Computational Geometry 48 (2015) 575–588

Contents lists available at ScienceDirect

Computational Geometry: Theory and Applications

www.elsevier.com/locate/comgeo

# Graph induced complex on point data

Tamal K. Dey *, Fengtao Fan, Yusu Wang

*Department of Computer Science and Engineering, The Ohio State University, Columbus, OH 43210, USA*

ARTICLE INFO

*Article history:*
Available online 14 April 2015

*Keywords:*
Point cloud data
Homology
Simplicial complex
Sparsification
Topological persistence

ABSTRACT

The efficiency of extracting topological information from point data depends largely on the complex that is built on top of the data points. From a computational viewpoint, the favored complexes for this purpose have so far been Vietoris–Rips and witness complexes. While the Vietoris–Rips complex is simple to compute and is a good vehicle for extracting topology of sampled spaces, its size becomes prohibitively large for reasonable computations. The witness complex on the other hand enjoys a smaller size because of a subsampling, but fails to capture the topology in high dimensions unless imposed with extra structure. We investigate a complex called the *graph induced complex* that, to some extent, enjoys the advantages of both. It works on a subsample but still retains the power of capturing the topology as the Vietoris–Rips complex. It only needs a graph connecting the original sample points from which it builds a complex on the subsample thus taming the size considerably. We show that, using the graph induced complex one can (i) infer the one dimensional homology of a manifold from a lean subsample, (ii) reconstruct a surface in three dimensions from a sparse subsample without computing Delaunay triangulations, (iii) infer the persistent homology groups of compact sets from a sufficiently dense sample. We provide experimental evidences in support of our theory.

Published by Elsevier B.V.

## 1. Introduction

Acquiring knowledge about a sampled space from a set of points has become a key problem in many areas of science and engineering. The sampled space could be a hidden manifold sitting in some high dimensions, or could be a compact subset of some Euclidean space. Topological information such as the rank of the homology groups, or their persistent behavior can divulge important features of the hidden space. Therefore, a considerable effort has ensued to extract topological information from point data in recent years [7,9,16,19]. With the advent of advanced technologies, the data is often generated in abundance. Mixed with the burden of high dimensionality, large data sets pose a challenge to the resource required to process them. As a result, some recent investigations have focused on how to use a lighter data structure or sparsify the input, which aids a faster computation, but still guarantees that the output inference is correct.

Point data by themselves do not have interesting topology. So, a foremost step in inferring topology from data is to impose a structure such as a simplicial complex onto it. The Delaunay, Čech, Vietoris–Rips, and witness complexes are some of the most commonly proposed complexes for this purpose. Among these, Vietoris–Rips (Rips in short) and witness complexes [9] have been favored because they can be constructed with simple computations. Rips complexes are easy to construct as they can be built from a graph by recognizing the cliques in it. However, the presence of simplices correspond-

* Corresponding author.
*E-mail addresses:* tamaldey@cse.ohio-state.edu (T.K. Dey), fanf@cse.ohio-state.edu (F. Fan), yusu@cse.ohio-state.edu (Y. Wang).

http://dx.doi.org/10.1016/j.comgeo.2015.04.003
0925-7721/Published by Elsevier B.V.

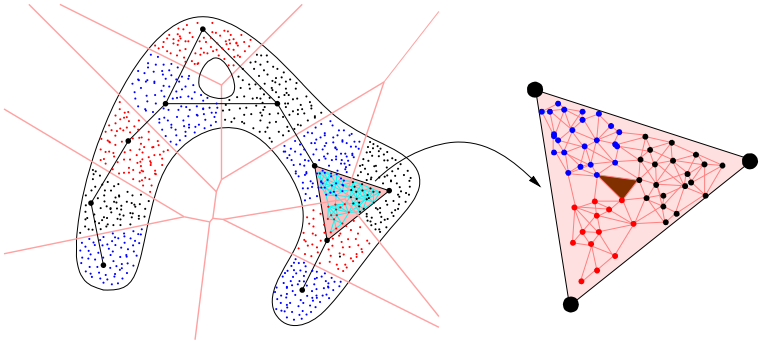


**Fig. 1.** A graph induced complex shown with bold vertices, edges, and a shaded triangle on the left. Input graph within the shaded triangle is shown on the right.

ing to all cliques makes its size quite large. Even in three dimensions with a few thousand points, the size of the Rips complex can be an obstacle, if not a stopper, for further processing. Witness complexes, on the other hand, have too few simplices to capture the topology of the sampled space in dimensions three or more [4]. To tackle this issue, Boissonnat et al. [4] suggested modifications to the original definition of witness complex [12]. This enlarges the witness complexes but makes it more complicated and costly to compute.

We investigate a new complex, a version of which was originally introduced in [18] for the application of sensor network routing. We set up a more general definition and call it the *graph induced complex*. We provide new theoretical understanding of the graph induced complex in terms of topology inference. In particular, we show that, when equipped with an appropriate metric, this complex can help deciphering the topology from data. It retains the simplicity of the Rips complex as well as the sparsity of the witness complex. Its construction resembles the sparsified Rips complex proposed in [24] and also the combinatorial Delaunay triangulation proposed in [6], but it does not build a Rips complex on the subsample and thus is sparser than the Rips complex with the same set of vertices. This fact makes a real difference in practice as our preliminary experiments show. The idea of graph induced complex also bears similarity to the geodesic Delaunay triangulation which was proposed to recover the topology of a bounded planar region (with holes) from point samples [23]. Our work investigates its theoretical properties and generalizes it to settings beyond the planar case.

Given a graph $G$ on a set of points $P$ equipped with a metric, one can build a graph induced complex on a subsample $Q \subseteq P$ as follows. A simplex is in the complex if and only if its vertex set $V \subseteq Q$ has the property that a set of points in $P$, each being closest to exactly one vertex in $V$, forms a clique in $G$. Fig. 1 shows a graph induced complex for a set of data points in the plane. Subsampled points are the darker vertices. Input points are grouped according to the proximity to the subsampled vertices (indicated with a Voronoi partition). The shaded triangle enlarged on the right-hand side is in the graph induced complex since there is a 3-clique in the input graph whose 3 vertices have 3 different closest point in the subsample. Observe that, in this example, the graph induced complex has the same homology as the sampled space.

Fig. 2 shows experimental results on two data sets, 40,000 sample points from a Klein bottle in $\mathbb{R}^4$ and 15,000 sample points from the primary circle of natural image data considered in $\mathbb{R}^{25}$ [1]. The graphs connecting any two points within $\alpha = 0.05$ unit distance for Klein bottle and $\alpha = 0.6$ unit distance for the primary circle were taken as input for the graph induced complexes. The 2-skeleton of the Rips complexes for these $\alpha$ parameters have 608,200 and 1,329,672,867 simplices respectively. These sizes are too large to carry out fast computations.

For comparison, we constructed the graph induced complex, sparsified Rips complex, and the witness complex on the same subsample determined by a parameter $\delta$. The parameter $\delta$ is also used in the graph induced complex (see definitions later) and the witness complex. The edges in the Rips complex built on the same *subsample* were of lengths at most $\alpha + 2\delta$ (justified by Proposition 2.8). We varied $\delta$ and observed the rank of the one dimensional homology group ($\beta_1$). The plots show that the graph induced complex captured $\beta_1$ correctly for a significantly wider range of $\delta$ (left plots) while its size remained comparable to that of the witness complex (right plots). In some cases, the graph induced complex could capture the correct $\beta_1$ with remarkably small number of simplices. For example, it had $\beta_1 = 2$ for Klein bottle when there were 278 simplices for $\delta = 0.7$ and 154 simplices for $\delta = 1.0$. In both cases $\beta_1$ for Rips and witness complexes did not match with that of the sampled spaces while the Rips complex had a much larger size ($\log_e$ scale plot) and the witness complex had comparable size. This illustrates why the graph induced complex can be a better choice than the Rips and witness complexes.

We establish three different results. First, we show that the one-dimensional homology group of surfaces in three dimensions can be determined by graph induced complexes. Even the surface itself can be reconstructed with some post-processing from a sparse subsample of a sample that could be excessively dense. Second, we show that, for higher-dimensional manifolds, one-dimensional homology can still be determined from graph induced complexes with a simple modification of the metric. Finally, we extend our results to other homology groups where we show that the persistent homology groups of a pair of graph induced complexes can determine the homology groups of compact spaces. Experimental results support our theory.

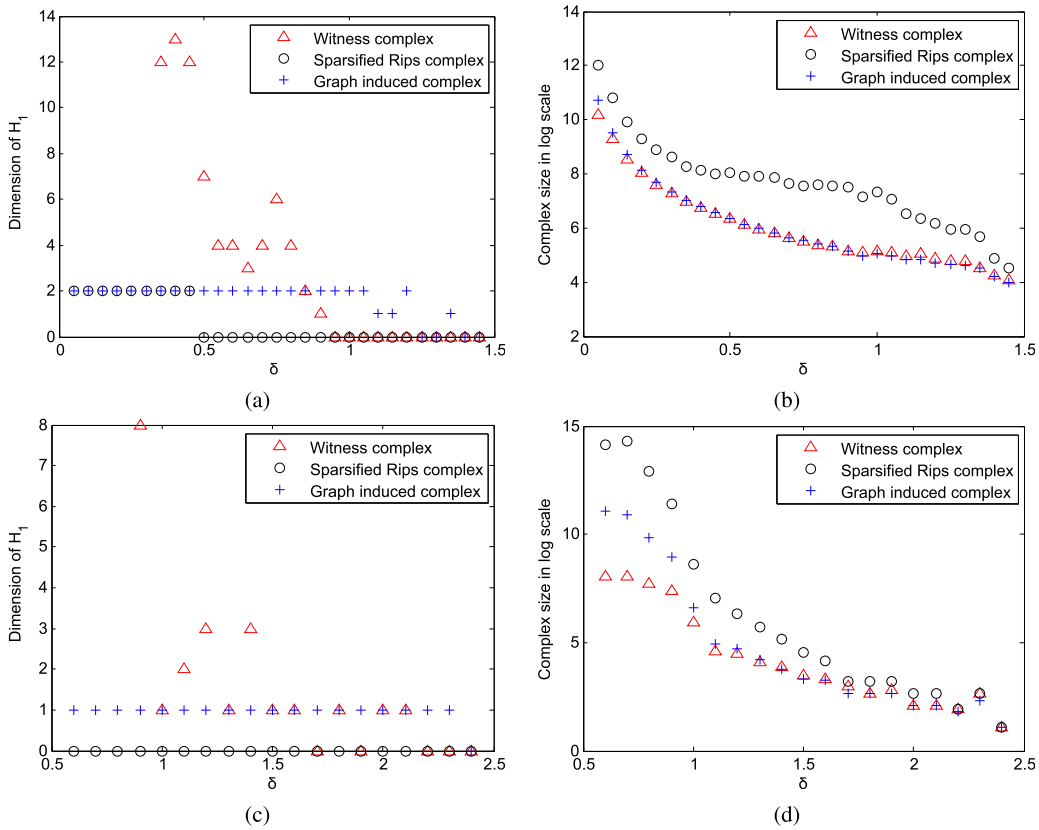



**Fig. 2.** Comparison results for Klein bottle in $\mathbb{R}^4$ (top row) and primary circle in $\mathbb{R}^{25}$ (bottom row). The estimated $\beta_1$ for three complexes are shown on the left, and their sizes are shown on log scale on the right.

## 2. Graph induced complex and preliminaries

First we define the graph induced complex in a more abstract setting which does not require a metric.

**Definition 2.1.** Let $G(V)$ be a graph with the vertex set $V$ and let $\nu : V \to V'$ be a vertex map where $\nu(V) = V' \subseteq V$. The graph induced complex $\mathcal{G}(G(V), V', \nu)$ is defined as the simplicial complex where a $k$-simplex $\sigma = \{v'_1, v'_2, \ldots, v'_{k+1}\}$ is in $\mathcal{G}(V, V', \nu)$ if and only if there exists a $(k+1)$-clique $\{v_1, v_2, \ldots, v_{k+1}\} \subseteq V$ so that $\nu(v_i) = v'_i$ for each $i \in \{1, 2, \ldots, k+1\}$. To see that it is indeed a simplicial complex, observe that a subset of a clique is also a clique.

Now we specialize the graph induced complex to the case where the vertices of the input graph come from a metric space.

**Definition 2.2.** A metric space $(X, d)$ is a pair where $X$ is a set and $d : X \times X \to \mathbb{R}_+$ is a distance function satisfying $d(x, y) \geq 0$, $d(x, y) = 0$ iff $x = y$, $d(x, y) = d(y, x)$, and $d(x, y) \leq d(x, z) + d(z, y)$.

**Definition 2.3.** Let $(P, d)$ be a metric space where $P$ is a finite point set and let $Q \subset P$ be a subset. Let $\nu_d : P \to Q$ denote the nearest point map where $\nu_d(p)$ is a point in $\operatorname{argmin}_{q \in Q} d(p, q)$. Given a graph $G(P)$ with vertex set $P$, we define its graph induced complex as $\mathcal{G}(G(P), Q, d) := \mathcal{G}(G(P), Q, \nu_d)$.

Among the many possible choices for $d$, we will focus on two cases where $d = d_E$, the Euclidean distance, and $d = d_G$, the graph distance induced by the graph $G(P)$ assuming its edges have non-negative weights. For any two vertices $p_1, p_2 \in P$, the distance $d_G(p_1, p_2)$ is the length of the shortest path between $p_1$ and $p_2$ in $G(P)$. We will describe the choices of the distance functions when necessary.

In our case, the point set $P$ will be a discrete subset of a compact smooth manifold $M \subset \mathbb{R}^n$ without boundary, or of a compact set $X \subset \mathbb{R}^n$. The graph $G(P)$ will be the graph $G^\alpha(P) = (P, E^\alpha)$ where $(p_1, p_2) \in E^\alpha$ if and only if $\|p_1 - p_2\| \leq \alpha$. The graph induced complex induced by $G^\alpha(P)$ on a subset $Q \subseteq P$ under a distance function $d$ will be the focus of our study. To emphasize the dependence on the parameter $\alpha$, we write it as $\mathcal{G}^\alpha(P, Q, d) := \mathcal{G}(G^\alpha(P), Q, d)$. Notice that $G^\alpha(P)$

is induced by Euclidean metric whereas $d$ is used to define the nearest point map in constructing $\mathcal{G}^\alpha(P, Q, d)$. As mentioned earlier, the graph induced complex bears a similarity to the complex considered in [18,23]. One may also draw a parallel between the graph induced complexes and the well-known witness complexes [12] where $P$ acts as a witness set $W$ and $Q$ acts as a landmark set $L \subseteq W$. However, the analogy does not extend any further since the construction of the witness complex and its variants [4] differs from that of the graph induced complex. For example, the original witness complex defined in [12] includes a $k$-simplex with vertex set in $L$ only if its vertices belong to the $k$-nearest neighbors of a point in $W$. In contrast the graph induced complex includes a $k$-simplex only if its vertices have nearest neighbors in $W$ that form a $k$-clique in a graph built on the vertices belonging to $W$. Similar to the witness complex, the graph induced complex builds upon a subsampling, but, unlike the witness complex, it enjoys some topological properties without any extra modifications such as weighting [4].

### 2.1. Sampling, homology, and sandwiching

As indicated before, the input point set $P$ is a sample of a smooth manifold $M$ or a compact set $X$ embedded in Euclidean space. We will also subsample $P$ according to a distance function $d$. Therefore, we define sampling in a more general context.

**Definition 2.4.** A finite set $P \subseteq X$ is an $\varepsilon$-sample of a metric space $(X, d)$, if for each point $x \in X$ there is a point $p \in P$ so that $d(x, p) \leq \varepsilon$. Additionally, $P$ is called $\delta$-sparse if $d(p_1, p_2) \geq \delta$ for any pair of points in $P$.

The point set $P$ does not have interesting topology by itself. We build simplicial complexes using $P$ as the vertex set to infer the topology of the sampled space $X$. Specifically, our goal is to infer the homology groups of a manifold or a compact set from which $P$ is sampled by computing the homology groups of a simplicial complex built with $P$ as vertices. Let $\mathsf{H}_r(\cdot)$ denote the $r$-dimensional homology group. It refers to the singular homology when the domain is a manifold or a compact set, and to the simplicial homology when it is a simplicial complex. Also, all homology groups are assumed to be defined over the finite field $\mathbb{Z}_2$.

Our main tool for topological inference rests on the relationship between the graph induced complexes and the Rips complexes that are known to capture information about the homology groups of spaces [3,21].

**Definition 2.5.** Given a point set $P \subseteq \mathbb{R}^n$ and a parameter $\alpha$, the Rips complex $\mathcal{R}^\alpha(P) = \mathcal{R}^\alpha(P, d_E)$ is a simplicial complex where a simplex $\sigma$ belongs to $\mathcal{R}^\alpha(P)$ if and only if all vertices of $\sigma$, drawn from $P$, are within $\alpha$ Euclidean distance of each other.

Notice that we define Rips complexes with Euclidean distances instead of general metrics which will be assumed throughout this paper. It is known that such Rips complexes capture the topology of a manifold $M$ if the parameters are chosen right [3,21]. We utilize this fact to infer $\mathsf{H}_1(M)$ by exploiting a sandwiching property of graph induced complexes by Rips complexes. To prove this fact, we recall the concept of contiguous maps from algebraic topology. Our main interest in this concept is the fact that two contiguous maps between two simplicial complexes induce the same homomorphism at the homology level.

**Definition 2.6.** (See [22].) Let $\mathcal{K}_1$ and $\mathcal{K}_2$ be two simplicial complexes connected by two simplicial maps $a : \mathcal{K}_1 \to \mathcal{K}_2$ and $b : \mathcal{K}_1 \to \mathcal{K}_2$. We say $a$ and $b$ are contiguous, if and only if for any simplex $\sigma \in \mathcal{K}_1$, the simplices $a(\sigma)$ and $b(\sigma)$ are faces of a common simplex in $\mathcal{K}_2$.

**Fact 2.7.** *(See [22].) If $a : \mathcal{K}_1 \to \mathcal{K}_2$ and $b : \mathcal{K}_1 \to \mathcal{K}_2$ are contiguous, then the induced homomorphisms $a_* : \mathsf{H}_r(\mathcal{K}_1) \to \mathsf{H}_r(\mathcal{K}_2)$ and $b_* : \mathsf{H}_r(\mathcal{K}_1) \to \mathsf{H}_r(\mathcal{K}_2)$ are equal.*

In our case two simplicial complexes will be $\mathcal{K}_1 = \mathcal{R}^\alpha(P)$ and $\mathcal{K}_2 = \mathcal{R}^\beta(P)$ for some $\beta > \alpha$. The map $a$ is an inclusion $\mathcal{R}^\alpha(P) \hookrightarrow \mathcal{R}^\beta(P)$. For the map $b$, we consider a simplicial map $h : \mathcal{R}^\alpha(P) \to \mathcal{G}^\alpha(P, Q, d)$ which composed with an inclusion $\mathcal{G}^\alpha(P, Q, d) \hookrightarrow \mathcal{R}^\beta(P)$ provides $b$. We elaborate on this construction.

The vertex sets of $\mathcal{R}^\alpha(P)$ and $\mathcal{G}^\alpha(P, Q, d)$ are $P$ and $Q$ respectively with a vertex map $\nu_d : P \to Q$ where $p \in P$ maps to one of its closest point(s) $\nu_d(p) \in Q$. Observe that $G^\alpha(P)$ is the 1-skeleton of $\mathcal{R}^\alpha(P)$. Therefore, the edges of a $(k+1)$-clique in $G^\alpha(P)$ constitute the 1-skeleton of a $k$-simplex in $\mathcal{R}^\alpha(P)$ and vice versa. The vertex map $\nu_d$ extends to a simplicial map $h : \mathcal{R}^\alpha(P) \to \mathcal{G}^\alpha(P, Q, d)$ where a $k$-simplex $\{p_1, p_2, \cdots, p_{k+1}\}$ in $\mathcal{R}^\alpha(P)$ is mapped to a simplex (of dimension at most $k$) with the vertex set $\{\nu_d(p_i)\}$. To see that $h$ is well defined, observe that any subset of the $(k+1)$-clique $\{p_1, p_2, \cdots, p_{k+1}\}$ is also a clique in $G^\alpha(P)$ and hence $\{\nu_d(p_i)\}$ is a simplex in $\mathcal{G}^\alpha(P, Q, d)$. The following result is used later.

**Proposition 2.8.** *Let $(P, d)$ be a metric space where $P \subset \mathbb{R}^n$ is a finite set and for every pair $p_1, p_2 \in P$, $d(p_1, p_2)$ is at least the Euclidean distance $\|p_1 - p_2\|$. Let $Q$ be a $\delta$-sample of $(P, d)$. We have the sequence*

$$\mathcal{R}^{\alpha}(P) \xrightarrow{h} \mathcal{G}^{\alpha}(P, Q, d) \xhookrightarrow{j} \mathcal{R}^{\alpha+2\delta}(P)$$

*where $j$ is an inclusion and $j \circ h$ is contiguous to the inclusion $i : \mathcal{R}^{\alpha}(P) \hookrightarrow \mathcal{R}^{\alpha+2\delta}(P)$. Hence, $j_* \circ h_* = i_*$.*

**Proof.** The map $h$ is well-defined as we detailed before. We observe that $\mathcal{G}^{\alpha}(P, Q, d) \subseteq \mathcal{R}^{\alpha+2\delta}(P)$ because, by triangle inequality, any edge $(q_1, q_2)$ of a simplex $\sigma \in \mathcal{G}^{\alpha}(P, Q, d)$ satisfies $d(q_1, q_2) \leq \alpha + 2\delta$. Since $d(q_1, q_2) \geq \|q_1 - q_2\|$ by assumption, the edge $(q_1, q_2)$ and hence the simplex $\sigma$ are in $\mathcal{R}^{\alpha+2\delta}(P)$. It follows that the inclusion map $j$ is well-defined.

To prove the contiguity, consider a simplex $\sigma$ in $\mathcal{R}^{\alpha}(P)$. We need to show that the vertices of $\sigma$ and $h(\sigma)$ span a simplex in $\mathcal{R}^{\alpha+2\delta}(P)$. Clearly, all vertices of $\sigma$ are within $\alpha$ distance of each other. By definition of $h$, all vertices of $h(\sigma)$ are within distance $\alpha + 2\delta$. Let $u$ be a vertex of $\sigma$ and $h(v)$ be a vertex of $h(\sigma)$ where $v$ is a vertex of $\sigma$. Then the Euclidean distance $\|u - h(v)\|$ is at most $\|u - v\| + \|v - h(v)\| \leq \alpha + \delta$. Therefore, all vertices of $\sigma$ and $h(\sigma)$ are within $\alpha + 2\delta$ distance. Hence, the simplex $\sigma$ and $h(\sigma)$ are faces of a common simplex in $\mathcal{R}^{\alpha+2\delta}(Q)$ proving the claim of contiguity. □

*Computing graph induced complexes* One may wonder how to efficiently construct the graph induced complexes in practice. Our experiments show that the following procedure runs quite efficiently in practice. It takes advantage of computing nearest neighbors within a range and, more importantly, computing cliques only in a sparsified graph.

Let the ball $B(q, \delta)$ in metric $d$ be called the $\delta$-cover for the point $q$. A graph induced complex $\mathcal{G}^{\alpha}(P, Q, d)$ where $Q$ is a $\delta$-sparse $\delta$-sample can be built easily by identifying $\delta$-covers with a rather standard iterative algorithm similar to the greedy (farthest point) iterative algorithm of [20]. Let $Q_i = \{q_1, \ldots, q_i\}$ be the point set sampled so far from $P$. We maintain the invariants (i) $Q_i$ is $\delta$-sparse and (ii) every point $p \in P$ that are in the union of $\delta$-covers $\bigcup_{q \in Q_i} B(q, \delta)$ have their closest point $\nu(p) = \operatorname{argmin}_{q \in Q_i} d(p, q)$ in $Q_i$ identified. To augment $Q_i$ to $Q_{i+1} = Q_i \cup \{q_{i+1}\}$, we choose a point $q_{i+1} \in P$ that is outside the $\delta$-covers $\bigcup_{q \in Q_i} B(q, \delta)$. Certainly, $q_{i+1}$ is at least $\delta$ units away from all points in $Q_i$ thus satisfying the first invariant. For the second invariant, we check every point $p$ in the $\delta$-cover of $q_{i+1}$ and update $\nu(p)$ to be $q_{i+1}$ if its distance to $q_{i+1}$ is smaller than the distance $d(p, \nu(p))$. At the end, we obtain a sample $Q \subseteq P$ whose $\delta$-covers cover the entire point set $P$ and thus is a $\delta$-sample of $(P, d)$ which is also $\delta$-sparse.

Next, we construct the simplices of $\mathcal{G}^{\alpha}(P, Q, d)$. This requires identifying cliques in $G^{\alpha}(P)$ that have vertices with different closest points in $Q$. We delete every edge $pp'$ from $G^{\alpha}(P)$ where $\nu(p) = \nu(p')$. Then, we determine every clique $\{p_1, \ldots p_k\}$ in the remaining sparsified graph and include the simplex $\{\nu(p_1), \ldots, \nu(p_k)\}$ in $\mathcal{G}^{\alpha}(P, Q, d)$. The main saving here is that many cliques of the original graph are removed before it is processed for clique computation. We use the recently proposed simplex tree which computes cliques efficiently both time and space-wise [5].

## 3. Surface point data

In this section, we infer the geometry and topology of a surface through the graph induced complex. Let $M$ be a smooth, compact, connected surface embedded in $\mathbb{R}^3$ that has no boundary. We assume that $M$ has positive reach $\rho = \rho(M)$ which is the minimum distance of $M$ to its medial axis. Let $P$ be an $\varepsilon$-sample of the metric space $(M, d_E)$ where $d_E$ is the Euclidean distance. Consider the graph induced complex $\mathcal{G}^{\alpha}(P, Q, d_E)$. In this section, the subset $Q \subset P$ is assumed to be a $\delta$-sparse $\delta$-sample of $(P, d_E)$.

Our result in this section is that under certain conditions on $\alpha$, $\varepsilon$ and $\delta$, $\mathcal{G}^{\alpha}(P, Q, d_E)$ captures the homology of $M$ and contains the restricted Delaunay triangulation $\mathrm{Del}|_M Q$ as defined below. The sparsity of $Q$ turns out to be a crucial condition in the argument.

**Definition 3.1.** Let $\mathrm{Del}\, Q$ denote the Delaunay triangulation of a point set $Q \subset \mathbb{R}^3$. The restricted Delaunay triangulation of $Q$ with respect to a manifold $M \subset \mathbb{R}^3$, denoted $\mathrm{Del}|_M Q$, is defined to be the subcomplex of $\mathrm{Del}\, Q$ formed by all Delaunay simplices whose Voronoi duals intersect $M$.

### 3.1. Topological inference from $\mathcal{G}^{\alpha}(P, Q, d_E)$

Consider the sequence $\mathcal{R}^{\alpha}(P) \xrightarrow{h} \mathcal{G}^{\alpha}(P, Q, d) \xhookrightarrow{j} \mathcal{R}^{\alpha+2\delta}(P)$ in Proposition 2.8. When $P$, an $\varepsilon$-sample of $(M, d_E)$, is sufficiently dense, it is known that $i_* : \mathsf{H}_1(\mathcal{R}^{\alpha}(P)) \to \mathsf{H}_1(\mathcal{R}^{\beta}(P))$ is an isomorphism for appropriate $\alpha$ and $\beta$. The homomorphism $h_*$ is injective if $i_*$ is an isomorphism since $i_* = j_* \circ h_*$. If we can show that $h_*$ is also surjective, then $h_*$ is an isomorphism. We now show that $h_*$ is indeed surjective for $\mathsf{H}_1$-homology and hence information about $\mathsf{H}_1(M)$ can be obtained by computing $\mathsf{H}_1(\mathcal{G}^{\alpha}(P, Q, d_E))$. First, we observe the following. Let $P_q \subseteq P$ be the points who have $q \in Q$ as the closest point. Notice that by the definition of $h$, $h(P_q) = \{q\}$. To prove that $h_*$ is surjective, it is sufficient to prove that the preimage of each 1-cycle in $\mathcal{G}^{\alpha}(P, Q, d_E)$ contains a 1-cycle of $\mathcal{R}^{\alpha}(P)$. This, in turn, is true if the 1-skeleton of $\mathcal{R}^{\alpha}(P_q)$ is connected.

**Proposition 3.2.** *If the 1-skeleton of $\mathcal{R}^{\alpha}(P_q)$ is connected for all $q \in Q$, then $h_*$ is surjective.*

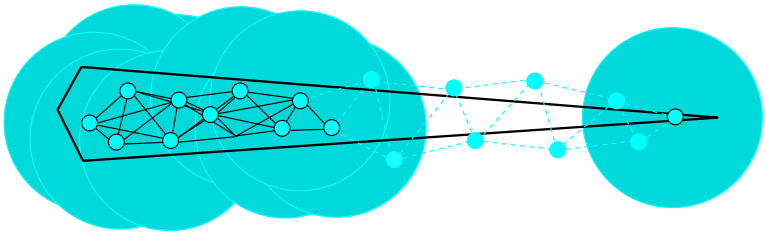


**Fig. 3.** In a long thin Voronoi cell, $\mathcal{B}_q$ may be disconnected and may not contain $M_q$.

**Proof.** We show that the chain map $h_\#$ induced by the simplicial map $h$ is surjective. It follows that the homomorphism $h_*$ induced at the homology level is also surjective. Let $c = q_0q_1 + q_1q_2 + \cdots + q_kq_0$ be any 1-cycle in $\mathcal{G}^\alpha(P, Q, d_E)$. The edges $q_{i-1}q_i$ and $q_iq_{i+1}$ have edges, say $p_{i-1}p'_i$ and $p_ip'_{i+1}$ respectively, in their preimage under $h$ in $\mathcal{R}^\alpha(P)$. Consider a path $\gamma_i$ between $p_i$ and $p'_i$ in $\mathcal{R}^\alpha(P_{q_i})$ where $h(p_i) = h(p'_i) = q_i$. Such a path exists because $\mathcal{R}^\alpha(P_q)$ is connected for all $q \in Q$. We have a 1-cycle

$$c' = p_0p'_1 + \gamma_1 + p_1p'_2 + \gamma_2 + p_2p'_3 + \cdots + \gamma_k + p_kp'_0 + \gamma_0$$

in $\mathcal{R}^\alpha(P)$ so that $h_\#(c') = c$. This shows that $h_*$ is surjective in the first homology group. □

The 1-skeleton of $\mathcal{R}^\alpha(P_q)$ is connected if the union of balls $\mathcal{B}_q = \bigcup_{P_q} B(p, \frac{\alpha}{2})$ is connected because an edge $p_1p_2$ is in $\mathcal{R}^\alpha(P_q)$ if the respective balls $B(p_1, \frac{\alpha}{2})$ and $B(p_2, \frac{\alpha}{2})$ intersect. Let $V_q$ be the Voronoi cell of $q$ in the Voronoi diagram $\mathrm{Vor}\, Q$. Let $M_q = V_q \cap M$ be the restricted Voronoi region. It turns out (we will prove it later in Proposition 3.4) that if $M_q$ is contained in $\mathcal{B}_q$ and $M_q$ is connected, then $\mathcal{B}_q$ is connected. It may seem a priori that $\mathcal{B}_q$ would contain $M_q$ if $P$ is a dense sample. Unfortunately, that is not true as Fig. 3 illustrates. To avoid such a case, we require that the Voronoi cells do not subtend very small angles between their facets which is ensured by the $\delta$-sparsity of $Q$. Proposition 3.3 below uses $\delta$-sparsity in a subtle way to prepare for the proof that $\mathcal{B}_q$ contains $M_q$. This result will also be used later to show that the graph induced complex $\mathcal{G}^\alpha(P, Q, d_E)$ in fact contains the restricted Delaunay triangulation $\mathrm{Del}|_M\, Q$.

For a simplex $\sigma \in \mathrm{Del}|_M\, Q$, we call a ball $B(c, r)$ a *surface Delaunay ball* of $\sigma$ if $c \in M$ and its boundary contains the vertices of $\sigma$.

**Proposition 3.3.** *Let $P$ be an $\varepsilon$-sample of $(M, d_E)$, and $Q$ a $\delta$-sparse $\delta$-sample of $(P, d_E)$. Let $\sigma \in \mathrm{Del}|_M\, Q$ be a restricted Delaunay triangle or edge with a vertex $q \in Q$. Let $c$ be the center of a surface Delaunay ball of $\sigma$. If $8\varepsilon \le \delta \le \frac{2}{27}\rho(M)$, then there is a point $p \in P$ so that $p \in B(c, 4\varepsilon)$ and $q$ is the closest point to $p$ among all points in $Q$.*

**Proof.** See Appendix A for the proof. □

Now, we are ready to prove that $M_q$ is contained in the union of balls $\bigcup_{P_q} B(p, \frac{\alpha}{2})$.

**Proposition 3.4.** *If $\alpha \ge 12\varepsilon$ and $8\varepsilon \le \delta \le \frac{2}{27}\rho(M)$, then $M_q \subset \bigcup_{\{p \in P_q\}} B(p, \frac{\alpha}{2})$ which implies that $\mathcal{R}^\alpha(P_q)$ is path connected if $M_q$ is path connected.*

**Proof.** Since $P$ is an $\varepsilon$-sample of $M$, $\forall x \in M$, there exists a point $p \in B(x, \varepsilon)$ where $p \in P$. Let $P'_q = P \cap (\bigcup_{x \in M_q} B(x, \varepsilon))$. Then, we have $M_q \subset \bigcup_{p \in P'_q} B(p, \varepsilon)$ for if $x \in M_q$, there exists $p \in P'_q$ with $p \in B(x, \varepsilon)$ requiring $x \in B(p, \varepsilon)$. On the other hand, recall that $P_q = M_q \cap P$. Hence if $p \in P'_q \setminus P_q$, then $B(p, \varepsilon)$ contains some boundary point $x \in \partial M_q$. The point $x$ belongs to a Voronoi facet in the Voronoi diagram of $Q$ and hence $B(x, \|q - x\|)$ is a surface Delaunay ball. By Proposition 3.3, we can find a point $u \in P_q$ such that $\|u - x\| \le 4\varepsilon$. Thus, $B(p, \varepsilon) \subset B(u, (4+2)\varepsilon)$. Taking $\alpha \ge 12\varepsilon$, we get that $M_q \subset \bigcup_{p \in P'_q} B(p, \varepsilon) \subset \bigcup_{p \in P_q} B(p, \frac{\alpha}{2})$.

Since every ball in $\{B(p, \varepsilon) | p \in P_q\}$ intersects $M_q$, we have that $\mathcal{B}_q = \bigcup_{\{p \in P_q\}} B(p, \frac{\alpha}{2})$ is path connected if $M_q$ is path connected. On the other hand $\mathcal{R}^\alpha(P_q)$ is path connected if $\mathcal{B}_q$ is path connected proving the claim. □

We can now present the main result of this subsection.

**Theorem 3.5.** *Let $P$ be an $\varepsilon$-sample of a smooth compact surface $M$ embedded in $\mathbb{R}^3$, and $Q \subseteq P$ a $\delta$-sparse $\delta$-sample of $(P, d_E)$. For $12\varepsilon \le \alpha \le \frac{2}{27}\rho$ and $8\varepsilon \le \delta \le \frac{2}{27}\rho$, the map $h_* : \mathrm{H}_1(\mathcal{R}^\alpha(P)) \to \mathrm{H}_1(\mathcal{G}^\alpha(P, Q, d_E))$ is an isomorphism where $h : \mathcal{R}^\alpha(P) \to \mathcal{G}^\alpha(P, Q, d_E)$ is the simplicial map induced by the nearest point map $\nu_{d_E} : P \to Q$.*

**Proof.** Since $\delta \le 0.18\rho$, we can assume each restricted Voronoi cell $M_q$ to be path connected [13]. This together with the lower bound on $\alpha$ imply that $\mathcal{R}^\alpha(P_q)$ is connected for each $q \in Q$ thanks to Proposition 3.4. Consequently, Proposition 3.2 establishes that $h_*$ is surjective.

From Proposition 4.1 of [16] and its proof, we obtain the following: for any $4\varepsilon \le r \le 2r \le r' \le \sqrt{\frac{3}{5}}\rho$,

$$\mathsf{H}_1(\mathcal{R}^r(P)) \cong \mathsf{H}_1(\mathcal{R}^{r'}(P)) \cong \mathsf{H}_1(M) \tag{1}$$

where the first isomorphism is induced by the canonical inclusion $i : \mathcal{R}^r(P) \hookrightarrow \mathcal{R}^{r'}(P)$. Our assumption on the ranges of $\alpha$ and $\delta$ implies the required conditions that $4\varepsilon \le \alpha \le \frac{1}{3}\sqrt{\frac{3}{5}}\rho$ and $4\varepsilon \le \delta \le \frac{1}{3}\sqrt{\frac{3}{5}}\rho$. We claim that $i_* : \mathsf{H}_1(\mathcal{R}^\alpha(P)) \to \mathsf{H}_1(\mathcal{R}^{\alpha+2\delta}(P))$ induced by the inclusion $i : \mathcal{R}^\alpha(P) \to \mathcal{R}^{\alpha+2\delta}(P)$ is an isomorphism.

First, note that this claim follows easily from (1) if $\alpha \le \delta$ by setting $r = \alpha$ and $r' = \alpha + 2\delta$. Now assume that $\delta \le \alpha$. Consider the following sequence:

$$\mathcal{R}^\delta(P) \overset{i_1}{\hookrightarrow} \mathcal{R}^\alpha(P) \overset{i}{\hookrightarrow} \mathcal{R}^{\alpha+2\delta}(P) \overset{i_2}{\hookrightarrow} \mathcal{R}^{3\alpha}(P).$$

By Eq. (1), we have that the composition of inclusions $i \circ i_1 : \mathcal{R}^\delta(P) \to \mathcal{R}^{\alpha+2\delta}(P)$ induces an isomorphism at the homology level. Hence $i_*$ is necessarily surjective. On the other hand, the composition of inclusions $i_2 \circ i : \mathcal{R}^\alpha(P) \to \mathcal{R}^{3\alpha}(P)$ induces an isomorphism at the homology level. Hence $i_*$ is necessarily injective. Putting these two together, we have that $i_*$ is indeed an isomorphism. Therefore $h_*$ is injective by Proposition 2.8. It then follows that $h_*$ is an isomorphism as claimed. □

Notice that the lower bound on $\delta$ in Theorem 3.5 is not restricted by $\alpha$. This means that one can have a dense input graph for a large $\alpha$ whose connectivity does not restrict the size of the subsample.

In the next subsection, we show two examples of surface data where the graph induced complex has the correct $\mathsf{H}_1$-homology with a considerably fewer simplices than the $\nu$-witness complex, a modified witness complex suggested in [4] for capturing the topology correctly.

### 3.2. Reconstruction of $M$ using $\mathcal{G}^\alpha(P, Q, d_E)$

In this subsection, we observe that the graph induced complexes can also be used for surface reconstruction. It is known that if $P$ is dense and $T$ is a simplicial complex with vertex set $P$ which satisfies the following conditions, a simplicial manifold can be extracted from $T$ that is homeomorphic to $M$ [2,13]. The conditions are: (i) $T$ is embedded in $\mathbb{R}^3$, (ii) all triangles in $T$ have small circumradii compared to reach and (iii) $T$ contains the restricted Delaunay triangulation. We show that $\mathcal{G}^\alpha(P, Q, d_E)$ contains the restricted Delaunay triangulation. We then prune $\mathcal{G}^\alpha(P, Q, d_E)$ so that conditions (i) and (ii) are satisfied, but none of the restricted Delaunay triangles are deleted in the process which then ensures condition (iii).

**Theorem 3.6.** *For $8\varepsilon \le \delta \le \frac{2}{27}\rho$ and $\alpha \ge 8\varepsilon$, we have that $\mathrm{Del}|_M Q \subseteq \mathcal{G}^\alpha(P, Q, d_E)$ where $P$ is an $\varepsilon$-sample of $(M, d_E)$ and $Q \subseteq P$ is a $\delta$-sparse $\delta$-sample of $(P, d_E)$.*

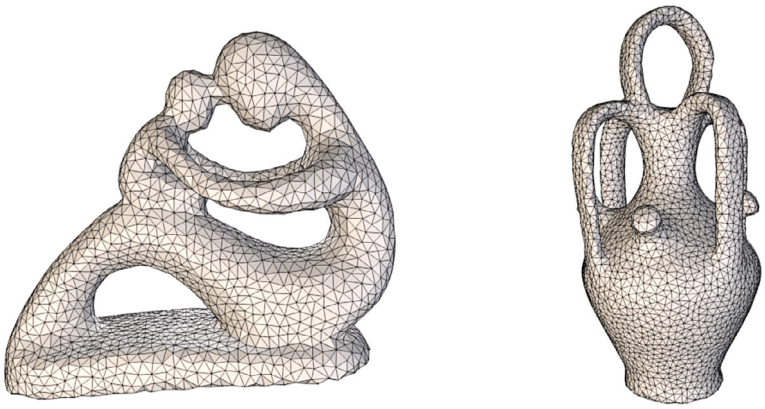
**Proof.** We will show that if $8\varepsilon \le \delta \le \frac{2}{27}\rho$ and $\alpha \ge 8\varepsilon$, then any triangle $\sigma \in \mathrm{Del}|_M Q$ is in $\mathcal{G}^\alpha(P, Q, d_E)$. The theorem follows from this.

Let $\sigma = \{q_1, q_2, q_3\}$, and $c$ the center of a surface Delaunay ball of $\sigma$. By Proposition 3.3, there exists a point $p_i \in P$ in $B(c, 4\varepsilon)$ so that $q_i$ is the closest point in $Q$ to $p_i$ for $i = 1, 2, 3$. It turns out that the interior of bounded cones used in the proof of Proposition 3.3 for $q_1$, $q_2$ and $q_3$ are disjoint. Hence each point $p_i$ found in $B(c, 4\varepsilon)$ corresponding to $q_i$ is distinct from the other two. Therefore, if $\alpha \ge 8\varepsilon$, the vertices $p_1, p_2$ and $p_3$ form a clique in $G^\alpha(P)$ and hence the triangle $\sigma$ is in $\mathcal{G}^\alpha(P, Q, d_E)$. □

The complex $\mathcal{G}^\alpha(P, Q, d_E)$ may have intersecting triangles. We prune $\mathcal{G}^\alpha(P, Q, d_E)$ to eliminate all such pairwise intersections while leaving the restricted Delaunay triangles in the complex. This ensures that the resulting complex embeds in $\mathbb{R}^3$ and still contains the restricted Delaunay triangulation. Our simple observation is that if two intersecting triangles $t_1$ and $t_2$ do not intersect in a common face, one can decide *locally* which of the two can possibly be in a Delaunay triangulation.

**Observation 3.7.** *If $V$ is the vertex set of two intersecting triangles $t_1$ and $t_2$ whose intersection is not a common face of both, then at least one of $t_1$ and $t_2$ is not in $\mathrm{Del}\, V$. The triangle which is not in $\mathrm{Del}\, V$ cannot be in $\mathrm{Del}\, P$ where $V \subseteq P$.*

One can check locally the Delaunay condition for $t_1$ and $t_2$ and decide to throw away at least one triangle which is not in $\mathrm{Del}\, V$. This takes only constant time since $V$ contains at most 6 vertices. Notice that no restricted Delaunay triangle can be thrown away by this process. After repeatedly pruning away one of the pairwise intersecting triangles, we arrive at a complex that embeds in $\mathbb{R}^3$ and contains the restricted Delaunay triangulation $\mathrm{Del}|_M Q$. Next, we prune all triangles that have circumradius more than $2\delta$. Again, since the surface Delaunay ball of each restricted Delaunay triangle has circumradius at most $\delta + \varepsilon \le 2\delta$, one is ensured that no restricted Delaunay triangle is eliminated. Assuming $\delta$ to be sufficiently small,



(a) Fertility model (b) Botijo model

**Fig. 4.** Reconstructed surfaces for Fertility and Botijo models.

a sharp edge pruning and a walk on the outside of the resulting complex as described in [2,13] provides the reconstructed surface. The output surface has one nice property that the triangles have bounded aspect ratios since their circumradii are at most $2\delta$ and their edge lengths are at least $\delta$ ($Q$ is $\delta$-sparse).

**Theorem 3.8.** *Let $M \subset \mathbb{R}^3$ be a smooth, compact, and connected surface. If $8\varepsilon \le \delta \le \frac{2}{27}\rho$, $\alpha \ge 8\varepsilon$, $P$ is an $\varepsilon$-sample of $(M, d_E)$, and $Q \subseteq P$ is a $\delta$-sparse $\delta$-sample of $(P, d_E)$, then a triangulation $T \subseteq \mathcal{G}^\alpha(P, Q, d_E)$ of $M$ can be computed where each triangle in $T$ has a bounded aspect ratio.*

We observe in experiments that surfaces can be reconstructed from a very sparse subsample with this strategy. Fig. 4 presents two examples for surface reconstruction. The original sample $P$ has 1,575,055 points for the Fertility model and 1,049,892 points for Botijo model. The input graphs for the graph induced complex are constructed by connecting two points within distance of $\alpha = 0.45$ for Fertility and $\alpha = 1.0$ for Botijo. The 2-skeleton of the Rips complex built on the input graph has 45,788,607 simplices for Fertility and 91,264,091 simplices for Botijo. The subsample $Q$ consists of 3007 points for Fertility with $\delta = 3.68$, and 4659 points for Botijo with $\delta = 4.0$. The graph induced complex $\mathcal{G}^\alpha(P, Q, d_E)$ built on the subsample has: 3007 vertices, 9178 edges, 6304 triangles, 139 tetrahedra and no other higher dimensional simplices for Fertility; 4659 vertices, 14,709 edges, 10,755 triangles, 718 tetrahedra, 5 4-dimensional simplices, and no other higher dimensional simplices for Botijo. The reconstructed surfaces from $\mathcal{G}^\alpha(P, Q, d_E)$ are shown in Fig. 4. For Fertility, it has 3007 vertices, 9039 edges and 6026 triangles; for Botijo, it has 4659 vertices, 14,001 edges and 9334 triangles. Evidently, the graph induced complex has only a few more simplices compared to the reconstructed surface.

For a comparison, we also constructed the $\nu$-witness complex suggested in [4] which also contains the restricted Delaunay triangulation $\mathrm{Del}|_M Q$ with $\nu = (1, 6, 6, 4)$. The $\nu$-witness complex for Fertility has 3007 vertices, 35,687 edges, 119,237 triangles and 19,874 tetrahedra; the $\nu$-witness complex for Botijo has 4659 vertices, 54,648 edges, 180,936 triangles and 29,654 tetrahedra. The graph induced complex has much smaller size, but still captures $\beta_1$ ($\beta_1 = 8$ for the Fertility, and $\beta_1 = 10$ for the Botijo).

## 4. Point data for more general domains

In this section, we consider domains beyond surfaces in $\mathbb{R}^3$.

### 4.1. Manifolds

Let $M$ be a $k$-manifold embedded in $\mathbb{R}^n$ and let $P$ be a discrete sample of $(M, d_E)$. We observe that the overall setup in Section 3.1 for inferring $\mathsf{H}_1$-homology from the graph induced complex generalizes easily to higher dimensions. The inclusion map $\mathcal{R}^\alpha(P) \hookrightarrow \mathcal{R}^{\alpha+2\delta}(P)$ still induces an isomorphism at the homology level if $\alpha$ and $\delta$ are chosen appropriately. In that case, the map $h_* : \mathsf{H}_1(\mathcal{R}^\alpha(P)) \to \mathsf{H}_1(\mathcal{G}^\alpha(P, Q, d))$ remains injective by the same argument as before. The main trouble arises when we try to prove that it is also surjective. Observe that, to prove the surjectivity of $h_*$, we used the fact that the restricted Voronoi cell $M_q = V_q \cap M$ in $\mathrm{Vor}\, Q$ is connected (Proposition 3.4). Unfortunately, this is not necessarily true in high dimensions given the counterexamples in [4,10]. To overcome this impediment we change the distance function replacing the Euclidean distance with the graph distance while building $\mathcal{G}^\alpha(P, Q, d)$. Specifically, we still consider $G^\alpha(P)$ to be the graph connecting points in $P$ with Euclidean distance $\alpha$ or less, but take $Q$ to be a $\delta$-sparse $\delta$-sample of $(P, d_G)$ where the graph distance $d_G = d_{G^\alpha(P)}$ is defined with the Euclidean lengths as the edge weights. Then, we consider $\mathcal{G}^\alpha(P, Q, d_G)$.

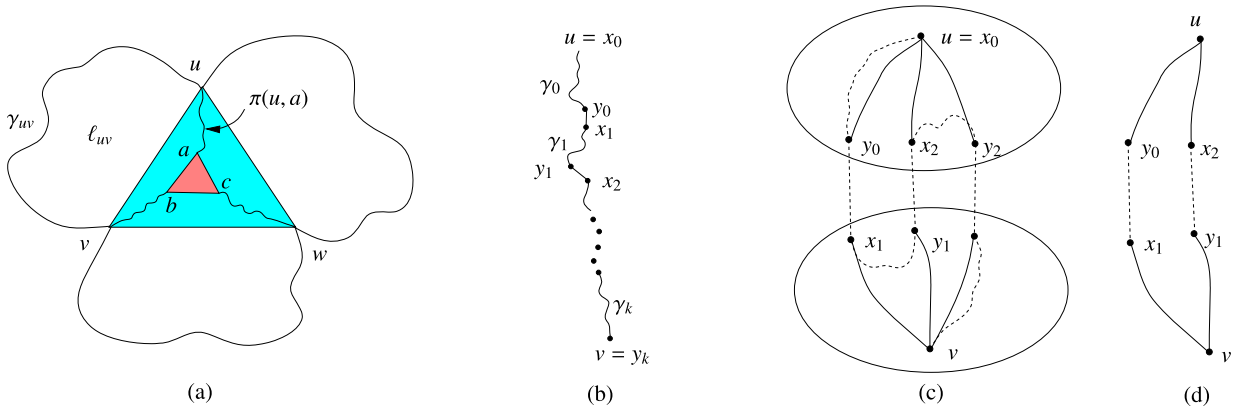



**Fig. 5.** (a) $\gamma_{uv}$ makes a cycle with $\pi(u,a)$, $\pi(v,b)$ and $ab$, (b) $\gamma_{uv}$ as a sum of unicolored chains and bicolored edges, (c) converting $\gamma_{uv}$ (shown dotted) to $\hat{\gamma}_{uv}$, (d) a diamond of $\hat{\gamma}_{uv}$.

As before, let $P_q \subseteq P$ be the set of points nearest to a point $q \in Q$ with respect to $d_G$. Observe that any point in $P$ that is on the shortest path between $q$ and a point $p \in P_q$ also belongs to $P_q$. This immediately allows us to claim that the 1-skeleton of $\mathcal{R}^\alpha(P_q)$ is connected, which was needed to claim that $h_*$ is surjective.

**Proposition 4.1.** *$\mathcal{R}^\alpha(P_q)$ is connected, and thus $h_*$ is surjective.*

**Theorem 4.2.** *Let $P$ be an $\varepsilon$-sample of an embedded smooth and compact manifold $M$ with reach $\rho$, and $Q$ a $\delta$-sample of $(P, d_G)$. For $4\varepsilon \le \alpha, \delta \le \frac{1}{3}\sqrt{\frac{3}{5}}\rho$, the map $h_* : \mathsf{H}_1(\mathcal{R}^\alpha(P)) \to \mathsf{H}_1(\mathcal{G}^\alpha(P, Q, d_G))$ is an isomorphism where $h : \mathcal{R}^\alpha(P) \to \mathcal{G}^\alpha(P, Q, d_G)$ is the simplicial map induced by the nearest point map $\nu_{d_G} : P \to Q$.*

### *4.2. A leaner subsampling for* $\mathsf{H}_1$

In this subsection we show that the subsample $Q$ can be made leaner. The main insight is that we can define a feature size larger than the reach which permits us to subsample more sparsely with respect to this larger feature size. Gao et al. [23] considered a similar feature size for the same reason of requiring sparser sampling for a two dimensional shape. Here we show that such a sparser sample is also adequate for determining $\mathsf{H}_1$ of manifolds in high dimensions. Our experimental results in Fig. 2 suggest that one can obtain information about $\mathsf{H}_1$ from a very sparse sample in practice.

Let $\mathcal{K}$ be a simplicial complex with non-negative weights on its edges. We define *homological loop feature size* as

$$\mathrm{hlfs}(\mathcal{K}) = \begin{cases} \frac{1}{2}\inf\{|c| : c \text{ is non-null-homologous 1-cycle in } \mathcal{K}\} \\ \infty \text{ if no such } c \text{ exists.} \end{cases}$$

This feature size is very similar to the *systolic feature size* $\mathrm{sfs}(X, d)$ of a compact metric space $(X, d)$ [23] which is the length of the shortest non-contractible loop in $X$. Our definition of hlfs when applied to a metric space $(X, d)$ becomes larger than or equal to $\mathrm{sfs}(X, d)$. Notice that every loop of $\mathcal{K}$ with length less than $2\mathrm{hlfs}(\mathcal{K})$ is null-homologous in $\mathcal{K}$.

Let $Q \subseteq P$ be a $\delta$-sample of $(P, d_G)$ as before but with $\delta \le \frac{1}{2}\mathrm{hlfs}(\mathcal{R}^\alpha(P)) - \frac{1}{2}\alpha$. Let $h : \mathcal{R}^\alpha(P) \to \mathcal{G}^\alpha(P, Q, d_G)$ be the simplicial map as defined earlier. We aim to show that the induced homomorphism $h_*$ on the first homology is injective. Since we use graph distances, Proposition 4.1 remains valid and hence $h_*$ remains to be surjective. However, we cannot claim $i_* : \mathsf{H}_1(\mathcal{R}^\alpha(P)) \to \mathsf{H}_1(\mathcal{R}^{\alpha+2\delta}(P))$ is an isomorphism because $\delta$ could be larger than required. Thus, we cannot use $i_*$ to infer that $h_*$ is injective as before. Nevertheless, we can prove the following result using a different approach.

**Theorem 4.3.** *If $Q$ is a $\delta$-sample of $(P, d_G)$ for $\delta < \frac{1}{2}\mathrm{hlfs}(\mathcal{R}^\alpha(P)) - \frac{1}{2}\alpha$, then $h_* : \mathsf{H}_1(\mathcal{R}^\alpha(P)) \to \mathsf{H}_1(\mathcal{G}^\alpha(P, Q, d_G))$ is an isomorphism.*

**Proof.** We only need to show that $h_*$ is injective, as its surjectivity follows from Proposition 4.1. To show the injectivity, it suffices to show that $h_*$ has a trivial kernel. Let $\sigma$ be any triangle in $\mathcal{G}^\alpha(P, Q, d_G)$. If under the chain map $h_\#$, every cycle in the preimage of the boundary cycle $\partial\sigma$ is null homologous, then every null homologous cycle in $\mathcal{G}^\alpha(P, Q, d_G)$ has only null homologous cycles in its preimage. This is true due to the fact that a bounded cycle is a sum of boundaries of triangles and the chain map $h_\#$ is surjective (see the proof of Proposition 3.2). Below, we show that under the chain map $h_\#$, every cycle in the preimage of the boundary cycle of any triangle indeed is null homologous. It would then follow that the kernel of $h_*$ is trivial.

Let $\gamma$ be any cycle in the preimage of $\partial\sigma$. We have $\gamma \in \sum_{uv \in \partial\sigma} h_\#^{-1}(uv)$ where $uv$ be any edge of $\sigma$. Let $\gamma_{uv}$ be any maximal subpath of $\gamma$ so that $h_\#(\gamma_{uv}) = uv$ (Fig. 5(a)). For each such $\gamma_{uv}$, we construct a cycle $\ell_{uv}$ so that $\ell_{uv}$ is null

homologous and $\gamma$ is homologous to $\sum \ell_{uv}$. Therefore, showing $\gamma$ is null homologous reduces to showing that every $\ell_{uv}$ is null-homologous.

We construct $\ell_{uv}$ as follows. By the construction of $\mathcal{G}^\alpha(P, Q, d_G)$, there is a triangle $abc \in \mathcal{R}^\alpha(P)$ such that $h(abc) = \sigma$ with $h(a) = u$, $h(b) = v$. Consider the shortest paths $\pi(u, a)$ and $\pi(v, b)$ in $G^\alpha(P)$ from $u$ to $a$ and from $v$ to $b$ respectively. Observe that all vertices in $\pi(u, a)$ and $\pi(v, b)$ are mapped to $u$ and $v$ respectively by $h$ since we are using the graph-induced distance $d_G$ to construct $\mathcal{G}^\alpha(P, Q, d_G)$. Take $\ell_{uv}$ to be the chain $\pi(u, a) + ab + \pi(v, b) + \gamma_{uv}$; refer to Fig. 5(a). With this choice, we have $\gamma = \sum \ell_{uv} + \partial(abc)$ and hence $\gamma$ is homologous to $\sum \ell_{uv}$ as promised. To prove $\ell_{uv}$ null-homologous, we construct a homologous path $\hat{\gamma}_{uv}$ to $\gamma_{uv}$ which gives a homologous cycle $\hat{\ell}_{uv}$ to $\ell_{uv}$. We then prove that $\hat{\ell}_{uv}$ is null-homologous.

Call an edge $e = (x, y)$ in $\mathcal{R}^\alpha(P)$ *bicolored* if its two end-points are mapped to two distinct vertices by $h$; otherwise, $e$ is *unicolored*. A 1-chain from $\mathcal{R}^\alpha(P)$ is *unicolored* if it has only unicolored edges. For simplicity, we assume that vertices from $\gamma_{uv}$ are all contained in $h^{-1}(u) \cup h^{-1}(v)$ because $\gamma_{uv}$ is always homologous to a path containing vertices mapped only to $u$ or $v$. In this case, $\gamma_{uv}$ can be decomposed into a set of bicolored edges $\{y_0x_1, y_1x_2, \ldots, y_{k-1}x_k\}$ together with a set of unicolored chains $\{\gamma_0, \gamma_1, \ldots, \gamma_k\}$ such that $\partial\gamma_i = x_i + y_i$. In particular, $\gamma_{uv}$ can be written as

$$\gamma_{uv} = \gamma_0 + y_0x_1 + \gamma_1 + y_1x_2 + \cdots + y_{k-1}x_k + \gamma_k. \tag{2}$$

See Fig. 5(b) for an illustration.

**Claim 4.4.** *Let $\gamma_i$ be a unicolored chain with two boundary points $x_i, y_i$ so that for any simplex $\tau \in \gamma_i$, $h(\tau) = u$. Then, $\gamma_i$ is homologous to the chain $\hat{\gamma}_i = \pi(x_i, u) + \pi(u, y_i)$.*

Given the chain $\gamma_{uv}$ with subchains $\gamma_i$ as in Eq. (2), we convert it to a homologous chain

$$\hat{\gamma}_{uv} = \hat{\gamma}_0 + y_0x_1 + \hat{\gamma}_1 + \cdots + y_{k-1}x_k + \hat{\gamma}_k.$$

Replace $\gamma_{uv}$ with $\hat{\gamma}_{uv}$ in $\ell_{uv}$ to obtain a homologous cycle $\hat{\ell}_{uv}$; refer to Fig. 5. Observe that $\hat{\ell}_{uv}$ is the sum of cycles (diamonds) that have two unicolored chains and two bicolored edges as shown in Fig. 5(d). Such a cycle $c$ has length at most $4\delta + 2\alpha$. This is because each unicolored chain in $c$ has at most two shortest paths of the form $\pi(u, x_i)$ and $\pi(u, y_i)$ (or $\pi(v, x_i)$ and $\pi(v, y_i)$) that have lengths $2\delta$ or less ($Q$ is a $\delta$-sample of $(P, d_G)$), and the two bicolored edges have lengths at most $2\alpha$ in $G^\alpha(P)$. The cycle $c$ is null homologous because its length is

$$|c| \le 4\delta + 2\alpha < 2\mathrm{hlfs}(\mathcal{R}^\alpha(P)), \quad \text{given that } \delta < \frac{1}{2}\mathrm{hlfs}(\mathcal{R}^\alpha(P)) - \frac{1}{2}\alpha.$$

It follows that $\hat{\ell}_{uv}$ is null homologous.

We only need to show Claim 4.4 to finish the proof. Let $x_i = p_0, p_1, \cdots, p_m = y_i$ be the sequence of vertices on the unicolored chain (path) $\gamma_i$. Consider the shortest paths $\pi(p_i, u)$ for each $p_i$ on this path. The length of the cycle $z_i = \pi(u, p_i) + p_ip_{i+1} + \pi(u, p_{i+1})$ is at most $2\delta + \alpha$ for each $i \in [0, m-1]$. Therefore, it is null homologous by our assumption. We have $\gamma_i + \hat{\gamma}_i = \sum_{i-0}^{m-1} z_i = 0$. Therefore, $\gamma_i$ and $\hat{\gamma}_i$ are homologous. □

Notice that, we can use Theorem 4.3 to compute $\mathsf{H}_1(M)$ for a manifold $M$ from a much leaner subsample than predicted by Theorem 3.5. It is known that, for an $\varepsilon$-sample $P$ of $M$, the shortest non-trivial 1-cycle in $\mathcal{R}^\alpha(P)$ has a length which is at least a constant times the length of the shortest non-trivial 1-cycle of $\mathsf{H}_1(M)$; see Theorem 3.11 in [15] for details. This means that a $\delta$-sample of $P$ where $\delta$ is at most some constant fraction of the shortest non-trivial 1-cycle of $M$ is sufficient to build the graph induced complex from $\mathcal{R}^\alpha(P)$ to infer $\mathsf{H}_1(M)$.

### 4.3. Point data for compact sets

So far we have focused on $\mathsf{H}_1$-homology. In this section we extend the domain to compact subspaces of Euclidean spaces and consider homology groups of all dimensions. This generality comes at the expense of additional computations. Unlike previous approaches that allow us to infer the $\mathsf{H}_1$-homology of the sampled manifold by computing directly the same for the graph induced complexes, now we need to compute the persistent homology [17] induced by simplicial maps. The well-known algorithms for computing persistent homology [17] work for maps induced by inclusions. In a contemporary paper [14], we present an algorithm that can compute the persistent homology induced by simplicial maps.

Let $X \subset \mathbb{R}^n$ be a compact set and $X^\lambda$ be its offset with $\lambda > 0$. Since it is difficult to compute $\mathsf{H}_k(X)$ from a sample [9], we aim for computing the homology groups $\mathsf{H}_k(X^\lambda)$ for the offset $X^\lambda$. Let $\mathrm{wfs}(X)$ denote the *weak feature size* which is defined as the smallest positive critical value of the distance function to $X$ [8]. We prove that the persistent homology of the graph induced complex defined with either Euclidean or graph distance $d$ provides the correct homology of $X^\lambda$ where $0 < \lambda < \mathrm{wfs}(X)$. Specifically, the image of $h_* : \mathsf{H}_k(\mathcal{G}^\alpha(P, Q, d)) \to \mathsf{H}_k(\mathcal{G}^{\alpha'}(P, Q', d))$ induced by a simplicial map $h : \mathcal{G}^\alpha(P, Q, d) \to \mathcal{G}^{\alpha'}(P, Q', d)$ becomes isomorphic to $\mathsf{H}_k(X^\lambda)$ for appropriate $\alpha$ and $\alpha'$. It is worth noting that the lower bound on $\alpha$ for which we prove this result depends only on the density $\varepsilon$ of the input $P$ and not on the density of the

subsampled set $Q$. This is in contrast with a similar result for witness complexes presented in [9]. We recall the following result from [9].

**Proposition 4.5.** *If the sequence of homomorphisms $A \to B \to C \to D \to E \to F$ between finite dimensional vector spaces satisfies that* $\operatorname{rank}(A \to F) = \operatorname{rank}(C \to D)$, *then* $\operatorname{rank}(B \to E) = \operatorname{rank}(C \to D)$.

Let $Q$ and $Q'$ be subsamples of $P$ where $Q$ is a $\delta$-sparse $\delta$-sample and $Q'$ is a $\delta'$-sparse $\delta'$-sample for $\delta' > \delta$. Consider the interleaving sequence between the graph induced and Rips complexes,

$$\begin{tikzcd}
\mathcal{R}^{\alpha}(P) \arrow[r, hook, "i_1"] \arrow[dr, "h_1"'] & \mathcal{R}^{\alpha+2\delta}(P) \arrow[r, hook, "i_2"] & \mathcal{R}^{4(\alpha+2\delta)}(P) \arrow[r, hook, "i_3"] \arrow[d, "h_2"] & \mathcal{R}^{4(\alpha+2\delta)+2\delta'}(P) \\
& \mathcal{G}^{\alpha}(P, Q, d) \arrow[u, hook, "j_1"] \arrow[r, dotted, "h"] & \mathcal{G}^{4(\alpha+2\delta)}(P, Q', d) \arrow[ur, hook, "j_2"'] &
\end{tikzcd} \tag{3}$$

where $i_1, i_2, i_3, j_1$ and $j_2$ are inclusions and $h = h_2 \circ i_2 \circ j_1$. By Proposition 2.8, $h_1$ and $h_2$ are simplicial maps. Therefore, $h$ is also a simplicial map as composition of simplicial maps. In particular, $h$ is the simplicial map induced by the vertex map that maps each point $q \in Q$ to its closest point $q' \in Q'$ in $Q'$. We prove that $\operatorname{im} h_* \cong \mathsf{H}_k(X^{\lambda})$ where $h_* : \mathsf{H}_k(\mathcal{G}^{\alpha}(P, Q, d)) \to \mathsf{H}_k(\mathcal{G}^{4(\alpha+2\delta)}(P, Q', d))$ and $\varepsilon$, $\alpha$ and $\delta$ fall in appropriate ranges.

**Theorem 4.6.** *Let $X \subset \mathbb{R}^n$ be a compact space. Let $0 < \varepsilon < \frac{1}{9}\operatorname{wfs}(X)$ and $P$ be an $\varepsilon$-sample of $(X, d_E)$. Let $Q$ be a $\delta$-sparse $\delta$-sample of $(P, d)$ and $Q'$ be a $\delta'$-sparse $\delta'$-sample of $(P, d)$, where $d$ is either Euclidean or graph distance and $\delta' > \delta$.*

*If $2\varepsilon \le \alpha \le \frac{1}{4}(\operatorname{wfs}(X) - \varepsilon)$ and $(\alpha + 2\delta) + \frac{1}{2}\delta' \le \frac{1}{4}(\operatorname{wfs}(X) - \varepsilon)$, then $\operatorname{im} h_* \cong \mathsf{H}_k(X^{\lambda})$ $(0 < \lambda < \operatorname{wfs}(X))$ where $h_* : \mathsf{H}_k(\mathcal{G}^{\alpha}(P, Q, d)) \to \mathsf{H}_k(\mathcal{G}^{4(\alpha+2\delta)}(P, Q', d))$ is induced by $h$ in diagram (3).*

**Proof.** The diagram (3) is not commutative in general. However, it is commutative at the homology level. Proposition 2.8 makes the two triangles at the left and right commutative. The middle square commutes by definition of $h$. Now consider the sequence,

$$\mathsf{H}_k(\mathcal{R}^{\alpha}(P)) \xrightarrow{h_{1*}} \mathsf{H}_k(\mathcal{G}^{\alpha}(P, Q, d)) \xrightarrow{j_{1*}} \mathsf{H}_k(\mathcal{R}^{\alpha+2\delta}(P)) \tag{4}$$

$$\xrightarrow{i_{2*}} \mathsf{H}_k(\mathcal{R}^{4(\alpha+2\delta)}(P)) \xrightarrow{h_{2*}} \mathsf{H}_k(\mathcal{G}^{4(\alpha+2\delta)}(P, Q', d)) \xrightarrow{j_{2*}} \mathsf{H}_k(\mathcal{R}^{4(\alpha+2\delta)+2\delta'}(P))$$

Consider the sequence of inclusions at the upper level of the diagram (3). Since $\alpha \ge 2\varepsilon$ and $(\alpha + 2\delta) + \frac{1}{2}\delta' \le \frac{1}{4}(\operatorname{wfs}(X) - \varepsilon)$, we have that

$$\operatorname{im}\,(i_3 \circ i_2 \circ i_1)_* \cong \mathsf{H}_k(X^{\lambda}) \text{ and } \operatorname{im}\,(i_2)_* \cong \mathsf{H}_k(X^{\lambda})$$

by Theorem 3.6 of [9]. Considering the diagram (3) and the sequence in (4) we have

$$\mathsf{H}_k(X^{\lambda}) \cong \operatorname{im}\,(i_{3*} \circ i_{2*} \circ i_{1*}) \cong \operatorname{im}((j_{2*} \circ h_{2*}) \circ i_{2*} \circ (j_{1*} \circ h_{1*})) \cong \operatorname{im} i_{2*} \tag{5}$$

Letting $h = h_2 \circ i_2 \circ j_1$, the rightmost isomorphism in (5) allows us to claim that $\operatorname{im} h_* \cong \mathsf{H}_k(X^{\lambda})$ by applying Proposition 4.5 to the sequence (4). □

## 5. Conclusions

In this work, we investigated the graph induced complex that can be built upon a given point cloud data and a suitable graph connecting them. This complex, to some extent, has the advantages of both Rips and witness complexes. We have identified several of its topological properties that can evidently be useful in extracting information from point data even in high dimensions.

In Section 4.3, we showed how to infer the homology groups of a compact set using the persistent homology of a pair of graph induced complexes constructed with two values of $\delta$. One can consider a filtration of $\mathcal{G}^{\alpha}(P, Q, d)$ with $Q$ sparsified for increasing values of $\alpha$ and $\delta$. Then, one can obtain a persistence diagram [11] out of this "full filtration" using our recently proposed algorithm for computing the topological persistence for filtrations connected with simplicial maps [14]. The algorithm will collapse vertices progressing through the filtration and hence will keep the size of the complex in question contained. Relating this persistence diagram to that of a filtration obtained by a Rips filtration is an interesting question. We have addressed this question in a subsequent work [14].

Finding other applications where the graph induced complex becomes useful also remains open for further investigations.

**Acknowledgements**

We acknowledge the support of the NSF grants CCF 1318595, CCF 1319406, CCF 1116258 and thank all the referees for their valuable comments.

## Appendix A. Proof of Proposition 3.3

First, we present an elementary geometric result that we need to use in the proof. Let $\mathsf{C}(o, \vec{v}, \alpha) \in \mathbb{R}^3$ denote the cone with apex $o$, axis in the direction of $\vec{v}$ and aperture $2\alpha$.

**Claim A.1.** *Given a ball $B$ with radius $r$ and center $c$, let $q$ be an arbitrary point on the boundary of $B$. Consider the two nested cones $\mathsf{C}_1 = \mathsf{C}(c, \overrightarrow{cq}, \alpha)$ and $\mathsf{C}_2 = \mathsf{C}(c, \overrightarrow{cq}, 2\alpha)$ with the same axis. Let $p$ be any point from the intersection of the ball and the inner cone; that is, $p \in B \cap \mathsf{C}_1$. Let $x$ be an arbitrary point from the boundary of $B$ outside the outer cone; that is, $x \notin \mathsf{C}_2$, and $x \in \partial B$. Then we have that $\|p - q\| < \|p - x\|$.*

**Proof.** Denote $\alpha_q := \angle pcq$ and $\alpha_x := \angle pcx$. Because $x$ is outside of the outer-cone with aperture $4\alpha$, and $p$ is inside of the inner cone of aperture $2\alpha$, we have that we have $\alpha_q < \alpha < \alpha_x$. Now consider the triangle $\triangle pcq$. By the Cosine Law, we have that

$$\|p - q\|^2 = \|p - c\|^2 + \|c - q\|^2 - 2\|p - c\| \cdot \|c - q\| \cos(\alpha_d) = \|p - c\|^2 + r^2 - 2r\|p - c\| \cos(\alpha_q).$$

Similarly, consider the $\triangle pox$, and we have

$$\|p - x\|^2 = \|p - c\|^2 + \|c - x\|^2 - 2\|p - c\| \cdot \|c - x\| \cos(\alpha_p) = \|p - c\|^2 + r^2 - 2r\|p - c\| \cos(\alpha_x).$$

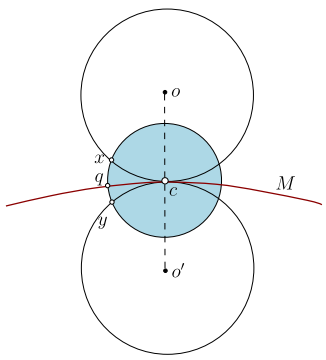
Since $0 \le \alpha_q < \alpha_x \le \pi$, we have $\|p - q\| < \|p - x\|$. $\square$

Now consider the surface Delaunay ball $B_c = B(c, r)$ that passing through the vertices of the simplex $\sigma$ and containing no other points from $Q$. Recall that $q$ is an arbitrary vertex of $\sigma$. Since all other vertices of $\sigma$ are at least $\delta$-Euclidean distance away from $q$, we then have that the intersection of $B_c$ with the cone $\mathsf{C}(c, \overrightarrow{cq}, 2\arcsin\frac{\delta}{2r})$ contains no point from $Q$ other than $q$. By applying Claim A.1 with $\alpha = \arcsin\frac{\delta}{2r}$, we then obtain that

**Corollary A.2.** *If there exists a point $p \in P$ such that $p \in B_c \cap \mathsf{C}(c, \overrightarrow{cq}, \arcsin\frac{\delta}{2r}) \cap M$, then $q$ must be the closest point to $p$ among all points in $Q$.*

In what follows, we will show that a point $p \in P$ satisfying the conditions in Corollary A.2 as well as that $p \in B(c, 4\varepsilon)$ indeed exists when $8\varepsilon \le \delta \le \frac{2}{27}\rho(M)$. This will then prove the proposition. Specifically, we will first identify a sample point $p \in P$, and then we will show that $p$ satisfies the requirements of the proposition.

**Identifying a point $p \in P$.** Let $B_o = B(m, \rho)$ and $B_{o'} = B(m', \rho)$ be two balls tangent to $M$ at $c$; assume without loss of generality that $B_o$ is inside of $M$ and $B_{o'}$ is outside. Locally around $c$, the surface $M$ is sandwiched between $B_o$ and $B_{o'}$. Now consider the plane $\mathcal{P} = \mathrm{span}\{o, o', q\}$; note that $c$ also lies in $\mathcal{P}$. Denote $B_{c,\mathcal{P}} = B_c \cap \mathcal{P}$, $B_{o,\mathcal{P}} = B_o \cap \mathcal{P}$ and $B_{o',\mathcal{P}} = B_{o'} \cap \mathcal{P}$. Let $x$ be the intersection point of $B_{c,\mathcal{P}}$ and $B_{o,\mathcal{P}}$ that is on the same side of the line passing through $oo'$ as the point $q$. Similarly, let $y$ be the intersection point of $B_{c,\mathcal{P}}$ with $B_{o',\mathcal{P}}$ on the same side of the line $oo'$ as $q$. Obviously, $q$ lines on the arc $\widehat{xy}$ that avoids $B_{o,\mathcal{P}}$ and $B_{o',\mathcal{P}}$. See the right figure for an illustration where the shaded region is $B_{c,\mathcal{P}}$. Set $\theta := \angle xcy$; easy to see that $\theta = \angle xoc = \angle yo'c$. Hence we have that $\sin\frac{\theta}{2} = \frac{\|c-x\|}{2\rho} = \frac{r}{2\rho}$.



Now consider the segment $cq$ and the point $w \in cq$ such that the length of $cw$ is a value $\ell$ which we will set later. As $w$ is contained in the cone with apex $c$ and aperture $\theta$ in the plane $\mathcal{P}$, the ball $B(w, \ell \sin\theta)$ will intersect both segment $cx$ and $cy$, thus intersecting both $B_o$ and $B_{o'}$. Since $B_o$ is inside the surface $M$ and $B_{o'}$ outside, it follows that $B(w, \ell \sin\theta) \cap M \neq \emptyset$. Pick any point $p' \in B(w, \ell \sin\theta) \cap M$. By the $\varepsilon$-sampling condition of $P$, there must exist a sample point $p \in P$ such that $\|p - p'\| \leq \varepsilon$. In other words, there is a sample point $p \in P$ such that $p \in B(w, \ell \sin\theta + \varepsilon)$.

*The requirements on p* We now need to show that the parameter $\ell$ can be chosen such that the point $p$ satisfies all the requirements from the proposition. In particular, we need the following:

**C-1** $p \in B(c, 4\varepsilon)$; and
**C-2** $q$ is the closest point to $p$ among all points in $Q$.

Now set $\tau = \arcsin \frac{\ell \sin\theta + \varepsilon}{\ell}$. Obviously, the ball $B(w, \ell \sin\theta + \varepsilon)$ (and thus the point $p$) is contained inside the cone $\mathrm{C}(c, \overrightarrow{cq}, \tau)$. Observe that by Corollary A.2, condition (C-2) is satisfied if (C-2.a) $B(w, \ell \sin\theta + \varepsilon) \in B(c, r)$ (implying that $p \in M$), and (C-2.b) $\tau \leq \arcsin \frac{\delta}{2r}$ (implying that $p$ is contained in the inner cone $\mathrm{C}(c, \overrightarrow{cq}, \arcsin \frac{\delta}{2r})$).

*The existence of a valid $\ell$* What remains is to find a value for $\ell$ so that (C-1), (C-2.a), and (C-2.b) are all satisfied simultaneously. Note that since $\|w - c\| = \ell$, we have that $\|p - c\| \leq \ell + \ell \sin\theta + \varepsilon$. Hence condition (C-1) is satisfied if $\ell + \ell \sin\theta + \varepsilon \leq 4\varepsilon$. Since $\sin\theta \leq 2 \sin\frac{\theta}{2} = \frac{r}{\rho}$, (C-1) holds as long as the following inequality holds.

$$\ell \leq \frac{3\varepsilon}{1 + \frac{r}{\rho}}. \tag{A.1}$$

Since $\delta \geq 8\varepsilon$, if (C-1) holds, then we have that $\|p - c\| \leq 4\varepsilon \leq \frac{\delta}{2} \leq r$, which implies (C-2.a). Now consider condition (C-2.b), which holds if $\frac{\ell \sin\theta + \varepsilon}{\ell} \leq \frac{\delta}{2r}$. Since $\delta/2 \leq r \leq \delta + \varepsilon$ and $8\varepsilon \leq \delta < 2\rho/27$, we have that

$$\frac{r}{\rho} \leq \frac{\delta + \varepsilon}{\rho} \leq \frac{\delta + \delta/8}{\rho} < \frac{1}{4} \leq \frac{\delta}{4\delta} \leq \frac{\delta}{2r}.$$

That is, $\frac{\delta}{2r} - \sin\theta \geq \frac{\delta}{2r} - \frac{r}{\rho} > 0$ (recall that $\sin\theta \leq 2\sin\theta = r/\rho$). Hence condition (C-2.b) holds if

$$\ell \geq \frac{\varepsilon}{\frac{\delta}{2r} - \frac{r}{\rho}} \left( \geq \frac{\varepsilon}{\frac{\delta}{2r} - \sin\theta} \right). \tag{A.2}$$

Putting (A.1) and (A.2) together, we have that as long as the value $\ell$ satisfies the following inequality

$$\frac{\varepsilon}{\frac{\delta}{2r} - \frac{r}{\rho}} \leq \ell \leq \frac{3\varepsilon}{1 + \frac{r}{\rho}}, \tag{A.3}$$

conditions (C-1) and (C-2) are satisfied, and there exists a point $p \in P$ as stated in the proposition. Given that $8\varepsilon \leq \delta < 2\rho/27$, we can show that valid $\ell$ exists. For example, for $\ell = \frac{36\varepsilon}{13}$, inequality in Eq. (A.3) holds as

$$\frac{\varepsilon}{\frac{\delta}{2r} - \frac{r}{\rho}} \leq \frac{\varepsilon}{\frac{\delta}{2(\delta+\varepsilon)} - \frac{\delta+\varepsilon}{\rho}} \leq \frac{\varepsilon}{\frac{4\delta}{9\delta} - \frac{\frac{9}{8}\delta}{\rho}} < \frac{\varepsilon}{\frac{4}{9} - \frac{\frac{9}{8} \cdot \frac{2\rho}{27}}{\rho}} = \frac{36\varepsilon}{13} = \ell$$

and

$$\frac{3\varepsilon}{1 + \frac{r}{\rho}} \geq \frac{3\varepsilon}{1 + \frac{\delta+\varepsilon}{\rho}} \geq \frac{3\varepsilon}{1 + \frac{\frac{9}{8}\delta}{\rho}} > \frac{3\varepsilon}{1 + \frac{\frac{9}{8} \cdot \frac{2\rho}{27}}{\rho}} = \frac{36\varepsilon}{12} = \ell.$$

## References


[1] H. Adams, G. Carlsson, On the nonlinear statistics of range image patches, SIAM J. Imaging Sci. 2 (2009) 110–117.
[2] N. Amenta, S. Choi, T.K. Dey, N. Leekha, A simple algorithm for homeomorphic surface reconstruction, Int. J. Comput. Geom. Appl. (2002) 125–141.
[3] D. Attali, A. Lieutier, D. Salinas, Vietoris–Rips complexes also provide topologically correct reconstructions of sampled shapes, in: Proceedings of the 27th Annual ACM Symposium on Computational geometry, SoCG '11, ACM, New York, NY, USA, 2011, pp. 491–500.
[4] J.D. Boissonnat, L.J. Guibas, S.Y. Oudot, Manifold reconstruction in arbitrary dimensions using witness complexes, Discrete Comput. Geom. (2009) 37–70.
[5] J.-D. Boissonnat, C. Maria, The simplex tree: an efficient data structure for general simplicial complexes, in: ESA, 2012, pp. 731–742.
[6] G. Carlsson, V. de Silva, Topological approximation by small simplicial complexes, Preprint, 2003.
[7] F. Chazal, L.J. Guibas, S.Y. Oudot, P. Skraba, Scalar field analysis over point cloud data, Discrete Comput. Geom. 46 (4) (2011) 743–775.
[8] F. Chazal, A. Lieutier, The "lambda-medial axis", Graph. Models 67 (4) (2005) 304–331.
[9] F. Chazal, S. Oudot, Towards persistence-based reconstruction in Euclidean spaces, in: Proc. 24th ACM Sympos. on Comput. Geom., 2008, pp. 232–241.
[10] S.-W. Cheng, T.K. Dey, E.A. Ramos, Manifold reconstruction from point samples, in: SODA '05: Proc. 16th Ann. ACM–SIAM Sympos. Discrete Algorithms, 2005, pp. 1018–1027.

[11] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, Discrete Comput. Geom. 37 (1) (2007) 103–120.
[12] V. de Silva, G. Carlsson, Topological estimation using witness complexes, in: Proc. Sympos. Point-Based Graphics, 2004, pp. 157–166.
[13] T.K. Dey, Curve and Surface Reconstruction: Algorithms with Mathematical Analysis, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, New York, NY, USA, 2006.
[14] T.K. Dey, F. Fan, Y. Wang, Computing topological persistence for simplicial maps, in: Proc. 30th Annu. ACM Sympos. Comput. Geom., 2014.
[15] T.K. Dey, J. Sun, Y. Wang, Approximating cycles in a shortest basis of the first homology group from point data, Inverse Probl. 27 (12) (2011) 124004.
[16] T.K. Dey, Y. Wang, Reeb graphs: approximation and persistence, Discrete Comput. Geom. 49 (1) (2013) 46–73.
[17] H. Edelsbrunner, J. Harer, Computational Topology: An Introduction, Amer. Math. Soc., Providence, Rhode Island, 2009.
[18] Q. Fang, J. Gao, L.J. Guibas, V. de Silva, L. Zhang, Glider: gradient landmark-based distributed routing for sensor networks, in: INFOCOM, 2005, pp. 339–350.
[19] R. Ghrist, Barcodes: the persistent topology of data, Bull. Am. Math. Soc. 45 (2008) 61–75.
[20] T. Gonzalez, Clustering to minimize the maximum inter-cluster distance, Theor. Comput. Sci. 38 (1985) 293–306.
[21] J.-C. Hausmann, On the Vietoris–Rips complexes and a cohomology theory for metric spaces, in: Proc. Conf. Prospects in Topology, 1994, pp. 175–188.
[22] J.R. Munkres, Elements of Algebraic Topology, Addison–Wesley Publishing Company, Menlo Park, 1984.
[23] S. Oudot, L.J. Guibas, J. Gao, Y. Wang, Geodesic Delaunay triangulations in bounded planar domains, ACM Trans. Algorithms 6 (4) (2010).
[24] D. Sheehy, Linear-size approximations to the Vietoris–Rips filtration, Discrete Comput. Geom. 49 (4) (2013) 778–796.