaved, then the relative modules $\{(F_\alpha, A_\alpha)\}_{\alpha \in \mathbb{R}}$ and $\{(G_\alpha, B_\alpha)\}_{\alpha \in \mathbb{R}}$ are ε -interleaved, where $\varepsilon = \max\{\varepsilon_1, \varepsilon_2\}$.*

Proof Sketch. Without loss of generality, assume $\varepsilon_1 = \varepsilon_2 = \varepsilon$. Each pair, $\{(F, A)\}$ and $\{(G, B)\}$, gives rise to a long exact sequence. The two sequences are related by interleaving maps yielding the commutative diagram in Figure 6.

To prove that the interleavings between individual modules imply an interleaving between $\{(F, A)\}$ and $\{(G, B)\}$, we would need some careful diagram chasing at the chain level. That is, we need to prove each of the four diagrams (reviewed in Fig. 1) needed for interleaving commutes, i.e. diagrams in Fig. 7 commute. The key issue is that although each row is exact, maps between persistence modules do not split — therefore we may have one persistent relative class without a persistent class in either component filtrations. The full details of the proof (with digram chasing arguments) are given in Appendix A. \square

$$\begin{array}{ccc}
 \mathbb{H}_n(F_\alpha, A_\alpha) & \longrightarrow & \mathbb{H}_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon}) \\
 & \searrow & \nearrow \\
 & \mathbb{H}_n(G_{\alpha+\varepsilon}, B_{\alpha+\varepsilon}) & \\
 & & \\
 & \mathbb{H}_n(F_{\alpha+\varepsilon}, A_{\alpha+\varepsilon}) & \\
 & \nearrow & \searrow \\
 \mathbb{H}_n(G_\alpha, B_\alpha) & \longrightarrow & \mathbb{H}_n(G_{\alpha+2\varepsilon}, B_{\alpha+2\varepsilon})
 \end{array}$$

Figure 7: Commuting diagrams for ε -interleaved persistence modules.

With Theorem 3.1 in hand, we can begin to prove the main result of the section. We would like to construct a persistence module (based upon Vietoris-Rips filtration) that interleaves with the α -filtration $\{(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r)\}$. The straightforward approach is to consider $\{(L_\alpha, L_\alpha - \text{int } B_r)\}$, as illustrated in Figure 8. Such a construction is possible with careful geometric considerations through interleaving with an intermediate complex described below. We obtain our main result by proving the following key steps:

1. $\{L_\alpha\}$ and $\{\mathbb{X}_\alpha\}$ are ε -interleaved.
2. $\{L_\alpha - \text{int } B_r\}$ and $\{\mathbb{X}_\alpha - \text{int } B_r\}$ are ε -interleaved.

$$\begin{array}{ccccccccc}
H_n(A_\alpha) & \xrightarrow{i_n^\alpha} & H_n(F_\alpha) & \xrightarrow{j_n^\alpha} & H_n(F_\alpha, A_\alpha) & \xrightarrow{k_n^\alpha} & H_{n-1}(A_\alpha) & \xrightarrow{i_{n-1}^\alpha} & H_{n-1}(F_\alpha) \\
\downarrow \phi_n^\alpha & & \downarrow f_n^\alpha & & \downarrow \mu_n^\alpha & & \downarrow \phi_{n-1}^\alpha & & \downarrow f_{n-1}^\alpha \\
H_n(B_{\alpha+\varepsilon}) & \xrightarrow{p_n^{\alpha+\varepsilon}} & H_n(G_{\alpha+\varepsilon}) & \xrightarrow{q_n^{\alpha+\varepsilon}} & H_n(G_{\alpha+\varepsilon}, B_{\alpha+\varepsilon}) & \xrightarrow{r_n^{\alpha+\varepsilon}} & H_{n-1}(B_{\alpha+\varepsilon}) & \xrightarrow{p_{n-1}^{\alpha+\varepsilon}} & H_{n-1}(G_{\alpha+\varepsilon}) \\
\downarrow \psi_n^{\alpha+\varepsilon} & & \downarrow g_n^{\alpha+\varepsilon} & & \downarrow \nu_n^{\alpha+\varepsilon} & & \downarrow \psi_{n-1}^{\alpha+\varepsilon} & & \downarrow g_{n-1}^{\alpha+\varepsilon} \\
H_n(A_{\alpha+2\varepsilon}) & \xrightarrow{i_n^{\alpha+2\varepsilon}} & H_n(F_{\alpha+2\varepsilon}) & \xrightarrow{j_n^{\alpha+2\varepsilon}} & H_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon}) & \xrightarrow{k_n^{\alpha+2\varepsilon}} & H_{n-1}(A_{\alpha+2\varepsilon}) & \xrightarrow{i_{n-1}^{\alpha+2\varepsilon}} & H_{n-1}(F_{\alpha+2\varepsilon})
\end{array}$$

Figure 6: Commuting diagrams for the long exact sequence involving two pairs of filtrations.

3. (Nerve Lemma for $\{L_\alpha - \text{int } B_r\}$) For $\alpha < r$, $\mathcal{N}(L_\alpha - \text{int } B_r)$ is homotopic to $L_\alpha - \text{int } B_r$.
4. For a fixed α , we give an algorithm that compute the 0- and 1-skeleton of $\mathcal{N}(L_\alpha - \text{int } B_r)$. Then we build a Vietoris-Rips complex based on these 0- and 1-skeletons by filling in the high-dimensional co-faces. This is in fact a flag complex, or equivalently, the clique complex of the 1-skeleton of $\mathcal{N}(L_\alpha - \text{int } B_r)$. As α -varies, we refer to such a filtration as $\mathcal{F}_\alpha(L)$. We show that $\{L_\alpha - \text{int } B_r\}$ is α -interleaved with $\mathcal{F}_\alpha(L)$.
5. We now arrive at the main result (Theorem 3.2): $\{(\mathcal{R}_\alpha(L), \mathcal{F}_\alpha(L))\}$ is $(\alpha + 2\varepsilon)$ -interleaved with the α -filtration $\{(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r)\}$.

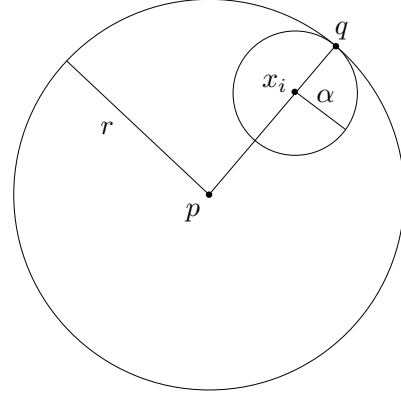


Figure 9: The boundary case for constructing the 0-skeleton of $\mathcal{N}(L_\alpha - \text{int } B_r)$, where $B_\alpha(x_i) - \text{int } B_r$ is nonempty and $\alpha + d(x_i, p) \geq r$.

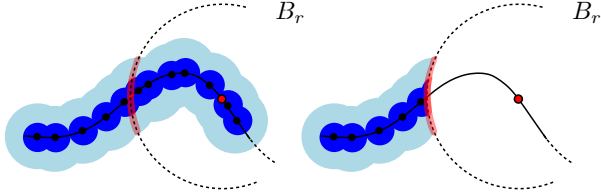


Figure 8: Illustration of filtrations, $\{L_\alpha\}$ (left) and $\{L_\alpha - \text{int } B_r\}$ (right).

For a fixed α , now we give the geometric constructions of the 0- and 1-skeleton of $\mathcal{N}(L_\alpha - \text{int } B_r)$, or equivalently, the 0- and 1-skeleton of $\mathcal{F}_\alpha(L)$.

To construct the 0-skeleton (with a fixed α), for a vertex $x_i \in L$ to be in the complex $\mathcal{N}(L_\alpha - \text{int } B_r)$, it is required that the ball of radius α centered at x_i has non-empty intersection with the underlying space $L_\alpha - \text{int } B_r$. Formally, let $B_\alpha(x_i)$ be the ball of radius α centered at x_i and let p denote the center of the ball $B_r = B_r(p)$ defining the local neighborhood. There is a corresponding vertex v_i in the nerve if and only if $B_\alpha(x_i) - \text{int } B_r \neq \emptyset$.

As α varies, if x_i lies outside B_r , i.e. $d(x_i, p) \geq r$, then the 0-skeleton is nonempty for all $\alpha > 0$. If x_i lies inside B_r , this set is non-empty for $\alpha > r - d(x_i, p)$. The boundary case is shown in Figure 9.

Therefore, to build a filtration of the 0-skeleton as α increases, we introduce a vertex v_i for every point $x_i \in L$ with the following auxiliary function, which represents the minimum value of α when $B_\alpha(x_i) - \text{int } B_r \neq \emptyset$ and hence a vertex enters the filtration. The auxiliary function, denoted by g , for a point $x_i \in L$, is given by

$$g(x_i) = \begin{cases} 0 & d(x_i, p) \geq r \\ r - d(x_i, p) & d(x_i, p) < r \end{cases}$$

The filtration we consider is the sublevel set filtration (at the simplicial level) of g .

To construct the 1-skeleton, we define the auxiliary function g for the edges. We consider the situation when the intersection of two balls (centered at x_1 and x_2 respectively) outside of the local neighborhood is non-empty, i.e. $B_\alpha(x_i) \cap B_\alpha(x_j) - \text{int } B_r(p) \neq \emptyset$. Denote the midpoint of the line segment connecting the two points

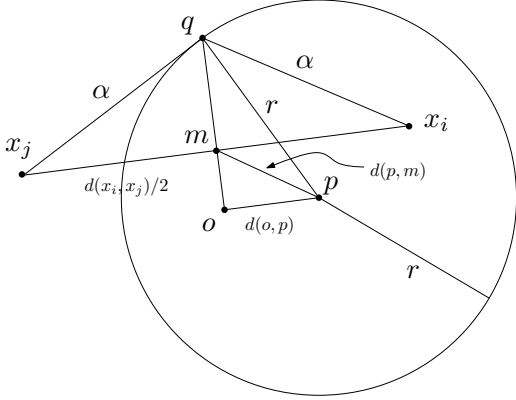


Figure 10: This figure illustrates the computation of α , that is, the time an edge $x_i x_j$ enters the filtration, $g(x_i x_j)$. All edges are labelled by known distances. We compute $d(q, o)$ based on the triangle opq , and $d(m, o)$ based on the triangle omp , whose difference gives us $d(q, m)$. α can then be computed using the right triangle $x_j q m$.

by m , where

$$m = \frac{x_i + x_j}{2}.$$

The earliest an edge $x_i x_j$ can enter the filtration is $\alpha > d(x_i, x_j)/2$. If m lies outside the ball B_r , $g(x_i x_j) = d(x_i, x_j)/2$. If m lies inside B_r , we must compute the closest point outside the ball. Consider, the Figure 10 where the line segment op is parallel to $x_i x_j$ so we can compute

$$d(o, p) = \left| \frac{(p - m)^T (x_i - x_j)}{d(x_i, x_j)} \right|.$$

Using basic geometry we get the following closed form expression for the filtration function g on the edges.

$$g(x_i x_j) = \begin{cases} d(x_i, x_j)/2, & \text{if } d(m, p) \geq r \\ \left(\frac{(\sqrt{r^2 - d(o, p)^2} - \sqrt{d(m, p)^2 - d(o, p)^2})^2}{+d(x_i, x_j)^2/4} \right)^{\frac{1}{2}} & \text{o.w.} \end{cases}$$

With g defined on the 0- and 1-skeleton, it extends to higher-dimensional simplices iteratively as the maximum value of its faces, e.g., triangles takes the maximum of its bounding edges, tetrahedra its bounding triangles and so on. Now we arrive at our main result:

THEOREM 3.2. *Suppose L is an ε -sample of \mathbb{X} . For $\alpha < r$, the relative module $\{(\mathcal{R}_\alpha(L), \mathcal{F}_\alpha(L))\}$ is $(\alpha + 2\varepsilon)$ -interleaved with the α -filtration $\{(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r)\}$.*

Proof Sketch. Recall the key steps involved in proving the main theorem. Step 1 and 2 are relative straightforward assuming L is an ε -sample of \mathbb{X} . Step 3 and 4

requires careful geometric constructions. We defer the technical details involving each step to Appendix A. We describe the key ingredients for arriving at the interleaving result (i.e. Step 5).

Step 1 gives an ε -approximation of the filtration $\{\mathbb{X}_\alpha\}$ and passing to the Vietoris-Rips complex introduces an additional multiplicative factor of 2. We have, on the chain level, the following sequence connected by injective maps,

$$C_i(\mathbb{X}_\alpha) \rightarrow C_i(L_{\alpha+\varepsilon}) \rightarrow C_i(\mathcal{R}_{2(\alpha+\varepsilon)}(L)),$$

$$C_i(\mathcal{R}_\alpha(L)) \rightarrow C_i(L_\alpha) \rightarrow C_i(\mathbb{X}_{(\alpha+\varepsilon)}).$$

Therefore, the interleaving parameter between the filtration $\{\mathbb{X}_\alpha\}$ and the Vietoris-Rips complex filtration $\{\mathcal{R}_\alpha(L)\}$ is $\alpha + 2\varepsilon$.

A similar construction gives us the same bound for the filtration $\{\mathbb{X}_\alpha - \text{int } B_r\}$. That is,

$$C_i(\mathbb{X}_\alpha - \text{int } B_r) \rightarrow C_i(L_{\alpha+\varepsilon} - \text{int } B_r) \rightarrow C_i(\mathcal{F}_{\alpha+\varepsilon}(L)),$$

$$C_i(\mathcal{F}_\alpha(L)) \rightarrow C_i(L_{2\alpha} - \text{int } B_r) \rightarrow C_i(\mathbb{X}_{2(\alpha+\varepsilon)} - \text{int } B_r).$$

Applying Theorem 3.1 gives the interleaving constant of $\alpha + 2\varepsilon$. \square

4 Approximating Local Homology: r -Filtration

4.1 r -Filtration approximation. In this section, we describe approximating local homology with respect to a fixed point x at multi-scale via r -filtration (Fig. 3). We fix a thickening parameter α and drop it from the notation, using only \mathbb{X} . Consider the following filtration, for $r \geq s \geq t$,

$$(4.3) \quad \begin{aligned} \dots &\rightarrow \mathbf{H}(\mathbb{X}, \mathbb{X} \cap B^r) \rightarrow \mathbf{H}(\mathbb{X}, \mathbb{X} \cap B^s) \rightarrow \\ &\rightarrow \mathbf{H}(\mathbb{X}, \mathbb{X} \cap B^t) \rightarrow \dots, \end{aligned}$$

Now we endow the space \mathbb{X} with a function $g : \mathbb{X} \rightarrow \mathbb{R}$, which is the Euclidean distance to a fixed point x , $g(y) = d(x, y) = d_x(y)$. g could be viewed as the restriction onto the space \mathbb{X} , of a Euclidean distance function to a point x , $d_x : \mathbb{R}^d \rightarrow \mathbb{R}$, that is, $g = d_x|_{\mathbb{X}}$. The function g is 1-Lipschitz and we see that $\mathbb{X} \cap B^r = g^{-1}[r, \infty)$, the superlevel set of g . The above sequence becomes,

$$(4.4) \quad \begin{aligned} \dots &\rightarrow \mathbf{H}(\mathbb{X}, g^{-1}[r, \infty)) \rightarrow \mathbf{H}(\mathbb{X}, g^{-1}[s, \infty)) \rightarrow \\ &\rightarrow \mathbf{H}(\mathbb{X}, g^{-1}[t, \infty)) \rightarrow \dots \end{aligned}$$

This is the relative persistence module of g . Now let $f = -g : \mathbb{X} \rightarrow \mathbb{R}$, f is also 1-Lipschitz. Sequence (4.5)

holds the same information as sequence (4.4) assuming tame functions ⁷, according to the Extended Persistence Symmetry Corollary [8] (that is, the ordinary persistence diagram of a function f equals the relative persistence diagram of $-f$ up to a dimension shift and central reflection),

$$(4.5) \quad \begin{aligned} \dots \rightarrow \mathbf{H}(f^{-1}(-\infty, a]) \rightarrow \mathbf{H}(f^{-1}(-\infty, b]) \rightarrow \\ \rightarrow \mathbf{H}(f^{-1}(-\infty, c]) \dots, \end{aligned}$$

where $a \leq b \leq c$, which corresponds to the persistence module of f based on its sublevel sets. Since the filtrations in sequence (4.5) and sequence (4.4) hold the same information, we can translate the diagram and recover the information for the original r -filtration (sequence (4.3)).

The key insight is that in this case, the r -filtration amounts to studying the persistent homology of a function on the space — the distance function to a point, which is a particularly nice function, i.e. 1-Lipschitz. In this section, we give results under strong assumptions on the space \mathbb{X} with some further discussions deferred to Section 4.2.

We introduce a strong assumption on homotopy between a pair of spaces, which requires that points are only moved a bounded amount in the homotopy. Two subsets of Euclidean space, \mathbb{X} and \mathbb{Y} are δ -homotopy equivalent, if there exists two functions $i : \mathbb{X} \rightarrow \mathbb{Y}$ and $h : \mathbb{Y} \rightarrow \mathbb{X}$ such that $h \circ i$ is homotopic to the identity map $id_{\mathbb{X}}$, $i \circ h$ is homotopic to $id_{\mathbb{Y}}$, $d(p, h \circ i(p)) \leq \delta$ ($\forall p \in \mathbb{X}$) and $d(q, i \circ h(q)) \leq \delta$ ($\forall q \in \mathbb{Y}$). Although weaker assumptions have been used in approximation results for a space, in this setting, we must approximate a sublevel set filtration. An example of when a space as a whole can be well-approximated but the approximation fails for the sublevel set, along with other consequences of such an assumption are discussed further in Section 4.2. In our context, the map i is typically the canonical inclusion map, therefore if $d(q, i \circ h(q)) \leq \delta$ ($\forall q \in \mathbb{Y}$) then $d(p, h \circ i(p)) \leq \delta$ ($\forall p \in \mathbb{X}$). Then we refer to $h : \mathbb{Y} \rightarrow \mathbb{X}$ as the δ -homotopy equivalence between \mathbb{Y} and \mathbb{X} , where $d(p, h(p)) \leq \delta$.

The first step in approximating the r -filtration is relating the sublevel set filtration of a 1-Lipschitz function $f : \mathbb{X} \rightarrow \mathbb{R}$ on the space \mathbb{X} , and the sublevel set filtration of a corresponding function $f_\varepsilon : \mathbb{X}_\varepsilon \rightarrow \mathbb{R}$ on \mathbb{X}_ε . These filtrations together with maps induced by space inclusions form the (homology) persistence module of f and f_ε , respectively. Recall that f is the negative of d_x restricted to \mathbb{X} , $f = -d_x|_{\mathbb{X}}$. Likewise, $f_\varepsilon = -d_x|_{\mathbb{X}_\varepsilon}$.

⁷It is unclear whether this holds in the case of non-tame functions which could arise as a consequence of a pathological underlying space.

Since there is an inclusion $\mathbb{X} \hookrightarrow \mathbb{X}_\varepsilon$, it follows that $f = f_\varepsilon|_{\mathbb{X}}$. For the rest of the section we use the following notation for sublevel sets: $F(a) = f^{-1}(-\infty, a]$, $F_\varepsilon(a) = f_\varepsilon^{-1}(-\infty, a]$, for every $a \in \mathbb{R}$. The persistence module of f and f_ε are represented as $\{\mathbf{H}(F(a))\}_a$ and $\{\mathbf{H}(F_\varepsilon(a))\}_a$ respectively.

LEMMA 4.1. *Suppose \mathbb{X} and \mathbb{X}_ε are δ -homotopy equivalent through the canonical inclusion map $i : \mathbb{X} \rightarrow \mathbb{X}_\varepsilon$ and the map $h : \mathbb{X}_\varepsilon \rightarrow \mathbb{X}$. Then the persistence modules of f and f_ε , that is, $\{\mathbf{H}(F(a))\}_a$ and $\{\mathbf{H}(F_\varepsilon(a))\}_a$, are δ -interleaved.*

Proof. Consider the following sequence of maps:

$$F(\alpha) \xrightarrow{i'} F_\varepsilon(\alpha + \varepsilon) \xrightarrow{h'} F(\alpha + \varepsilon + \delta).$$

We define the map $i' = i|_{F(\alpha)}$ and show that i' is well-defined. $\forall p \in F(\alpha)$, by definition, we have $f(p) \leq \alpha$ and $f = f_\varepsilon|_{\mathbb{X}}$, therefore $f_\varepsilon(p) \leq \alpha$. This implies that $p \in F_\varepsilon(\alpha) \subseteq F_\varepsilon(\alpha + \varepsilon)$, therefore, i' is a well-defined inclusion, which induces inclusion on the homology level, $i_* : \mathbf{H}(F(\alpha)) \rightarrow \mathbf{H}(F_\varepsilon(\alpha + \varepsilon))$.

We define $h' = h|_{F_\varepsilon(\alpha + \varepsilon)}$, and we need to show that h' is well-defined, that is, the image of h' lies in $F(\alpha + \varepsilon + \delta)$. $\forall p \in F_\varepsilon(\alpha + \varepsilon)$, by definition, we have $f_\varepsilon(p) \leq \alpha + \varepsilon$. Since $f_\varepsilon = -d_x|_{\mathbb{X}_\varepsilon}$, then $-d(x, p) \leq \alpha + \varepsilon$. Combining with $d(p, h(p)) \leq \delta$, we have

$$f(h(p)) := -d(x, h(p)) \leq d(p, h(p)) - d(x, p) \leq \alpha + \varepsilon + \delta.$$

This implies that $h(p) \in F(\alpha + \varepsilon + \delta)$. Therefore h' is well-defined. In addition, based on our assumption that \mathbb{X} and \mathbb{X}_ε are homotopy equivalent through maps i and h , this implies that h' is a homotopy equivalence, which induces an isomorphism h_* on the homology level,

$$h_* : \mathbf{H}(F_\varepsilon(\alpha + \varepsilon)) \rightarrow \mathbf{H}(F(\alpha + \varepsilon + \delta)).$$

In order to show persistence modules $\{\mathbf{H}(F(a))\}_a$ and $\{\mathbf{H}(F_\varepsilon(a))\}_a$ are $\max(\delta, \varepsilon)$ -interleaved, it is easy to verify that the four diagrams in Fig. 1 commute based on the linear maps i_* and h_* . Finally, by definition, $\varepsilon \leq \delta$, that is, the boundary of the offset must be moved at least ε . Therefore, $\max(\delta, \varepsilon) = \delta$ and we conclude that the persistence modules are δ -interleaved. \square

The next step is to relate the above filtrations to the union of balls on the samples. For notational convenience we define the union of balls centered around points with a function value less than some threshold $a \in \mathbb{R}$ as $U_\varepsilon(a) = \cup_{p \in L, f(p) \leq a} B_\varepsilon(p)$, where $a \in \mathbb{R}$ and $a \leq 0$. Since $U_\varepsilon(a)$ contains Euclidean balls which are convex, the Nerve Lemma holds, that is, its

nerve $\mathcal{N}(U_\varepsilon(a))$, which corresponds to the Čech complex $\mathcal{C}_\varepsilon(a)$, and $U_\varepsilon(a)$ are homotopy equivalent. As a varies, these complexes together with the maps induced by inclusions form a persistence module $\{\mathbf{H}(\mathcal{C}_\varepsilon(a))\}_a$. Similarly we define the corresponding Vietoris-Rips complex and its persistence module as $\mathcal{R}_\varepsilon(a)$ and $\{\mathbf{H}(\mathcal{R}_\varepsilon(a))\}_a$ respectively.

LEMMA 4.2. *Suppose \mathbb{X} and \mathbb{X}_ε are δ -homotopy equivalent through the canonical inclusion map $i : \mathbb{X} \rightarrow \mathbb{X}_\varepsilon$ and the map $h : \mathbb{X}_\varepsilon \rightarrow \mathbb{X}$ for $\varepsilon \leq \delta$. Suppose L is an ε -sample of \mathbb{X} . Then the persistence modules $\{\mathbf{H}(F(a))\}_a$ of f and $\{\mathbf{H}(\mathcal{C}_\varepsilon(a))\}_a$ are $(\varepsilon + \delta)$ -interleaved.*

Proof. The proof is nearly identical to the proof of Lemma 4.1. Consider the following sequence:

$$F(\alpha) \xrightarrow{i'} U_\varepsilon(\alpha + \varepsilon) \xrightarrow{h'} F(\alpha + 2\varepsilon + \delta)$$

We define the map $i' = i|_{F(\alpha)}$ and show i' is well-defined. $\forall p \in F(\alpha)$, by definition, $f(p) = -d(x, p) \leq \alpha$. Since L is an ε -sample of \mathbb{X} , there exists $q \in L$ such that $p \in B_\varepsilon(q)$, that is, $d(p, q) \leq \varepsilon$. Combining the above inequalities, we obtain $f(q) = -d(x, q) \leq -d(x, p) + d(p, q) \leq \alpha + \varepsilon$, implying that $p \in U_\varepsilon(\alpha + \varepsilon)$.

For map h' , since $U_\varepsilon(\alpha + \varepsilon) \subseteq F_\varepsilon(\alpha + 2\varepsilon)$, based on the results in Lemma 4.1 that the map $F_\varepsilon(\alpha + 2\varepsilon) \rightarrow F(\alpha + 2\varepsilon + \delta)$ is well-defined, we can define $h' = h|_{U_\varepsilon(\alpha + \varepsilon)}$. Following similar argument in Lemma 4.1, $\{\mathbf{H}(U_\varepsilon(a))\}_a$ is $(\varepsilon + \delta)$ -interleaved with $\{\mathbf{H}(F(a))\}_a$. By the Nerve Lemma, the union of balls⁸ is homotopic to the Čech complex for all a , leading to $\mathbf{H}(\mathcal{C}_\varepsilon(a)) \cong \mathbf{H}(U_\varepsilon(a))$. Furthermore, the homotopy commutes with inclusion ([10], Lemma 3.4). \square

THEOREM 4.1. *Suppose h is δ -homotopy equivalence between $\mathbb{X}_{2\varepsilon}$ and \mathbb{X} for $\delta \geq 2\varepsilon$, and L is an ε -sample of \mathbb{X} . Then the Vietoris-Rips module $\{\mathbf{H}(\mathcal{R}_{2\varepsilon}(a))\}_a$ is $(2\varepsilon + \delta)$ -interleaved with the r -filtration $\{\mathbf{H}(F(a))\}_a$.*

Proof. Suppose $\mathbb{X}_{2\varepsilon}$ and \mathbb{X} are δ -homotopy equivalent through the canonical inclusion map $i : \mathbb{X} \rightarrow \mathbb{X}_{2\varepsilon}$ and the map $h : \mathbb{X}_{2\varepsilon} \rightarrow \mathbb{X}$. We can construct the following commutative diagram in Figure 11.

First we consider the top and bottom rows in the diagram. The 1st map is an inclusion on the space level. The 2nd and 5th maps are homotopy equivalences based on the Nerve Lemma (which induces isomorphisms on the homology level). The 3rd and 4th maps are inclusions based on interleaving between Čech and Vietoris-Rips complexes, i.e. $\mathcal{C}_\varepsilon \subseteq \mathcal{R}_{2\varepsilon} \subseteq \mathcal{C}_{2\varepsilon}$. Second,

⁸Note the definition of the union of balls filtration – it precisely equals to the lower star filtration of the Čech complex.

all the vertical maps between the top and bottom rows are inclusions. Finally, we define the connecting map $U_{2\varepsilon}(\alpha + \varepsilon) \rightarrow F(\alpha + 3\varepsilon + \delta)$ as $h' = h|_{U_{2\varepsilon}(\alpha + \varepsilon)}$. To show h' is well-defined, $\forall p \in U_{2\varepsilon}(\alpha + \varepsilon)$, $f(p) \leq \alpha + 3\varepsilon$, since h' is a δ -homotopy, $h'(p)$ has a function value at most $\alpha + 3\varepsilon + \delta$, therefore $h'(p) \in F(\alpha + 3\varepsilon + \delta)$.

From the above commutative diagram, we consider the following maps between spaces: $F(\alpha) \rightarrow \mathcal{R}_{2\varepsilon}(\alpha + \varepsilon) \rightarrow F(\alpha + 3\varepsilon + \delta)$. This leads to a factor of $2\varepsilon + \delta$ in the interleaving between persistence modules $\{\mathbf{H}(\mathcal{R}_{2\varepsilon}(a))\}_a$ and $\{\mathbf{H}(F(a))\}_a$ and since $\delta \geq 2\varepsilon$, we can simplify this to 2δ . \square

4.2 Discussion on the r -filtration. In this section, we describe the assumption of δ -homotopy equivalence between a pair of spaces. This is a strong assumption since it is a homotopic equivalence but also requires that points are only moved a bounded amount in the homotopy. This is to quantify the distortion introduced in situations such as the one illustrated in Fig. 12. Here we define the Euclidean distance function to the point p as $d_p(x) := d(p, x) = \|x - p\|$. We study the local homology of any point $p \in \mathbb{R}^d$, hence p can be chosen to be any point in \mathbb{R}^d , including any point in \mathbb{X} .

Now consider d_p restricted to \mathbb{X} and \mathbb{X}_α . Although the part of \mathbb{X} shown is well-approximated via a retract from \mathbb{X}_α , it is insufficient to guarantee that we could closely approximate the persistence module of $d_p|_{\mathbb{X}}$ through that of $d_p|_{\mathbb{X}_\alpha}$. The homology changes when the two pieces shown merge in the sublevel set of d_p . The parameter at which the change occurs differs in the filtrations on \mathbb{X} and \mathbb{X}_α by δ , or in other words, the persistence diagrams of these filtrations differs by at least δ .

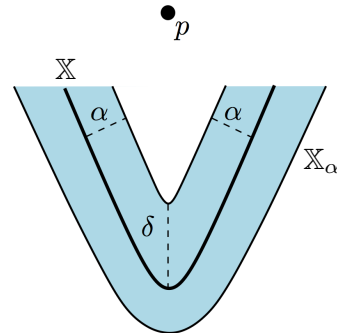


Figure 12: The space \mathbb{X} , its α -offset \mathbb{X}_α and a point p . Consider the distance function d_p to p . This is an example where the offset \mathbb{X}_α and space \mathbb{X} are homotopy equivalent but the persistence diagrams of the functions $d_p|_{\mathbb{X}}$ and $d_p|_{\mathbb{X}_\alpha}$ are potentially far apart.

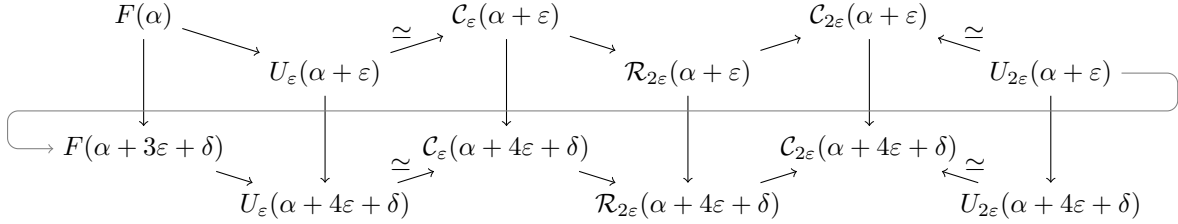


Figure 11: The commutative diagram interleaving the r -filtrations for the sublevel set filtration F , the filtration of the function, the Čech complex \mathcal{C} and the Vietoris-Rips complex \mathcal{R} .

The problem of approximating a sublevel set filtration of a function on a space has been studied before. The setting is closely related to the results of [12]. In [12], there is an approximation guarantee between a sublevel set filtration of a c -Lipschitz function on a space and an image persistence filtration on two nested Vietoris-Rips complexes with an appropriately chosen parameter. There are numerous requirements to apply such results, which we outline here.

The first requirement is that we have access to geodesic distances or some provable approximation of it. While the geodesic distance can be inferred from the Euclidean distance in certain cases, this can be a difficult problem depending on how our space is embedded. The second requirement is that the space has positive convexity radius. While this is generally a safe assumption for manifolds; for spaces where local homology yields interesting information, such as stratified spaces, this measure can often be zero (i.e. a cone has zero convexity radius). If, however such requirement is satisfied, we can apply the results in [12] directly. The resulting algorithm is to build the underlying simplicial complex using geodesic distances, which given a sufficiently dense sampling relative to the convexity radius, gives an approximation for *any* c -Lipschitz function. Since distance functions are 1-Lipschitz, the approximation results follow.

This highlights a key obstacle in stating sampling results for function filtrations as well as an open problem we discuss below: in terms of which measures should we state sampling results? Is there a global geometric measure which is meaningful for stratified spaces? Are there weaker conditions than δ -homotopy for approximating sublevel set behavior? As pointed out above, geometric measures, such as *reach* or *convexity radius* can be zero even for nice spaces. It would be preferable to use quantifiers such as *homological feature size* [14]. Research in these directions is left for future work.

5 Algorithms

With the theory developed in the previous sections, we outline the algorithms for computing the different filtrations. In a nutshell, for the r -filtration, we could reduce the computation of persistent local homology to standard persistence on the sample points [22, 33]. On the other hand, for the α -filtration, using Theorem 3.2, we can compute relative persistent homology of the filtrations built on the sample points using the algorithm described in [29].

We begin with the the r -filtration. As input we take a point set L which is an ε -sample of a space \mathbb{X} such that \mathbb{X} and its offset $\mathbb{X}_{2\varepsilon}$ are homotopic, and the point p at which we want to study the local homology. We construct a 2ε -Vietoris-Rips complex and compute the function defined by the distance to p for all points in the point sample, i.e. $f(x) = d(x, p)$, $\forall x \in L$. Here we assume that we can choose ε such that Theorem 4.1 holds and that we obtain a 4ε -approximation of the sublevel set filtration. By the Extended Persistence Symmetry Corollary [8], this gives us an approximation of the filtration in Equation 4.3. Therefore, we use the standard persistence algorithm on the Vietoris-Rips complex filtered by f [22, 33] and we correct for reflection and dimension shifts. This is described in [17], however we recount it here for completeness. We substitute birth and death times (i.e. reflection) and increase the dimension of all infinitely persistent classes by one. Correctness is proven in [8].

For the α -filtration, as input we again take a point set L which is an ε -sample of a space \mathbb{X} , a point p and radius r at which we want to study the local homology. We again consider the two component filtrations which make up the relative persistence of the pair. The filtration on the whole space is the standard Vietoris-Rips filtration up to parameter r . For the other filtration, we modify the parameter values according to the formulas given in Section 3.

We can then compute the relative homology directly using the methods of [29]. To compute the approxima-

tion to the α -filtration we construct the Vietoris-Rips filtration on the point sample L up to parameter r (which is given as input). Now we give a bit more technical details on the algorithm described in [29]. Recall that persistence may be viewed as homology over $k[t]$ where k is a field and the grading in t represents the time when simplices enter the filtration. Furthermore, finitely generated persistence modules may be represented by a presentation of generators and relations. For standard persistence, we obtain

$$B(X) \rightarrow Z(X) \rightarrow H(X) \rightarrow 0$$

where $B(X)$ and $Z(X)$ are freely generated graded modules over $k[t]$ representing the boundaries and cycles respectively. That is, they represent the basis chains of the boundaries and the cycles (and these chains are graded in t). To compute the relative persistent homology of a pair of spaces $A \subseteq X$, we compute the following presentation

$$C(A) \oplus B(X) \rightarrow Z(X) \rightarrow H(X, A) \rightarrow 0$$

where $C(A)$ represents the graded chain space of A . Here we have the implicit assumption that the filtrations are compatible such that when a simplex enters A it also enters X or has already entered X . Formally, any chain in $C(A)$ is expressible in $C(X)$ with a possible multiplication by t . Therefore, we can find the graded linear map which sends a the span of $C(A) \oplus B(X)$ into $Z(X)$. The Smith-Normal form of this map is the barcode.

To speed up the algorithm, we can remove the rows and columns corresponding to simplices which lie outside the ball $B_r(p)$ in $C(A)$ and $C(X)$. More precisely, we remove a simplex if and only if all of its vertices lie outside $B_r(p)$. This is an observation on the above presentation, since if a simplex lies completely outside the ball it is immediately quotiented out (since for these simplices we do not change the filtration value). Also note that this does not correspond to a simplicial complex, since, for example, an edge with one vertex in the ball and one outside will only have the row that corresponds to the vertex outside the ball removed.

The algorithm can be further improved by observing that since we only compute the filtration up to r , only the points in the sample which lie at most $2r$ from x need to be considered, rather than the entire sample. This is due to the fact that no simplex will exist in the Vietoris-Rips filtration with an edge longer than $2r$. Alternatively, this could be proved using excision. This implies that the size of the complex we need to consider depends on r . Namely, if we have a dense sample near p we can choose a small r to keep the complexity manageable.

6 Discussion

Local homology and relative homology are common tools in algebraic topology. In this paper, we recounted two different multi-scale notions of local homology: the α - and r -filtrations. We show that both can be well-approximated using Vietoris-Rips complexes based on a finite sample of the space and therefore efficiently computed. We also prove a novel technical result involving interleaving between relative persistence modules derived from interleaving between absolute persistence modules. Several open questions remain: Are there better geometric measures to describe the sampling conditions in approximating local homology? Could a similar sampling theory be developed for witness complexes? Under what conditions on the space are the underlying filtrations we study tame?

Our work was motivated by stratification learning. We prove in this paper that the multi-scale versions of local homology, vehicles for describing local structures of stratified spaces, could be approximated with Vietoris-Rips complexes. In higher dimensions, the size of the Vietoris-Rips complexes could be manageable if the topological space has fairly low intrinsic dimension and the number of sampled points is small. However, although Vietoris-Rips complexes are simple to compute and preserve the topology of the underlying space, they do not scale well with higher intrinsic dimension or large number of sampled points. In this case, we would rely on fast and efficient construction and approximation of Vietoris-Rips [31, 1, 32, 28]. In particular, the work in [28] provides a linear-size approximation of the Vietoris-Rips filtration across all scales for large data sets. The recently introduced Graph Induced Complex (GIC) [19] also offers an alternative approximation scheme to make Vietoris-Rips approximation in high-dimension computable in practice. It would be interesting to extend the results of this paper to other complexes such as GIC or witness complexes.

Furthermore, the results in this paper could be applied to any applications where local or relative homology computations are relevant, i.e. for future directions, the approximation of the Conley index or well groups [13], which also include computation of relative homology groups.

7 Acknowledgements

This work was funded in part by the EU project TOPOSYS (FP7-ICT-318493-STREP), DOE/SciDAC DESC0007446, NSF OCI-0906379, NSF OCI-0904631, DOE/NEUP 120341, DOE/MAPD DESC000192, and DOE/Codesign P01180734. The authors would also like to thank the anonymous reviewers for many useful comments to improve the readability of the paper.

References

- [1] Dominique Attali, André Lieutier, and David Salinas. Efficient data structure for representing and simplifying simplicial complexes in high dimensions. *Proceedings 27th Annual Symposium on Computational Geometry*, pages 501–509, 2011.
- [2] Dominique Attali, André Lieutier, and David Salinas. Vietoris-Rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry: Theory and Applications*, 46(4):448–465, 2013.
- [3] Paul Bendich. *Analyzing Stratified Spaces Using Persistent Versions of Intersection and Local Homology*. PhD thesis, Duke University, 2008.
- [4] Paul Bendich, David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. Inferring local homology from sampled stratified spaces. *Proceedings 48th Annual IEEE Symposium on Foundations of Computer Science*, pages 536–546, 2007.
- [5] Paul Bendich and John Harer. Persistent intersection homology. *Foundations of Computational Mathematics*, 11:305–336, 2011.
- [6] Paul Bendich, Bei Wang, and Sayan Mukherjee. Local homology transfer and stratification learning. *Proceedings 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1355–1370, 2012.
- [7] J. Boissonnat and A. Ghosh. Manifold reconstruction using tangential delaunay complexes. *Proceedings 26th Annual Symposium on Computational Geometry*, pages 324–333, 2010.
- [8] Gunnar Carlsson, Vin de Silva, and Dmitriy Morozov. Zigzag persistent homology and real-valued functions. *Proceedings 25th Annual Symposium on Computational Geometry*, pages 247–256, 2009.
- [9] Frédéric Cazals and Joachim Giesen. Delaunay triangulation based surface reconstruction: Ideas and algorithms. In *Effective Computational Geometry for Curves and Surfaces*, pages 231–273, 2006.
- [10] F. Chazal and S. Oudot. Towards persistence-based reconstruction in Euclidean spaces. *Proceedings 24th Annual Symposium on Computational Geometry*, pages 232–241, 2008.
- [11] Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. *Proceedings 25th Annual Symposium on Computational Geometry*, pages 237–246, 2009.
- [12] Frédéric Chazal, Leonidas J. Guibas, Steve Y. Oudot, and Primoz Skraba. Analysis of scalar fields over point cloud data. *Proceedings 20th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1021–1030, 2009.
- [13] Frédéric Chazal, Primoz Skraba, and Amit Patel. Computing well diagrams for vector fields on \mathbb{R}^n . *Applied Mathematics Letters*, 25(11):1725–1728, 2012.
- [14] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37:103–120, 2007.
- [15] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Dmitriy Morozov. Persistence homology for kernels, images and cokernels. *Proceedings 20th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1011–1020, 2009.
- [16] V. de Silva and G. Carlsson. Topological estimation using witness complexes. *Symposium on Point-Based Graphics*, pages 157–166, 2004.
- [17] Vin de Silva, Dmitriy Morozov, and Mikael Vejdemo-Johansson. Dualities in persistent (co)homology. *Inverse Problems*, 27(12), 2011.
- [18] Tamal K. Dey. *Curve and Surface Reconstruction*. Cambridge University Press, 2007.
- [19] Tamal K. Dey, Fengtao Fan, and Yusu Wang. Graph induced complex on point data. *Proceedings 29th Annual Symposium on Computational Geometry*, pages 107–116, 2013.
- [20] H. Edelsbrunner and E. Mucke. Three-dimensional alpha shapes. *Proceedings workshop on Volume Visualization*, pages 75–82, 1992.
- [21] Herbert Edelsbrunner and John Harer. *Computational Topology*. American Mathematical Society, 2010.
- [22] Herbert Edelsbrunner, David Letscher, and Afra J. Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28:511–533, 2002.
- [23] M Goresky and R MacPherson. Intersection homology i. *Topology*, 19:135–162, 1982.
- [24] Mark Goresky and Robert MacPherson. *Stratified Morse Theory*. Springer-Verlag, 1988.
- [25] Marvin J. Greenberg and John R. Harper. *Algebraic Topology A First Course*. Addison-Wesley, 1981.
- [26] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.
- [27] James R. Munkres. *Elements of algebraic topology*. Addison-Wesley, 1984.
- [28] Donald R. Sheehy. Linear-size approximations to the Vietoris-Rips filtration. *Discrete & Computational Geometry*, 49(4):778–796, 2013.
- [29] Primoz Skraba and Mikael Vejdemo-Johansson. Persistence modules: Algebra and algorithms. arXiv:1302.2015, 2013.
- [30] Shmuel Weinberger. The topological classification of stratified spaces. *Chicago Lectures in Mathematics*, 1994.
- [31] Afra J. Zomorodian. Fast construction of the Vietoris-Rips complex. *Computers & Graphics*, 34(3):263–271, 2010.
- [32] Afra J. Zomorodian. The tidy set: a minimal simplicial set for computing homology of clique complexes. *Proceedings 26th Annual Symposium on Computational Geometry*, pages 257–266, 2010.
- [33] Afra J. Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33:249–274, 2005.

A Detailed Proofs for Approximating Local Homology at Multi-scale via α -filtration

A.1 Lemma 3.1 We first prove the following lemma:

Lemma 3.1. Assuming that spaces \mathbb{X}_α and $\mathbb{X}_\alpha - \text{int } B_r$ form a *good pair*, then $H(\mathbb{X}_\alpha \cap B_r, \mathbb{X}_\alpha \cap \partial B_r) \cong H(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r)$.

Proof. First we recall several theorems related to excisions. Let $\mathbb{Y}, \mathbb{U}, \mathbb{A}$ be topological spaces. The inclusion map of pairs $(\mathbb{Y} - \mathbb{U}, \mathbb{A} - \mathbb{U}) \rightarrow (\mathbb{Y}, \mathbb{A})$ is called an *excision* if it induces a homology isomorphism. In this case, one says that \mathbb{U} can be excised. We will make use of the following two results about excision ([25]).

THEOREM A.1. (EXCISION THEOREM) ([25], *Theorem 15.1, page 82*) *If the closure of \mathbb{U} is contained in the interior of \mathbb{A} , that is, $\text{cl } \mathbb{U} \subseteq \text{int } \mathbb{A}$, then \mathbb{U} can be excised.*

THEOREM A.2. (EXCISION EXTENSION) ([25], *Theorem 15.2, page 82*) *Suppose $\mathbb{V} \subset \mathbb{U} \subset \mathbb{A}$ and (i) \mathbb{V} can be excised; (ii) $(\mathbb{Y} - \mathbb{U}, \mathbb{A} - \mathbb{U})$ is a deformation retract of $(\mathbb{Y} - \mathbb{V}, \mathbb{A} - \mathbb{V})$. Then \mathbb{U} can be excised. fig*

In our context, let $\mathbb{Y} = \mathbb{X}_\alpha$, $\mathbb{A} = \mathbb{X}_\alpha - \text{int } B_r$, $\mathbb{U} = \mathbb{X}_\alpha - B_r$. Therefore $\mathbb{Y} - \mathbb{U} = \mathbb{X}_\alpha \cap B_r$ and $\mathbb{A} - \mathbb{U} = \mathbb{X}_\alpha \cap \partial B_r$. However, since $\text{cl } \mathbb{U}$ needs not be contained in $\text{int } \mathbb{A}$, so we must define a suitable $\mathbb{V} \subset \mathbb{U}$. One direct way is to choose some small enough positive δ and a neighborhood \mathbb{I} , such that we define, $\mathbb{I} = \mathbb{X}_\alpha \cap \partial B_r \cap \text{cl } \mathbb{U}$, $\mathbb{I}_\delta = \{x \in \text{cl } \mathbb{U} \mid d_{\mathbb{I}}(x) \leq \delta\}$, and $\mathbb{V} = \mathbb{U} - \mathbb{I}_\delta$, where $d_{\mathbb{I}}(x)$ is the Euclidean distance from x to the set \mathbb{I} .

The existence of this δ follows from the assumption that the pair $(\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r) := (\mathbb{Y}, \mathbb{A})$ form a *good pair*. This is a technical condition which implies the existence of a neighborhood of $\mathbb{Y} - \mathbb{U}$ (i.e. $\mathbb{Y} - \mathbb{V}$) that deformation retracts to $\mathbb{Y} - \mathbb{U}$. It is then straightforward to verify that $\mathbb{V} \subset \mathbb{U} \subset \mathbb{A}$ satisfies the hypotheses of Theorem A.2.

Therefore the chain map $k : C(\mathbb{Y}, \mathbb{A}) \rightarrow C(\mathbb{Y} - \mathbb{U}, \mathbb{A} - \mathbb{U})$ is an excision. It is defined as $k = r_\# \circ s$, where $r_\#$ is the chain map induced by the retraction $r : (\mathbb{Y} - \mathbb{V}, \mathbb{A} - \mathbb{V}) \rightarrow (\mathbb{Y} - \mathbb{U}, \mathbb{A} - \mathbb{U})$, and s is the chain-homotopy inverse of the chain map included by the inclusion of pairs $(\mathbb{Y} - \mathbb{V}, \mathbb{A} - \mathbb{V}) \rightarrow (\mathbb{Y}, \mathbb{A})$, $s : C(\mathbb{Y}, \mathbb{A}) \rightarrow C(\mathbb{Y} - \mathbb{V}, \mathbb{A} - \mathbb{V})$. \square

A.2 Theorem 3.1 We describe our long and technical proof of Theorem 3.1 based on diagram chasing. We first need the following lemma that comes from the short exact sequences of a pair ([27], page 140).

LEMMA A.1. *The quotient map on the chain level commutes. That is, for compatible maps $A \rightarrow B$ and $X \rightarrow Y$ there is a map $(X, A) \rightarrow (Y, B)$ such that the diagram in Fig. 13 is commutative.*

Proof. The assumption of compatibility ensures the left square commutes. Note that i, j must be injective maps and in all the case we consider f and g are also injective, which is sufficient for compatibility. To define h we note that $\text{im } h = \text{im } g / (\text{im } (g \circ i) \oplus \text{im } j)$. To show that the right square commutes ($h \circ q = r \circ g$), we note that any class in $\text{im } (r \circ g)$ must be in $\text{im } q$ by exactness and the assumption that the left square commutes ($g \circ i = j \circ f$). Since it is not in $\text{im } i$ or map to $\text{im } j$, it is in $\text{im } h$. Alternatively, any class in $\text{im } (h \circ q)$ must have a lift to $C(Y)$ since r is a surjection. This must be in $\text{im } g$ by the definition of h , which concludes the proof. \square

Theorem 3.1. If we have two compatible filtrations interleaved with two other compatible filtrations, the relative filtration is also interleaved. Formally, if compatible persistence modules $\mathcal{F} = \{F_\alpha\}_{\alpha \in \mathbb{R}}$ and $\mathcal{G} = \{G_\alpha\}_{\alpha \in \mathbb{R}}$ are ε_1 -interleaved, $\mathcal{A} = \{A_\alpha\}_{\alpha \in \mathbb{R}}$ and $\mathcal{B} = \{B_\alpha\}_{\alpha \in \mathbb{R}}$ are ε_2 -interleaved, then the relative modules $\{(F_\alpha, A_\alpha)\}_{\alpha \in \mathbb{R}}$ and $\{(G_\alpha, B_\alpha)\}_{\alpha \in \mathbb{R}}$ are ε -interleaved, where $\varepsilon = \max\{\varepsilon_1, \varepsilon_2\}$.

Proof. We begin with a list of notations. Suppose $\{F\}$ and $\{G\}$ are compatible and are ε -interleaved with homomorphisms $\{f_\alpha : H(F_\alpha) \rightarrow H(G_{\alpha+\varepsilon})\}$ and $\{g_\alpha : H(G_\alpha) \rightarrow H(F_{\alpha+\varepsilon})\}$. Suppose $\{A\}$ and $\{B\}$ are also compatible and ε -interleaved, with homomorphisms $\{\phi_\alpha : H(A_\alpha) \rightarrow H(B_{\alpha+\varepsilon})\}$ and $\{\psi_\alpha : H(B_\alpha) \rightarrow H(A_{\alpha+\varepsilon})\}$. For relative homology to be well-defined, we have injective maps at chain level, for simplicity, we further require $A_\alpha \hookrightarrow F_\alpha$ and $B_\alpha \hookrightarrow G_\alpha$.

We would like to prove that $\{(F, A)\}$ and $\{(G, B)\}$ are also interleaved, and we could construct their corresponding homomorphisms, $\{\mu_\alpha : H(F_\alpha, A_\alpha) \rightarrow H(G_{\alpha+\varepsilon}, B_{\alpha+\varepsilon})\}$ and $\{\nu_\alpha : H(G_\alpha, B_\alpha) \rightarrow H(F_{\alpha+\varepsilon}, A_{\alpha+\varepsilon})\}$.

To prove the result, we pass to the stack of long exact sequences in Fig. 14. First, we explain the notation. A map, i.e. ϕ_n^α , represents a map that maps n -dimensional homology groups of A_α to some other homology groups. We note that all the squares in this diagram (Fig. 14) commute based on Lemma A.1, and by assumption the two component filtrations are interleaved, so the first, second, fourth and fifth columns commute with the maps induced by inclusion. For example, the map induced by inclusion $\text{im } (H_n(F_\alpha) \rightarrow H_n(F_{\alpha+2\varepsilon}))$ equals $\text{im } (g_n^{\alpha+\varepsilon} \circ f_n^\alpha)$. Commutativity implies interleaving in some of the cases. We prove

$$\begin{array}{ccccccccc}
0 & \longrightarrow & C_n(A) & \xrightarrow{i} & C_n(X) & \xrightarrow{q} & C_n(X, A) & \longrightarrow & 0 \\
& & \downarrow f & & \downarrow g & & \downarrow h & & \\
0 & \longrightarrow & C_n(B) & \xrightarrow{j} & C_n(Y) & \xrightarrow{r} & C_n(Y, B) & \longrightarrow & 0
\end{array}$$

Figure 13: Commuting diagrams on the chain level.

$$\begin{array}{cccccccccccc}
H_n(A_\alpha) & \xrightarrow{i_n^\alpha} & H_n(F_\alpha) & \xrightarrow{j_n^\alpha} & H_n(F_\alpha, A_\alpha) & \xrightarrow{k_n^\alpha} & H_{n-1}(A_\alpha) & \xrightarrow{i_{n-1}^\alpha} & H_{n-1}(F_\alpha) & & & \\
\downarrow \phi_n^\alpha & & \downarrow f_n^\alpha & & \downarrow \mu_n^\alpha & & \downarrow \phi_{n-1}^\alpha & & \downarrow f_{n-1}^\alpha & & & \\
H_n(B_{\alpha+\varepsilon}) & \xrightarrow{p_n^{\alpha+\varepsilon}} & H_n(G_{\alpha+\varepsilon}) & \xrightarrow{q_n^{\alpha+\varepsilon}} & H_n(G_{\alpha+\varepsilon}, B_{\alpha+\varepsilon}) & \xrightarrow{r_n^{\alpha+\varepsilon}} & H_{n-1}(B_{\alpha+\varepsilon}) & \xrightarrow{p_{n-1}^{\alpha+\varepsilon}} & H_{n-1}(G_{\alpha+\varepsilon}) & & & \\
\downarrow \psi_n^{\alpha+\varepsilon} & & \downarrow g_n^{\alpha+\varepsilon} & & \downarrow \nu_n^{\alpha+\varepsilon} & & \downarrow \psi_{n-1}^{\alpha+\varepsilon} & & \downarrow g_{n-1}^{\alpha+\varepsilon} & & & \\
H_n(A_{\alpha+2\varepsilon}) & \xrightarrow{i_n^{\alpha+2\varepsilon}} & H_n(F_{\alpha+2\varepsilon}) & \xrightarrow{j_n^{\alpha+2\varepsilon}} & H_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon}) & \xrightarrow{k_n^{\alpha+2\varepsilon}} & H_{n-1}(A_{\alpha+2\varepsilon}) & \xrightarrow{i_{n-1}^{\alpha+2\varepsilon}} & H_{n-1}(F_{\alpha+2\varepsilon}) & & &
\end{array}$$

Figure 14: Commuting diagrams for the long exact sequence involving two pairs of filtrations. This is identical to Figure 6 for convinience.

the following triangle commutes (Fig. 7) through four claims.

Claim 1: if a relative class is in $\text{im}(H_n(F_\alpha, A_\alpha) \rightarrow H_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon}))$, and it is in $\text{im} j_n^\alpha$ and $\text{im} j_n^{\alpha+2\varepsilon}$, then it is in $\text{im} q_n^{\alpha+\varepsilon}$.

If a relative class γ in $\text{im}(H_n(F_\alpha, A_\alpha) \rightarrow H_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon}))$ is in $\text{im} j_n^\alpha$ and $\text{im} j_n^{\alpha+2\varepsilon}$, then by the interleaving, it must be in $H_n(G_{\alpha+\varepsilon})$. Therefore suppose γ is not in $\text{im} q_n^{\alpha+\varepsilon}$, it must have a preimage in $H_n(B_{\alpha+\varepsilon})$. Since γ is in $\text{im} j_n^{\alpha+2\varepsilon}$, it does not have a preimage in $H_n(A_{\alpha+2\varepsilon})$. This would imply that the lower left square does not commute ($g_n^{\alpha+\varepsilon} \circ p_n^{\alpha+\varepsilon} \neq j_n^{\alpha+2\varepsilon} \circ \psi_n^{\alpha+\varepsilon}$). That is a contradiction, therefore it must be in $\text{im} q_n^{\alpha+\varepsilon}$.

Claim 2: If a relative class is in $\text{im}(H_n(F_\alpha, A_\alpha) \rightarrow H_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon}))$, and it is in $\text{cok} j_n^\alpha$ and $\text{cok} j_n^{\alpha+2\varepsilon}$, it must be in $\text{cok} q_n^{\alpha+\varepsilon}$. If the relative class γ in $\text{im}(H_n(F_\alpha, A_\alpha) \rightarrow H_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon}))$ is in $\text{cok} j_n^\alpha$ and $\text{cok} j_n^{\alpha+2\varepsilon}$, then by exactness γ maps into $\text{im} k_n^\alpha$ and $\text{im} k_n^{\alpha+2\varepsilon}$, that is, it maps to a non-trivial element in $H_{n-1}(A_\alpha)$ and $H_{n-1}(A_{\alpha+2\varepsilon})$. By the interleaving between A and B , it must also map to an element of $H_{n-1}(B_{\alpha+\varepsilon})$. Furthermore, it must be in $\ker i_{n-1}^\alpha$. Therefore suppose γ is not in $\text{cok} q_n^{\alpha+\varepsilon}$ (or equivalently, $\text{im} r_n^{\alpha+\varepsilon}$ or $\ker p_{n-1}^{\alpha+\varepsilon}$), it must map to a class in $H_{n-1}(G_{\alpha+\varepsilon})$, which implies that the top right square does not commute ($f_{n-1}^\alpha \circ i_{n-1}^\alpha \neq p_{n-1}^{\alpha+\varepsilon} \circ \phi_{n-1}^{\alpha+\varepsilon}$) leading to a contradiction.

We now show that commutativity is not a suffi-

cient argument. Consider a persistent relative class in $H_n(F_\alpha, A_\alpha) \rightarrow H_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon})$ such that it is in $\text{im} j_n^\alpha$ and $\text{cok} j_n^{\alpha+2\varepsilon}$. Alternatively, it may be in $\text{cok} j_n^\alpha$ and $\text{im} j_n^{\alpha+2\varepsilon}$. In these cases, we may map this class to zero the middle row and still maintain the commutativity of the diagram (although this implies the relative filtrations are not interleaved). This problem stems from the fact that the maps between persistent modules do not *split* (The relative persistence module does not split into direct sum of the image and cokernel in the long exact sequence).

Claim 3: If the relative class is in $\text{im}(H_n(F_\alpha, A_\alpha) \rightarrow H_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon}))$, then it is not possible that it is in $\text{im} j_n^\alpha$ and $\text{cok} j_n^{\alpha+2\varepsilon}$ at the same time.

First we handle the case where the relative class is in $\text{im} j_n^\alpha$ and $\text{cok} j_n^{\alpha+2\varepsilon}$ by showing this cannot occur. Since it is in $\text{im} j_n^\alpha$ at the chain level, there is a cycle representative in $Z_n(F_\alpha)$. Since this maps to a cycle representative in $Z_n(F_{\alpha+2\varepsilon})$, this implies that the cycle is in the boundary. However, looking at the relevant part of the short exact sequence shown in Fig. 15.

The cycle representative in $C_n(F_{\alpha+2\varepsilon})$ lifts to some element in $C_{n+1}(F_{\alpha+2\varepsilon})$. Now by assumption, there is still some cycle representative in $C_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon})$. By commutativity, the bounding element in $C_{n+1}(F_{\alpha+2\varepsilon})$ must also map to a bounding element of the cycle representative in $C_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon})$, meaning it cannot be a relative homology class. If on the other hand, the

$$\begin{array}{ccc}
C_{n+1}(F_{\alpha+2\alpha}) & \xrightarrow{q} & C_{n+1}(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon}) \\
\downarrow \partial & & \downarrow \partial \\
C_n(F_{\alpha+2\alpha}) & \xrightarrow{q} & C_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon})
\end{array}$$

Figure 15: Short exact sequence on chain level.

cycle representative in $C_n(F_{\alpha+2\varepsilon})$ is in the kernel of the quotient map, a relative homology class would appear one dimension up. This is the case we deal with next.

Claim 4: If the relative class is in $\text{im}(H_n(F_\alpha, A_\alpha) \rightarrow H_n(F_{\alpha+2\varepsilon}, F_{\alpha+2\varepsilon}))$, and it is in $\text{cok } j_n^\alpha$ and $\text{im } j_n^{\alpha+2\varepsilon}$, then it must be in $\text{im } q_n^{\alpha+\varepsilon}$ or $\text{cok } q_n^{\alpha+\varepsilon}$ (i.e. it must be $H_n(G_{\alpha+\varepsilon}, B_{\alpha+\varepsilon})$).

For a relative class in $\text{cok } j_n^\alpha$, there is a cycle representative in $C_{n-1}(A_\alpha)$ of the corresponding class $H_{n-1}(A_\alpha)$ which by the injectivity of the interleaving, maps to a cycle in $C_{n-1}(B_{\alpha+\varepsilon})$ and $C_{n-1}(A_{\alpha+2\varepsilon})$. Further, since it is in $\text{cok } j_n^\alpha$, it follows that it maps to a bounded cycle in $C_{n-1}(F_\alpha)$ (and by injectivity) the corresponding cycle representatives in $C_{n-1}(G_{\alpha+\varepsilon})$ and $C_{n-1}(F_{\alpha+2\varepsilon})$ are also bounded. Since this relative class is assumed to be in $\text{im } j_n^{\alpha+2\varepsilon}$, it follows that the cycle representative in $C_{n-1}(A_{\alpha+2\varepsilon})$ is now bounded, with the pre-boundary mapping to a cycle in $C_n(F_{\alpha+2\varepsilon})$. This follows from a chain level understanding of the exactness of the bottom row. Take the representative $(n-1)$ -cycle in $C_{n-1}(A_\alpha)$ denoted by a and map it into $C_{n-1}(F_\alpha)$. $i_{n-1}^\alpha(a)$ has a pre-boundary in $C_n(F_\alpha)$ which maps to the cycle representative of the relative class in $C_n(F_\alpha, A_\alpha)$. This is just the connecting homomorphism construction. If we map this relative cycle representative to $C_n(F_{\alpha+2\varepsilon}, A_{\alpha+2\varepsilon})$, since the class is in $j_n^{\alpha+2\varepsilon}$, it lifts to a non-trivial cycle in $C_n(F_{\alpha+2\varepsilon})$. This cycle is precisely the image of the pre-boundary of a in $C_{n-1}(A_{\alpha+2\varepsilon})$ mapped to $C_n(F_{\alpha+2\varepsilon})$ plus the pre-boundary of the image of $i_{n-1}^\alpha(a)$ in $C_{n-1}(F_{\alpha+2\varepsilon})$.

There are two case to consider. If $\phi_{n-1}^\alpha(a)$ is a non-trivial cycle, then there is a homology class in $\ker p_{n-1}^{\alpha+\varepsilon}$ and by exactness, a corresponding class in the $\text{cok } q_n^{\alpha+\varepsilon}$.

If $\phi_{n-1}^\alpha(a)$ maps to a bounded cycle, then by the same reasoning as above, the pre-boundary of this cycle in $C_n(B_n^{\alpha+\varepsilon})$ must map to a non-trivial cycle in $C_n(G_n^{\alpha+\varepsilon})$. Hence there is a corresponding class in $\text{im } q_n^{\alpha+\varepsilon}$. Proving the claim.

Following the above four claims, we've shown the triangle in Fig. 7 commutes. Fig. 7 equals the trapezoid in Fig. 16(a) by setting $\alpha' = \alpha$. It follows that the trapezoid in Fig. 16(a) commutes based on similar diagram chasing argument.

The other diagrams in Fig. 16 follow similar proofs. For example, to show that the diagram in Fig. 16(d) commutes, the argument goes through in precisely the same way, on diagrams shown in Fig. 17 and Fig. 18.

This shows that the two commute and hence we conclude that the relative filtrations are interleaved. \square

A.3 Theorem 3.2 To prove Theorem 3.2, first we prove a collection of lemmas (A.2, A.3, A.4, A.5, A.8).

LEMMA A.2. *If L is an ε -sample of \mathbb{X} then $\{\mathbb{X}_\alpha\}$ is ε -interleaved with $\{L_\alpha\}$.*

Proof. Given that L is an ε -sample of \mathbb{X} , by definition, $L \subseteq \mathbb{X}$, this implies that (a) $L_\alpha \subseteq \mathbb{X}_\alpha$ and (b) $L_{\alpha+\varepsilon} \subseteq \mathbb{X}_{\alpha+\varepsilon}$. Subsequently, we would prove by the triangle inequality that, (c) $\mathbb{X}_\alpha \subseteq L_{\alpha+\varepsilon}$. Combining (a), (b) and (c), we have,

$$L_\alpha \subseteq \mathbb{X}_\alpha \subseteq L_{\alpha+\varepsilon} \subseteq \mathbb{X}_{\alpha+\varepsilon}.$$

By the special case of ε -interleaving, we have $L_\alpha \subseteq \mathbb{X}_{\alpha+\varepsilon}$ and $\mathbb{X}_\alpha \subseteq L_{\alpha+\varepsilon}$, therefore the persistent homology modules of $\{\mathbb{X}_\alpha\}$ and $\{L_\alpha\}$ is ε -interleaved.

Now we prove that the inclusion in (c) holds. For any point $p \in \mathbb{X}_\alpha$, let $q = \arg \min_{x \in \mathbb{X}} d(p, x)$, therefore by definition of \mathbb{X}_α , $d(p, q) \leq \alpha$. Since $q \in \mathbb{X}$ and L is an ε -sample, let $s = \arg \min_{z \in L} d(p, z)$, by definition of L , $d(q, s) \leq \varepsilon$. By triangle inequality, $d(p, s) \leq d(p, q) + d(q, s) \leq \alpha + \varepsilon$. Therefore $p \in L_{\alpha+\varepsilon}$. \square

LEMMA A.3. *The nerve of L_α , $\mathcal{N}(L_\alpha)$, is homotopic to L_α .*

Proof. This is an application of the Nerve Theorem. Since these are Euclidean balls in Euclidean space, they are all convex as are all their intersections. They are hence contractible and the Nerve Theorem applies. \square

LEMMA A.4. *$\{L_\alpha - \text{int } B_r\}$ is ε -interleaved with $\{\mathbb{X}_\alpha - \text{int } B_r\}$.*

Proof. The proof follows directly from Lemma A.2, since $\{L_\alpha\}$ and $\{\mathbb{X}_\alpha\}$ are ε -interleaved, removing the same set from both spaces does not change this. \square

LEMMA A.5. *For $\alpha < r$, the nerve of $L_\alpha - \text{int } B_r$ is homotopic to the union of balls L_α with $\text{int } B_r$ removed.*

Proof. By removing $\text{int } B_r$, the balls and their intersections are no longer convex. To apply the Nerve theorem, we show that the condition $\alpha < r$ ensures that they are

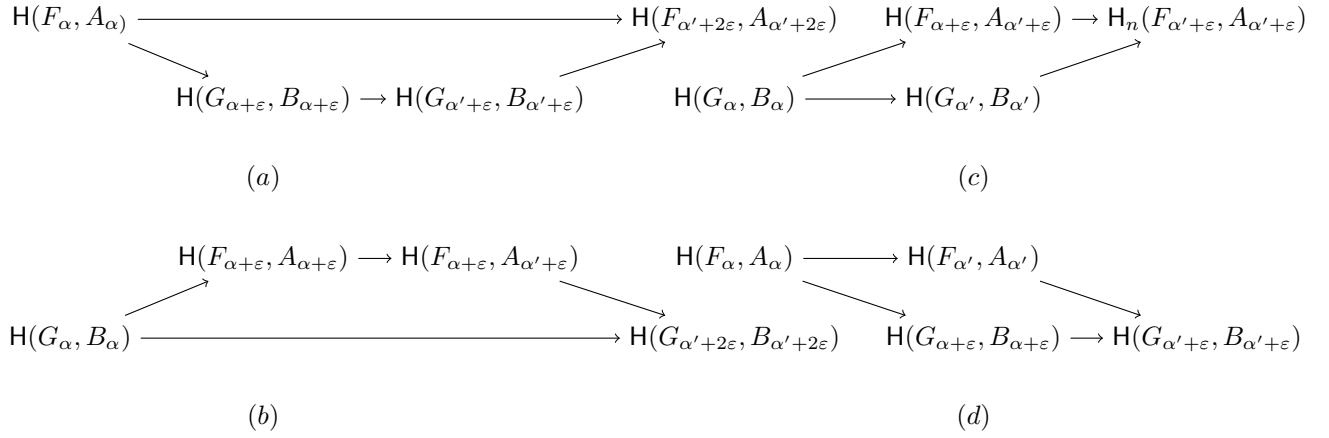


Figure 16: Commuting diagrams for ε -leaving of the pairs.

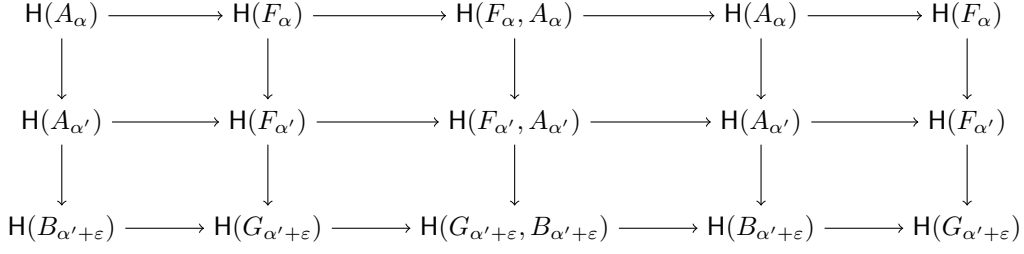


Figure 17: Commuting diagrams for Fig. 16 (d) top path.

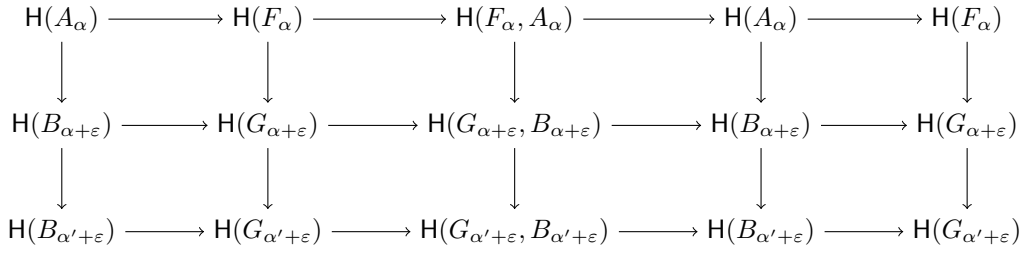


Figure 18: Commuting diagrams for Fig. 16 (d) bottom path.

still contractible. This is only an outline of the proof. The goal is to prove that from any intersection there is a homotopy to a convex body and hence all the intersections are contractible.

Take an arbitrary intersection. If it does not intersect $\text{int } B_r$, it is convex. If it does, then take the tangent plane to B_r at a point on the boundary within the intersection. Clearly the half-plane which does not contain B_r intersected with the intersection is convex and hence contractible.

The rest of the intersection can be retracted to the tangent plane, which we prove by giving an explicit deformation retract. The tangent plane will be referred to as $T(s)$ (the tangent plane at point s).

First, we define a deformation retract before we remove $B_r(x)$ (Figure 19(a)). We consider a straight-line homotopy to s . Since the space is convex, this retract is *within the intersection*. Now for any point p , consider the point where the retract intersects $B_r(x)$ and denote it by q . To define the retract, we alter the

path from the straight line ps to the following path (Figure 19(b)): from p to q , it coincides with straight-line ps ; and from q to s , it is the geodesic along the boundary of the ball, $\partial B_r(x)$. This path is continuous on the set outside $B_r(x)$ and to ensure it is a valid retract, we must show that for any intersection, this geodesic lies in the intersection.

By definition, the intersection must be contained in all the balls of radius α . Without loss of generality, we only consider the points q to s . Any ball which contains both q and s , and whose radius is smaller than r , must also contain the geodesic (Figure 19(c)). When the center of the ball lies outside $B_r(x)$, all points along the geodesic are closer to the center than s and q and hence must also be in the ball (Figure 19(d)).

In particular, for balls centered inside $B_r(x)$ (Figure 19(c)), consider the sphere centered at that point which passes through q and s . The sphere is a subspace of the ball since q and s are in the ball and this implies that the sphere has a higher curvature than $B_r(x)$. This means the length of the arc along the sphere is longer than the geodesic, which implies that it lies within the ball. Hence, the non-convex part is contractible as well. \square

We now prove that the algorithm described in Section 3 accurately computes the parameter values of the filtered Vietoris-Rips complex for the 0 and 1-skeleton.

LEMMA A.6. *Let B_r denote a ball of radius r centered at point p . For a point x_i , let $g(x_i)$ denote the infimum of parameter α such that $B_\alpha(x_i) - \text{int } B_r \neq \emptyset$. Then,*

$$g(x_i) = \begin{cases} 0 & d(x_i, p) \geq r \\ r - d(x_i, p) & d(x_i, p) < r \end{cases}$$

Proof. If x_i lies outside B_r , that is, $d(x_i, p) \geq r$, then $B_0(x_i) - \text{int } B_r$ is the point at x_i and hence nonempty. If x_i lies within B_r , then for $\alpha = r - d(x_i, p)$, $B_\alpha(x_i)$ intersects the boundary of B_r . Therefore, $B_\alpha(x_i) - \text{int } B_r \neq \emptyset$. This is illustrated in Figure 9. \square

LEMMA A.7. *Let B_r denote a ball of radius r centered at point p . For two points x_i and x_j , let the function $g(x_i, x_j)$ denotes the infimum of α such that $B_\alpha(x_i) \cap B_\alpha(x_j) - \text{int } B_r \neq \emptyset$. Let m denote their midpoint, i.e.*

$$m = \frac{x_i + x_j}{2}.$$

If $d(m, p) \geq r$, then $g(x_i, x_j) = d(x_i, m)$. Otherwise, if

$d(m, p) < r$, then

$$g(x_i, x_j) = \sqrt{\left(\sqrt{r^2 - d(o, p)^2} - \sqrt{d(m, p)^2 - d(o, p)^2} \right)^2 + d(x_i, x_j)^2 / 4}$$

where

$$d(o, p) = \left| \frac{(p - m)^T (x_i - x_j)}{d(x_i, x_j)} \right|.$$

Proof. First, if $d(m, p) \geq r$, then the first time $B_\alpha(x_i) \cap B_\alpha(x_j) \neq \emptyset$ is at $\alpha = d(x_i, x_j)/2$ at the point m . Since m lies outside $\text{int } B_r$ (i.e. $d(m, p) \geq r$), we conclude that the intersection is still non-empty with $\text{int } B_r$ removed. For $d(m, p) < r$ for $\alpha = d(x_i, x_j)/2$ although the intersection $B_\alpha(x_i) \cap B_\alpha(x_j)$ is non-empty, $B_\alpha(x_i) \cap B_\alpha(x_j) - \text{int } B_r$ remains empty.

To find the first non-empty intersection, we must find the closest equidistant point to x_i and x_j which lies on the boundary of B_r . Suppose we are in d -dimensional Euclidean space. The space of equidistant points is given by a $(d-1)$ -hyperplane which is determined by the point m and the normal vector $x_i - x_j$. The boundary of a B_r is a $(d-1)$ -sphere and their intersection forms a $(d-2)$ -sphere, as m lies on the hyperplane and inside B_r .

To find the closest point on the $(d-2)$ -sphere, we can perform the following bit of geometry. Denote the center of the $(d-2)$ -sphere as o . The closest point on the sphere is the point at a distance r (on the boundary of B_r) on the line going through o and m .

The first step is to compute $d(o, p)$. This is given by normalized inner product

$$d(o, p) = \left| \frac{(p - m)^T (x_i - x_j)}{d(x_i, x_j)} \right|$$

Furthermore, using the Pythagorean theorem on the construction in Figure 10, we can compute the time of the first non-empty intersection as

$$\alpha = \sqrt{\left(\sqrt{r^2 - d(o, p)^2} - \sqrt{d(m, p)^2 - d(o, p)^2} \right)^2 + d(x_i, x_j)^2 / 4}$$

concluding the proof. \square

Lemmas A.6 and A.7 describe the filtration of the 0- and 1-skeleton of the nerve of the union of balls with B_r removed. Lemma A.5 tells us that the Nerve Lemma still applies for $\alpha < r$ and so we have a faithful topological representation. We now recount the construction of the Vietoris-Rips filtration \mathcal{F}_α .

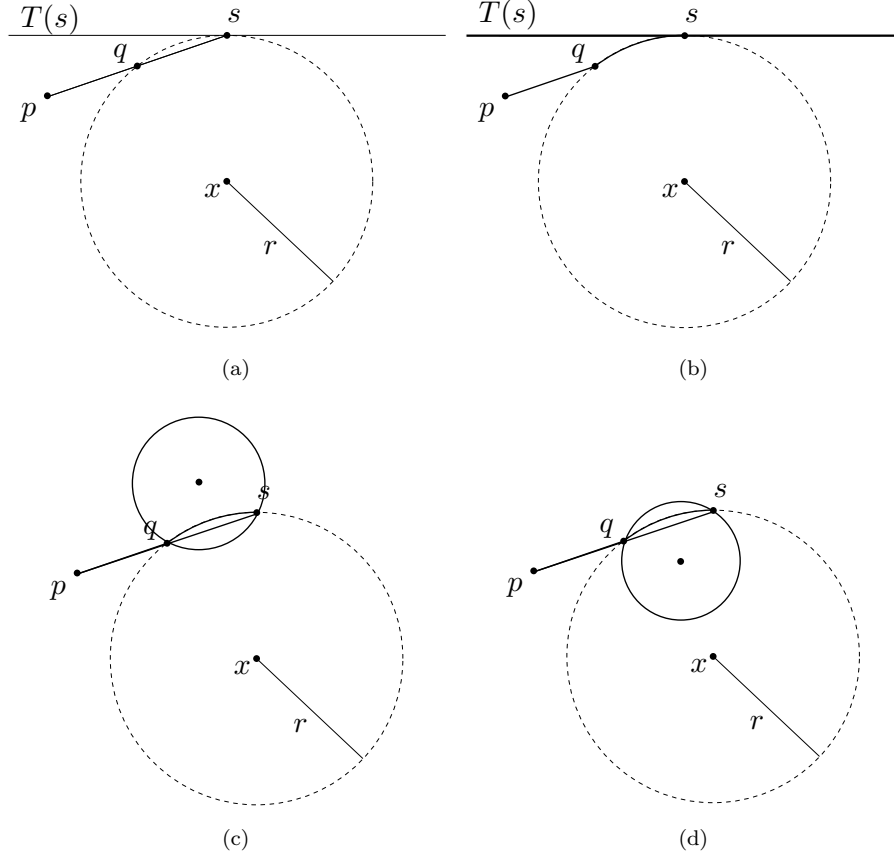


Figure 19: (a) The layout of the points p, q, s along with the deformation retract. (b) The modified deformation retract. (c) The situation when the center of the ball is outside $B_r(p)$. (d) The situation the center of the ball is inside $B_r(p)$ and the radius of the ball is less than r .

The Vietoris-Rips filtration \mathcal{F}_α is based on the 0- and 1-skeleton of $\mathcal{N}(L_\alpha - \text{int } B_r)$. By construction, for a fixed α , the 0- and 1-simplices of $\mathcal{F}_\alpha(L)$ are identical to the 0- and 1-skeleton of $\mathcal{N}(L_\alpha - \text{int } B_r)$, and $\mathcal{F}_\alpha(L)$ by definition is the clique complex of the 1-skeleton of $\mathcal{N}(L_\alpha - \text{int } B_r)$.

Formally, for dimensions 0 and 1 and a fixed α , we could define such a filtration directly at the chain level,

$$C_0(\mathcal{F}_\alpha(L)) = C_0(\mathcal{N}(L_\alpha - \text{int } B_r)),$$

$$C_1(\mathcal{F}_\alpha(L)) = C_1(\mathcal{N}(L_\alpha - \text{int } B_r)),$$

and construct the clique complex based on the 1-skeleton of the nerve, which gives us higher-dimensional chain complexes.

LEMMA A.8. $\{L_\alpha - \text{int } B_r\}$ is α -interleaved with $\mathcal{F}_\alpha(L)$.

Proof. For $i \leq 1$, $C_i(\mathcal{F}_\alpha) = C_i(\mathcal{N}(L_\alpha - \text{int } B_r))$ follows by construction. For higher dimensions e.g. $i > 1$ we

have the following sequence,

$$C_i(\mathcal{N}(L_\alpha - \text{int } B_r)) \rightarrow C_i(\mathcal{F}_\alpha(L)) \rightarrow C_i(\mathcal{N}(L_{2\alpha} - \text{int } B_r)),$$

where the maps are injective. The first injective chain map is induced from inclusion on the space level (i.e. 0- and 1-skeleton) and follows from the fact that the Vietoris-Rips complex fills in all possible co-faces⁹. The second injective chain map is induced as follows: if we have k points such that, for each pair of points, the corresponding balls centered at these points have intersection that lies partially outside B_r , then at parameter 2α , the corresponding balls centered at these points will have a $(k - 1)$ -way intersection. \square

⁹To get the standard Vietoris-Rips inclusions we would have to scale the edges by 2 to take into account the standard definition of the Vietoris-Rips filtration.

Theorem 3.2 For $\alpha < r$, $\{(\mathcal{R}_\alpha(L), \mathcal{F}_\alpha(L))\}$ is $(\alpha + 2\varepsilon)$ -interleaved with $\{\mathbb{X}_\alpha, \mathbb{X}_\alpha - \text{int } B_r\}$.

Proof. Lemma A.2 and Lemma A.3 give an ε -approximation of the filtration $\{\mathbb{X}_\alpha\}$ and passing to the Vietoris-Rips complex introduces an additional multiplicative factor of 2. We have, on the chain level, the following sequence connected by injective maps,

$$C_i(\mathbb{X}_\alpha) \rightarrow C_i(L_{\alpha+\varepsilon}) \rightarrow C_i(\mathcal{R}_{2(\alpha+\varepsilon)}(L)),$$

$$C_i(\mathcal{R}_\alpha(L)) \rightarrow C_i(L_\alpha) \rightarrow C_i(\mathbb{X}_{(\alpha+\varepsilon)}).$$

Therefore, the interleaving parameter between the filtration $\{\mathbb{X}_\alpha\}$ and the Vietoris-Rips complex filtration $\{\mathcal{R}_\alpha(L)\}$ is $\alpha + 2\varepsilon$.

Similar construction gives us the same bound for the filtration $\{\mathbb{X}_\alpha - \text{int } B_r\}$. That is, based on Lemma A.5, Lemma A.6, Lemma A.7, and Lemma A.8 we have,

$$C_i(\mathbb{X}_\alpha - \text{int } B_r) \rightarrow C_i(L_{\alpha+\varepsilon} - \text{int } B_r) \rightarrow C_i(\mathcal{F}_{\alpha+\varepsilon}(L)),$$

$$C_i(\mathcal{R}_\alpha(L)) \rightarrow C_i(L_{2\alpha} - \text{int } B_r) \rightarrow C_i(\mathbb{X}_{2(\alpha+\varepsilon)} - \text{int } B_r).$$

Applying Theorem 3.1 gives the result. \square