# Towards Stratified Space Learning: 2-complexes

Yossi Bokor Bleile

Department of Mathematical Sciences, Aalborg University, Fredrik Bajers Vej 7K, Aalborg Øst, 9220, Denmark.

Contributing authors: yossib@math.aau.dk;

**Abstract**

In this paper, we consider a simple class of stratified spaces – 2-complexes. We present an algorithm that learns the abstract structure of an embedded 2-complex from a point cloud sampled from it. We use tools and inspiration from computational geometry, algebraic topology, and topological data analysis and prove the correctness of the identified abstract structure under assumptions on the embedding.

**Keywords:** Stratified space learning, embedded spaces, applied topology, computational geometry

**MSC Classification:** 55N31 , 68T09 , 51-08

## 1 Introduction

Recent developments in technology have led to a dramatic increase in the quantity and complexity of data we can collect. These increases require new methods to enable efficient discovery and modelling of the structures underlying them. As the dimension in which we can observe data increases, it becomes more important to be able to reduce the dimensionality of large amounts of data. Some of the difficulties can be addressed by expanding the class of structures we can identify. In Bokor et al. (2021), the authors removed the assumption that the dimension is constant and presented an algorithm for learning the simplest class of stratified spaces – graphs. A stratified space is a space described by gluing together (manifold) pieces, called strata. There are no restrictions placed upon each stratum's dimension, and the gluing can give rise to a variety of interesting and complex local structures. We extend their work to the identification of the abstract structure underlying a 2-complex.

As observed in Bokor et al. (2021), manifold learning can be used to detect and model structures underlying data sets. A variety of approaches and algorithms exist to learn manifold structures from (noisy) samples, see Cheng et al. (2005), Dey (2007), Dey and Wang (2014). These methods often place assumptions on the manifold and the sampling procedure, generally in the form of restrictions on curvatures, as well as on the density of the sample and the type of noise. The assumptions on curvature are not satisfied by data sets arising in many applications, in particular geospatial data sets arising from person and vehicle movement in transportation networks. We make a second step towards in expanding the set of allowable underlying structures to include stratified spaces of dimension 2. Bendich et al. (2010) focuses on an algorithm to identify when two points have been sampled from the same stratum of a stratified space, but does not present a method for detecting what the dimension of this piece is, or what the global structure is. Stolz et al. (2020) present an algorithm for detecting samples of two intersecting manifolds, which is a first approximation of splitting a space into stratified pieces, and it comes with experimental verification but no theoretical guarantees. In Aanjaneya et al. (2012), the focus is on reconstructing a metric on a graph, with the input consisting of intrinsic distance on the metric graph, the associated theoretical guarantees are about the lengths of the edges in the metric, instead of relating to the geometry of the embedding. In particular, they do not need to consider vertices of degree 2, as in their setting these are points on an edge. Chazal et al. (2009) presents a method for sampling and reconstructing compact sets in Euclidean space, with a similar focus on samples with bounded Hausdorff noise and a sufficient density. They guarantee a homotopy equivalence under sufficient conditions but do not present a method for learning the stratified structure. Bendich et al. (2007) present an algorithm using persistent homology to assess the local homology of a sample at a particular point, using Delauney triangulations, which comes with a great computational cost.

## 1.1 Contribution

This paper describes an algorithm for learning the abstract structure underlying an embedded 2-complex, and provides theoretical guarantees in terms of the geometric embedding that the sample has come from. In particular, the algorithm can be used to learn the number of cells of each dimension, and how they piece together. The output of this algorithm can then be used as a starting point to learn the particular embedding the sample came from. Previous work has focussed on using persistent homology to approximate the local homology at a point, which comes with significant computational overheads. We avoid this by approximating the local homology at each point using a fixed approximation scale, relating to the geometric conditions on the embedding. The algorithm easily works in parallel, which significantly reduces the run time on large data sets. While the algorithm only applies to 2-complexes, many data sets arising from applications are 2-dimensional and it provides a foundation for further developments to increase/remove the dimensionality assumption. We acknowledge that from certain perspectives, moving from graphs to 2-complexes is a small step, yet there are many technical and geometric details involved in guaranteeing the accuracy of the structure learnt even for 2-complexes, and this is the limiting factor for removing the dimensionality assumption at this stage.

This article begins with Section 2, containing definitions of the main objects and tools we use throughout the article. After this, Section 3 consists of geometric lemmas used in Section 4, which considers the local geometry and topology we use to partition the sample $P$. Finally, Section 5 presents algorithms for recovering the abstract structure. Section 5 contains a sequence of lemmas (Lemmas 5.9 to 5.24), which cover cases used in Theorem 5.25, also known as the 'Big Theorem' of this article.

## 2 Definitions and Notations

We begin with some definitions and notations we use throughout this article. We begin with the following definition of complex, following Definition 2.4 Carlsson (2014).

**Definition 2.1** (Abstract Complex, Definition 2.4 Carlsson (2014))**.** *An* abstract simplicial complex *$X$ consists of a pair $(V(X), \Sigma(X))$, with $V(X)$ a finite set, and $\Sigma(X)$ a subset of the power set of $V(X)$, such for all $\sigma \in \Sigma(X)$ and $\emptyset \neq \tau \subseteq \sigma$, we have $\tau \in \Sigma(X)$. We call $V(X)$ the vertices, and $\Sigma(X)$ the* simplices *of $X$.*

For ease of notation and to avoid confusion later in this paper, we will use the following specialised definition for abstract simplicial complexes with top dimension 2.

**Definition 2.2** (Abstract 2-Complex)**.** *An* abstract 2-complex *$X$ consists of*

1. *a set $V = V(X)$ of vertices,*
2. *a set $E = \{ \sigma \in \Sigma(X) \,|\, \sigma$ contains 2 unique elements$\}$ of edges,*
3. *a set $T = \{ \sigma \in \Sigma(X) \,|\, \sigma$ contains 3 unique elements$\}$ of triangles,*

*and an incidence operator $\mathcal{I}$, which acts as follows: for any pair of cells $\sigma, \tau \in X$*

$$\mathcal{I}(\sigma, \tau) = \begin{cases} 1 & \text{if } \sigma \subsetneq \tau \\ 0 & \text{otherwise} \end{cases}$$

We restrict ourselves to linear embeddings of 2-complexes $X$ in $\mathbb{R}^n$ for some $n \geq 3$.

**Definition 2.3** (Linear embedding of 2-complex)**.** *Fix $n \geq 3$, then a linear embedding of a 2-complex $X$ in $\mathbb{R}^n$, $(X, \Theta)$, consists of an abstract 2-complex $X$ and a map*

$$\Theta : X \to \mathbb{R}^n$$

*such that*

1. *on vertices $v \in V$, $\Theta$ is injective,*
2. *on edges $\{u, v\} \in E$, $\Theta$ is defined by linear interpolation on $\Theta(u)$ and $\Theta(v)$: $\Theta(\{ u, v \}) = \overline{uv}$ is the line segment between $\Theta(u)$ and $\Theta(v)$,*
3. *on triangles $\{ u, v, w \} \in E$, $\Theta$ is defined by linear interpolation on $\Theta(u)$, $\Theta(v)$ and $\Theta(w)$: $\Theta(\{ u, v, w \}) = \triangle uvw$ is the triangle with vertices $\Theta(u)$, $\Theta(v)$ and $\Theta(w)$, and $\Theta(u)$, $\Theta(v)$, $\Theta(w)$ are no co-linear,*
4. *for any two cells $\sigma, \tau$ of $X$, we have $\Theta(\sigma) \cap \Theta(\tau) = \Theta(\sigma \cap \tau)$.*

   *We restrict our attention to embedded 2-complexes $|X|_\Theta$ such that*

5. *if a vertex $v$ is in the boundary of precisely two edges $\{v, u_1\}$ and $\{v, u_2\}$, then $\angle u_1 v u_2 \neq \pi$,*

6. *if an edge* $\{v_0, v_1\}$ *is in the boundary of precisely two triangles* $\{v_0, v_1, u_1\}$ *and* $\{v_0, v_1, u_2\}$*, then* $v_0, v_1, u_1, u_2$ *are not co-planar.*

*We denote the image of* $\Theta$ *in* $\mathbb{R}^n$ *by* $|X|_\Theta$*.*

We often talk about the *boundary* of a cell.

**Definition 2.4** (Cell boundary)**.** *Let* $X$ *be an abstract* 2*-complex, and take a cell* $\sigma \in X$*. Then the* boundary *of* $\tau$*,* $\partial\tau$*, consists of the cells* $\sigma \in X$ *such that* $\mathcal{I}(\sigma, \tau) = 1$*.*

An important property of a cell $\sigma \in X$, is whether it is *locally maximal* or not.

**Definition 2.5** (Locally maximal cell)**.** *Let* $\sigma$ *be a cell in a* 2*-complex. We say* $\sigma$ *is* locally maximal *if there is no cell* $\tau \in X, \tau \neq \sigma$ *with* $\sigma \subset \tau$*. That is, there is no cell* $\tau$ *with* $\sigma$ *in the boundary of* $\tau$*.*

**Remark 1.** *Consider two cells* $\sigma, \tau$ *in a complex* $X$*, we say* $\sigma$ is a face of $\tau$ *if* $\sigma$ *is in the boundary of* $\tau$*, and we say* $\sigma$ is a co-face of $\tau$ *if* $\tau$ *is in the boundary of* $\sigma$*.*

We can represent the incidence relationships of cells in $X$ in a weighted graph $B$.

**Definition 2.6** (Incidence graph)**.** *Take an abstract* 2*-complex* $X$*. The* incidence graph $B$ *of* $X$ *is the weighted graph with*

1. *a weight* 0 *node* $n_v$ *for each vertex* $v$ *of* $X$*,*
2. *a weight* 1 *node* $n_e$ *for each edge* $e = \{u, v\}$ *of* $X$*,*
3. *a weight* 2 *node* $n_t$ *for each triangle* $t = \{u, v, w\}$ *of* $X$*,*
4. *an edge between a weight* 2 *node* $n_t$ *and weight* 1 *node* $n_e$ *if* $e \subset t$*,*
5. *an edge between a weight* 2 *node* $n_t$ *and weight* 0 *node* $n_v$ *if* $v \in t$*,*
6. *an edge between a weight* 1 *node* $n_e$ *and weight* 0 *node* $n_v$ *if* $v \in e$*.*

Abusing notation, we usually write $|X|$ instead of $|X|_\Theta$ or $(X, \Theta)$, use $v$ to denote both the abstract vertex and its embedded location $\Theta(v)$, $\overline{uv}$ to denote both the abstract edge and the embedded image $\Theta(\{\, u, v \,\})$, and $\triangle uvw$ to denote both the abstract triangle and the embedded image $\Theta(\{\, u, v, w \,\})$. Whether we are referring to an element of the abstract 2-complex or its image in $\mathbb{R}^n$ should be clear from the context.

We use the following conventions in this article. Given two points $x, y \in \mathbb{R}^n$, $\|x-y\|$ is the standard Euclidean distance between $x$ and $y$, for a point $x \in \mathbb{R}^n$ and a set $Y \subset \mathbb{R}^n$, we set

$$d(x, Y) := \inf_{y \in Y} \|x - y\|,$$

and for two sets $X, Y \subset \mathbb{R}^n$, we set

$$d(X, Y) := \min \left\{ \inf_{x \in X} d(x, Y), \inf_{y \in Y} d(y, X) \right\},$$

$$d_H(X, Y) := \max \left\{ \sup_{x \in X} d(x, Y), \sup_{y \in Y} d(y, X) \right\}.$$

We also consider thickenings of a subset $X$: we let

$$X^\alpha := \{p \in \mathbb{R}^n \mid d(p, H) \leq \alpha\}.$$

In proofs towards the end of this article, we use the *weak feature size* of $X$ to allow us to construct isomorphism, which was introduced in Chazal and Lieutier (2007) as the infimum of the positive critical values of the distance function of $X$.

At various moments in the algorithm, we consider the *diameter* of a set of points $X$. The *diameter of* $X$, $\mathcal{D}(X)$, is the maximum distance between any pair of points $x, y \in X$:

$$\mathcal{D}(X) := \max_{x,y \in X} \|x - y\|.$$

We use $B_r(p)$ to denote the ball of radius $r$ centred at a point $p \in \mathbb{R}^n$, by $\partial B_R(p)$ we mean the boundary of such a ball, and let

$$\mathbb{S}^k = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$$

denote the standard $k$-sphere. We also regularly consider points in a *spherical shell.*

**Definition 2.7.** *Fix $a < b$, and let $y$ be a point in $\mathbb{R}^n$. The* spherical shell of radii $a$ and $b$ *centered at $p$, $S_a^b(p)$ is the set*

$$\{q \in \mathbb{R}^n \mid a \leq \|q - p\| \leq b\}.$$

We consider dihedral angles between two half-planes.

**Definition 2.8.** *Let $H_1, H_2$ be two half-planes with a common boundary line $L$. Then, the* dihedral angle $\alpha$ *between $H_1$ and $H_2$ is the angle formed by two vectors $v_1 \in H_1$ and $v_2 \in H_2$ originating from the same point $x \in L$ such that both $v_1$ and $v_2$ are perpendicular to $L$.*

We work with $\varepsilon$-samples $P$ of the embedded 2-complex $|X|$.

**Definition 2.9** ($\varepsilon$-sample)**.** *Let $|X| \subset \mathbb{R}^n$ be an embedded 2-complex. An* $\varepsilon$-sample $P$ *of $|X|$ is a finite subset of $\mathbb{R}^n$ such that $d_H(|X|, P) \leq \varepsilon$.*

In Bokor et al. (2021) the authors use the threshold graph on a set of points, which we will also use.

**Definition 2.10** (Threshold graph, Definition 3.1 Bokor et al. (2021))**.** *Let $P \subset \mathbb{R}^N$ be a finite collection of points, and fix $r > 0$. The* graph at threshold $r$ on $P$, $\mathfrak{G}_r(P)$, *is the graph with vertices $p \in P$, and edges $(p, q)$ if $\|p - q\| \leq r$.*

The objects we consider in this article are 2-dimensional, and so we also use *Čech* complexes.

**Definition 2.11** (Čech Complex)**.** *Let $P \subset \mathbb{R}^n$ be a finite set of points. The* Čech complex at scale $\delta$, $\check{\mathcal{C}}_\delta(P)$ *is the complex with $j$-cells $\{v_i\}_{i=0}^j$ such that the intersection $\bigcap_{i=0}^j B_\delta(v_i)$ is non-empty.*

Now, we formalise the aim of this article. Given an $\varepsilon$-sample $P$ of some linearly embedded 2-complex $|X|$, we want to recover the abstract structure of the 2-complex $X$. To do this, we need to learn the number of vertices, the number of edges, and the number of triangles, as well as the incidence relations between them. We achieve this by first deciding for each $p \in P$ if it is near a cell that is not locally maximal, or far away from all cells which are not locally maximal. This partitions $P$ into two subsets which intuitively are $P_{NLM}$ containing samples $p$ near non-locally maximal cells, and $P_{LM}$ containing samples $p$ only near locally maximal cells. Rigorous definitions of $P_{NLM}$ and $P_{LM}$ are in Definition 4.6. Part of this process involves approximating the local homology at each $p \in P$ using a radius $r$. This requires a choice of scale at which to approximate $|X|$ from $P$. Unlike in Bokor et al. (2021), the relationship between clusters in $P_{NLM}$ and $P_{LM}$ to vertices, edges and triangles is not direct. We can, however, still infer the incidence operator.

**Remark 2.** *In this paper, we use* local homology *in very restrictive settings. It is a very generally construction: for a space $X$, the local homology of $X$ at a point $x \in X$ is the relative homology $H(X, X \setminus \{x\})$.*

# 3 Geometric Lemmas

We provide some geometric lemmas as motivation for the definitions of local structures and the geometric assumptions we place on the embeddings of a 2-complex. There are two parts to the definition of the local structure of a point cloud $P$ at a sample $p$: the first is a topological condition relating to the homology of the samples in a spherical shell around $p$, and the second relates to the geometry of these samples. The geometric lemmas in this section allow us to distinguish between points near cells that are not locally maximal and those that are only near locally maximal cells when the topological structure of $P$ at $p$ does not, see Section 4. The proofs of the lemmas in this section can be found in Appendix A.

We begin with a helpful lemma that bounds the distance between a point in a spherical shell within $\varepsilon$ of a ray and the point in the ray in the middle of the shell.

**Lemma 3.1.** *Let $L \subset \mathbb{R}^n$ be a ray originating at a point $z$, and fix*

$$R \geq 14\varepsilon > 0.$$

*Let $P \subset \mathbb{R}^n$ have $d_H(P, L) \leq \varepsilon$ and take $p \in \mathbb{R}^n$ with*

$$\|p - z\| \leq \frac{R}{2}.$$

*Let $x$ be the point in $L$ with $\|x - p\| = R$. Then for all $q \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$*

$$\|q - x\| \leq \sqrt{2}\varepsilon.$$

Next, Lemma 3.2, which motivates part 3 in Definition 4.4. The lemma considers the distances between triples of points in $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap H^\varepsilon$ for some point $p \in H^\varepsilon$, where $H^\varepsilon$ is the thickening of a plane $H$ by $\varepsilon$, with $\varepsilon > 0$.

**Lemma 3.2.** *Consider an affine 2-hyperplane $H \subset \mathbb{R}^n$ and fix*

$$R \geq 14\varepsilon \geq 0.$$

*Let $P \subset \mathbb{R}^n$ be such that $d_H(P, H) \leq \varepsilon$, and take $p$ with $d(p, H) \leq \varepsilon$. Then, for all $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$, there exists $q_2 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ with*

$$\|q_2 - q_1\| \geq 2\sqrt{R^2 - \varepsilon^2} - (1 + \sqrt{2})\varepsilon.$$

Now that we have a geometric property to test if a point $p$ and the samples in $S_{R-\varepsilon}^{R+\varepsilon}(p)$ are from a subset of a plane. We want to understand what conditions need to be placed on points near an edge in two triangles to guarantee this property does not hold. In particular, Lemma 3.3 motivates part 4 of Definition 4.5.

For ease of reading, we let

$$\Psi(\varepsilon, R) = \arccos\left(\frac{(R+2\varepsilon)^2 + \left(\frac{3R}{2} - \varepsilon\right)^2 - \left(2\sqrt{R^2-\varepsilon^2} - \left(2+2\sqrt{2}\right)\varepsilon\right)^2}{2(R+2\varepsilon)\left(\frac{3R}{2} - \varepsilon\right)}\right).$$

The following lemma motivates the conditions we place on the dihedral angle between two triangles with a common boundary edge $\overline{uv}$ (of degree 2). This allows us to guarantee that the geometry of the samples in $S_{R-\varepsilon}^{R+\varepsilon}(p)$ for a sample $p$ *near* $\overline{uv}$ is not the same as the geometry of samples in $S_{R-\varepsilon}^{R+\varepsilon}(p)$ when $p$ is near a triangle but *far away* from its boundary.

**Lemma 3.3.** *Consider two affine* 2*-half-planes* $H_1, H_2 \subset \mathbb{R}^n$ *whose boundaries are equal, say* $L$*, and fix* $R \geq 14\varepsilon > 0$*. Let* $\alpha$ *be the dihedral angle between* $H_1$ *and* $H_2$*. Let* $P$ *be a set of points such that* $d_H(P, H_1 \cup H_2) \leq \varepsilon$*. Further, take* $p$ *such that* $d(p, H_1) \leq \varepsilon$*. If*

$$d(L,p) \leq \frac{R}{2} - 2\varepsilon$$

*and*

$$\alpha \in (0, \Psi(\varepsilon, R))$$

*then there exist* $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ *such that for all* $q_2 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(1+\sqrt{2}\right)\varepsilon.$$

Next, we investigate the geometry of points near a ray and half-plane, to develop a test for points near not locally maximal cells.

There are several local structures that have the same topological structure: they consist of two connected components with no 1-cycles. In Bokor et al. (2021), the authors used the angle between the centroids of the two connected components to distinguish between points near a degree 2 vertex and points near the *interior* of an edge. Unfortunately, this test is not sufficient after introducing triangles. If we first check for the presence of triangles, we can again use the inner-product test. To test for the presence of triangles, we examine the diameters of the two connected components.

So, we first bound the diameter of a set of samples only near a line.

**Lemma 3.4** (Diameter of points near ray)**.** *Let* $L \subset \mathbb{R}^n$ *be a ray originating at a point* $z$*, and fix* $R > 14\varepsilon > 0$*. Let* $P \subset \mathbb{R}^n$ *have* $d_H(P, L) \leq \varepsilon$ *and take* $p \in \mathbb{R}^n$ *with* $d(L,p) \leq \varepsilon$ *and* $\|p - z\| \leq \frac{R-\varepsilon}{2}$*. Then* $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{3\varepsilon}{2}}$ *has* 1 *connected component* $c$*, and the diameter is less than* $2\sqrt{2}\varepsilon$*.*

The previous lemma bounds the diameter of a connected component containing points with $\varepsilon$ of an edge, that are within $S_{R-\varepsilon}^{R+\varepsilon}(p)$ for a sample $p$ near a vertex in the boundary of this edge. We need to guarantee that if $p$ is near the interior of an edge, it does not fail the diameter test. To ensure this, we obtain the following as a corollary of Lemma 3.4.

**Corollary 1.** *Let $L \subset \mathbb{R}^n$ be a line, and fix $R > 3\varepsilon > 0$. Let $P \subset \mathbb{R}^n$ have $d_H(P, L) \leq \varepsilon$ and take $p \in \mathbb{R}^n$ with $d(L, p) \leq \varepsilon$ and*

$$\|p - z\| \leq \frac{R - \varepsilon}{2}.$$

*Then $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{3\varepsilon}{2}}$ has 2 connected components $c_1, c_2$, and their diameters are less than $2\sqrt{2}\varepsilon$.*

*Proof.* First note that $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap L$ consists of two connected components, $C_1, C_2$, and the distance between them is $R - \varepsilon$. Hence, we can apply Lemma 3.4, to $C_1$ and $C_2$ individually, obtaining a connected component for each, say $c_1$ and $c_2$. Further, the diameters of $c_1$ and $c_2$ are less than $2\sqrt{2}\varepsilon$. □

The following lemma guarantees that if there are samples in $S_{R-\varepsilon}^{R+\varepsilon}(p)$ that are within $\varepsilon$ of a triangle, the diameter test fails.

**Lemma 3.5.** *Let $L_1, L_2 \subset \mathbb{R}^n$ be two rays originating at the same point $z$ with the angle $\alpha$ between in the interval*

$$\left[\frac{\pi}{6}, \pi\right),$$

*and fix $R \geq 14\varepsilon > 0$. Let $T$ be the set between $L_1$ and $L_2$. Take $p \in \mathbb{R}^n$ with $d(T, p) \leq \varepsilon$ and $\|p - x\| \leq \frac{R-\varepsilon}{2}$, and $P \subset \mathbb{R}^n$ with $d_H(P, T) \leq \varepsilon$. Then, there exist points $q_1, q_2$ in $P$ with $\|q_1 - p\|, \|q_2 - p\| \in [R - \varepsilon, R + \varepsilon]$ such that $\|q_1 - q_2\| > 2\sqrt{2}\varepsilon$, and $q_1, q_2$ are path connected. Furthermore, the connected component containing $q_1$ and $q_2$ has diameter bigger than $2\sqrt{2}\varepsilon$.*

# 4 Local Structures

To identify the abstract structure of the 2-complex, the algorithm in Section 5 first partitions the sample $P$ into sets $P_{LM}$, containing samples that are only near locally maximal cells, and $P_{NLM}$, containing samples near cells that are not locally maximal. The decision tree for if a point is in $P_{NLM}$ or $P_{LM}$ is summarised in Figure 1. After this, we further partition $P_{LM}$ and $P_{NLM}$ to infer the number of cells and their dimensions, as well as the incidence operator.

Take an embedded 2-complex $|X| \subset \mathbb{R}^n$, fix (an appropriate) $0 < \varepsilon \leq R$ and take $p \in \mathbb{R}^n$ with $d(|X|, p) \leq \varepsilon$. Consider the topological and geometric structure of $|X|$ in a neighbourhood of $p$, beginning with $B_R(p) \cap |X|$. If $B_R(p) \cap |X|$ is disconnected, we restrict to the connected component $C_p$ containing $\mathrm{proj}_{|X|}(p)$. Then, we consider $\partial B_R(p) \cap C_p$. Let $\mathrm{proj}_{|X|}(p)$ be the projection of $p$ to $|X|$, and let $\sigma_p$ be the cell containing $\mathrm{proj}_{|X|}(p)$. If $\sigma_p$ is locally maximal and $d(|\partial\sigma_p|, p) > R$, then $\partial B_R(p) \cap C_p$ has one of the following structures:

1. $\partial B_R(p) \cap C_p$ is empty, in which case $\sigma_p$ is a locally maximal vertex,
2. $\partial B_R(p) \cap C_p$ is a pair of antipodal points, in which case $\sigma_p$ is a locally maximal 1-cell,
3. $\partial B_R(p) \cap C_p$ is homotopic to $\mathcal{S}^1$ lying in a plane, in which case $\sigma_p$ is a 2-cell.

The above structures consist of two parts: we examine the topological structure of $\partial B_R(p) \cap C_p$, and then look at its geometry. If $p$ is within $R$ of some cell $\tau_p$ (possibly $\tau_p = \sigma_p$) which is not locally maximal, then either the topological structure or the geometric structure is not one of the above cases. As such, we use a two-step process to decide if a given sample $p$ is within $R$ of some not locally maximal cell $\tau_p$: first, we examine the topological structure of $\partial B_R(p) \cap C_p$ by looking at its homology, and then if necessary, we consider its geometric structure. We let

$$\mathcal{H}_\bullet(p) := H_\bullet\left(\partial B_R(p) \cap C_p\right).$$

As we are restricting ourselves to 2-complexes, we focus on $\mathcal{H}_0(p)$ and $\mathcal{H}_1(p)$.

**Definition 4.1** (Local homology signature)**.** *Let $|X| \subset \mathbb{R}^n$ be an embedded 2-complex, and fix $R > \varepsilon > 0$. Take a point $p \in \mathbb{R}^n$ with $d(p, |X|) \leq \varepsilon$. The* local homology signature *of $|X|$ at $p$ is*

$$\mathrm{Sig}(p) := (|\mathcal{H}_0(p)|, |\mathcal{H}_1(p)|).$$

In the above cases, the local homology signature of $|X|$ at $p$ is as follows.

1. $\mathrm{Sig}(p) = (0, 0)$,
2. $\mathrm{Sig}(p) = (2, 0)$,
3. $\mathrm{Sig}(p) = (1, 1)$.

and so if $\mathrm{Sig}(p)$ is not equal to $(0,0), (2,0)$ or $(1,1)$, then $p$ is within $R$ of a cell $\tau_p$ which is not locally maximal. If $\mathrm{Sig}(p)$ is $(0,0)$ then $p$ is within $\varepsilon$ of a degree 0 vertex. Unfortunately, if $\mathrm{Sig}(p)$ is either $(2,0)$ or $(1,1)$, we need to examine the geometric structure of $\partial B_R(p) \cap C_p$. When $\mathrm{Sig}(p) = (2,0)$, we can distinguish between the case where $\sigma_p$ is a locally maximal 1-cell and where $\sigma_p$ is a vertex of degree 2 as follows: let the two points in $\partial B_R(p) \cap C_p$ be $c_1$ and $c_2$. If $\sigma_p$ is a 1-cell, then $\angle c_1 p c_2 = \pi$, and other $\angle c_1 p c_2 \neq \pi$. When $\mathrm{Sig}(p) = (1,1)$ we need to distinguish between if $\sigma_q$ is a 2-cell, and if $\sigma_p$ is in the boundary of 2-cells. We can do so by checking if $\partial B_R(p) \cap C_p$ is contained in a plane: if it is, then $\sigma_p$ is a 2-cell, if not $\sigma_p$ is either an edge or a vertex that is not locally maximal.

Recall that we are working with an $\varepsilon$-sample $P$ of the embedded 2-complex $|X|$ instead of $|X|$. We want to approximate $\mathrm{Sig}(p)$ with $P$. As $P$ is an $\varepsilon$-sample, we can approximate $\partial B_R(p) \cap C_p$ by first considering the structure of $B_{R+\varepsilon}(p) \cap P$, then the structure of $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$. Before we define the $(\varepsilon, R)$-local structure of $P$ at $p$ (Definition 4.3), we need the following notation.

**Definition 4.2.** *Let $P \subset \mathbb{R}^n$ be a finite set of points. Then, $\mathrm{rk}_k^{\delta,\gamma}(P)$ is the rank of the map on the $k^{th}$ homology groups induced by the inclusion $P^\delta \hookrightarrow P^\gamma$.*

We can now formally define the $(\varepsilon, R)$-local structure of $P$ at $p$.

**Definition 4.3** ($(\varepsilon, R)$-local homology signature)**.** *Let $P \subset \mathbb{R}^n$ be an $\varepsilon$-sample of an embedded 2-complex $|X|$, and fix $R \geq 14\varepsilon$. Let $C_p^{\frac{3\varepsilon}{2}}$ be samples in the same connected component of threshold graph $\mathfrak{G}_{3\varepsilon}\left(B_{R+\varepsilon}(p) \cap P\right)$ as $p$. The* $(\varepsilon, R)$-local homology signature $\mathrm{Sig}_{\varepsilon,R}(p)$ *of $P$ at a sample $p$ is*

$$\mathrm{Sig}_{\varepsilon,R}(p) := \left(\mathrm{rk}_0^{\frac{3\varepsilon}{2},\frac{7\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right), \mathrm{rk}_1^{\frac{3\varepsilon}{2},\frac{7\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)\right).$$

We now define the types of local structures, beginning with *maximal* local structures.

**Definition 4.4** (Maximal $(\varepsilon, R)$-local structure)**.** *Let $P$ be an $\varepsilon$ sample of a linearly embedded 2-complex $|X|$ and fix $R \geq 14\varepsilon$. Let $C_p^{\frac{3\varepsilon}{2}}$ be the set of samples in the same connected component of $(B_{R+\varepsilon}(p) \cap P)^{\frac{3\varepsilon}{2}}$ as $p$. We say the* $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal *if any of the following hold:*

1. $\mathrm{Sig}_{\varepsilon,R}(p) = (0,0)$*, in which case we say that the* $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 0*,*
2. $\mathrm{Sig}_{\varepsilon,R}(p) = (2,0)$*, and the two connected components $c_1, c_2$ of $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)^{\frac{3\varepsilon}{2}}$ have diameters less than $5\varepsilon$ and mid-points $q_1$ and $q_2$ such that*
$$\langle q_1 - p, q_2 - p \rangle \leq -R^2 + 2R\varepsilon + 7\varepsilon^2,$$
*in which case we say that the* $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 1*,*
3. $\mathrm{Sig}_{\varepsilon,R}(p) = (1,1)$*, and for all $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P, \exists q_2 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ with*
$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(1 + \sqrt{2}\right)\varepsilon.$$
*in which case we say that the* $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 2*,*

Next, we define *not maximal* $(\varepsilon, R)$-local stuctures.

**Definition 4.5** (Not maximal $(\varepsilon, R)$-local structure)**.** *Let $P$ be an $\varepsilon$ sample of a linearly embedded 2-complex $|X|$ and fix $R \geq 14\varepsilon$. Let $C_p^{\frac{3\varepsilon}{2}}$ be the set of samples in the same connected component of $\check{C}_{\frac{3\varepsilon}{2}}(S_{R+\varepsilon}(p) \cap P)$ as $p$. We say that* the $(\varepsilon, R)$-local structure of $P$ at $p \in P$ is not maximal *if any of the following hold:*

1. $\mathrm{Sig}_{\varepsilon,R}(p) = (n,0)$ *for some $n \in \mathbb{Z}_{\geq 0}$, $n \neq 0, 2$,*
2. $\mathrm{Sig}_{\varepsilon,R}(p) = (1,n)$ *for some $n \in \mathbb{Z}_{\geq 0}$, $n \neq 1$,*
3. $\mathrm{Sig}_{\varepsilon,R}(p) = (2,0)$ *and letting two connected components of*
$$\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)^{\frac{3\varepsilon}{2}}$$
*be $c_1, c_2$, either* $\max\{\mathcal{D}(c_1), \mathcal{D}(c_2)\} \leq 2\sqrt{2}\varepsilon$ *and letting mid-points of $c_1, c_2$ be $q_1, q_2$*
$$\langle q_1 - p, q_2 - p \rangle > -R^2 + 2R\varepsilon + 7\varepsilon^2,$$
4. $\mathrm{Sig}_{\varepsilon,R}(p) = (1,1)$ *and there exists $q_1 \in P \cap S_{R-\varepsilon}^{R+\varepsilon}$ such that for all $q_2 \in P \cap S_{R-\varepsilon}^{R+\varepsilon}$*
$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(1 + \sqrt{2}\right)\varepsilon.$$

Having defined the two classes of $(\varepsilon, R)$-local structures, we can define our initial partition.

**Definition 4.6** ($P_{LM}$ and $P_{NLM}$)**.** *Let $P$ be an $\varepsilon$-sample of an embedded 2-complex $|X|$. We partition $P$ into two sets $P_{LM}$ and $P_{NLM}$ defined as*

$$P_{LM} := \{p \in P \mid \text{ the } (\varepsilon, R)\text{-local structure at of } P \text{ at } p \text{ is maximal.}\}$$
$$P_{NLM} := \{p \in P \mid \text{ the } (\varepsilon, R)\text{-local structure of } P \text{ at } p \text{ is not maximal.}\}$$

**Remark 3.** *For all $p \in P$, $P$ either has maximal $(\varepsilon, R)$-local structure at $p \in P$ or it does not. Hence, the partitioning of $P$ into $P_{LM}$ and $P_{NLM}$ defined in Definition 4.6 is disjoint.*

Recall that the samples we are working with can contain noise, and we use the homology of $\check{C}_{\frac{3\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)$ in the definition of $(\varepsilon, R)$-local structure. Hence, we place assumptions on $|X|$ to ensure that we correctly detect when samples are near cells that are not locally maximal. We place assumptions on the distances between any two vertices $u$ and $v$, the distance between an edge $\overline{uw}$ and a vertex $v \neq u, w$, the angle between any pair of edges with a common boundary vertex. Additionally, we place assumptions on the dihedral angle between any two 2-cells which have common boundary components. So that we can infer the incidence operator, we will require an upper bound on the relationship between $R$ and $\varepsilon$, and so we also restrict out choice of $R$ in terms of $\varepsilon$. We use the following notation in the decicision flow chart (Figure 1):

$$\beta = -R^2 + 2R\varepsilon + 7\varepsilon^2,$$
$$\gamma = 2\sqrt{R^2 - \varepsilon^2} - \left(1 + \sqrt{2}\right)\varepsilon.$$

To increase the readability of this article, we define the following functions.

**Definition 4.7.** *Fix $R > 14\varepsilon > 0$. We define the following functions:*

1.

$$\Psi_1(\varepsilon, R) = \arccos\left(\frac{\left(\frac{R}{2} - \varepsilon\right)^2 - 18\varepsilon^2}{\left(\frac{R}{2} - \varepsilon\right)^2}\right)$$
$$\geq \arccos\left(\frac{(R-\varepsilon)^2 - 18\varepsilon^2}{(R-\varepsilon)^2}\right) + 2\arcsin\left(\frac{2\varepsilon}{(R-\varepsilon)}\right)$$

2.

$$\Psi_2(\varepsilon, R) = \pi - \arctan\left(\frac{R + 3\varepsilon}{6\varepsilon}\right) + \arcsin\left(\frac{R^2 - 4R\varepsilon - 9\varepsilon^2}{(R+\varepsilon)\sqrt{R^2 + 6R\varepsilon + 34\varepsilon^2}}\right)$$

3.

$$\Psi_3(\varepsilon, R) = \arccos\left(\frac{(R + 2\varepsilon)^2 + \left(\frac{3R}{2} - \varepsilon\right)^2 - \left(2\sqrt{R^2 - \varepsilon^2} - \left(2 + 2\sqrt{2}\right)\varepsilon\right)^2}{2(R + 2\varepsilon)\left(\frac{3R}{2} - \varepsilon\right)}\right)$$

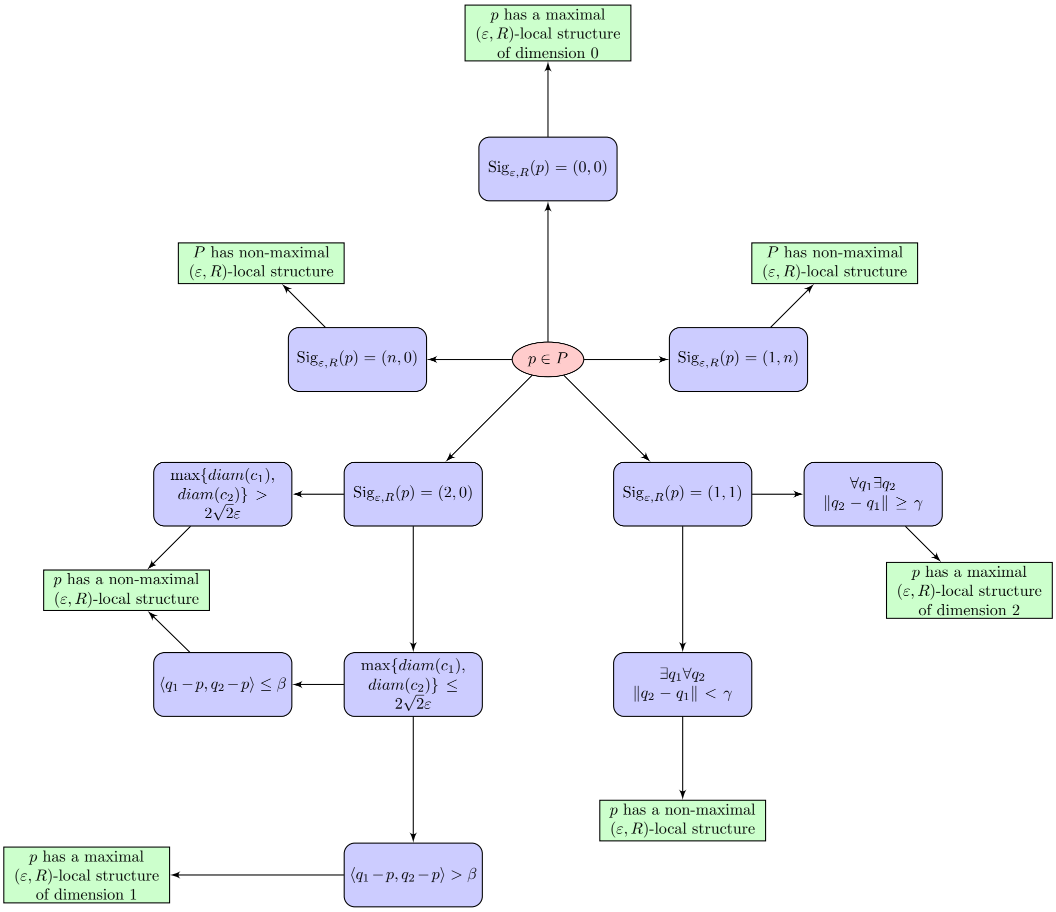

**Fig. 1**: Flow chart for determining if the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal or not. If maximal, what the dimension is.

To improve intuition of these functions, Figures 2 to 4 provide graphs of them. Note they are effectively a function of $\frac{R}{\varepsilon}$ as they are invariant to scaling both $R$ and $\varepsilon$ by the same amount.

We now state the assumptions we place on $|X|$.

**Assumption 1.** *Fix $R \geq 14\varepsilon > 0$. We restrict to embedded 2-complexes $|X| = (X, \pi)$ which satisfy the following.*

1. *For all vertices $u, v$,*
$$\|u - v\| > 6(R + \varepsilon).$$
2. *For a vertex $v$ and edge $\overline{uw}$ with $v \neq u, w$,*
$$d(\overline{uw}, v) > 6(R + \varepsilon).$$

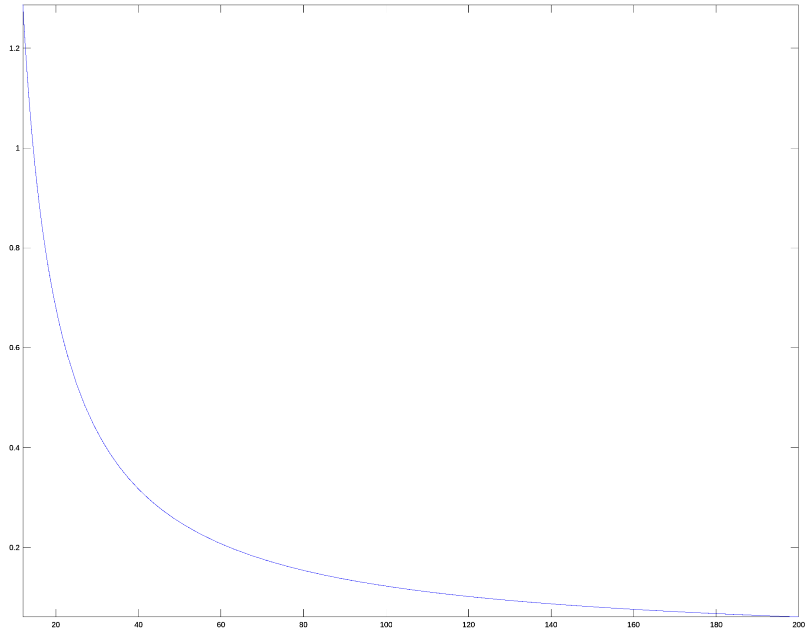


**Fig. 2**: Graph of $\Psi_1\left(1, \frac{R}{\varepsilon}\right)$.

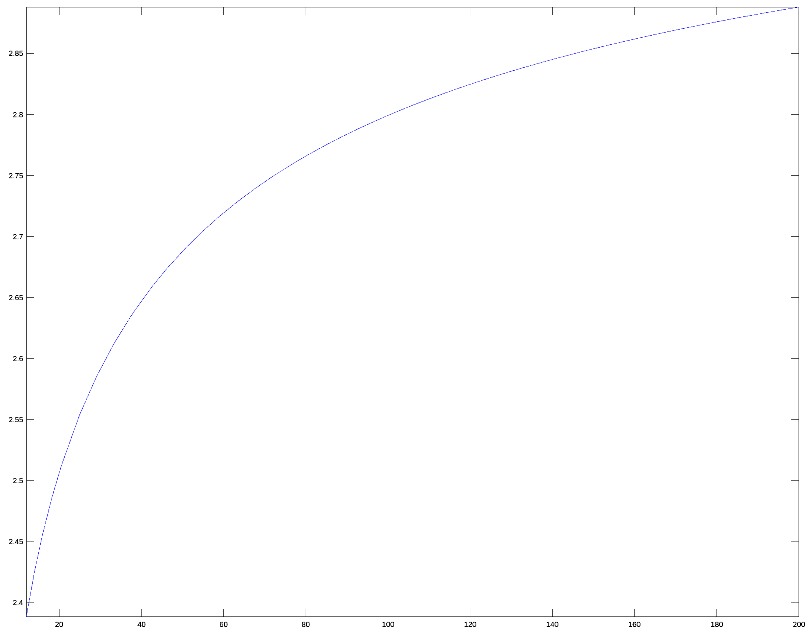


**Fig. 3**: Graph of $\Psi_2\left(1, \frac{R}{\varepsilon}\right)$.

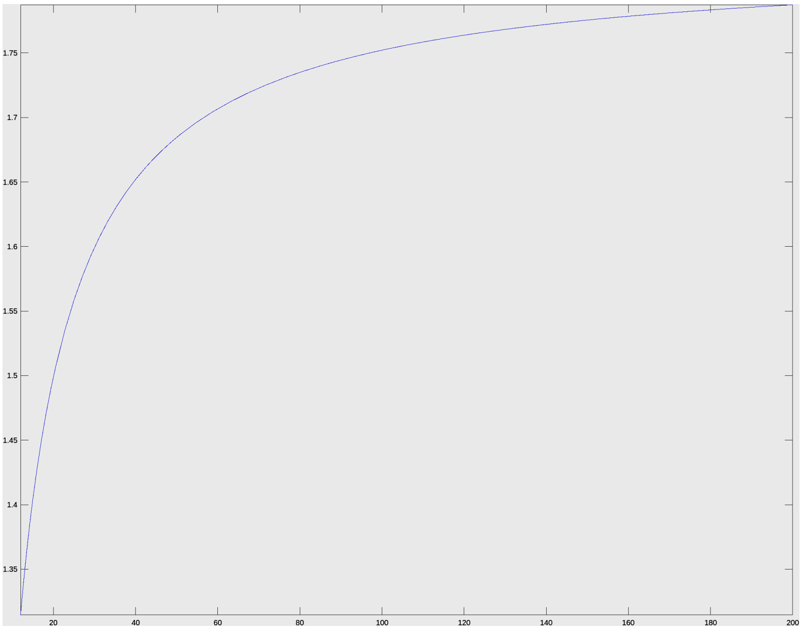


**Fig. 4**: Graph of $\Psi_3\left(1, \frac{R}{\varepsilon}\right)$.

*3. For a vertex $v$ and a triangle $\triangle uwx$ with $v \neq u, w, x$,*

$$d(\triangle uwx, v) > 6(R+\varepsilon).$$

*4. For an edge $\overline{uv}$ and a triangle $\triangle wxy$ with $v, u \neq w, x, y$,*

$$d(\triangle wxy, \overline{uv}) > 6(R+\varepsilon).$$

*5. For any triangle $\triangle uvw$,*

$$\angle uvw, \angle vwu, \angle wuv \geq \frac{\pi}{6}.$$

*6. For any pair of edges $\overline{uv}, \overline{xy}$ with no common vertex,*

$$d(\overline{uv}, \overline{xy}) > 6(R+\varepsilon).$$

*7. For any triangles $\triangle uwv, \triangle xyz$,*

$$d(\triangle uwv, \triangle xyz) > 6(R+\varepsilon).$$

*8. For any pair of edges $\overline{uv}, \overline{wv}$,*

$$\angle uvw \geq \Psi_1(\varepsilon, R).$$

9. *For all degree 2 vertices $v$ with edges $\overline{uv}, \overline{wv}$ and no triangle $\triangle uvw$,*

$$\angle uvw \leq \Psi_2(\varepsilon, R).$$

10. *For any pair of triangles $\triangle uvw_1, \triangle uvw_2$, the dihedral angle between them is bounded below by $\Psi_1(\varepsilon, R)$.*
11. *For any pair of triangles $\triangle uvw_1, \triangle uvw_2$, with $\overline{uv}$ of degree 2, the dihedral angle between them is bounded above by $\Psi_2(\varepsilon, R)$.*
12. *For any triangle $\triangle wwvw_2$ and edge $\overline{uv}$ the angle between $\overline{uv}$ and and ray $L$ in $\triangle w_1vw_2$ at $v$ is bounded below by $\Psi_1(\varepsilon, R)$ and the radius of the largest circle inscribed by $\triangle uvw$ is at least $2R + 3\varepsilon$.*
13. *For any vertex $v$ such that*

$$|H_0\left(B_R(v) \cap |X|\right)| = 1, \text{ and } |H_1\left(B_R(v) \cap |X|\right)| = 1,$$

*the angle between any two rays $L_1, L_2 \in |X|$ at $v$ is bounded above $\Psi_3(\varepsilon, R)$.*

**Remark 4.** *The reasons behind some of the conditions in Assumption 1 are relatively clear, while others are a bit more obscure. In particular, the roles of conditions 11 and 12 are not immediately clear. Condition 12 allows us to detect the vertex $v$ in our algorithms. In particular, it is used in Proposition 4.11 show that we obtain $\mathrm{Sig}_{\varepsilon,R} = (n, \bullet),\ n \geq 2$. Condition 13 allows us to detect which topologically looks similar to an edge of degree 2 or a triangle, and so we place restrictions on the formation of the* cone*, potentially with* fins*, so that we can detect the vertex (Proposition 4.11). This condition is equivalent to bounding the angle at $v$ of the convex hull which contains the triangles with vertex $v$.*

The following Propositions provide us with ‘regions’ near locally maximal $i$-cells $\sigma$ (for $i$ = 0, 1, 2), where we can guarantee that at any sample in this region, the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension $i$.

We begin with the region around a locally maximal vertex.

**Proposition 4.8.** *Let $v$ be a vertex of $|X| \subset \mathbb{R}^n$, which is locally maximal, and let $P$ be an $\varepsilon$-sample of $|X|$. Then, for all $p \in P$ with $\|p - v\| \leq 4\varepsilon$, the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 0.*

*Proof.* As $v$ is locally maximal, it is not in the boundary of any other cell, and from Assumption 1 for all vertices $u \neq v$, $\|u - v\| > 6(R+\varepsilon)$, for all edges $\overline{uw}$ with $v \neq u, w$,

$$d(\overline{uv}, v) > 6(R + \varepsilon),$$

and for all triangles $\triangle uwx$ with $v \neq u, w, x$,

$$d(\triangle uwx, v) > 6(R + \varepsilon).$$

Hence, any sample $p \in P$ within $4\varepsilon$ of $v$ is within $\varepsilon$ of $v$. Thus, $(B_{R+\varepsilon}(p) \cap P)^{\frac{3\varepsilon}{2}}$ consists of a single connected component, and $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P = \emptyset$.

Thus, $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$, $\mathrm{Sig}_{\varepsilon,R}(p) = (0, 0)$, and the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 0. ∎

Next, we bound the region near a locally maximal edge.

**Proposition 4.9.** *Let $\overline{uv}$ be an edge of $|X|\subset \mathbb{R}^n$, which is locally maximal, and let $P$ be an $\varepsilon$-sample of $|X|$. Then, for all $p \in P$ with $d(\overline{uv}, p) \leq \varepsilon$, and $\|p-u\|, \|p-v\| \geq \frac{3R}{2}+\varepsilon$, the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 1.*

*Proof.* By Assumption 1, for any vertex $w \neq u, v$

$$d(\overline{uv}, w) > 6(R + \varepsilon),$$

for any edge $\overline{wx}$, with $w, x \neq u, v$,

$$d(\overline{uv}, \overline{wx}) > 6(R + \varepsilon),$$

for any triangle $\triangle wxy$, with $w, x, y \neq u, v$,

$$d(\triangle wxy, \overline{uv}) > 6(R + \varepsilon),$$

and so the connected component $C_p^{\frac{3\varepsilon}{2}}$ of $(B_{R+\varepsilon}(p) \cap P)^{\frac{3\varepsilon}{2}}$ which contains $p$, contains only points $q \in P$ with $d(q, \overline{uv}) \leq \varepsilon$.

Hence, $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)$ consists of two connected components, $c_1$ and $c_2$. By Lemma 3.4, the diameters of $c_1$ and $c_2$ are less than $5\varepsilon$. Let $x_1$ and $x_2$ be the centroids of $c_1$ and $c_2$. Then, applying Lemma 2.1 in Bokor et al. (2021),

$$\langle x_1 - p, x_2 - p\rangle \leq -R^2 + 2R\varepsilon + 7\varepsilon^2,$$

so the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 1. □

Finally, we bound the region near (locally maximal) triangles.

**Proposition 4.10.** *Let $\triangle uvw$ be an triangle of $|X|\subset \mathbb{R}^n$, and let $P$ be an $\varepsilon$-sample of $|X|$. Then, for all $p \in P$ with $d(\triangle uvw, p) \leq \varepsilon$, and $d(\partial\triangle uvw, p) \geq \frac{3R}{2} + \varepsilon$, the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 2.*

*Proof.* From Assumption 1, for all triangles $\triangle xyz$, with $x, y, z \neq u, v, w$,

$$d(\triangle uwv, \triangle xyz) > 6(R + \varepsilon),$$

and hence the connected component $C_p^{\frac{3\varepsilon}{2}}$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(B_{R+\varepsilon}(p) \cap P)$ containing $p$, consists only of samples $q \in P$ with $d(q, \triangle uvw) \leq \varepsilon$, as the angle between triangles is bounded below (Assumption 1).

First, we need to show that $\mathrm{Sig}_{\varepsilon,R}(p) = (1, 1)$, after which Lemma 3.2 implies that for all $q_1 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$, there exists $q_2 \in S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$ such that

$$\|q_2 - q_1\| \geq 2\sqrt{R^2 - \varepsilon^2} - (1 + \sqrt{2})\varepsilon.$$

As $d(\partial\triangle uvw, p) > \frac{3R}{2} + \varepsilon$, we have the following inclusions

$$\begin{aligned} S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{3\varepsilon}{2}} \\ &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{5\varepsilon}{2}} \\ &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{7\varepsilon}{2}} \\ &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{9\varepsilon}{2}}. \end{aligned}$$

By the bounds in Assumption 1 on the distances between a triangle and cells not in its boundary, the weak feature size of $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw$ is greater than $5\varepsilon$, and so the inclusion maps induce isomorphisms

$$H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right) \cong H_\bullet\left(\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{5\varepsilon}{2}}\right) \cong H_\bullet\left(\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{9\varepsilon}{2}}\right).$$

The above homology factors through $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{3\varepsilon}{2}}$ and $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{7\varepsilon}{2}}$ so we have

$$\mathrm{rk}_\bullet^{\frac{3\varepsilon}{2},\frac{5\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right) = \left|H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)\right|,$$

and as

$$|H_0\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)| = 1, |H_1\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)| = 1,$$

it follows that $\mathrm{Sig}_{\varepsilon,R}(p) = (1,1)$. Now we apply Lemma 3.2 and conclude that the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 2. □

Now, we obtain the regions around not locally maximal $i$-cells $\sigma$ ($i = 0$, 1) in which we can guarantee that the $(\varepsilon, R)$-local structure of $P$ at a sample $p$ in this region is not locally maximal. Again, we begin with non-locally maximal vertices.

**Remark 5.** *As we have restricted our considerations to* 2*-complexes, every triangle* $\sigma$ *is locally maximal; hence, we need only to consider vertices and edges that are not locally maximal.*

**Proposition 4.11.** *Let* $v$ *be a vertex of* $|X| \subset \mathbb{R}^n$*, which is not locally maximal, and let* $P$ *be an* $\varepsilon$*-sample of* $|X|$*. Then, for all* $p \in P$ *with*

$$\|p - v\| \leq \frac{R}{2} - 2\varepsilon,$$

*the* $(\varepsilon, R)$*-local structure of* $P$ *at* $p$ *is not maximal.*

*Proof.* There are several cases we need to consider, which we can classify by the homology of $\partial B_R(v) \cap |X|$:

1. $|H_0(\partial B_R(v) \cap |X|)| = n$, $|H_1(\partial B_R(v) \cap |X|)| = 0$, $n \neq 2$,
2. $|H_0(\partial B_R(v) \cap |X|)| = 2$, $|H_1(\partial B_R(v) \cap |X|)| = 0$,

3. $|H_0(\partial B_R(v) \cap |X|)| = 1$, $|H_1(\partial B_R(v) \cap |X|)| = 1$,
4. $|H_0(\partial B_R(v) \cap |X|)| = 1$, $|H_1(\partial B_R(v) \cap |X|)| = n$, $n \geq 2$.

In each of these cases, the following argument holds. Let $C_p$ be the connected component of $B_{R+\varepsilon}(p) \cap |X|$ which contains the projection of $p$ to $|X|$, and let $C_p^{\frac{3\varepsilon}{2}}$ be the connected component of $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{3\varepsilon}{2}}$. As $P$ is a $\varepsilon$-sample of $|X|$, we have the following inclusions

$$\begin{aligned} S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{3\varepsilon}{2}} \\ &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{5\varepsilon}{2}} \\ &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{7\varepsilon}{2}} \\ &\hookrightarrow \left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap \triangle uvw\right)^{\frac{9\varepsilon}{2}}. \end{aligned}$$

By the bounds in Assumption 1 on

- the angle betwen edges at a common vertex,
- the distance between vertices,
- the angles between triangles with a common vertex or edge,
- the distance between vertices and cells they do not intersect with,

the weak feature size of $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}$ is greater than $5\varepsilon$, and we have the following isomorphism on homology induced by the inclusions above

$$H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right) \cong H_\bullet\left(\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right)^{\frac{5\varepsilon}{2}}\right) \cong H_\bullet\left(\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right)^{\frac{9\varepsilon}{2}}\right).$$

The above homology factors through $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{3\varepsilon}{2}}$ and $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{7\varepsilon}{2}}$ so we have

$$\mathrm{rk}_\bullet^{\frac{3\varepsilon}{2},\frac{7\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right) = \left|H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right)\right|.$$

As $\|p - v\| \leq \frac{R}{2} - 2\varepsilon$, we have

$$\left|H_\bullet\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|\right)\right| = |H_\bullet(\partial B_R(v) \cap |X|)|,$$

giving

$$\mathrm{rk}_\bullet^{\frac{3\varepsilon}{2},\frac{7\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right) = |H_\bullet(\partial B_R(v) \cap |X|)|.$$

**Case 1:** $|H_0(\partial B_R(v) \cap |X|)| = n$, $|H_1(\partial B_R(v) \cap |X|)| = 0$, $n \neq 2$

By the above, we have $\mathrm{Sig}(p) = (n, 0)$, $n \neq 2$, and so the $(\varepsilon, R)$-local structure of $P$ at $p$ is not maximal.

**Case 2:** $|H_0(\partial B_R(v) \cap |X|)| = 2$, $|H_1(\partial B_R(v) \cap |X|)| = 0$

By the above, we have $\mathrm{Sig}(p) = (2,0)$. Let $C_p^{2\varepsilon}$ be the connected component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)$ containing $p$.

Assume that $v$ is a face of some triangle $\triangle uvw$. Then by the bounds placed on angles between edges, and distances between edges without a common face, edges and vertices which are not faces, and vertices and triangles they are not a face of (see Assumption 1), and Lemma 3.5 at least one connected component in $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap C_p^{\frac{3\varepsilon}{2}}\right)^{\frac{3\varepsilon}{2}}$ has a diameter greater than $2\sqrt{2}\varepsilon$. Thus, the $(\varepsilon, R)$-local structure of $P$ at $p$ is not maximal.

If $v$ is only the face of edges, then by the bounds placed on angles between edges, and distances between edges without a common face, edges and vertices which are not faces, and vertices and triangles they are not a face of (see Assumption 1), both connected components come from two edges $uv$ and $wv$, Lemma 2.1, in Bokor et al. (2021) and Lemma 3.4 give that the $(\varepsilon, R)$-local structure of $P$ at $p$ is not maximal.

**Case 3:** $|H_0\left(\partial B_R(v) \cap |X|\right)| = 1$ $|H_1\left(\partial B_R(v) \cap |X|\right)| = 1$

Again, we have $\mathrm{Sig}(p) = (1,1)$ so there are at least three triangles having $v$ as a common vertex. Let $p_X$ be the closest point in $|X|$ to $p$, and let $x_1 \in \partial B_R(p) \cap |X|$ be colinear with $v$ and $p_X$, then there is $q_1 \in S_{R-\varepsilon}^{R+\varepsilon} \cap P$ with $\|q_1 - x_1\| \leq \varepsilon$.

Now take any $q_2 \in S_{R-\varepsilon}^{R+\varepsilon} \cap P$, and let $x_2$ be the point in $|X| \cap \partial B_R(p)$ closest to $q_2$. Then from Lemma 3.1

$$\|q_2 - x_2\| \leq \sqrt{2}\varepsilon.$$

Consider the rays $L_1, L_2$ from $v$ through $x_1, x_2$ respectively, and assume $d(p, L_1) \leq \varepsilon$.

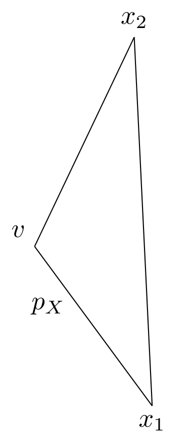


**Fig. 5**: $d(q_2, H_2) \leq \varepsilon$

We have

$$\|x_1 - v\| = \|x_1 - p_X\| + \|p_X - v\| \leq \frac{3R}{2} - 2\varepsilon,$$

$$\|x_2 - v\| \leq R + \varepsilon,$$

and so

$$\begin{aligned}\|x_2 - x_1\| &= \|x_2 - v\|^2 + \|x_1 - v\|^2 - 2\|x_2 - v\|\|x_1 - v\|^2 \cos \angle x_1 v x_2 \\ &\leq \left(\frac{3R}{2} - 2\varepsilon\right)^2 + (R+\varepsilon)^2 - \left(\frac{3R}{2} - 2\varepsilon\right)(R+\varepsilon)\cos x_1 v x_2.\end{aligned}$$

By condition 13 in Assumption 1 the angle between them is bounded above by $\Psi_3(\varepsilon, R)$, so

$$\|x_2 - x_1\| \leq 2\sqrt{R^2 - \varepsilon^2} - (1+\sqrt{2})\varepsilon,$$

and so

$$\|q_2 - q_1\| \leq 2\sqrt{R^2 - \varepsilon^2} - (2 + 2\sqrt{2})\varepsilon.$$

Thus, the $(\varepsilon, R)$-local structure of $P$ at $p$ is not maximal.

**Case 4:** $|H_0(\partial B_R(v) \cap |X|)| = 1$, $|H_1(\partial B_R(v) \cap |X|)| = n$, $n \geq 2$

By the argument at the start of this proof, $\mathrm{Sig}(p) = (1, n)$, $n \geq 2$ and so the $(\varepsilon, R)$-local structure of $P$ at $p$ is not maximal. □

Next, we bound the region near edges that are not locally maximal.

**Proposition 4.12.** *Let $\overline{uv}$ be an edge of $|X| \subset \mathbb{R}^n$, which is not locally maximal, and let $P$ be an $\varepsilon$-sample of $|X|$. Then, for all $p \in P$ with $d(\overline{uv}, p) \leq \frac{R}{2} - 2\varepsilon$, the $(\varepsilon, R)$-local structure of $P$ at $p$ is not maximal.*

*Proof.* If an edge $\overline{uv}$ is not locally maximal, then there is at least one triangle $\triangle uvw$.

We consider 3 cases:

1. there is a unique triangle $\triangle uvw$ with $\overline{uv}$ in the boundary,
2. there are exactly two triangles $\triangle uvw_1$ and $\triangle uvw_2$ with $\overline{uv}$ in their boundaries,
3. there are three or more triangles $\triangle uvw_1, \triangle uvw_2$ and $\triangle uvw_3$ with $\overline{uv}$ in their boundaries.

Recall that we restrict our attention to the connected components $C_p, C_p^{\frac{3\varepsilon}{2}}$ of $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap |X|$ and $\left(S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P\right)^{\frac{3\varepsilon}{2}}$ which contains $p$.

By the bounds in Assumption 1 on

- the angle betwen edges at a common vertex,
- the distance between edges that do not have a common face,
- the angles between triangles with a common edge,
- the distance between edges and cells they do not intersect with,

the weak feature size of $C_p$ is greater than $5\varepsilon$. Hence by the same argument as at the start of the poof of Proposition 4.11,

$$\mathrm{Sig}_{\varepsilon,R}(p) = \left(|H_0(\partial B_R(m) \cap |X|)|, |H_1(\partial B_R(m) \cap |X|)|\right).$$

Thus, in cases 1 and 3, we get $\text{Sig}(p) = (1, 0)$ and $\text{Sig}(p) = (1, n)$ for $n \geq 3$ respectively.

In case 2, we get $\text{Sig}(p) = (1, 1)$, and so need to check the geometric condition. By Lemma 3.3, there is a $q_1 \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$ such that for all $q_2 \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - (1 + \sqrt{2})\varepsilon,$$

and so the $(\varepsilon, R)$-local structure of $P$ at $p$ is not maximal.

Hence, in all 3 cases, the $(\varepsilon, R)$ local structure of $P$ at $p$ is not maximal. □

# 5 2-Complex Algorithm and Correctness

In this section, we present a set of algorithms, which together, recover the structure of $X$ from an $\varepsilon$-sample $P$ of an embedding $(X, \Theta) \subset \mathbb{R}^n$. Theorem 5.25 states that given an $\varepsilon$-sample $P$ of an embedded 2 complex $|X| = (X, \Theta_X) \subset \mathbb{R}^n$ satisfying Assumption 1, we can recover the structure of $X$ using this algorithm. There is a sequence of lemmas (Lemmas 5.9 to 5.24), which culminates in the 'big theorem' (Theorem 5.25). The proofs of the lemmas are in Appendix B.

The algorithm partitions $P$ into $P_{LM}$ and $P_{NLM}$, such that for each $p \in P_{LM}$ the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal, and for each $p \in P_{NLM}$ the $(\varepsilon, R)$-local structure of $P$ at $p$ is not maximal. We then detect the number of vertices, the number of edges, the number of triangles and the incidence operator. To obtain $P_{LM}$ and $P_{NLM}$, we use

$$\Delta_{\varepsilon,R} : P \to \{0, 1\},$$

see Algorithm 1.

Let $\mathcal{C}_p$ be the samples $q \in P$ in the connected component containing $p$ in the threshold graph

$$\mathcal{G}_p = \mathfrak{G}_{3\varepsilon}\left(B_{R+\varepsilon}(p) \cap P\right)$$

with $\|q - p\| \in [R - \varepsilon, R + \varepsilon]$. In the definitions of $(\varepsilon, R)$-local structure (Definitions 4.4 and 4.5), we used

$$\text{rk}_{\bullet}^{\frac{3\varepsilon}{2}, \frac{7\varepsilon}{2}}\left(S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P\right),$$

which by the Nerve Lemma (Corollary 4G.3 Hatcher (2000)) is equal to the rank, $\mathcal{RK}_{\bullet}$, of the map

$$H_{\bullet}\left(\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S^{R+\varepsilon}_{R-\varepsilon}(p) \cap \mathcal{C}_p\right)\right) \to H_{\bullet}\left(\check{\mathcal{C}}_{\frac{7\varepsilon}{2}}\left(S^{R+\varepsilon}_{R-\varepsilon}(p) \cap \mathcal{C}_p\right)\right)$$

induced by the inclusion

$$\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P\right) \hookrightarrow \check{\mathcal{C}}_{\frac{7\varepsilon}{2}}\left(S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P\right).$$

Hence, $\Delta_{\varepsilon,R}(p)$ returns 0 if the $(\varepsilon, R)$-local structure of $P$ at $P$ is not maximal, and returns 1 if it is maximal. Then,

$$P_{NLM} = \Delta^{-1}_{\varepsilon,R}(0)$$

and

$$P_{NLM} = \Delta_{\varepsilon,R}^{-1}(1).$$

**Remark 6.** *We can appeal to the Nerve Lemma, as the balls used in the construction of* $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p)\cap\mathcal{C}_p\right)$ *and* $\check{\mathcal{C}}_{\frac{7\varepsilon}{2}}\left(S_{R-\varepsilon}^{R+\varepsilon}(p)\cap\mathcal{C}_p\right)$ *lead us to* good covers *of* $\left(S_{R-\varepsilon}^{R+\varepsilon}(p)\cap P\right)^{\frac{3\varepsilon}{2}}$ *and* $\left(S_{R-\varepsilon}^{R+\varepsilon}(p)\cap P\right)^{\frac{7\varepsilon}{2}}$ *respectively. To see that these covers satisfy the 'every non-empty intersection is contractible' condition required to be a good cover, note that we are using the Čhech complex, rather than the Viertoris-Rips complex. Combining this with the linearity of the embedding and the assumptions placed on both* $\varepsilon$ *and* $R$*, we have covers that satisfy the Nerve Lemma.*

---

**Algorithm 1:** $\Delta_{\varepsilon,R}(p)$

---

**Data:** An $\varepsilon$-dense sample $P$ of an embedded 2-complex $|X|$, a point $p \in P$.
**Result:** 0 if the $(\varepsilon, R)$-local structure of $P$ at $p$ is not maximal,
1 if the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal.

```
begin
    G_p ⟵ {q ∈ P | ||p − q|| ≤ R + ε};
    connect q, q' ∈ G_p if ||q − q'|| ≤ 3ε;
    C_p ⟵ {q ∈ G_p | q is path connected to p in G_p};
    remove q ∈ C_p if ||p − q|| ≥ R − ε;
    if RK_0 = 0 and RK_1 = 0 then
        return 1
    else if RK_0 = 1 and RK_1 ≠ 1 then
        return 0
    else if RK_0 = 1 and RK_1 = 1 then
        if ∀q_1, q_2 ∈ C_p, ∃q_0 such that
          ||q_1 − q_0||, ||q_2 − q_0||, ||q_2 − q_1|| ∈ [√(3(R^2 − ε^2)), √3 R] then
            return 1
        else
            return 0
    else if RK_0 = 2 and RK_1 = 0 then
        if max {D(c_1), D(c_2)} ≤ 5ε then
            if ⟨q_1 − p, q_2 − p⟩ > −R^2 + 2Rε − 7ε^2 then
                return 1
            else
                return 0
        else
            return 0
    else if RK_0 = n, n ≠ 0, 1, 2 and RK_1 = 0 then
        return 0
```

---

After we have $P_{LM}$, we use the function

$$\mathfrak{D}_{\varepsilon,R}(p) : P_{LM} \to \{\, 0,\ 1,\ 2\,\},$$

see Algorithm 2 to determine what dimension of $(\varepsilon, R)$-local structure each sample in $P_{LM}$ has.

---

**Algorithm 2:** $\mathfrak{D}_{\varepsilon,R}(p)$

---

**Data:** An $\varepsilon$-dense sample $P$ of an embedded 2-complex $|X|$, a point $p \in P$ such that the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal.
**Result:** 0 if the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 0,
1 if the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 1,
2 if the $(\varepsilon, R)$-local structure of $P$ at $p$ is maximal of dimension 2.

```
begin
    G_p ⟵ {q ∈ P | ‖p − q‖ ≤ R + ε};
    connect q, q′ ∈ G_p if ‖q − q′‖ ≤ 3ε;
    C_p ⟵ {q ∈ G_p | q is path connected to p in G_p};
    remove q ∈ C_p if ‖p − q‖ ≤ R − ε;
    if RK_0 = 0 and RK_1 = 0 then
        return 0
    else if RK_0 = 2, n ≠ 0,1,2 and RK_1 = 0 then
        return 1
    else if RK_0 = 1, n ≠ 0,1,2 and RK_1 = 1 then
        return 2
```

---

Recall that our end goal is to learn the combinatorial structure of $X$. We begin by learning the number of triangles, locally maximal edges, and locally maximal vertices. Consider the following three subsets of $P_{LM}$:

$$P_{LM,2} = \{p \in P_{LM} \mid \mathfrak{D}_{\varepsilon,R}(p) = 2\},$$
$$P_{LM,1} = \{p \in P_{LM} \mid \mathfrak{D}_{\varepsilon,R}(p) = 1\},$$
$$P_{LM,0} = \{p \in P_{LM} \mid \mathfrak{D}_{\varepsilon,R}(p) = 0\}.$$

When partitioning $P$ into $P_{LM}$ and $P_{NLM}$, there is a *grey* region where a sample $p$ could be in either of these two sets. This presents a problem for learning the combinatorics of $X$ from the partitioning $P_{LM}$ and $P_{NLM}$. We can overcome this, by *cleaning* $P_{LM}$. In particular, we clean $P_{LM,2}$ and $P_{LM,1}$.

We begin by introducing the notion of a connected component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$ *spanning* an edge, and then introduce the notion of a connected component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$ *spanning* a triangle.

**Definition 5.1** (Spanning an edge)**.** *We say a connected component of* $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$ spans *a locally maximal edge* $\overline{uv}$ *if it contains a sample* $p$ *within* $\varepsilon$ *of the midpoint of* $\overline{uv}$*.*

**Definition 5.2** (Spanning a triangle)**.** *We say a connected component of* $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$ spans *a triangle* $\triangle uvw$ *if it contains a sample* $p$ *within* $\varepsilon$ *of the midpoint of* $\triangle uvw$*.*

We require some geometric conditions on when a connected component spans an edge or a triangle. For an edge, we will use the diameter of the connected component as a condition.

**Proposition 5.3.** *A connected component $C$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$* spans *a locally maximal edge $\overline{uv}$ if and only if $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$.*

Unfortunately, it is not immediately clear that such a test is suitable for detecting components that span triangles. For instance, consider a complex which consists of a single triangle, its three edges, and the three required vertices. While heuristically, it is unlikely to occur, the sampling could lead to 2 connected components $C_1, C_2 \in \check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$: one which is far away from the boundary of the triangle, and one that is *surrounded* by points in $P_{NLM}$, both with large diameters. In fact, the one we wish to say is spanning, say $C_1$, will have a smaller diameter than the other one, $C_2$. Note, however, that as $C_2$ does not contain a sample $p$ near the midpoint of $\triangle uvw$, if $\mathcal{D}(C_1) \leq \mathcal{D}(C_2)$, then $C_2$ contains a non-contractible loop. However, a sample $p \in P$ near the midpoint $m_{\triangle uvw}$ of a triangle $\triangle uvw$ is not near any samples $q \notin P_{LM,2}$, and so we can exploit this fact to obtain a geometric test.

**Proposition 5.4.** *A connected component $C$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}$* spans *a triangle $\triangle uvw$ if and only if there is a point $p \in C$ such that*

$$B_{\frac{R}{2}+\varepsilon}(p) \cap P \subset P_{LM,2}.$$

We now have geometric conditions for determining if a connected component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})/\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$ spans a triangle/edge respectively. Next, show that the locally maximal vertices of $X$ are in bijection with connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,0})$, the locally maximal edges of $X$ are in bijection with the spanning connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$, and that the triangles of $X$ are in bijection with the spanning connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$.

We begin with the locally maximal vertices.

**Proposition 5.5.** *The connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,0})$ are in bijection with the set $V_{LM}$ of locally maximal vertices of $X$.*

Next, we show that the edge spanning components are in bijection with the locally maximal edges.

**Proposition 5.6.** *The* spanning *components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$ are in bijection with the set $E_{LM}$ of locally maximal edges of $X$.*

Finally, we show that the spanning components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$ are in bijection with the triangles of $X$.

**Proposition 5.7.** *The* spanning *components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$ are in bijection with the set $T$ of triangles in $X$.*

Having identified the locally maximal cells $X_{LM}$ of $X$, we could learn the combinatorial structure of $X$ by identifying the structure of $X_{NLM}$ from $P_{NLM}$, and combining this with what we know about $X_{LM}$ from $P_{LM}$. The process in Bokor et al. (2021) could be applied, but this requires the existence of some $\widetilde{\varepsilon}$ such that $P_{NLM}$ is a $\widetilde{\varepsilon}$-sample of $X_{NLM}$ satisfying Assumptions 1 in Bokor et al. (2021) This would impose stricter assumptions than Assumption 1, but after ensuring these new assumptions are satisfied, works out of the box.

To avoid placing stricter assumptions on $|X|$, we use the idea of *witness points* to discover the combinatorics. For each sample $p \in P_{NLM}$, we can examine the spanning connected components $C_{LM}$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$ and $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$ such that $C_{LM} \cap B_{R+3\varepsilon}(p) \neq \emptyset$. In particular, we can use $\mathfrak{D}_{\varepsilon,R}(q)$ for some $q \in C_{LM}$, to determine of what dimension the local structure is maximal. If there is a $q$ in $C_{LM} \cap S_{R-\varepsilon}^{R+\varepsilon}(p)$ such that $\mathfrak{D}_{\varepsilon,R}(q) = 1$, then $p$ is near a vertex.

If there are no connected components $C_{LM}$ which are $(\varepsilon, R)$-locally maximal of dimension 1, then $p$ only *witnesses* samples $q \in P_{LM}$ such that the $(\varepsilon, R)$-local structure of $P$ at $q$ is maximal of dimension 2. Hence, we need to understand the combinatorics of $|X| \setminus (E_{LM} \cup V_{LM})$ where $E_{LM}$ is the set of locally maximal edges and $V_{LM}$ the set of locally maximal vertices.

In Assumption 1, we assumed that for any triangle $\triangle uvw$,

$$\angle uvw, \angle vwu, \angle wuv \geq \frac{\pi}{6}.$$

This means that for any sample $p \in P_{NLM}$ with $d(\partial\triangle uvw, p) < R+\varepsilon$ for some $\triangle uvw$, there is some sample $q \in P_{LM,2}$ with $d(\triangle uvw, q) \leq \varepsilon$ and $d(\partial\triangle uvw, p) \geq R+\varepsilon$, such that $\|q-p\| \leq \frac{2\sqrt{2}(R+2\varepsilon)}{\sqrt{3}-1}$. Further, $q$ is in a triangle spanning component $\mathcal{T}$.

Similarly, for any sample $p \in P_{NLM}$ with $d(\partial\overline{uv}, p) < \frac{3R}{2} + \varepsilon$ for some edge $\overline{uv}$, there is a sample $q \in P_{LM,1}$ with $d(\partial\overline{uv}, p) \geq \frac{3R}{2} + \varepsilon$ such that $\|q-p\| \leq \frac{2\sqrt{2}(R+2\varepsilon)}{\sqrt{3}-1}$. Further, $q$ is in an edge spanning component $\mathcal{E}$.

This leads us to say a sample $p \in P_{NLM}$ *witnesses* a spanning connected component $\mathcal{C}$ if

$$B_{\frac{2\sqrt{2}(R+2\varepsilon)}{\sqrt{3}-1}}(p) \cap \mathcal{C} \neq \emptyset.$$

For ease of reading, we set $\kappa = \frac{2\sqrt{2}}{\sqrt{3}-1}$.

**Definition 5.8** (Witnessing a spanning component)**.** *Let $P$ be an $\varepsilon$-sample $P$ of an embedded 2-complex $|X|$ satisfying Assumption 1. Then a sample $p \in P_{NLM}$* witnesses *an edge/triangle spanning component if*

$$B_{\kappa(R+\varepsilon)}(p) \cap \mathcal{C} \neq \emptyset.$$

To determine the final combinatorial structure of $X$, we look at the local neighbourhood of each $p \in P_{NLM}$ and look at both

$$B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$$
$$B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2}).$$

If

$$B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1}) \neq \emptyset$$

then we know that $p$ is *near* a vertex, and the spanning components $\mathcal{E}$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$ that $p$ witnesses, share a boundary vertex. Further, if

$$B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2}) \neq \emptyset$$

as well, then there are spanning components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$ that $p$ witnesses, which have a vertex in common with the edges.

If only

$$B_{(R+2\varepsilon)\kappa}(p) \cap \check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2}) \neq \emptyset$$

we examine how many spanning components $\mathcal{T}$ are seen by $p$, as well as if samples $p \in P_{NLM}$ that witness $\mathcal{T}$, also witness any other spanning components $\mathcal{T}'$. We use this information to partition $P_{NLM}$ into $\{P_i\}$ in Algorithm 5, with a final clean of the partitions, to account for some special cases. As $R \leq 16\varepsilon$, for all $p \in P_{NLM}$ there is some spanning connected component $\mathcal{C}$ such that $B_{\frac{R+\varepsilon}{\kappa}}(p) \cap \mathcal{C} \neq \emptyset$.

We then label each component $P_i$ as follows, from Algorithms 7 and 8:

- $-1$ if $P_i$ corresponds to 2 vertices,
- 0 if $P_i$ corresponds to a vertex,
- 1 if $P_i$ corresponds to a vertex and an edge,
- 2 if $P_i$ corresponds to two vertices and an edge,
- 3 if $P_i$ corresponds to just an edge,
- 4 if $P_i$ corresponds to two edges and a vertex,
- 5 if $P_i$ corresponds to three edges and two vertices,
- 6 if $P_i$ corresponds to three edges and a vertex,
- 7 if $P_i$ corresponds to three edges and three vertices,
- 8 if $P_i$ corresponds to three edges,
- 9 if $P_i$ corresponds to two edges,

---

**Algorithm 3:** Spanning triangle components

---

**Data:** Parameters $\varepsilon, R$ and $P_{LM,1}$.
**Result:** The set of triangle spanning components.
**begin**
  Initialise empty set $T$;
  Let $C$ be the set of connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$;
  **for** $\mathcal{T} \in C$ **do**
    **if** $\exists p \in \mathcal{C}$ *such that* $B_{R/2+\varepsilon}(p) \cap P \subset P_{LM,2}$ **then**
      Add $\mathcal{T}$ to $T$;
  **return** $T$

---

**Algorithm 4:** Spanning edge components

**Data:** Parameters $\varepsilon, R$ and $P_{LM,1}$.
**Result:** The set of triangle spanning components.
**begin**
    Initialise empty set $E$;
    Let $C$ be the set of connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$;
    **for** $\mathcal{E} \in C$ **do**
        **if** $\mathcal{D}(\mathcal{T}) \geq \frac{3R}{2} - 2\varepsilon$ **then**
            Add $\mathcal{E}$ to $E$;
    **return** $E$

**Algorithm 5:** Partitioning $P_{NLM}$

**Data:** An $\varepsilon$-dense sample $P$ of an embedded 2-complex $|X|$, partitioned into $P_{NLM}$, $P_{LM,0}$, $P_{LM,1}$, $P_{LM,2}$.
**Result:** A partition $\{P_i\}$ of $P_{NLM}$, and for each $P_i$, two sets $S_E(P_i)$, $S_T(P_i)$.
**begin**
    For each $p \in P_{NLM}$, find all the edge spanning components $\mathcal{E}$ such that $\mathcal{E} \cap B_{(R+2\varepsilon)\kappa}(p) \neq \emptyset$, and place them in $S_E(p)$;
    Find all the triangle spanning components $\mathcal{T}$ such that $\mathcal{T} \cap B_{(R+2\varepsilon)\kappa}(p) \neq \emptyset$, and place them in $S_T(p)$;
    Partition $P_{NLM}$ into $\{ P_i \}$ such that for each $p, q \in P_i$, $S_E(p) = S_E(q)$ and $S_T(p) = S_T(q)$;
    Assign $S_E(P_i)$ and $S_T(P_i)$ to each $P_i$;
    **for** $P_i$ *and* $P_j$ *with* $S_E(P_j) \subseteq S_E(P_i)$ *and* $S_T(P_j) \subseteq S_T(P_i)$ **do**
        **if** $S_E(P_j), S_T(P_j) \neq \emptyset$ **then**
            Merge $P_j$ into $P_i$ with labels $S_E(P_i)$, $S_T(P_i)$;
        **else if** $|S_T(P_j)| \geq 2$ *and* $\forall p \in P_j$ *such that* $\mathrm{Sig}_{\varepsilon,R}(p) = (n, 0)$, $n \in \mathbb{Z}_{\geq 0}$ **then**
            Merge $P_j$ into $P_i$ with labels $S_E(P_i)$, $S_T(P_i)$;
    **return** $\{P_i\}$, and $S_E(P_i)$, $S_T(P_i)$ for each $P_i$

**Algorithm 6:** Order $\{P_i\}$

**Data:** An $\varepsilon$-dense sample $P$ of an embedded 2-complex $|X|$, partition $\{P_i\}$ of $P_{NLM}$ with two sets $S_E(P_i)$, $S_T(P_i)$ for each $P_i$ and partitions of $P_{LM,0}$, $P_{LM,1}$, $P_{LM,2}$.
**Result:** Two sets $P^1, P^2 \subset \{ P_i \}$.

```
begin
    Initialise empty P^1 and P^2;
    for P_i ∈ {P_i} do
        if S_E(P_i) ≠ ∅ then
            Add P_i to P^1
        else if ∃p ∈ P_i such that Sig(p) ≠ (1, n) then
            Add P_i to P^1
        else if |S_T(P_i)| ≠ 1 then
            Add P_i to P^1
        else
            Add P_i to P^2
    return P^1, P^2
```

**Algorithm 9:** Number of triangles, edges and vertices.

**Data:** An $\varepsilon$-dense sample $P$ of an embedded 2-complex $|X|$, partitions of $P_{NLM}$, $P_{LM,0}$, $P_{LM,1}$, $P_{LM,2}$ and the labelled list $C$ from Algorithm 8.
**Result:** The triangles, edges, and vertices in $X$.

```
begin
    Initialise an empty weighted graph B;
    ∀ spanning components 𝒯 of P_{LM,2}, add weight 2 node to B, labelled with 𝒯;
    ∀ spanning components ℰ of P_{LM,1}, add weight 1 node to B, labelled with ℰ;
    ∀ components 𝒱 of P_{LM,0}, add weight 0 node to B, labelled with 𝒱;
    for P_i ∈ C do
        if P_i has label −1 then
            Add 2 weight 0 nodes to B, labelled with P_i;
        else if P_i has label 0 then
            Add weight 0 node to B, labelled with P_i;
        else if P_i has label 1 then
            Add 2 weight 0 nodes to B, labelled with P_i;
            Add weight 1 node to B, labelled with P_i;
        else if P_i has label 2 then
            Add weight 0 node to B, labelled with P_i;
            Add weight 1 node to B, labelled with P_i;
        else if P_i has label 3 then
            Add two weight 0 nodes to B, labelled with P_i;
            Add weight 1 node to B, labelled with P_i;
        else if P_i has label 4 then
            Add weight 1 node to B, labelled with P_i;
        else if P_i has label 5 then
            Add weight 0 node to B, labelled with P_i;
            Add two weight 1 nodes to B, labelled with P_i;
        else if P_i has label 6 then
            Add two weight 0 nodes to B, labelled with P_i;
            Add three weight 1 nodes to B, labelled with P_i;
        else if P_i has label 7 then
            Add three weight 0 nodes to B, labelled with P_i;
            Add three weight 1 nodes to B, labelled with P_i;
```

**Algorithm 7:** Classification of $P^1$

```
Data: An ε-dense sample P of an embedded 2-complex |X|, P^1, and
      partitions of P_NLM, P_LM,0, P_LM,1, P_LM,2.
Result: A labeled list C, where the label for P_i is −1 if P_i corresponds to 2
        vertices, 0 if P_i corresponds to a vertex, 1 if P_i corresponds to a
        vertex and an edge, 2 if P_i corresponds to two vertices and an edge,
        3 if P_i corresponds to just an edge.
begin
    Initialise empty list C;
    for P_i ∈ P^1 do
        if |S_E(P_i)| = 1 and S_T(P_i) = ∅ then
            if ℰ ∉ S_E(P_j) ∀P_j ≠ P_i then
                Add P_i to C with label −1;
            else if ∃P_j ≠ P_i such that ℰ ∈ S_E(P_j) then
                Add P_i to C with label 0;
        else if S_E(P_i) ≠ ∅ then
            Add P_i to C with label 0;
        else
            for 𝒯 ∈ S_T(P_i) do
                Let LN(𝒯) = {P_k | 𝒯 ∈ S_T(P_k)}
            Let N(P_i) = ⋂_{𝒯 ∈ S_T(P_i)} LN(𝒯);
            if N(P_i) = {P_i, P_k} then
                Add P_i to C with label 1;
                Add P_k to C with label 0, unless P_k is already in C;
            else if N(P_i) = {P_i, P_k, P_l} then
                Add P_i to C with label 3;
                Add P_k to C with label 0, unless P_k is already in C;
                Add P_l to C with label 0, unless P_l is already in C;
    if ∃P_i ∈ P^1 \ C then
        Add P_i to C with label 2;
    return C
```

The following lemmas together show that Algorithms 5, 7 and 8 correctly partition $P_{NLM}$ and label the partitions $P_i$ appropriately.

**Lemma 5.9.** *Let $\overline{uv}$ be a locally maximal edge of $X$, such that $u, v$ are only faces of $\overline{uv}$. Then, there is a unique partition $P_1$ of $P_{NLM}$ which witnesses $\mathcal{E}$, where $\mathcal{E}$ is the edge spanning component corresponding to $\overline{uv}$. Further, $P_1$ is assigned label $-1$ by Algorithms 7 and 8.*

**Lemma 5.10.** *Let $\overline{uv}$ be a locally maximal edge of $X$, such that $u$ and/or $v$ is the face of some locally maximal cell $\sigma \in X$, $\sigma \neq \overline{uv}$. Then, there are partitions $P_1, P_2$*

**Algorithm 8:** Classification of $P^2$

**Data:** An $\varepsilon$-dense sample $P$ of an embedded 2-complex $|X|$, $P^2$, and partitions of $P_{NLM}$, $P_{LM,0}$, $P_{LM,1}$, $P_{LM,2}$, a labelled list $C$ obtained from Algorithm 7.

**Result:** A labelled list $C$.

```
begin
    for P_i ∈ P^2 do
        if P_i ∉ C then
            Let LN = {P_k | T ∈ S_T(P_k)};
            if LN ∩ P^2 = {P_i, P_k, P_l} then
                Add P_i, P_k, P_l to C with label 3;
            else if LN ∩ P^2 = {P_i, P_k} then
                Add P_i to C with label 3;
                Add P_l to C with label 4;
            else if LN ∩ P^2 = {P_i} then
                if LN = {P_i} then
                    Add P_i to C with label 7;
                else if LN = {P_i, P_k} and P_k has label 0 then
                    Add P_i to C with label 5;
                else if LN = {P_i, P_k} and P_k has label 2 then
                    Add P_i to C with label 4;
                else if LN = {P_i, P_k, P_l} and P_k has label 0, P_l label 1 then
                    Add P_i to C with label 4;
                else if LN = {P_i, P_k, P_l} and P_k has label 1, P_l label 2 then
                    Add P_i to C with label 3;
                else if LN = {P_i, P_k, P_l} and P_k has label 0, P_l label 0 then
                    Add P_i to C with label 6;
                else if LN = {P_i, P_k, P_l, P_j} and P_k, P_l, P_j have label 0 then
                    Add P_i to C with label 8;
                else if LN = {P_i, P_k, P_l, P_j, P_m} and P_k, P_l, P_j have label 0
                 and P_m has label 3 then
                    Add P_i to C with label 9;
    return C
```

*of $P_{NLM}$, which witness $\mathcal{E}$, where $\mathcal{E}$ is the edge spanning component corresponding to $\overline{uv}$. Further, $P_1$ and $P_2$ are assigned label 0 by Algorithms 7 and 8.*

**Lemma 5.11.** *Let $\triangle uvw$ be a triangle of $X$, such that for all locally maximal cells $\sigma \in X$ with $\sigma \neq \triangle uvw$, we have*

$$u,\, v,\, w \notin \sigma.$$

*Then, there is a unique partition $P_1$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1$ is given label 7 by Algorithms 7 and 8.*

**Lemma 5.12.** *Let $\triangle uvw$ be a triangle of $X$, such that there is some locally maximal cell $\sigma \in X$ with $\sigma \neq \triangle uvw$, such that $v \in \sigma$, without loss of generality, and for all locally maximal $\tau \in X$, $\tau \neq \sigma, \triangle uvw$, either $\triangle uvw \cap \tau = v$ or $\triangle uvw \cap \tau = \emptyset$.*

*Then, there are exactly two partitions $P_1, P_2$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1$ is given label 0 and $P_2$ label 5 by Algorithms 7 and 8.*

**Lemma 5.13.** *Let $\triangle uvw$ be a triangle of $X$, such that there is some locally maximal cell $\sigma \in X$ with $\sigma \neq \triangle uvw$, such that $v \in \sigma$, without loss of generality, and for all locally maximal $\tau \in X$, $\tau \neq \sigma, \triangle uvw$, either $\triangle uvw \cap \tau = \overline{uv}$ or $\triangle uvw \cap \tau = \emptyset$.*

*Then, there are exactly two partitions $P_1, P_2$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1$ is given label 0 and $P_2$ label 5 by Algorithms 7 and 8.*

**Lemma 5.14.** *Let $\triangle uvw$ be a triangle of $X$, such that there are some locally maximal cells $\sigma_1 \neq \sigma_2 \in X$ with $\sigma_1, \sigma_2 \neq \triangle uvw$, such that*

$$\sigma_1 \cap \triangle uvw = v$$
$$\sigma_2 \cap \triangle uvw = u$$

*and for all other locally maximal cells $\tau \in X$, either*

1. *$\tau \cap \triangle uvw = v$,*
2. *$\tau \cap \triangle uvw = u$,*
3. *$\tau \cap \triangle uvw = \emptyset$.*

*Then, there are exactly three partitions $P_1, P_2, P_2$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1, P_2$ are given label 0 and $P_3$ label 6 by Algorithms 7 and 8.*

**Lemma 5.15.** *Let $\triangle uvw$ be a triangle of $X$, such that there are some locally maximal cells $\sigma_1 \neq \sigma_2 \in X$ with $\sigma_1, \sigma_2 \neq \triangle uvw$, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = v$$

*and for all other locally maximal cells $\tau \in X$, either*

1. *$\tau \cap \triangle uvw = \overline{uv}$,*
2. *$\tau \cap \triangle uvw = v$,*
3. *$\tau \cap \triangle uvw = \emptyset$.*

*Then, there are exactly three partitions $P_1, P_2, P_2$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1$ has label 0, $P_2$ label 1 and $P_3$ label 4 by Algorithms 7 and 8.*

**Lemma 5.16.** *Let $\triangle uvw$ be a triangle of $X$, such that there are some locally maximal cells $\sigma_1 \neq \sigma_2 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = w$$

*and for all other locally maximal cells $\tau \in X$, either*

1. $\tau \cap \triangle uvw = \overline{uv}$,
2. $\tau \cap \triangle uvw = w$,
3. $\tau \cap \triangle uvw = \emptyset$.

*Then, there are exactly three partitions $P_1, P_2, P_2$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1$ has label 0, $P_2$ label 2 and $P_3$ label 9 by Algorithms 7 and 8.*

**Lemma 5.17.** *Let $\triangle uvw$ be a triangle of $X$, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = u$$
$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = w$$

*and for all other locally maximal cells $\tau \in X$, either*

1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \emptyset$.

*Then, there are exactly four partitions $P_1, P_2, P_3, P_4$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1, P_2$ and $P_3$ are labelled with 0 and $P_4$ with 8 by Algorithms 7 and 8.*

**Lemma 5.18.** *Let $\triangle uvw$ be a triangle of $X$, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = w$$

*and for all other locally maximal cells $\tau \in X$, either*

1. $\tau \cap \triangle uvw = \overline{uv}$,
2. $\tau \cap \triangle uvw = v$,

*3.* $\tau \cap \triangle uvw = w$,

*4.* $\tau \cap \triangle uvw = \emptyset$.

*Then, there are exactly four partitions* $P_1, P_2, P_3, P_4$ *of* $P_{NLM}$ *which witness* $\mathcal{T}$, *where* $\mathcal{T}$ *is the edge spanning component corresponding to* $\triangle uvw$. *Further,* $P_1$ *is labelled with* 1, $P_2, P_3$ *with* 0 *and* $P_4$ *with* 9 *by Algorithms 7 and 8.*

**Lemma 5.19.** *Let* $\triangle uvw$ *be a triangle of* $X$*, such that there are some locally maximal cells* $\sigma_1, \sigma_2, \sigma_3 \in X$ *with* $\sigma_i \neq \triangle uvw$ *and* $\sigma_i \neq \sigma_j$ *for* $i \neq j$*, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = u$$
$$\sigma_3 \cap \triangle uvw = v$$

*and for all other locally maximal cells* $\tau \in X$*, either*

*1.* $\tau \cap \triangle uvw = \overline{uv}$,

*2.* $\tau \cap \triangle uvw = u$,

*3.* $\tau \cap \triangle uvw = v$,

*4.* $\tau \cap \triangle uvw = \emptyset$.

*Then, there are exactly four partitions* $P_1, P_2, P_3, P_4$ *of* $P_{NLM}$ *which witness* $\mathcal{T}$, *where* $\mathcal{T}$ *is the edge spanning component corresponding to* $\triangle uvw$. *Further,* $P_1$ *is labelled with* 3, $P_2, P_3$ *with* 0 *and* $P_4$ *with* 4 *by Algorithms 7 and 8.*

**Lemma 5.20.** *Let* $\triangle uvw$ *be a triangle of* $X$*, such that there are some locally maximal cells* $\sigma_1, \sigma_2, \sigma_3 \in X$ *with* $\sigma_i \neq \triangle uvw$ *and* $\sigma_i \neq \sigma_j$ *for* $i \neq j$*, such that*

$$\sigma_1 \cap \triangle uvw = \overline{uv}$$
$$\sigma_2 \cap \triangle uvw = \overline{vw}$$
$$\sigma_3 \cap \triangle uvw = v$$

*and for all other locally maximal cells* $\tau \in X$*, either*

*1.* $\tau \cap \triangle uvw = \overline{uv}$,

*2.* $\tau \cap \triangle uvw = \overline{vw}$,

*3.* $\tau \cap \triangle uvw = v$,

*4.* $\tau \cap \triangle uvw = \emptyset$.

*Then, there are exactly four partitions* $P_1, P_2, P_3, P_4$ *of* $P_{NLM}$ *which witness* $\mathcal{T}$, *where* $\mathcal{T}$ *is the edge spanning component corresponding to* $\triangle uvw$. *Further,* $P_1$ *is labelled with* 0, $P_2, P_3$ *with* 1, *and* $P_3$ *with* 3 *by Algorithms 7 and 8.*

**Lemma 5.21.** *Let* $\triangle uvw$ *be a triangle of* $X$*, such that there are some locally maximal cells* $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \in X$ *with* $\sigma_i \neq \triangle uvw$ *and* $\sigma_i \neq \sigma_j$ *for* $i \neq j$*, such that*

$$\sigma_1 \cap \triangle uvw = u$$

$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = w$$
$$\sigma_4 \cap \triangle uvw = \overline{uv}$$

*and for all other locally maximal cells $\tau \in X$, either*

1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \overline{uv}$
5. $\tau \cap \triangle uvw = \emptyset$.

*Then, there are exactly five partitions $P_1, P_2, P_3, P_4, P_5$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1, P_2, P_3$ are labelled with 0, and $P_4$ with 8 by Algorithms 7 and 8.*

**Lemma 5.22.** *Let $\triangle uvw$ be a triangle of $X$, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3, \sigma_4 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = u$$
$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = \overline{vw}$$
$$\sigma_4 \cap \triangle uvw = \overline{uv}$$

*and for all other locally maximal cells $\tau \in X$, either*

1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \overline{uv}$
5. $\tau \cap \triangle uvw = \overline{vw}$
6. $\tau \cap \triangle uvw = \emptyset$.

*Then, there are exactly five partitions $P_1, P_2, P_3, P_4, P_5$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1, P_2$ are labelled with 0, $P_3$ with 1, and $P_4, P_5$ with 3 by Algorithms 7 and 8.*

**Lemma 5.23.** *Let $\triangle uvw$ be a triangle of $X$, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = u$$
$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = w$$
$$\sigma_4 \cap \triangle uvw = \overline{uv}$$
$$\sigma_5 \cap \triangle uvw = \overline{vw}$$

*and for all other locally maximal cells $\tau \in X$, either*

1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \overline{uv}$
5. $\tau \cap \triangle uvw = \overline{vw}$
6. $\tau \cap \triangle uvw = \emptyset$.

*Then, there are exactly five partitions $P_1, P_2, P_3, P_4, P_5, P_6$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1, P_2, P_3$ are labelled with 0, $P_4, P_5, P_6$ with 3 by by Algorithms 7 and 8.*

**Lemma 5.24.** *Let $\triangle uvw$ be a triangle of $X$, such that there are some locally maximal cells $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6 \in X$ with $\sigma_i \neq \triangle uvw$ and $\sigma_i \neq \sigma_j$ for $i \neq j$, such that*

$$\sigma_1 \cap \triangle uvw = u$$
$$\sigma_2 \cap \triangle uvw = v$$
$$\sigma_3 \cap \triangle uvw = w$$
$$\sigma_4 \cap \triangle uvw = \overline{uv}$$
$$\sigma_5 \cap \triangle uvw = \overline{vw}$$
$$\sigma_6 \cap \triangle uvw = \overline{uw}$$

*and for all other locally maximal cells $\tau \in X$, either*

1. $\tau \cap \triangle uvw = u$,
2. $\tau \cap \triangle uvw = v$,
3. $\tau \cap \triangle uvw = w$,
4. $\tau \cap \triangle uvw = \overline{uv}$,
5. $\tau \cap \triangle uvw = \overline{vw}$,
6. $\tau \cap \triangle uvw = \overline{uw}$,
7. $\tau \cap \triangle uvw = \emptyset$.

*Then, there are exactly six partitions $P_1, P_2, P_3, P_4, P_5, P_6$ of $P_{NLM}$ which witness $\mathcal{T}$, where $\mathcal{T}$ is the edge spanning component corresponding to $\triangle uvw$. Further, $P_1, P_2, P_3$ are labelled with 0, $P_4, P_5, P_6$ with 3 by Algorithms 7 and 8.*

**Theorem 5.25.** *Let $P$ be an $\varepsilon$-sample of an embedded 2-complex $|X| \subset \mathbb{R}^n$ satisfying Assumption 1, and let $B$ be the graph obtained from Algorithm 9.*

*Then, we can complete $B$ to be the incidence graph of $X$, to recover the abstract structure.*

*Proof.* From Propositions 5.5 to 5.7, we correctly identify the locally maximal components of $X$. It remains to show that we correctly learn the number of not locally maximal cells, and the incidence relationship.

For a locally maximal edge, we need to identify two vertices as its faces. To do so, we must identify which partition(s) of $P_{NLM}$ correspond to these vertices.

Take a spanning edge component $\mathcal{E}$. Then there is some locally maximal edge $\overline{uv}$ corresponding to $\mathcal{E}$. There are two cases to consider:

A: $\overline{uv}$ is disconnected from every other part of $X$, B: $\overline{uv}$ is not disconnected every other part of $X$.

Case A: From Propositions 4.8 to 4.12 and Assumption 1, there is a single partition $P_i \subset P_{NLM}$ which contains points $p$ such that $\mathcal{E} \cap B_{(R+\varepsilon)/\kappa+3\varepsilon}(p) \neq \emptyset$. Hence, $P_i$ contains samples $p$ such that either $\|v-p\| \leq \frac{3R}{2}+\varepsilon$ or $\|u-p\| \leq \frac{3R}{2}+\varepsilon$, and $P_i$ corresponds to $u$ and $v$. In this case, $P_i$ is labelled with $-1$ in Algorithm 7. This occurs only when $\overline{uv}$ is disconnected from the rest of $|X|$; hence, we infer the two boundary vertices.

Case B: As $\overline{uv}$ is not disconnected, there is some locally maximal cell $\sigma \in X$, $\sigma \neq \overline{uv}$ such that either $u$ or $v$ is a vertex of $\sigma$. Without loss of generality, let $v \in \sigma$. For the vertices $u$ and $v$ let the set of locally maximal faces they see be $S(u)$ and $S(v)$, respectively. As $X$ is a 2-complex, and $\overline{uv}$ a locally maximal edge, $\sigma \notin S(u)$. Hence, there are two partitions, $P_u, P_v$, which correspond to the vertices $u$ and $v$, respectively. In this case, $P_u$ and $P_v$ are labelled with 0 in Algorithm 7.

We now need to examine how we identify the faces of triangles.

For a triangle spanning component $\mathcal{T}$, let $\mathcal{P}_{\mathcal{T}}$ be the set of partitions $P_i$ of $P_{NLM}$ such that $d(\mathcal{T}, P_i) \leq 3\varepsilon$. There are a few cases we need to consider to ensure we correctly recover the structure of $X$:

1. $|\mathcal{P}_{\mathcal{T}}|= 1$,
2. $|\mathcal{P}_{\mathcal{T}}|= 2$,
3. $|\mathcal{P}_{\mathcal{T}}|= 3$,
4. $|\mathcal{P}_{\mathcal{T}}|= 4$,
5. $|\mathcal{P}_{\mathcal{T}}|= 5$,
6. $|\mathcal{P}_{\mathcal{T}}|= 6$.

Let the weight 2 node labelled with $\mathcal{E}$ be $t$.

Case 1 $|\mathcal{P}_{\mathcal{T}}|= 1$: Let $P_1$ be the single partition in $\mathcal{P}_{\mathcal{T}}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ does not share any faces with another cell. Then, $P_1$ corresponds to three edges and three vertices and is correctly labelled with 7 by Algorithms 7 and 8. Let the corresponding weight 1 nodes of $B$ be $e_1, e_2, e_3$ and the weight 0 nodes be $v_1, v_2, v_3$. We add an edge between $t$ and $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

Case 2 $|\mathcal{P}_{\mathcal{T}}|= 2$: Let $\mathcal{P}_{\mathcal{T}} = \{P_1, P_2\}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares either a vertex, or an edge and two vertices with other triangles or locally maximal edges. Thus, either $P_1$ is labelled with 0 and $P_2$ with 5, or $P_1$ is labelled with 2 and $P_2$ with 4 by Algorithms 7 and 8.

If $P_1$ has label 0 and $P_2$ has label 5, we find the weight 0 node $v_1$ with label $P_1$ and the three weight 1 nodes $e_1, e_2, e_3$ and two weight 0 nodes $v_2, v_3$ with label $P_2$. Then, we add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If $P_1$ has label 2 and $P_2$ has label 4, we find the weight 1 note $e_1$ and two weight 0 node $v_1, v_2$ with label $P_1$, the two weight 1 nodes $e_2, e_3$ and one weight 0 nodes $v_3$

with label $P_2$. We add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

Case 3 $|\mathcal{P}_\mathcal{T}|= 3$: Let $\mathcal{P}_\mathcal{T} = \{P_1, P_2, P_3\}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares either two vertices, or two vertices and an edge with other triangles or locally maximal edges. Thus, either $P_1$ and $P_2$ are labelled with 0 and $P_2$ with 6; or $P_1$ is labelled with 0, $P_2$ with 1 and $P_3$ with 4; or $P_1$ is labelled 0, $P_2$ with 2 and $P_3$ with 9.

If $P_1, P_2$ have label 0 and $P_3$ has label 6, we find the weight 0 node $v_1$ with label $P_1$, the weight 0 node $v_2$ with label $P_2$, the three weight 1 nodes $e_1, e_2, e_3$ and the weight 0 node $v_3$ with label $P_3$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If $P_1$ has label 0, $P_2$ label 1 and $P_3$ label 4, we find the weight 0 node $v_1$ with label $P_1$, the weight 0 node $v_2$ with label $P_2$, weight 1 node $e_1$ with label $P_2$, the weight 0 node $v_3$ with label $P_3$, and the two weight 1 nodes $e_2, e_3$ with label $P_3$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If $P_1$ has label 0, $P_2$ label 2 and $P_3$ label 9, we find the weight 0 node $v_1$ with label $P_1$, the weight 0 node $v_2$ and weight 1 node $e_1$ with label $P_2$, and the weight 1 nodes $e_2, e_3$ and weight 0 node $v_3$ with label $P_3$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

Case 4 $|\mathcal{P}_\mathcal{T}|= 4$: Let $\mathcal{P}_\mathcal{T} = \{P_1, P_2, P_3, P_4\}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares three vertices, or three vertices and an edge, or three vertices and two edges with other triangles or locally maximal edges. Thus, either $P_1, P_2$ and $P_3$ are labelled with 0 and $P_4$ with 8; or $P_1$ is labelled with 1, $P_2, P_3$ with 0 and $P_3$ with 9; or $P_1$ with 3, $P_2, P_3$ with 0 and $P_4$ with 4; or $P_1$ is labelled with 0, $P_2, P_3$ with 1, and $P_3$ with 3 by Algorithms 7 and 8.

If $P_1, P_2, P_3$ have label 0 and $P_4$ has label 8, find the weight 0 node $v_1$ with label $P_1$, weight 0 node $v_2$ with label $P_2$, weight 0 node $v_3$ with label $P_3$, and the three weight 1 nodes $e_1, e_2, e_3$ with label $P_4$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If $P_1$ has label 1, $P_2, P_3$ have label 0, and $P_4$ has label 9, find the weight 0 node $v_1$ and weight 1 node $e_1$ with label $P_1$, weight 0 node $v_2$ with label $P_2$, weight 0 node $v_3$ with label $P_3$, and the two weight 1 nodes $e_2, e_3$ with label $P_4$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If $P_1$ has 3, $P_2, P_3$ label 0 and $P_4$ label 4;, find the weight 1 node $e_1$ with label $P_1$, weight 0 node $v_1$ with label $P_2$, weight 0 node $v_2$ with label $P_3$, and the two weight 1 nodes $e_2$ and weight 0 node $e_3$ with label $P_4$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

If $P_1$ has label 0, $P_2, P_3$ have label 1, and $P_4$ has label 3, find the weight 0 node $v_1$ with label $P_1$, weight 0 node $v_2$ and weight 1 node $e_1$ with label $P_2$, weight 0 node $v_3$ and weight 1 node $e_3$ with label $P_3$, and the two weight 1 nodes $e_2$ with label $P_4$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between the following pairs:

$$(e_1, v_1), (e_1, v_2), (e_2, v_2), (e_2, v_3), (e_3, v_3), (e_3, v_1).$$

Case 5 $|\mathcal{P}_\mathcal{T}|= 5$: Let $\mathcal{P}_\mathcal{T} = \{P_1, P_2, P_3, P_4, P_5\}$.

This can occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares three vertices and two edges; or three vertcies and one edge with other triangles or locally maximal edges. Thus, $P_1, P_2$ are labelled with 0, $P_3$ with 1 and $P_4, P_5$ with 3; or $P_1, P_2, P_3$ are labelled with 0, $P_4$ with 3 and $P_5$ with 9 by Algorithms 7 and 8.

If $P_1, P_2$ are labelled with 0, $P_3$ with 1 and $P_4, P_5$ with 3 we find the weight 0 node $v_1$ with label $P_1$, find the weight 0 node $v_2$ with label $P_2$, find the weight 1 node $e_1$ and weight 0 node $v_3$ with label $P_3$, find the two weight 1 nodes $e_2, e_3$ with label $P_4$, and the two weight 1 nodes $e_2, e_3$ with label $P_5$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between $e_i$ with label $P_i$ and $v_j$ with label $P_j$ if $d(P_i, P_j)$.

If $P_1, P_2, P_3$ are labelled with 0, $P_4$ with 3 and $P_5$ with 9 we find the weight 0 node $v_1$ with label $P_1$, find the weight 0 node $v_2$ with label $P_2$, find the weight 0 node $v_3$ with label $P_3$, find the weight 1 node $e_1$ with label $P_4$, and the two weight 1 nodes $e_2, e_3$ with label $P_5$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between $e_i$ with label $P_i$ and $v_j$ with label $P_j$ if $d(P_i, P_j)$.

Case 6 $|\mathcal{P}_\mathcal{T}|= 6$: Let $\mathcal{P}_\mathcal{T} = \{P_1, P_2, P_3, P_4, P_5, P_6\}$.

This can only occur if the triangle $\triangle uvw$ corresponding to $\mathcal{T}$ shares three vertices and two edges, or three vertices and three edges with other triangles or locally maximal edges. In either case, $P_1, P_2, P_3$ are labelled with 0, $P_4, P_5, P_6$ with 3 by Algorithms 7 and 8.

So we find the weight 0 node $v_1$ with label $P_1$, find the weight 0 node $v_2$ with label $P_2$, find the weight 0 node $v_3$ with label $P_3$, find the weight 1 node $e_1$ with label $P_4$, the weight 1 node $e_2$ with label $P_5$, and the weight 1 node $e_3$ with label $P_6$. Then add an edge between $t$ and each of $e_1, e_2, e_3, v_1, v_2, v_3$ and between $e_i$ with label $P_i$ and $v_j$ with label $P_j$ if $d(P_i, P_j)$.

In each of these 6 cases, we have connected the weight 2 node $t$ corresponding to the cell $\tau$ to each weight 1 node $e$ corresponding to an edge $\sigma_e$ of $\tau$, as well as to each weight 0 node $v$ corresponding to a vertex $\sigma_v$ of $\tau$. Further, in the process, we also connect the weight 1 node $e$ and weight 0 node $v$ if $\sigma_v$ is a vertex of $\sigma_e$.

We have shown that the weight 2 nodes of $B$ correspond bijectively to the triangles of $X$, the weight 1 nodes of $B$ correspond bijectively to the edges of $X$, and the weight 0 nodes of $B$ correspond bijectively to the vertices of $X$. We have also shown that for any pair of nodes $n_1, n_2$ with corresponding cells $\sigma_1, \sigma_2$, there an edge between them if and only if $\sigma_1 \subset \sigma_2$ or $\sigma_2 \subset \sigma_1$.

Hence, $B$ is the incidence graph of $X$. □

In this article, we have presented a method for learning the abstract structure $X$ underlying an embedded 2-simplicial complex $|X|= (X, \Theta)$ (satisfying Assumption 1) from an $\varepsilon$-sample $P$. For abstract 2-complexes, modelling the embedding is future work. In particular, to modelling embeddings that are not linear or where we allow for cells of dimension 2, which are not triangles (along the lines of CW-complexes), we need to develop the process for learning the faces of locally maximal cells further.

# 6 Future directions

There are several natural paths for the work in this article to be extended. In particular, removing the assumption that the maximal dimension of a cell in the complex is 2 is a direct next step. It is also natural to consider how to modify the algorithm to allow for non-linear embeddings, in particular using semi-algebraic sets, as well as what happens when the noise is not assumed to be Hausdorff. These directions form a sort of 'orthogonal' basis for future research, as they can be thought of as independent problems, but when combined present a rather significant development towards learning stratified spaces.

**Acknowledgments.** This work was undertaken during the author's PhD, which was supported by Australian Federal Government Grant, 2019-2022, 'Stratified Space Learning'. The author would like to thank Kate Turner, Chris Williams, Jonathan Spreer, Stephan Tillmann, Vanessa Robins, Vigleik Angeltveit, Martin Helmer, and James Morgan for very helpful discussions.

# Declarations

## Competing interests

The work in this paper was undertaken during my PhD at the Australian National University and the University of Sydney, which was supported by an Australian Federal Government Grant, 2019-2022, *Stratified Space Learning.*

# Appendix A Proofs of Geometric Lemmas

**Proof 1** (Proof of Lemma 3.1)**.** *Consider* $S^{R+\varepsilon}_{R-\varepsilon}(p) \cap L$ *say* $C$. *Consider a point* $q \in S^{R+\varepsilon}_{R-\varepsilon}(p)$ *with* $d(L, q) \leq \varepsilon$. *Let* $q_L$ *be the projection of* $q$ *to* $L$, $p_L$ *the projection of* $p$ *to* $L$.

*There are two cases we need to consider,*

1. $\|x - q_L\| \geq \|q_L - p_L\|$,
2. $\|x - q_L\| < \|q_L - p_L\|$.

*We begin with case 1.*


(A) When $p$ and $q$ are on the same side of $x$

(B) When $p$ and $q$ are on different sides of $x$

**Fig. A1**

*We want to bound $\|x-q\|$. Note that*

$$\begin{aligned}
\|q-x\|^2 &= \|q-q_L\|^2 + \|q_L-x\|^2,\\
\|q_L-x\| &= \|p_L-x\| - \|p_L-q_L\|,\\
\|p_L-q_L\|^2 &= \|q-p\|^2 - \left(\|p-p_L\| + \|q-q_L\|\right)^2,\\
\|p_L-x\|^2 &= \|x-p\|^2 - \|p_L-p\|^2.
\end{aligned}$$

*Hence,*

$$\begin{aligned}
&\|q-x\|^2\\
&= \|q-q_L\|^2 + (\|p_L-x\| - \|p_L-q_L\|)^2\\
&= \|q-q_L\|^2 + \left(\sqrt{\|q-p\|^2 - \|p_L-p\|^2} - \sqrt{\|q-p\|^2 - (\|p-p_L\| + \|q-q_L\|)^2}\right)^2\\
&= \|q-q_L\|^2 +\\
&\left(\sqrt{\|q-p\|^2 - \|p_L-p\|^2} - \sqrt{\|q-p\|^2 - \|p-p_L\|^2 - (\|q-q_L\|^2 + \|p-p_L\|\|q-q_L\|)}\right).
\end{aligned}$$

*Let*

$$\begin{aligned}
A &= \|q-p\|^2 - \|p-p_L\|^2,\\
B &= \|q-q_L\|^2 + \|p-p_L\|\|q-q_L\|.
\end{aligned}$$

*As*

$$\begin{aligned}
\|q-p\| &\le R,\\
\|p-p_L\| &\le \frac{R}{2},\\
\|q-q_L\| &\le \varepsilon,
\end{aligned}$$

*we have*

$$\begin{aligned}
A &> (R-\varepsilon)^2 - \varepsilon^2\\
B &< 3\varepsilon^2
\end{aligned}$$

*and so* $A > \frac{4B}{3}$. *Then*

$$\begin{aligned}
\frac{AB}{3} &> \frac{4B^2}{9} \\
A^2 - AB &> A^2 - \frac{4AB}{3} + \frac{4B^2}{9} \\
\sqrt{A(A-B)} &> A - \frac{2B}{3} \\
-2\sqrt{A(A-B)} &< -2A + \frac{4B}{3} \\
2A - B - 2\sqrt{A(A-B)} &< \frac{B}{3} \\
\left(\sqrt{A} - \sqrt{A-B}\right)^2 &< \frac{B}{3}
\end{aligned}$$

*Recall* $A > \frac{4B}{3}$, *thus*

$$\begin{aligned}
\|q - x\|^2 &= \|q - q_L\| + \left(\sqrt{A} - \sqrt{A-B}\right)^2 \\
&\leq \varepsilon^2 + \frac{B}{3} \\
&\leq 2\varepsilon
\end{aligned}$$

*A similar calculation in case 2 gives a smaller bound, so*

$$\|q - x\| \leq \sqrt{2}\varepsilon.$$

**Proof 2** (Proof of Lemma 3.2)**.** *First, let* $p_H$ *be the projection of* $p$ *to* $H$*, and note that* $\|p_H - p\| \leq \varepsilon$. *Take* $q_1 \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$. *Let* $x_1$ *be the point in* $\partial B_R(p) \cap H$ *closest to* $q_1$*, and* $q_H$ *the projection of* $q_1$ *to* $H$. *Note that* $p_H, q_H, x_1$ *are co-linear, lying on the ray* $L$ *from* $p_H$*, and* $\|q_1 - q_H\| \leq \varepsilon$. *By Lemma 3.1,* $\|q_1 - x_1\| \leq \sqrt{2}\varepsilon$.

*As* $H \cap \partial B_R(p)$ *is a circle with radius* $\sqrt{R^2 - \|p_H - p\|^2}$*, there is a point* $x_2 \in H \cap \partial B_R(p)$ *such that* $\|x_2 - x_1\| = 2\sqrt{R^2 - \|p_H - p\|^2}$. *As* $d_H(p, H) \leq \varepsilon$*, we have*

$$\|x_2 - x_1\| \geq 2\sqrt{R^2 - \varepsilon^2},$$

*and as* $d_H(P, H) \leq \varepsilon$*, there is* $q_1 \in P$ *with* $\|q_1 - x_1\| \leq \varepsilon$. *Hence*

$$\|q_2 - q_1\| \geq 2\sqrt{R^2 - \varepsilon^2} - (1 + \sqrt{2})\varepsilon.$$

**Proof 3** (Proof of Lemma 3.3)**.** *First, let* $H_1'$ *be the half plane containing* $H_1$ *with bounding line* $L'$ *such that* $D(L, L') = \varepsilon$*,* $p_H$ *be the projection of* $p$ *onto* $H_1'$ *and* $p_L$ *the projection of* $p$ *to* $L$. *Then take* $x_1 \in H_1$ *such that* $\|p - x_1\| = R$ *and* $p_H, p_L$ *and* $x_1$ *are co-linear. Take* $q_1 \in P$ *with* $\|q_1 - x_1\| \leq \varepsilon$*, so* $q_1 \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$.

*Let $q_2$ be a point in $S_{R-\varepsilon}^{R+\varepsilon}(p) \cap P$. There are two cases to consider: $d(q_2, H_1') \leq \varepsilon$ and $d(q_2, H_2) \leq \varepsilon$.*

*If $d(q_2, H_1') \leq \varepsilon$, take $x_2 \in \partial B_R(p) \cap H_1'$ such that $x_2, p_H$ and the projection of $q_2$ to $H_1'$ are co-linear. Then by Lemma 3.1 $\|q_2 - x_2\| \leq \sqrt{2}\varepsilon$.*

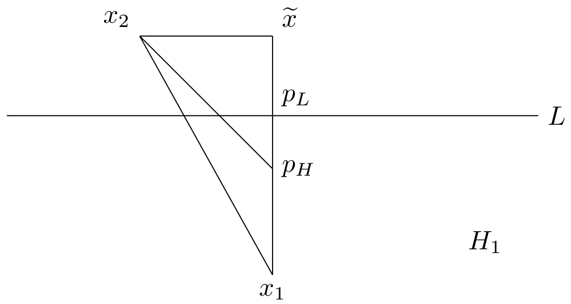


**Fig. A2**: Understanding the behaviour of points near the common boundary of two half-planes

*Consider the triangle formed by $x_1, p_h, x_2$. By assumption,*

$$\|\widetilde{x} - p_H\| < \frac{R}{2} < R - 7\varepsilon,$$
$$\|x_2 - p_H\| = \|x_1 - p_H\| \leq R.$$

*Let $\widehat{R} = \sqrt{R^2 - \|p_H - p\|^2}$. Then*

$$\begin{aligned}
\|\widetilde{x} - p_H\| &< \widehat{R} \\
\|\widetilde{x} - p_H\| &< \widehat{R} - 6\varepsilon \\
2\widehat{R}\|\widetilde{x} - p_H\| &< 2\widehat{R}^2 - 12\widehat{R}\varepsilon \\
2\widehat{R}^2 + 2\widehat{R}\|\widetilde{x} - p_H\| &< 4\widehat{R}^2 - 4(1+\sqrt{2})\widehat{R}\varepsilon + (1+\sqrt{2})\varepsilon \\
2\widehat{R}^2 + 2\widehat{R}\left(\frac{\|\widetilde{x} - p_H\|}{\widehat{R}}\right) &< \left(2\widehat{R} - (1+\sqrt{2})\varepsilon\right)^2.
\end{aligned}$$

*Further,*

$$\begin{aligned}
2\widehat{R}^2 + 2\widehat{R}\left(\frac{\|\widetilde{x} - p_H\|}{\widehat{R}}\right) &= \|x_1 - p_H\|^2 + \|x_2 - p_H\|^2 + 2\|x_1 - p_H\|\|x_2 - p_h\| \cos\angle x_2 p_h \widetilde{x} \\
&= \|x_1 - p_H\|^2 + \|x_2 - p_H\|^2 - 2\|x_1 - p_H\|\|x_2 - p_h\| \cos\angle x_2 p_h x_1 \\
&= \|x_2 - x_1\|^2,
\end{aligned}$$

*so*

$$\|x_2 - x_1\| < \sqrt{R^2 - \|p_H - p\|^2} - (1+\sqrt{2})\varepsilon.$$

*which implies*

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - (2+2\sqrt{2})\varepsilon.$$

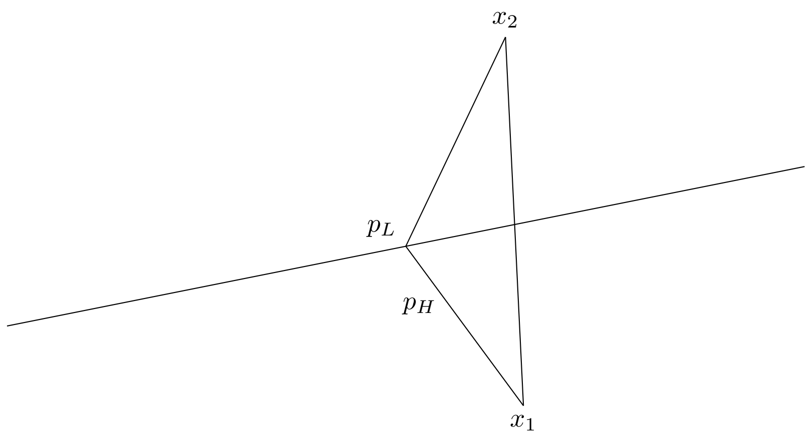


**Fig. A3**: $d(q_2, H_2) \leq \varepsilon$

*Now assume* $d(q_2, H_2) \leq \varepsilon$. *Let* $H_2'$ *be the half-plane which contains* $H_2$ *and has boundary* $L'$ *with* $d(L, L') = \varepsilon$. *As* $d(q_2, H_2) \leq \varepsilon$, *then there is* $x_2 \in \partial B_R(p) \cap H_2'$ *with* $\|q_2 - x_2\| \leq \sqrt{2}\varepsilon$. *Hence,*

$$\|x_1 - x_2\| \geq 2\sqrt{R^2 - \varepsilon^2} - (2+2\sqrt{2})\varepsilon.$$

*If* $x_2 \in H_2' \setminus H_2$, *then by a similar argument to above,*

$$\|x_1 - x_2\| \geq 2\sqrt{R^2 - \varepsilon^2} - (2+2\sqrt{2})\varepsilon.$$

*If* $x_2 \in H_2 \subsetneq H_2'$, *by the cosine rule we have*

$$\|x_2 - x_1\|^2 = \|x_2 - p_L\|^2 + \|x_1 - p_L\|^2 - 2\|x_2 - p_L\|\|x_1 - p_L\| \cos \angle x_1 p_L x_2.$$

*Note* $\|x_1 - p_L\| = \|x_1 - p_H\| + \|p_H - p_L\|$, *and* $\|x_2 - x_1\|$ *is bounded above by the case when*

$$\begin{aligned} \angle x_1 p_L x_2 &= \alpha, \\ \|x_2 - p_L\| &= R + 2\varepsilon, \end{aligned}$$

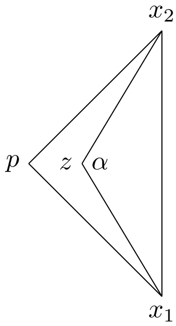


**Fig. A4**: Bounding the diameter of a set of points

$$\|x_1 - p_L\| = \|x_1 - p_H\| + \|p_H - p_L\| = \frac{3R}{2} + \varepsilon.$$

*Hence, we have*

$$\|x_2 - x_1\| < (R + 2\varepsilon)^2 + \left(\frac{3R}{2} + \varepsilon\right)^2 - (R + 2\varepsilon)\left(\frac{3R}{2} + \varepsilon\right)\cos\alpha.$$

*By assumption, $\alpha \in (0, \Psi(\varepsilon, R))$, and so*

$$\|x_2 - x_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(2 + 2\sqrt{2}\right)\varepsilon,$$

*which implies that*

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(1 + \sqrt{2}\right)\varepsilon.$$

*Hence, there is a $q_1 \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$ such that for all $q_2 \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$*

$$\|q_2 - q_1\| < 2\sqrt{R^2 - \varepsilon^2} - \left(1 + \sqrt{2}\right)\varepsilon.$$

**Proof 4** (Proof of Lemma 3.4)**.** *By Lemma 3.1, every $q \in S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P$ is with in $\sqrt{2}\varepsilon$ of the point $x$ in $L$ with $\|x - p\| = R$. Hence, $\left(S^{R+\varepsilon}_{R-\varepsilon}(p) \cap P\right)^{\frac{3\varepsilon}{2}}$ consists of a single connected component and it has diameter less than $2\sqrt{2}\varepsilon$.*

**Proof 5** (Proof of Lemma 3.5)**.** *As $\|p - z\| \leq \frac{R-\varepsilon}{2}$, the intersection $S^{R+\varepsilon}_{R-\varepsilon}(p) \cap T$ is not empty, connected, and $\mathcal{H}_1\left(S^{R+\varepsilon}_{R-\varepsilon}(p) \cap T\right) = 0$. Further, the intersections $S^{R+\varepsilon}_{R-\varepsilon}(p) \cap L_1$ and $S^{R+\varepsilon}_{R-\varepsilon}(p) \cap L_2$ are also connected.*

*Now, let $x_1$ be the point on $L_1$ with $\|q_1 - p\| = R$ and let $x_2$ be the point on $L_2$ with $\|x_2 - p\| = R$. As $S^{R+\varepsilon}_{R-\varepsilon}(p) \cap T$ is path connected, $x_1$ and $x_2$ are path connected in $T$.*

*Consider the triangle $\triangle x_1 p x_2$, we have*

$$\begin{aligned}\|x_1 - x_2\|^2 &= \|x_1 - z\|^2 + \|x_2 - z\|^2 - 2\|x_1 - z\|\|q_2 - z\| \cos\alpha \\ &\geq \left(R - \frac{R-\varepsilon}{2}\right)^2 + \left(R - \frac{R-\varepsilon}{2}\right)^2 - 2\left(R - \frac{R-\varepsilon}{2}\right)^2 \cos\alpha \\ &= 2\left(\frac{R+\varepsilon}{2}\right)^2 (1-\cos\alpha).\end{aligned}$$

*Now, as $d_H(P,T) \leq \varepsilon$, there are points $q_1, q_2 \in P$ with*

$$\|q_1 - x_1\|, \|q_2 - x_2\| \leq \varepsilon.$$

*Then by the triangle inequality*

$$\begin{aligned}\|q_1 - q_2\|^2 &= 2\left(\frac{R+\varepsilon}{2}\right)^2 (1-\cos\alpha) - 2\varepsilon \\ &> 2\sqrt{2}\varepsilon, \quad \text{as } \alpha \in \left[\frac{\pi}{6}, \pi\right)\end{aligned}$$

# Appendix B Proof of Correctness Lemmas

**Proof 6** (Proof of Proposition 5.3)**.** *Let $C$ be a connected component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$ which spans a locally maximal edge $\overline{uv}$, with midpoint $m_{uv}$. Then, there is a sample $p_m \in C$ such that $\|p_m - m_{uv}\| \leq \varepsilon$.*

*To show that $\mathcal{D}(C) \geq \frac{9R}{2}$, we show that there are two points $x_u, x_y \in \overline{uv}$ such that*

1. *$\|u - x_u\| > \frac{3R}{2} + 2\varepsilon$,*
2. *$\|v - x_v\| > \frac{3R}{2} + 2\varepsilon$,*
3. *$\|x_u - x_v\| \geq \frac{3R}{2}$.*

*Without loss of generality, we show that $x_u$ exists, and*

$$\|x_u - m_{uv}\| \geq \frac{3R}{4} + \varepsilon.$$

*By Assumption 1, $\|u-v\| \geq 6(R+\varepsilon)$. As $\overline{uv}$ is a line segment, for all $\eta \in [0, \frac{9R}{4}+3\varepsilon]$ there is a point $x_\eta \in \overline{uv}$ such that $\|x_\eta - u\| = \eta$. Letting $\eta = \frac{3R}{2} + 2\varepsilon$, there is a point, namely $x_u$ such that $\|x_u - u\| = \frac{3R}{2} + 2\varepsilon$. As $P$ is an $\varepsilon$-sample, there is a sample $p_u$ such that $\|x_u - p_u\| \leq \varepsilon$, and hence $\|p_u - u\| > \frac{3R}{2} + \varepsilon$. Thus, the $(\varepsilon, R)$-local structure of $P$ at $p_u$ is maximal of dimension 1.*

*We can repeat this argument for all $\eta \in [\frac{3R}{2} + 2\varepsilon, \frac{9R}{4} + 3\varepsilon]$, and obtain a path of points $x_\eta \in \overline{uv}$ and samples $p_\eta \in P$ connecting $p_u$ to $p_m$.*

*This also holds when we replace $u$ with $v$, and hence we have $p_u$ and $p_v$. Finally, we have*

$$\|p_u - p_v\| \geq \|x_u - x_v\| - \|p_u - x_u\| - \|p_v - x_v\|$$

$$\geq \frac{3R}{2} - 2\varepsilon,$$

*and hence* $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$.

*Now, we show that if* $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$*, then* $C$ *spans some locally maximal edge.*

*If* $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$*, then there are points* $p, q \in C$ *with*

$$\|p - q\| \geq \frac{3R}{2} - 2\varepsilon.$$

*As* $P$ *is an* $\varepsilon$*-sample of* $|X|$*, there are points* $x_p, x_q \in |X|$*, with*

$$\|x_p - p\|,\ \|x_q - q\| \leq \varepsilon.$$

*Let* $m_{pq}$ *be the midpoint of* $x_p$ *and* $x_q$*. As* $p$ *and* $q$ *are in the same connected component of* $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$*, we know there is a sequence of points* $\{q_i\}_{i=0}^{m}$ *with* $q_0 = p$*,* $q_m = q$ *and for all* $0 < i \leq m$*,* $\|q_i - q_{i-1}\| \leq 3\varepsilon$*. Again,* $P$ *is an* $\varepsilon$*-sample of* $|X|$*, and as* $q_i \in P_{LM,1}$*,* $\forall 0 \leq i \leq m$*, for each* $q_i$ *there is some* $x_i \in |X|$ *which is on a locally maximal edge, and* $\|q_i - x_i\| \leq \varepsilon$*. From Assumption 1 and Proposition 4.9, there is a locally maximal edge, say* $\overline{uv}$ *such* $x_i \in \overline{uv}, \forall 0 \leq i \leq m$*. Let the midpoint of* $\overline{uv}$ *be* $x_{uv}$*.*

*We now split into two cases:*

*I there is some* $i$ *such that* $x_i = x_{uv}$*,*

*II for all* $i$ *we have* $x_i \neq x_{uv}$*.*

*Case I: The connected component* $C$ *is a spanning connected component, as it contains a sample which is within* $\varepsilon$ *of the midpoint* $x_{uv}$ *of the locally maximal edge* $\overline{uv}$*.*

*Case II: As no* $q_i$ *is within* $\varepsilon$ *of* $m_{uv}$*, we know that* $q_i\ \forall 0 \leq i \leq m$ *are on the same side of* $\overline{uv}$*. That is, for all* $q_i$*, without loss of generality,*

$$\|q_i - x_{uv}\| \leq \|q_i - u\| \geq \frac{3\sqrt{3}}{2}R + 3\varepsilon$$
$$\|q_m - v\| \geq \frac{3R}{2} + \varepsilon.$$

*Further, assume that*

$$\|q_0 - m_{uv}\| \leq \|q_m - x_{uv}\|.$$

*There is another sequence of points* $\{x'_j\}_{j=0}^{m'}$ *in* $\overline{uv}$ *with* $x'_0 = x_m$ *and* $x'_{m'} = x_{uv}$*, and for* $0 < j \leq m'$

$$\|x'_j - x'_{j-1}\| \leq \varepsilon.$$

*Then, there exists* $q'_j \in P$ *with*

$$\|q'_j - x'_j\| \leq \varepsilon$$

$$\|q'j - q'_{j-1}\| \leq \varepsilon \, \forall 0 < j \leq m'$$
$$\|q'_j - v\| \geq \frac{3R}{2} + \varepsilon.$$

*By Assumption 1 and Proposition 4.9, $q'_j \in P_{LM,1}$ for all $0 \leq j \leq m'$. Hence, each $q'_j$ is in the same connected component $C$ as $q_m$.*

*Thus, $C$ contains a sample $q'_{m'}$ which is within $\varepsilon$ of the midpoint of the locally maximal edge $\overline{uv}$. Hence, $C$ is a spanning connected component.*

*Thus a component $C$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$ spans a locally maximal edge $\overline{uv}$ if and only if $\mathcal{D}(C) \geq \frac{3R}{2} - 2\varepsilon$.*

**Proof 7** (Proof of Proposition 5.4)**.** *First, let $C$ be a connected component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}$ which spans some triangle $\triangle uvw$ with midpoint $m$. As $P$ is an $\varepsilon$-sample of $X$, there is a sample $p_m \in P$ with $\|p_m - m\| \leq \varepsilon$. As the radius of the inscribed circle of $\triangle uvw$ is at least $2R + 3\varepsilon$, $m$ is at least $2R + 3\varepsilon$ from $\partial\triangle uvw$. Thus, $d(p_m, \partial\triangle uvw) \geq 2R + 2\varepsilon$.*

*Hence, for all $q \in B_{\frac{R}{2}+2\varepsilon}(p) \cap P$, $d(q, \partial\triangle uvw) \geq \frac{3R}{2} + \varepsilon$, and so $q \in P_{LM,2}$.*

*Now, take $p \in P_{LM,2}$ such that $B_{\frac{R}{2}+\varepsilon}(p) \cap P \subset P_{LM,2}$. Then, there is some triangle $\triangle uvw$ with $d(\triangle uvw, p) \leq \varepsilon$. As $p \in P_{LM,2}$, we know that $d(\partial\triangle uvw, p) > \frac{R}{2} - \varepsilon$. By assumption, for all $q \in B_{\frac{R}{2}+\varepsilon}(p) \cap P$, we have $d(\partial\triangle uvw, q) > \frac{R}{2} - \varepsilon$. Recall that $P$ is an $\varepsilon$-sample of $|X|$, so there is a point $x \in X$ such that $\|p - x\| \leq \varepsilon$. As $\triangle uvw$ is convex, and every $B_{\frac{R}{2}+\varepsilon}(p) \cap P \subset P_{LM,2}$, we have*

$$d(\partial\triangle uvw, x) \geq \frac{R}{2} + 2\varepsilon + \frac{R}{2} - 2\varepsilon = R.$$

*Hence, is a point $y \in B_{\frac{R}{2}+2\varepsilon}(p) \cap \triangle uvw$ with*

$$d(\partial\triangle uvw, y) \geq \frac{R}{2} + 2\varepsilon.$$

*and a sample $q \in B_{\frac{R}{2}+2\varepsilon}(p) \cap P_{LM,2}$ with $\|q - y\| \leq \varepsilon$.*

*Now, we can construct a sequence of points $\{y_i\}_{i=0}^m \subset \triangle uvw$ such that $\|y_i - y_{i-1}\| \leq \varepsilon$ for $1 \leq i \leq m$, and $y_0 = x$, $y_m = y$. Further, for each $y_i$ there is a $q_i \in P$ with $\|q_i - y_i\| \leq \varepsilon$, and $q_i \in P_{LM,2}$. Note, that this means $p$ and $q_m$ are in the same connected component $C$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$.*

*Finally, we construct a similar sequence of points $\{\widetilde{y}_j\}_{j=0}^{\widetilde{m}}$ in $|X|$ from $y$ to $m_{\triangle uvw}$ with $\widetilde{y}_0 = y$, $\widetilde{y}_{\widetilde{m}} = m_{\triangle uvw}$. Again, for each $\widetilde{y}_j$, there is a $\widetilde{q}_j \in P$ with $\|\widetilde{y}_j - \widetilde{q}_j\| \leq \varepsilon$ and $\widetilde{q}_j \in P_{LM,2}$. Hence, the $\widetilde{q}_j$ are in the same connected component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$, and further, this connected component is $C$.*

**Proof 8** (Proof of Proposition 5.5)**.** *Let $V_{LM}$ be the set of locally maximal vertices of $X$. Let $v$ be a locally maximal vertex, then by Proposition 4.8, $\forall p \in P$ with $\|p - v\| \leq 4\varepsilon$, $p \in P_{LM,0}$. In fact, by Assumption 1, any $p \in P$ with $\|p - v\| \leq 4\varepsilon$ is actually within $\varepsilon$ of $v$. Hence, every $p \in P_{LM,0}$ within $\varepsilon$ of $v$ are in the same connected component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,0})$.*

*Now, take a connected component $C$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,0})$. Each $p \in C$ is within $\varepsilon$ of a locally maximal vertex $v_p$ of $X$. By Assumption 1, every locally maximal vertex $v$ is at least $5\varepsilon$ away from any other cell of $X$, and hence $\forall p \in C$, $v_p$ is the same.*

*Hence, the connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,0})$ correspond bijectively to the locally maximal vertices of $X$.*

**Proof 9** (Proof of Proposition 5.6)**.** *Let $E_{LM} \subset E$ be the set of locally maximal edges in $X$. By Proposition 5.3, a connected component $C$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$ spans an edge $\overline{uv}$ if and only if it contains a sample $p$ within $\varepsilon$ of the midpoint $m$ of $\overline{uv}$.*

*If a connected component $C$ is a spanning component, then there is some locally maximal edge $\overline{uv}$ with midpoint $m$ such that there is a sample $p \in C$ with $\|m-p\| \le \varepsilon$.*

*For any locally maximal $\overline{uv} \in E_{LM}$ with midpoint $m$, there is some sample $p \in P$ such that $\|m - p\| \le \varepsilon$. Then, by Assumption 1 and proposition 4.9, $p \in P_{LM,1}$, and so there is some spanning connected component $C_{\overline{uv}}$ in $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$.*

*Now, consider a locally maximal edge $\overline{uv'}$, $v' \neq v$, and take samples $p, q \in P_{LM,2}$ such that $d(\overline{uv}, p)$, $d(\overline{uv'}, q) \le \varepsilon$. By Assumption 1, $\|p - q\| > 6\varepsilon$, and so $p$ and $q$ are in different connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$.*

*Finally, consider a locally maximal edge $\overline{u'v'}$ such that $\overline{uv}$ and $\overline{u'v'}$ do not have a common vertex. Take samples $p, q \in P_{LM,2}$ such that*

$$d(\overline{uv}, p),\ d(\overline{u'v'}, q) \le \varepsilon.$$

*Again, by Assumption 1, $\|p - q\| > 6\varepsilon$, and so $p$ and $q$ are in different connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,1})$.*

*Hence, each connected component $C$ only consists of samples $p$ with $d(\overline{uv}, p) \le \varepsilon$ for a single locally maximal edge $\overline{uv}$.*

*Thus, the spanning connected components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$ are in bijection with the locally maximal edges of $|X|$.*

**Proof 10** (Proof of Proposition 5.7)**.** *From Proposition 5.4, a connected component $C$ of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$ spans a triangle $\triangle uvw$ if and only if it contains a sample $p$ within $\varepsilon$ of the midpoint $m$ of $\triangle uvw$.*

*As $P$ is a $\varepsilon$-sample of $|X|$, for every $\triangle uvw$ with midpoint $m$, there is a sample $p \in P$ such that $\|p - m\| \le \varepsilon$. Hence, there is a spanning connected component $C$ in $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$.*

*Now, consider $C$ a spanning component of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$. Then, as $P$ is a $\varepsilon$-sample, there is some $\triangle uvw$ with midpoint $m$ such that there is a sample $p \in C$ with $\|p-m\| \le \varepsilon$.*

*Consider two triangles $\triangle uvw$, $\triangle u'v'w'$, and take two samples $p, p' \in P_{LM,2}$ with*

$$d(\triangle uvw, p),\ d(\triangle u'v'w', p') \le \varepsilon.$$

*As $p, p' \in P_{LM,2}$, we know that*

$$d(\partial\triangle uvw, p),\ d(\partial\triangle u'v'w', p') > R + \varepsilon,$$

*and so by Assumption 1, $\|p - p'\| > 6\varepsilon$.*

*Hence, the spanning components of $\check{\mathcal{C}}_{\frac{3\varepsilon}{2}}(P_{LM,2})$ are in bijection with the triangles of $X$.*

**Proof 11** (Proof of Lemma 5.9)**.** *As $\overline{uv}$ is a locally maximal edge, there is a corresponding edge spanning component $\mathcal{E}$. As $u, v$ are not faces of any other cell $\sigma \in X$, by Assumption 1 and Propositions 4.9 and 4.11, the points $p \in P_{NLM}$ which witness $\mathcal{E}$ do not witness any other edge spanning component $\mathcal{E}'$ or any triangle spanning component $\mathcal{T}$.*

*Thus, there is a single partition $P_1$ of $P_{NLM}$ which contains all the samples $p$ that witness $\mathcal{E}$. By Assumption 1, there is no other partition $P_2$ of $P_{NLM}$ that witnesses $\mathcal{E}$. Hence, $P_1$ is assigned label $-1$.*

**Proof 12** (Proof of Lemma 5.10)**.** *As $\overline{uv}$ is a locally maximal edge, there is a corresponding edge spanning component $\mathcal{E}$. Without loss of generality, assume $v$ is the face of some locally maximal cell $\sigma \neq \overline{uv}$.*

*By Assumption 1 and Propositions 4.9 to 4.11, there are samples $p_u, pv_v \in P_{NLM}$ such that*

$$\|p_u - u\|, \|p_v - v\| \leq \varepsilon.$$

*Further, there is a spanning connected component $\mathcal{C}$ which $p_v$ also witnesses but $p_u$ does not witness. Hence, there are two partitions $P_v, P_u$ which witness $\mathcal{E}$. By assumption 1 and Algorithm 5, there are no other partitions which witness $\mathcal{E}$.*

*Hence, both $P_v$ and $P_u$ are labelled with $0$ by Algorithms 7 and 8.*

**Proof 13** (Proof of Lemma 5.11)**.** *Let $\mathcal{T}$ be the triangle spanning component that corresponds to $\triangle uvw$. By Assumption 1 and propositions 4.8, 4.10 and 4.12, the samples $p \in P_{NLM}$ that witness $\mathcal{T}$ do not witness any spanning connected component $\mathcal{C} \neq \mathcal{T}$. By assumption 1 and Algorithm 5 there is a unique connected component $P_1$ that witnesses $\mathcal{T}$.*

*As $P$ is an $\varepsilon$-sample of $|X|$, and from Crefprop:nlmedge,prop:lmvertex, there are samples $p_u, p_v, p_w, p_{uv}, p_{vw}, p_{uw} \in P_1$ such that*

$$\|p_u - u\|, \|p_v - v\|, \|p_w - w\| \leq \varepsilon,$$
$$d(\overline{uv}, p_{uv}), d(\overline{vw}, p_{vw}), d(\overline{uw}, p_{uw}) \leq \varepsilon.$$

*Hence, $P_1$ is assigned label $7$ by Algorithms 7 and 8.*

**Proof 14** (Proof of Lemma 5.12)**.** *Let $\mathcal{T}$ be the triangle spanning component that corresponds to $\triangle uvw$. By Assumption 1 and Propositions 4.8 to 4.10 and 4.12, any spanning connected component $\mathcal{C}$ witnessed by samples $p \in P_{NLM}$ that witness $\mathcal{T}$ corresponds to a locally maximal cell $\tau$ such that $\triangle uvw \cap \tau \neq \emptyset$.*

*We need to split into two cases:*

1. *there is a unique locally maximal cell $\tau \in X$ with $\triangle uvw \cap \tau = v$*
2. *there are at least two locally maximal cells $\tau, \sigma \in X$, $\tau \neq \sigma$ with $\triangle uvw \cap \tau = \triangle uvw \cap \sigma = v$.*

*<u>Case 1:</u> We assumed there was a unique locally maximal $\tau$ with $\triangle uvw \cap \tau = v$, and hence, by Propositions 5.6 and 5.7 there is some spanning component $\mathcal{C}_\tau$ which corresponds to $\tau$. with By Assumption 1 and Propositions 4.8 to 4.10 and 4.12, in Algorithm 5 there is a single partition $P_1$ of $P_{NLM}$ which witnesses $\mathcal{T}$ and $\mathcal{C}_\tau$, and*

*there is a unique partition $P_2$ which witnesses just $\mathcal{T}$. Further, $P_1$ is assigned label* 0 *and $P_2$ label* 5 *by Algorithms 7 and 8.*

*Case 2: From our assumptions, there are two locally maximal cells $\tau, \sigma \in X$, $\tau \neq \sigma$ such that*

$$\tau \cap \triangle uvw = v = \sigma \cap \triangle uvw.$$

*By Propositions 5.6 and 5.7 there is some spanning component $\mathcal{C}_\tau$ which corresponds to $\tau$, and some spanning component $\mathcal{C}_\sigma$ which corresponds to $\sigma$.*

*By Assumption 1 and from Algorithm 5, there is a single partition $P_1$ of $P_{NLM}$ which witnesses $\mathcal{T}, \mathcal{C}_\tau, \mathcal{C}_\sigma$, and no partitions which witness a subset of these spanning components. This holds, by induction, for any locally maximal cell $\tau' \in X, \tau' \neq \tau, \sigma$ with $\tau' \cap \triangle uvw = v$. Similarly, there is a single partition $P_2$ of $P_{NLM}$ which witnesses only $\mathcal{T}$. Further, $P_1$ is assigned label* 0 *and $P_2$ label* 5 *by Algorithms 7 and 8.*

**Proof 15** (Proof of Lemma 5.13)**.** *Let $\mathcal{T}$ be the triangle spanning component that corresponds to $\triangle uvw$. By Assumption 1 and Propositions 4.8 to 4.10 and 4.12, any spanning connected component $\mathcal{C}$ witnessed by samples $p \in P_{NLM}$ that witness $\mathcal{T}$ corresponds to a locally maximal cell $\tau$ such that $\triangle uvw \cap \tau \neq \emptyset$.*

*We need to split into two cases:*

1. *there is a unique locally maximal cell $\tau \in X$ with $\triangle uvw \cap \tau = \overline{uv}$*
2. *there are at least two locally maximal cells $\tau, \sigma \in X$, $\tau \neq \sigma$ with $\triangle uvw \cap \tau = \triangle uvw \cap \sigma = \overline{uv}$.*

*Case 1: We assumed there was a unique locally maximal $\tau$ with $\triangle uvw \cap \tau = \overline{uv}$, and hence, by Propositions 5.6 and 5.7 there is some spanning component $\mathcal{C}_\tau$ which corresponds to $\tau$. with By Assumption 1,Propositions 4.8 to 4.10 and 4.12, in Algorithm 5 there is a single partition $P_1$ of $P_{NLM}$ which witnesses $\mathcal{T}$ and $\mathcal{C}_\tau$, and there is a unique partition $P_2$ which witnesses just $\mathcal{T}$. Further, $P_1$ is assigned label* 1 *and $P_2$ label* 4 *by Algorithms 7 and 8.*

*Case 2: From our assumptions, there are two locally maximal cells $\tau, \sigma \in X$, $\tau \neq \sigma$ such that*

$$\tau \cap \triangle uvw = \overline{uv} = \sigma \cap \triangle uvw.$$

*By Propositions 5.6 and 5.7 there is some spanning component $\mathcal{C}_\tau$ which corresponds to $\tau$, and some spanning component $\mathcal{C}_\sigma$ which corresponds to $\sigma$.*

*By Assumption 1 and from Algorithm 5, there is a single partition $P_1$ of $P_{NLM}$ which witnesses $\mathcal{T}, \mathcal{C}_\tau, \mathcal{C}_\sigma$, and no partitions which witness a subset of these spanning components. This holds, by induction, for any locally maximal cell $\tau' \in X, \tau' \neq \tau, \sigma$ with $\tau' \cap \triangle uvw = v$. Similarly, there is a single partition $P_2$ of $P_{NLM}$ which witnesses only $\mathcal{T}$. Further, $P_1$ is assigned label* 1 *and $P_2$ label* 4 *by Algorithms 7 and 8.*

**Proof 16** (Proof of Lemma 5.14)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemma 5.12. By combining the arguments at the two shared vertices, there are three partitions $P_1, P_2, P_3$ from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2$ such that $P_1$ witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2$, and $P_2$ witnesses $\mathcal{C}_2$ but not $\mathcal{C}_1$. Further, $P_3$ only witnesses $\mathcal{T}$. Hence, $P_1, P_2$ are labelled with* 0 *and $P_3$ with* 6*.*

**Proof 17** (Proof of Lemma 5.15)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions $P_1, P_2, P_3$ from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2$ such that $P_1$ witnesses $\mathcal{C}_1$ and $\mathcal{C}_2$, and $P_2$ witnesses $\mathcal{C}_2$ but not $\mathcal{C}_1$. Further, $P_3$ only witnesses $\mathcal{T}$. Hence, $P_1$ is labelled with $0$, $P_2$ with $1$ and $P_3$ with $3$.*

**Proof 18** (Proof of Lemma 5.16)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions $P_1, P_2, P_3$ from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2$ such that $P_1$ witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2$, and $P_2$ witnesses $\mathcal{C}_2$ but not $\mathcal{C}_1$. Further, $P_3$ only witnesses $\mathcal{T}$. Hence, $P_1$ is labelled with $0$, $P_2$ with $2$ and $P_3$ with $9$.*

**Proof 19** (Proof of Lemma 5.17)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions $P_1, P_2, P_3, P_4$ from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni$ such that $P_1$ witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2, \mathcal{C}_3$, $P_2$ witnesses $\mathcal{C}_2$ but not $\mathcal{C}_1, \mathcal{C}_3$, and $P_2$ witnesses $\mathcal{C}_3$ but not $\mathcal{C}_1, \mathcal{C}_2$. Further, $P_4$ only witnesses $\mathcal{T}$. Hence, $P_1, P_2$ and $P_3$ are labelled with $0$ and $P_4$ with $8$.*

**Proof 20** (Proof of Lemma 5.18)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions*

$$P_1, P_2, P_3, P_4$$

*from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni$ such that $P_1$ witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2, \mathcal{C}_3$, $P_2$ witnesses $\mathcal{C}_1, \mathcal{C}_2$ but not $\mathcal{C}_3$, and $P_2$ witnesses $\mathcal{C}_3$ but not $\mathcal{C}_1, \mathcal{C}_2$. Further, $P_4$ only witnesses $\mathcal{T}$. Hence, $P_1, P_2$ and $P_3$ are labelled with $0$ and $P_4$ with $8$.*

**Proof 21** (Proof of Lemma 5.19)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are three partitions $P_1, P_2, P_3, P_4$ from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni$ such that $P_1$ witnesses $\mathcal{C}_1$ but not $\mathcal{C}_2, \mathcal{C}_3$, $P_2$ witnesses $\mathcal{C}_1, \mathcal{C}_2$ but not $\mathcal{C}_3$, and $P_2$ witnesses $\mathcal{C}_1, \mathcal{C}_3$ but not $\mathcal{C}_2$. Further, $P_4$ only witnesses $\mathcal{T}$. Hence, $P_1$ is labelled with $3$, $P_2, P_3$ with $0$ and $P_4$ with $4$.*

**Proof 22** (Proof of Lemma 5.20)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are four partitions $P_1, P_2, P_3, P_4$ from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni$ such that $P_1$ witnesses $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$, $P_2$ witnesses $\mathcal{C}_1, \mathcal{C}_2$ but not $\mathcal{C}_3$, and $P_3$ witnesses $\mathcal{C}_1, \mathcal{C}_3$ but not $\mathcal{C}_2$. Further, $P_4$ only witnesses $\mathcal{T}$. Hence, $P_1$ is labelled with $0$, $P_2, P_3$ with $1$, and $P_3$ with $3$.*

**Proof 23** (Proof of Lemma 5.21)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are four partitions $P_1, P_2, P_3, P_4$*

*from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni, \mathcal{C}_\ni$ such that $P_1$ witnesses $\mathcal{C}_1$ and not $\mathcal{C}_2, \mathcal{C}_3$, $P_2$ witnesses $\mathcal{C}_1$ and not $\mathcal{C}_2, \mathcal{C}_3$, and $P_3$ witnesses $\mathcal{C}_3$ but not $\mathcal{C}_1, \mathcal{C}_2$. Further, $P_4$ only witnesses $\mathcal{T}$. Hence, $P_1, P_2, P_3$ are labelled with 0, and $P_4$ with 8.*

**Proof 24** (Proof of Lemma 5.22)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments, there are five partitions*

$$P_1, P_2, P_3, P_4, P_5$$

*from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\ni, \mathcal{C}_\ni, \mathcal{C}_\triangle$ such that $P_1$ witnesses $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_\triangle$ and not $\mathcal{C}_3$, $P_2$ witnesses $\mathcal{C}_2$ and not $\mathcal{C}_1, \mathcal{C}_3, \mathcal{C}_4$, $P_3$ witnesses $\mathcal{C}_2, \mathcal{C}_3$ but not $\mathcal{C}_1, \mathcal{C}_4$, and $P_4$ witnesses $\mathcal{C}_4$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$. Further, $P_5$ only witnesses $\mathcal{T}$, and hence $P_4$ only witnesses $\mathcal{T}$. Hence, $P_1, P_2$ are labelled with 0, $P_3$ with 1, and $P_4, P_5$ with 3.*

**Proof 25** (Proof of Lemma 5.23)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are six partitions*

$$P_1, P_2, P_3, P_4, P_5, P_6$$

*from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4, \mathcal{C}_5$ such that $P_1$ witnesses $\mathcal{C}_1, \mathcal{C}_4$ and not $\mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5$, $P_2$ witnesses $\mathcal{C}_2, \mathcal{C}_4, \mathcal{C}_5$ and not $\mathcal{C}_1, \mathcal{C}_3$, $P_3$ witnesses $\mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5$ but not $\mathcal{C}_1, \mathcal{C}_4$, $P_4$ witnesses $\mathcal{C}_4$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5$, and $P_5$ witnesses $\mathcal{C}_5$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$. Further, $P_6$ only witnesses $\mathcal{T}$, and hence $P_1, P_2, P_3$ are labelled with 0, $P_4, P_5, P_6$ with 3.*

**Proof 26** (Proof of Lemma 5.24)**.** *Let $\mathcal{T}$ be the triangle spanning component which corresponds to $\triangle uvw$. Then, the proof is an adaption of the proof of Lemmas 5.12 and 5.13. By combining the arguments there are six partitions*

$$P_1, P_2, P_3, P_4, P_5, P_6$$

*from Algorithm 5 which witness $\mathcal{T}$, and there are spanning connected components $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4, \mathcal{C}_5, \mathcal{C}_6$ such that $P_1$ witnesses $\mathcal{C}_1, \mathcal{C}_4, \mathcal{C}_6$ and not $\mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5$, $P_2$ witnesses $\mathcal{C}_2, \mathcal{C}_4, \mathcal{C}_5$ and not $\mathcal{C}_1, \mathcal{C}_3, \mathcal{C}_6$, $P_3$ witnesses $\mathcal{C}_3, \mathcal{C}_5, \mathcal{C}_6$ but not $\mathcal{C}_1, \mathcal{C}_2\mathcal{C}_4$, $P_4$ witnesses $\mathcal{C}_4$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_5, \mathcal{C}_6$, and $P_5$ witnesses $\mathcal{C}_5$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4, \mathcal{C}_6$, and $P_6$ witnesses $\mathcal{C}_6$ but not $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4, \mathcal{C}_5$. Hence $P_1, P_2, P_3$ are labelled with 0, $P_4, P_5, P_6$ with 3.*

## References


Bokor, Y., Turner, K., Williams, C.: Reconstructing linearly embedded graphs: A first step to stratified space learning. Foundations of Data Science (2021)

Cheng, S., Dey, T.K., Ramos, E.: Manifold Reconstruction from Point Samples. Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (2005)

Dey, T.K.: Curve and Surface Reconstruction. Cambridge University Press, Cambridge (2007)

Dey, T.K., Wang, Y.: Dimension Detection with Local Homology. Proceedings of the 26th Canadian Conference on Computational Geometry (2014)

Bendich, P., Wang, B., Mukherjee, S.: Towards Stratification Learning through Homology Inference. AAAI Fall Symposium on Manifold Learning and its Applications (2010)

Stolz, B., Tanner, J., Harrington, H., Nanda, V.: Geometric anomaly detection in data. Proceedings of the National Academy of Sciences (2020)

Aanjaneya, M., F., C., D., C., M., G., Guibas, L., Morozov, D.: Metric graph reconstruction from noisy data. International Journal of Computational Geometry & Applications **22**, 205–325 (2012) https://doi.org/10.1142/S0218195912600072

Chazal, F., Cohen-Steiner, D., Lieutier, A.: A sampling theory for compact sets in euclidean space. Discrete & Computational Geometry **41**(3), 461–479 (2009) https://doi.org/10.1007/s00454-009-9144-8

Bendich, P., Cohen-Steiner, D., Edelsbrunner, H., Harer, J., Morozov, D.: Inferring local homology from sampled stratified spaces. 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), 536–546 (2007)

Bokor, Y.: Geometric and topological shape analysis: Investigating and summarising the shape of data. PhD thesis, Mathematical Science Institute, ANU Colleges of Science, The Australian National University (2023)

Carlsson, G.: Topological pattern recognition for point cloud data. Acta Numerica (2014) https://doi.org/10.1017/S0962492914000051

Chazal, F., Lieutier, A.: Stability and computation of topological invariants of solids in $\mathbb{R}^n$. Discrete & Computational Geometry **37**(4), 601–617 (2007) https://doi.org/10.1007/s00454-007-1309-8

Hatcher, A.: Algebraic Topology. Cambridge University Press, Cambridge (2000)