Inferring Local Homology from Sampled Stratified Spaces [*]

Paul Bendich[†]   David Cohen-Steiner[‡]   Herbert Edelsbrunner[§]   John Harer[¶]   Dmitriy Morozov[‖]

Abstract

We study the reconstruction of a stratified space from a possibly noisy point sample. Specifically, we use the vineyard of the distance function restricted to a 1-parameter family of neighborhoods of a point to assess the local homology of the stratified space at that point. We prove the correctness of this assessment under the assumption of a sufficiently dense sample. We also give an algorithm that constructs the vineyard and makes the local assessment in time at most cubic in the size of the Delaunay triangulation of the point sample.

Keywords. Topological data analysis, local homology, persistence, stratified spaces, simplicial complexes, power diagrams, Delaunay triangulations, algorithms.

1 Introduction

Much of the progress in today's experimental sciences is predicated on the ability to collect larger sets of more accurate measurements faster. Each data element is a tuple interpreted as a point in a space of the appropriate dimension. The resulting point set is often referred to as a *point cloud* so we are reminded that there is a lot of accumulated data. The main task is to detect patterns and to infer properties of the measured process from the structure of the point cloud.

Motivation. A common phenomenon in experimental measurements is that the data appears to describe a space whose intrinsic dimension is significantly smaller than that of the ambient space $\mathbb{R}^n$. This statement needs some clarification since every finite point set is, by definition, zero-dimensional. What we mean is that there is a relatively simple subset $\mathbb{X} \subseteq \mathbb{R}^n$ of dimension $m \ll n$ such that all data points lie on or near this subset. The reason for this phenomenon is perhaps self-inflicted by our inability to make sense of processes that depend on a large number of independent parameters. The problem of reconstructing this subset (or one such subset from the class of possibilities) is often referred to as *manifold learning* [4]. The name betrays the tacit assumption that the subset is thought to have the topological structure of a manifold. In other words, it is locally homeomorphic to $\mathbb{R}^m$, or possibly to the m-dimensional half-space if we allow the manifold to have boundary. In case this assumption is grossly false it is suggested that these violations are artifacts of the mapping into the ambient space.

Similar to [18], we take a deliberate departure from the manifold assumption. While their methods are statistical in nature, we use local homology to recognize locations where the assumption is violated. Specifically, we consider samplings of spaces that are partitioned into strata, each a manifold of dimension m or less, and we focus on the characterization of how these strata connect to each other. Stratified spaces can be described relatively compactly while significantly generalizing the class of spaces beyond manifolds. They include smooth images of manifolds into $\mathbb{R}^n$ and permit different local dimensions at different locations, but they do not include sets of fractal dimension. Since our goal is to deal with scientific data, the inability to describe fractal behavior might be considered a serious drawback. We argue otherwise, delegating the expression of fractal or chaotic behavior to the multi-scale description of the data. While this is not yet addressed in this paper, we lay the foundations by parametrizing all our results in terms of scale parameters.

Results and prior work. Stratified spaces have been studied extensively in the mathematical literature [16, 27]. Particularly relevant to the line of work started in this paper is the development of intersection homology [15] and of persistence for intersection homology [3]. There is a striking paucity in computational studies of the reconstruction of

[†]Department of Mathematics, Duke University, Durham, North Carolina, USA.

[‡]INRIA, 2004 Route des Lucioles, BP93, Sophia-Antipolis, France.

[§]Departments of Computer Science and Mathematics, Duke University, Durham, and Geomagic, Research Triangle Park, North Carolina, USA.

[¶]Department of Mathematics and Section in Computational Biology and Bioinformatics, Duke University, Durham, North Carolina, USA.

[‖]Department of Computer Science, Duke University, Durham, North Carolina, USA.

stratified spaces from point samples which testifies for the technical difficulties caused by the presence of singularities. In general, the reconstruction of spaces from point samples is an important topic in a number of fields, each putting its own emphasis on the subject. Computer graphics and visualization stresses fast algorithms inspired by work in numerical analysis and image processing and focuses on data that describes surfaces in $\mathbb{R}^3$ [20, 26]. Computational geometry favors combinatorial algorithms based on Delaunay triangulations [10] and provides proofs of correctness under assumptions of dense sampling [1, 5]. Machine learning uses statistical methods to study high-dimensional data that describes relatively low-dimensional manifolds [2, 25]. Finally, topological data analysis relies on algebraic methods to reveal the topological structure of high-dimensional data [9]. In this paper we combine characteristics found in computational geometry and in topological data analysis:

- we turn the algebraic concept of local homology into a multi-scale notion by constructing extended series of homology groups;

- we describe the $(\alpha|r)$-vineyard, which we introduce as a practical tool for studying the local homology of a sampled space;

- we explain how the $(\alpha|r)$-vineyard expresses the local homology of the sampled space at a point and prove the relation under the assumption of a dense sampling;

- we give an algorithm that computes the $(\alpha|r)$-vineyard of a point cloud in time cubic in the number of simplices in the Delaunay triangulation.

The main technical ingredients used to obtain these results are from the theory of persistent homology [12, 28] and weighted Delaunay triangulations [11]. In particular, we make extensive use of the stability of extended persistence diagrams [6, 7] and the construction of vineyards [8] from a filtration of alpha shapes [13].

Outline. Section 2 introduces homology groups and related structures associated with a topological space. Section 3 defines the $(\alpha|r)$-vineyard, a 1-parameter family of persistence diagrams. Section 4 describes the class of spaces we consider in our reconstruction effort. Section 5 proves that a sufficiently dense sampling on the space permits the determination of the local homology of the space at a point. Section 6 explains an algorithm for constructing the $(\alpha|r)$-vineyard and Section 7 proves its correctness. Section 8 discusses the contributions of this paper and mentions extensions.

2 Homological Structures

In this section, we introduce a variety of homological structures associated with a topological space.

Spaces, groups, and maps. Let Y be a subset of $\mathbb{R}^n$. Its subspace topology is obtained by intersecting Y with all open sets in the Euclidean topology of $\mathbb{R}^n$. The homology functor maps the topological space Y to a series of *homology groups*, $\mathsf{H}_p(Y)$, one for each dimension p. In this paper, we use $\mathbb{Z}/2\mathbb{Z}$-coefficients; see Munkres [24] or other texts in algebraic topology for background information on homology. It will often be convenient to suppress the dimension from the notation which we will do by writing

$$\mathsf{H}(Y) \quad = \quad (\ldots, \mathsf{H}_p(Y), \mathsf{H}_{p+1}(Y), \ldots)$$

for the infinite series of homology groups. For negative dimensions and for dimensions beyond n these homology groups are necessarily trivial. Let Y' be another subset of $\mathbb{R}^n$ with induced subspace topology. If $Y \subseteq Y'$ then Y is a subspace of Y' and inclusion induces a homomorphism $\mathsf{H}_p(Y) \to \mathsf{H}_p(Y')$ between homology groups of the same dimension. As an example consider the space Y in Figure 1. It has a single hole marked by the dashed circle that surrounds it. This circle generates a non-trivial class $\gamma \in \mathsf{H}_1(Y)$. In contrast, the same circle bounds in Y' which implies that the homomorphic image of γ in $\mathsf{H}_1(Y')$ is 0. As with groups we combine the homomorphisms to form a series that maps $\mathsf{H}(Y)$ to $\mathsf{H}(Y')$ componentwise.
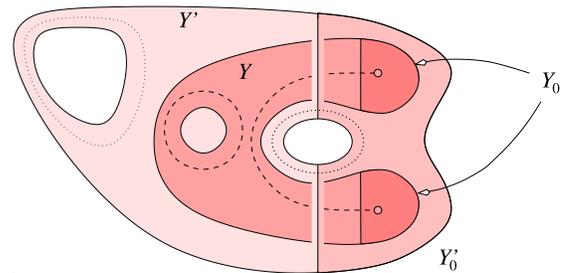

Figure 1: The dimension 1 homology groups of Y (dashed closed curve) and of Y' (dotted closed curves) have rank 1 and 2 but the map induced by inclusion has only rank 0. The relative homology groups of (Y, Y_0) (dashed curves) and of (Y', Y_0') (dotted curves) both have rank 2 but the map induced by inclusion of pairs has only rank 1.

We also consider pairs of topological spaces $Y_0 \subseteq Y$. The homology functor maps this pair to the *relative homology groups*, $\mathsf{H}_p(Y, Y_0)$. As before we get one group for each dimension p and we write $\mathsf{H}(Y, Y_0)$ for the series. Each relative homology group consists of classes generated by relative cycles, that is, chains in Y whose boundary is either empty or contained in Y_0. For example in Figure 1, the dimension 1 relative homology group of the pair (Y, Y_0) is generated by two classes, the absolute class from before and the relative class generated by the dashed curve whose ends lie in Y_0. Let $Y_0' \subseteq Y'$ be another pair of topological spaces. We write $(Y, Y_0) \subseteq (Y', Y_0')$ if $Y \subseteq Y'$ and $Y_0 \subseteq Y_0'$. In this case, inclusion induces again a homomorphism $\mathsf{H}_p(Y, Y_0) \to \mathsf{H}_p(Y', Y_0')$. In the example in Figure

2

1, the image of the absolute class is zero, as before, and the image of the relative class is another relative class of the pair (Y', Y_0'), namely the one generated by the dotted cycle surrounding the hole on the right in Y'. As before, we simplify notation by considering the series of homomorphisms mapping $\mathsf{H}(Y, Y_0)$ to $\mathsf{H}(Y', Y_0')$ componentwise.

Distance, filtrations, and diagrams. Let $d_Y : \mathbb{R}^n \to \mathbb{R}$ be the *distance function* defined by $d_Y(x) = \inf_{y \in Y} \|x - y\|$. For each real number α the corresponding *sublevel set* consists of all points at distance at most α, $Y_\alpha = d_Y^{-1}[0, \alpha]$; see Figure 2. Clearly $Y_\alpha \subseteq Y_{\alpha'}$ whenever $\alpha \leq \alpha'$. We call
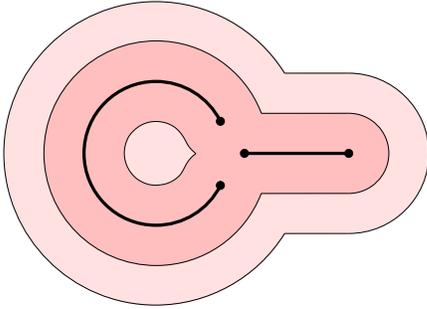

Figure 2: The space Y consists of the brace on the left and the stick to its right. It is also the first sublevel set in the picture, $Y = d_Y^{-1}[0, 0]$. The second sublevel set merges the two components and creates a hole. The third sublevel set fills the hole.

α an *absolute homological regular value* if there is a sufficiently small $\varepsilon > 0$ such that the maps between homology groups induced by the inclusion $Y_{\alpha-\delta} \subseteq Y_{\alpha+\delta}$ form a series of isomorphisms for every $0 < \delta < \varepsilon$. Otherwise, α is an *absolute homological critical value*. We also define *superlevel sets* $Y^\alpha = d_Y^{-1}[\alpha, \infty)$, using them to define pairs $(\mathbb{R}^n, Y^\alpha)$. We call α a *relative homological regular value* if there is a sufficiently small $\varepsilon > 0$ such that the maps between relative homology groups induced by the inclusion $(\mathbb{R}^n, Y^{\alpha+\delta}) \subseteq (\mathbb{R}^n, Y^{\alpha-\delta})$ form a series of isomorphisms for every $0 < \delta < \varepsilon$. Otherwise, α is a *relative homological critical value*. Following [6] we call a function *tame* if it has finitely many (absolute and relative) homological critical values and its sublevel sets have finite rank (absolute and relative) homology groups.

We consider the extended sequence of absolute and relative homology groups as introduced in [7],

$$0 \to \mathsf{H}(Y_\alpha) \to \ldots \to \mathsf{H}(\mathbb{R}^n)$$
$$\to \mathsf{H}(\mathbb{R}^n, Y^\alpha) \to \ldots \to 0, \qquad (1)$$

where α increases from 0 to infinity going up during the first half and then decreases from infinity to 0 coming down during the second half of the sequence. A homology class γ of Y_α is *born at* Y_α if it is not in the image of the map $\mathsf{H}(Y_{\alpha-\delta}) \to \mathsf{H}(Y_\alpha)$ for any $\delta > 0$. Furthermore, a homology class γ born at Y_α *dies entering* $Y_{\alpha'}$ if the image of γ

in $\mathsf{H}(Y_{\alpha'-\delta})$ is not in the image of $\mathsf{H}(Y_{\alpha-\delta})$, for any $\delta > 0$, but the image of γ in $\mathsf{H}(Y_{\alpha'})$ is. Similarly, a class can be born at $(\mathbb{R}^n, Y^\alpha)$ and die entering $(\mathbb{R}^n, Y^{\alpha'})$. It is also possible that a class is born going up and dies coming down. Following [7], we record the evolution of homology classes by drawing a point (α, α') for each class that is born at Y_α or $(\mathbb{R}^n, Y^\alpha)$ and dies entering $Y_{\alpha'}$ or $(\mathbb{R}^n, Y^{\alpha'})$. The resulting multiset of points defined by the classes in the p-th homology groups is referred to as the *dimension p persistence diagram*, $\mathrm{Dgm}_p(d_Y)$. As usual we drop the dimension to denote the series of persistence diagrams, $\mathrm{Dgm}(d_Y)$. We will distinguish between classes born and dying going up, classes born going up and dying coming down, and classes born and dying coming down. The corresponding three types are referred to as *ordinary*, *extended*, and *relative points* and they make up the *ordinary*, *extended*, and *relative subdiagrams* of the persistence diagram.
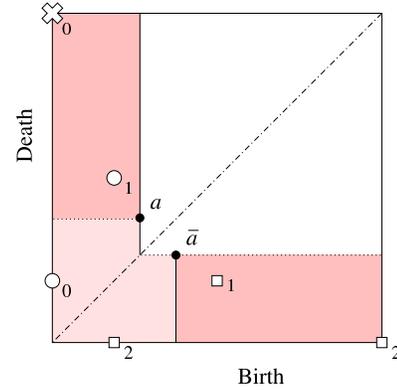

Figure 3: The persistence diagrams of the distance function in Figure 2. We draw time of birth from left to right and time of death from bottom to top, each ranging from 0 to ∞. The white dots, crosses, and squares represent the ordinary, extended, and relative points in the diagrams of dimension given by the subscripts. The shading shows quadrants and half-planes defined by the points $a = (\xi, \zeta)$ and $\bar{a} = (\zeta, \xi)$.

To recover information from the diagrams we count points in subregions. Given $a = (\xi, \zeta)$ we count the classes that are alive during the entire interval, on the way up or the way down. Assuming $\xi \leq \zeta$ this number is

$$
\begin{aligned}
\#_p^a(d_Y) &= \mathrm{rank}\left(\mathrm{im}\left(\mathsf{H}_p(Y_\xi) \to \mathsf{H}_p(Y_\zeta)\right)\right) \\
&+ \mathrm{rank}\left(\mathrm{im}\left(\mathsf{H}_p(\mathbb{R}^n, Y^\zeta) \to \mathsf{H}_p(\mathbb{R}^n, Y^\xi)\right)\right) \\
&- \mathrm{rank}\left(\mathrm{im}\left(\mathsf{H}_p(Y_\xi) \to \mathsf{H}_p(\mathbb{R}^n, Y^\xi)\right)\right).
\end{aligned}
$$

Of course we get a number for each dimension p and we write $\#^a(d_Y)$ for the series. To cover the other case, when $\zeta \leq \xi$, we set $\#^a(d_Y) = \#^{\bar{a}}(d_Y)$ with $\bar{a} = (\zeta, \xi)$. As illustrated in Figure 3, $\#^a$ counts the ordinary points in the upper left quadrant, $[0, \xi] \times (\zeta, \infty)$, the relative points in the lower right quadrant, $[\zeta, \infty) \times [0, \xi)$, as well as the extended points in the union of the left and lower half-planes, $[0, \xi] \times \mathbb{R} \cup \mathbb{R} \times [0, \xi)$. The resulting series for the example

illustrated in Figures 2 and 3 is then $\#^a = (\ldots, 1, 2, 1, \ldots)$, where we only show the numbers for dimensions $p = 0, 1, 2$.

Local homology. Let $z \in \mathbb{R}^n$ be a point and $d_z : \mathbb{R}^n \to \mathbb{R}$ the distance function defined by $d_z(x) = \|x - z\|$. We write $B_r = d_z^{-1}[0, r]$ and $B^r = d_z^{-1}[r, \infty)$ for the sublevel and superlevel sets defined by r. Fix $\alpha \geq 0$ and consider Y_α, the space of points at distance at most α from Y. The sublevel sets and superlevel sets restricted to Y_α are $Y_\alpha \cap B_r$ and $Y_\alpha \cap B^r$. Traditionally, the *local homology groups* at a point z in a space Y_α are defined to be the relative homology groups $\mathsf{H}(Y_\alpha, Y_\alpha - z)$ [24]. In words, a *local cycle* γ at z is a chain whose boundary misses z. Additionally, the boundary of γ must also miss some small open set $Y_\alpha \cap \mathrm{int}\, B_r$ containing z, that is, γ belongs to $\mathsf{H}(Y_\alpha, Y_\alpha - (Y_\alpha \cap \mathrm{int}\, B_r)) = \mathsf{H}(Y_\alpha, Y_\alpha \cap B^r)$. Now for any $s < r$, we have a map $\mathsf{H}(Y_\alpha, Y_\alpha \cap B^r) \to \mathsf{H}(Y_\alpha, Y_\alpha \cap B^s)$ induced by inclusion and excision. Our local cycle γ must lie in the image of this map for all possible choices of s. As a consequence, we see that the above definition of local homology at a point z is equivalent to the *direct limit*, $\lim_{r \to 0} \mathsf{H}(Y_\alpha, Y_\alpha \cap B^r)$. To make this a multi-scale concept we consider again the extended sequence of homology groups,

$$0 \to \mathsf{H}(Y_\alpha \cap B_r) \to \ldots \to \mathsf{H}(Y_\alpha)$$
$$\to \mathsf{H}(Y_\alpha, Y_\alpha \cap B^r) \to \ldots \to 0, \qquad (2)$$

where r first increases from 0 to infinity and then decreases from infinity back to 0. As before, we record the evolution of homology classes using the thus defined series of persistence diagrams, $\mathrm{Dgm}(d_z | Y_\alpha)$. The relative subdiagrams contain the information most directly relevant to estimating the local homology at z.

Discontinuity in α. The extended sequence of homology groups (2) provides a feasible approach to assessing local homology if the space, Y_α, is fixed. In the context of this paper, we assume that the space has not been reconstructed, and we examine it at various scales by varying α. A drawback of the above construction is that the diagrams are not continuous in α. To see this let $0 < a' < a''$ be the distance thresholds of the three sublevel sets shown in Figure 2. Let z be the right endpoint of the stick in Y. For $\alpha = a'$ we have a one-dimensional homology class, γ, that is born going up and dies coming down with r. The class is represented by an off-diagonal point in the extended subdiagram of $\mathrm{Dgm}_1(d_z | Y_{a'})$. In contrast, there are no one-dimensional classes for $\alpha = 0$ and for $\alpha = a''$. The class γ first appears when α reaches half the distance between the ends of the brace and the left endpoint of the stick. Right from the start, the representing point in the diagram is some distance away from the diagonal. Later, γ disappears when α reaches the radius of the brace and the representing point in the diagram merges into the diagonal.

3 The $(\alpha | r)$-Vineyard

In this section, we introduce the main algebraic tool we use to study sampled stratified spaces.

Two filtrations. To obtain a continuous expression of the 2-parameter variation, we swap the order and vary α to construct the diagrams. Let $z \in \mathbb{R}^n$, as before, but now fix $r > 0$. Recall that $d_Y : \mathbb{R}^n \to \mathbb{R}$ is the distance function defined by $Y \subseteq \mathbb{R}^n$. The sublevel and superlevel sets of its restrictions to the ball of radius r around z are $Y_\alpha \cap B_r$ and $Y^\alpha \cap B_r$. Going first up with α from 0 to ∞ and then down from ∞ to 0 we get

$$0 \to \mathsf{H}(Y_\alpha \cap B_r) \to \ldots \to \mathsf{H}(B_r)$$
$$\to \mathsf{H}(B_r, Y^\alpha \cap B_r) \to \ldots \to 0, \qquad (3)$$

and we write $\mathrm{Dgm}(d_Y | B_r)$ for the series of persistence diagrams that records the evolution of the homology classes in the sequence. The notion of local homology suggests we modify the filtration (3) and take the homology of Y_α within B_r relative to Y_α within the sphere ∂B_r. Constructing the extended sequence by first going up with α from 0 to ∞ and then down from ∞ to 0, we get

$$0 \to \mathsf{H}(Y_\alpha \cap B_r, Y_\alpha \cap \partial B_r) \to \ldots \to \mathsf{H}(B_r, \partial B_r)$$
$$\to \mathsf{H}(B_r, \partial B_r \cup (Y^\alpha \cap B_r)) \to \ldots \to 0. \qquad (4)$$

The evolution of the homology classes is again recorded in the series of persistence diagrams, which we denote as $\mathrm{Dgm}(d_Y | (B_r, \partial B_r))$. In the rest of the paper we need a mild assumption on Y, namely that the restrictions of its distance function to balls and to ball-sphere pairs are tame.

Equivalence of diagrams. We now show that the two diagrams contain the same information. Specifically, we establish isomorphisms between the homology groups in (3) and (4) and show that the corresponding pairings are dual and thus give the same diagrams. To shorten the notation and clarify the relations we set $X = Y_\alpha \cap B_r$, decompose its boundary $\partial X = F \cup G$ where $F = Y_\alpha \cap Y^\alpha \cap B_r$ and $G = \partial B_r \cap Y_\alpha$, and set $A = F \cap G$. Generically, X is an n-manifold with boundary, F and G are $(n-1)$-manifolds with boundary, and A is an $(n-2)$-manifold without boundary. Assuming tameness of Y, we use excision to rewrite (3) and (4), running them anti-parallel against each other:

$$\begin{array}{ccccc}
\to & \mathsf{H}_{n-p}(X) & \to & \mathsf{H}_{n-p}(X, F) & \to \\
& \otimes & & \otimes & \\
\leftarrow & \mathsf{H}_p(X, \partial X) & \leftarrow & \mathsf{H}_p(X, G) & \leftarrow \\
& \downarrow & & \downarrow & \\
& \mathbb{Z}/2\mathbb{Z} & & \mathbb{Z}/2\mathbb{Z} &
\end{array}$$

By Lefschetz duality, the first vertical pairing is perfect. This means that the paired groups are isomorphic and the persistence pairs in the first half of the two sequences are the same [7]. The other vertical pairing is also perfect, but it takes a little more effort to prove this.

ISOMORPHISM LEMMA. For every dimension p, the intersection pairing on X induces a perfect pairing

$$\mathsf{H}_{n-p}(X, F) \otimes \mathsf{H}_p(X, G) \to \mathbb{Z}/2\mathbb{Z}.$$

PROOF. First notice that by excision, the relative homology groups $\mathsf{H}_p(\partial X, G)$ and $\mathsf{H}_p(F, A)$ are isomorphic. Next consider the exact cohomology sequence of the pair (X, F), shown in the bottom row in Table 1, and the exact homology sequence of the triple $(X, \partial X, G)$,

$$\to \mathsf{H}_{p+1}(X, \partial X) \to \mathsf{H}_p(\partial X, G) \to \mathsf{H}_p(X, G)$$
$$\to \mathsf{H}_p(X, \partial X) \to \mathsf{H}_{p-1}(\partial X, G) \to$$

Replacing $\mathsf{H}_p(\partial X, G)$ by $\mathsf{H}_p(F, A)$ we get the diagram in Table 1. Here each vertical arrow is the Poincaré-Lefschetz duality map defined by $\gamma \to f_\gamma$ where $f_\gamma(\delta) = \gamma \cdot \delta$, the intersection number between the two classes. It is not difficult to check that this diagram commutes. The two vertical maps on the left and the two vertical maps on the right are isomorphisms by Poincaré-Lefschetz duality. The Steenrod Five-Lemma then tells us that the center vertical map is also an isomorphism [24]. Finally we note that the Poincaré duality map $\mathsf{H}_p(X, G) \to \mathsf{H}^{n-p}(X, F)$ being an isomorphism implies that the intersection pairing is perfect. Indeed, every non-zero $\gamma \in \mathsf{H}_p(X, G)$ has at least one $\delta \in \mathsf{H}_p(X, G)$ such that $f_\gamma(\delta) = \gamma \cdot \delta \neq 0$. The claim follows. $\qquad \Box$

We see that the pairings between the groups in (3) and (4) are perfect and all diagrams commute. It follows that if we use the superscript T to denote reflection across the diagonal we have $\mathrm{Dgm}_p(d_Y | B_r) = \mathrm{Dgm}_{n-p}^T(d_Y | (B_r, \partial B_r))$ for all dimensions p and all radii r.

Vineyard. An important result is that the diagrams of the filtrations in (3) and (4) vary continuously with r. It is convenient to show this for (3) again using the assumption of tameness. To prove continuity we need a notion of distance between diagrams. This is provided by the *bottleneck distance* which is the infimum, over all perfect matchings, of the supremum L_∞-distance, over all pairs of matched points [6]. To extend this notion to series of persistence diagrams we let d_B be the supremum bottleneck distance between corresponding diagrams.

STABILITY LEMMA. Let $Y \subseteq \mathbb{R}^n$ and $z \in \mathbb{R}^n$ such that the restriction of $d_Y : \mathbb{R}^n \to \mathbb{R}$ to any ball and any ball-sphere pair centered at z is tame. Then the bottleneck distance between the series of persistence diagrams for two radii $r \leq r'$ is $d_B(\mathrm{Dgm}(d_Y | B_r), \mathrm{Dgm}(d_Y | B_{r'})) \leq r' - r$.

PROOF. Letting $f, g : \mathbb{R}^n \to \mathbb{R}$ be defined by $f(x) = d_Y(z + rx)$ and $g(x) = d_Y(z + r'x)$, the restrictions of d_Y correspond to the restrictions of f and g to the unit ball. Changing coordinates does not affect the diagrams. Since d_Y is a distance function we have

$$
\begin{aligned}
|f(x) - g(x)| &= |d_Y(z + rx) - d_Y(z + r'x)| \\
&\leq \|x\| (r' - r).
\end{aligned}
$$

Since $\|x\| \leq 1$ the difference between the two functions is $\|f - g\|_\infty \leq r' - r$. The extension of the Stability Theorem in [6] to extended persistence as described in [7] implies the claim. $\qquad \Box$

The stability of the persistence diagram suggests we vary r within $[0, \infty)$ and describe the homology in the neighborhood of $z \in \mathbb{R}^n$ by the resulting 1-parameter family of persistence diagrams. Stacking up the diagrams in $\mathbb{R}^3$ using r as the third coordinate, each point sweeps out a curve which we refer to as a *vine*. Together the vines form a collection of curves which we refer to as the *vineyard* of the two distance functions; see [8]. Specifically, we denote the vineyard obtained by stacking up the dimension p persistence diagrams by $\mathrm{Vnrd}_p(d_Y | d_z)$ and the series of vineyards by $\mathrm{Vnrd}(d_Y | d_z)$. On occasion we call this the series of $(\alpha | r)$-*vineyards* thus emphasizing that the diagrams are obtained by varying the threshold α for the distance to Y while fixing the threshold r for the distance to z, and the vines are obtained by varying r. This series of vineyards is the main technical ingredient in our approach to understanding sampled stratified spaces.

4 Spaces

In this section, we introduce the type of topological spaces to which our methods apply. They are not limited to manifolds.

Stratification. Recall that a topological m-manifold is a space $\mathbb{M}$ such that every point $z \in \mathbb{M}$ has a neighborhood homeomorphic to $\mathbb{R}^m$. If a space fails to be a manifold, it is because of the existence of singular points where no such neighborhood exists. For example, the figure-8 is not a 1-manifold; the singular crossing point has no neighborhood homeomorphic to $\mathbb{R}$. On the other hand, every other point has such a neighborhood; in other words the figure-8 minus the crossing point is a 1-manifold and the crossing point itself is a 0-manifold. In general, a *stratification* of a topological space $\mathbb{X}$ is a filtration by closed subsets,

$$\emptyset = \mathbb{X}_{-1} \subseteq \mathbb{X}_0 \subseteq \ldots \subseteq \mathbb{X}_{m-1} \subseteq \mathbb{X}_m = \mathbb{X},$$

such that $\mathbb{X}_i - \mathbb{X}_{i-1}$ is a (possibly empty) i-manifold for each i. The set $\mathbb{X}_i - \mathbb{X}_{i-1}$ is called the *i-stratum* and its components are the *dimension i pieces* of $\mathbb{X}$.

Local structure. The above definition does not require that the points on a piece have similar neighborhoods outside the piece. Such requirements are usually added by extra conditions. Although there are many different conditions that might be added, each with its own subtleties [21], the following will do for our purpose. A stratified space $\mathbb{X} \subseteq \mathbb{R}^n$ with a stratification as above is a *cs-space* if every point $x \in \mathbb{X}_i - \mathbb{X}_{i-1}$ has a neighborhood in $\mathbb{X}$ homeomorphic to the product of an open i-ball in $\mathbb{X}_i - \mathbb{X}_{i-1}$ and the open

$$\begin{array}{ccccccccc}
\rightarrow & \mathsf{H}_{p+1}(X, \partial X) & \rightarrow & \mathsf{H}_p(F, A) & \rightarrow & \mathsf{H}_p(X, G) & \rightarrow & \mathsf{H}_p(X, \partial X) & \rightarrow & \mathsf{H}_{p-1}(F, A) & \rightarrow \\
& \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & \\
\rightarrow & \mathsf{H}^{n-p-1}(X) & \rightarrow & \mathsf{H}^{n-p-1}(F) & \rightarrow & \mathsf{H}^{n-p}(X, F) & \rightarrow & \mathsf{H}^{n-p}(X) & \rightarrow & \mathsf{H}^{n-p}(F) & \rightarrow
\end{array}$$

Table 1: Commuting diagram with isomorphisms between the terms in the exact homology sequence of the triple $(X, \partial X, G)$ at the top and the exact cohomology sequence of the pair (X, F) on the bottom.

cone on a compact topological space. The homeomorphism is assumed to take the product of the open i-ball and the cone point to the intersection of the neighborhood with $\mathbb{X}_i$. This is illustrated in Figure 4 where $\mathbb{X}$ is a torus with one of the meridian circles pinched to a point and a disk stretched taut across its tunnel. If we remove the boundary circle of
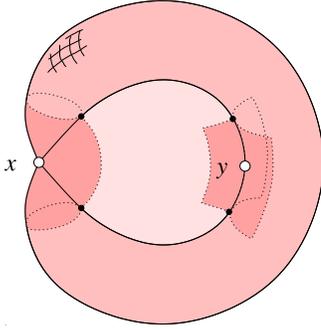

Figure 4: A 2-dimensional cs-space. The neighborhood of the pinch point x and of another point y along the boundary circle of the disk are highlighted. The two neighborhoods are not homeomorphic but they have the same homotopy type and therefore isomorphic homology groups.

the disk, we disconnect $\mathbb{X}$ and obtain a 2-manifold. The removed circle itself is a 1-manifold. However, the local structure is not uniform throughout the circle as the pinch point has a different neighborhood than any other point on the circle. Clearly, the pinch point is distinguishable from any other point on the circle.

For a cs-space, the cone in the definition of the neighborhood at a point x depends only on the piece that contains x. Since this piece is itself a manifold, the open balls are also homeomorphic. Hence the condition on the neighborhoods enforces exactly the requirement that each point in a piece has the same local structure in $\mathbb{X}$.

Connection with local homology. We note that the filtration in the definition of the stratified space is not unique. However, there is a natural coarsest filtration [17] which consists of the components in the partition of $\mathbb{X}$ defined by calling points x and y equivalent if there exist neighborhoods of x and y and a homeomorphism between these neighborhoods that maps x to y. Any cs-space meets this condition already; for the coarsest filtration, we just impose the converse.

Now if two points x and y have such neighborhoods, then their local homology groups are also the same. It is the con-

trapositive of this statement that we hope to use in finding the best stratification of point cloud data. Of course, our approach must be adapted to the persistence framework of homology outlined in the previous section. The goal is to classify points in terms of the shapes of their vineyards. The next section aims at quantifying what exactly this should mean.

5   Local Homology Inference

In this section, we prove that even with rather mild assumptions on the sampling of a space it is possible to infer its local homology. Perhaps more important than the guaranteed recognition is the interpretation of our result as describing the set of spaces that can possibly give rise to the sample.

Sample. The data we consider is a finite set of points, $U \subseteq \mathbb{R}^n$. It will be convenient to index the points in this set as u_i. We assume that U is sampled from or near a compact space $\mathbb{X} \subseteq \mathbb{R}^n$. For example, $\mathbb{X}$ may be a compact stratified space but the existence of a stratification will play no role in what we prove in this section. It will, however, be important that the diagram of the restricted distance functions of $\mathbb{X}$ be stable. We therefore assume that $d_{\mathbb{X}}|B_r$ is tame for every $z \in \mathbb{R}^n$ and every ball B_r centered at z. As mentioned earlier, this is a rather mild assumption whose violation requires infinitely many oscillations, like in the topologist's sine curve [24, p. 168], or similar phenomena. Recall that the space $\mathbb{X}$ is unknown and the main question we ask is how much we can find out about $\mathbb{X}$ under what assumptions relating U with $\mathbb{X}$.

Throughout this paper we use a constant $\varepsilon > 0$ that quantifies the relation between $\mathbb{X}$ and its sample. More precisely, we call U an ε-*approximation* of $\mathbb{X}$ if the Hausdorff distance between U and $\mathbb{X}$ is at most ε. Equivalently, U is contained in $\mathbb{X}_\varepsilon = d_{\mathbb{X}}^{-1}[0, \varepsilon]$ and, symmetrically, $\mathbb{X}$ is contained in $U_\varepsilon = d_U^{-1}[0, \varepsilon]$. It follows that the maximum difference between the distance functions defined by U and by $\mathbb{X}$ is at most ε. The converse is also true. Therefore U is an ε-approximation of $\mathbb{X}$ iff $\|d_U - d_{\mathbb{X}}\|_\infty \leq \varepsilon$.

Resolution. When we refer to the local homology at a point z we consider the family of balls B_r centered at z and for each r we study the sequence of homology groups

$$\begin{aligned}
0 \to \mathsf{H}(Y_\alpha \cap B_r) &\to \ldots \to \mathsf{H}(B_r) \\
&\to \mathsf{H}(B_r, Y^\alpha \cap B_r) \to \ldots \to 0,
\end{aligned} \tag{5}$$

where Y is either U or $\mathbb{X}$; see sequence (3) in Section 3. For each radius $r > 0$ we thus consider the series of persistence diagrams $\mathrm{Dgm}(d_Y | B_r)$. The only non-trivial homology group of B_r is H_0 which has rank one. There is therefore only one extended point in this series tracking the first component that appears in the filtration. To determine the local homology of $\mathbb{X}$ at a point z from the sample U it is necessary that the points sample all relevant features of the space fine enough to be recognized. To make this precise, we consider the homological critical values of the distance function of $\mathbb{X}$ restricted to the ball.

DEFINITION. A radius r *resolves* $\mathbb{X}$ *at* z *to* ε if the smallest positive absolute or relative homological critical value of $d_{\mathbb{X}}$ restricted to B_r exceeds 3ε.
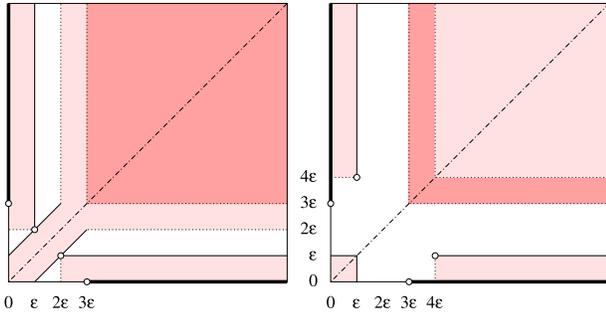

Figure 5: Left: the dark regions contain the persistence diagram of $d_{\mathbb{X}}$ for every radius $r \in R_{\mathbb{X}}(\varepsilon)$ and the light regions expand them to contain the persistence diagram of d_U for every radius $r \in R'_U(\varepsilon)$. Right: the light regions contain the persistence diagram of d_U for every radius $r \in R''_U(\varepsilon)$ and the dark regions contain the persistence diagrams of the distance function of U_ε.

For a radius r that resolves $\mathbb{X}$ to ε there are no births and no deaths in the interval $(0, 3\varepsilon]$. In other words, the corridors separating the two boldface segments from the dark square in Figure 5, left, are empty. It follows that everything born at $\alpha = 0$ lives for a while and if it dies on the way up, as α increases, then it dies strictly after 3ε. Symmetrically, everything that dies at $\alpha = 0$ must have lived for a while and if it was born on the way down, as α decreases, then it was born strictly before 3ε. Radii that have this property are of special interest, so we define $R_{\mathbb{X}}(\varepsilon)$ as the set of radii r for which the points in $\mathrm{Dgm}(d_{\mathbb{X}} | B_r)$ all lie in the dark portion of Figure 5, left, which includes the vertical segment with lower endpoint $(0, 3\varepsilon)$, the horizontal segments with left endpoint $(3\varepsilon, 0)$, and the quadrant $(3\varepsilon, \infty)^2$.

Inference. By slightly extending the notation introduced earlier, we write $\#^a(d_{\mathbb{X}} | B_r)$ for the series of point counts of the corresponding diagrams in the region illustrated in Figure 3. For example, if a is the origin, 0, then $\#^a(d_{\mathbb{X}} | B_r)$ counts the points on the horizontal Birth-axis and the vertical Death-axis.

LOCAL HOMOLOGY INFERENCE THEOREM. Let $\varepsilon > 0$, $\mathbb{X}$ a compact space, U an ε-approximation of $\mathbb{X}$, and z a point in $\mathbb{R}^n$. Then $\#^0(d_{\mathbb{X}} | B_r) = \#^{(\varepsilon, 2\varepsilon)}(d_U | B_r)$ for every radius $r \in R_{\mathbb{X}}(\varepsilon)$.

PROOF. We will prove $R_{\mathbb{X}}(\varepsilon) \subseteq R'_U(\varepsilon)$, where the latter set consists of all radii r for which the points in $\mathrm{Dgm}(d_U | B_r)$ all lie in the shaded portion of Figure 5, left, which expands the dark regions and the diagonal by ε in the vertical as well as the horizontal direction. We will see that this containment of sets implies the claimed equation.

Since $r \in R_{\mathbb{X}}(\varepsilon)$ we have $\#^0(d_{\mathbb{X}} | B_r) = \#^a(d_{\mathbb{X}} | B_r)$ for every $a \in [0, 3\varepsilon]^2$. Since $\|d_U - d_{\mathbb{X}}\|_\infty \leq \varepsilon$, the Stability Theorem of extended persistence implies a bijection such that each point in $\mathrm{Dgm}(d_{\mathbb{X}} | B_r)$ lies within L_∞-distance ε from its corresponding point in $\mathrm{Dgm}(d_U | B_r)$. This implies that all points of $\mathrm{Dgm}(d_U | B_r)$ lie inside the ε-expanded region depicted in Figure 5, left. This region consists of three disjoint subregions, one expanding the vertical segment, one expanding the horizontal segment, and the third expanding the quadrant that contains the remaining points of $\mathrm{Dgm}(d_{\mathbb{X}} | B_r)$ as well as the diagonal. By disjointness of the three subregions, the points of $\mathrm{Dgm}(d_{\mathbb{X}} | B_r)$ in the two segments cannot map to any points other than the ones in the subregions that expand them. The points of $\mathrm{Dgm}(d_U | B_r)$ in these two subregions are counted by $\#^a(d_U | B_r)$ with $a = (\varepsilon, 2\varepsilon)$. This implies the claimed equality. ⌞⌟

Inverse. Recall that U is known but $\mathbb{X}$ is not. The way we hope to use the Local Homology Inference Theorem is that we identify radii r for which the white corridors in Figure 5, left, are empty. For each such r there is a chance that it belongs to $R_{\mathbb{X}}(\varepsilon)$ and if it does we know the local homology of $\mathbb{X}$ for this radius r. The trouble is that we can generally not be sure that r really belongs to $R_{\mathbb{X}}(\varepsilon)$. However, we can further restrict the regions that contain the points of $\mathrm{Dgm}(d_U | B_r)$ so that they imply the existence of a space $\mathbb{X}$ for which U is an ε-approximation and r is in $R_{\mathbb{X}}(\varepsilon)$. Let $R''_U(\varepsilon)$ be the set of radii r for which the points in $\mathrm{Dgm}(d_U | B_r)$ are contained in the light shaded region in Figure 5, right.

INVERSE LHI THEOREM. Let $\varepsilon > 0$, U a subset of $\mathbb{R}^n$, and z a point in $\mathbb{R}^n$. Then there exists a compact space $\mathbb{X} \subseteq \mathbb{R}^n$ for which U is an ε-approximation and $R''_U(\varepsilon) \subseteq R_{\mathbb{X}}(\varepsilon)$.

PROOF. Set $\mathbb{X} = U_\varepsilon$ and note that U is an ε-approximation of $\mathbb{X}$. The distance function defined by $\mathbb{X}$ is $d_{\mathbb{X}}(x) = \max\{0, d_U(x) - \varepsilon\}$. It follows that each birth and each death happens at 0 or ε earlier than before. The corresponding transformation of persistence diagrams is a shift by ε down and a shift by ε to the left, except that a movement stops before the point enters the negative regions of birth or of death. If $r \in R''_U(\varepsilon)$ then all points in the diagrams of $\mathbb{X}$ lie on the two segments and the quadrant that define $R_{\mathbb{X}}(\varepsilon)$. ⌞⌟

Multi-scale example. Observe that the Local Homology Inference Theorem describes the relationship between the persistence diagrams of $\mathbb{X}$ and of U for a fixed radius r. It is difficult to know ahead of time which value of r is most appropriate and in many situations it is not even desirable to make a choice. We cope with this difficulty by examining the persistent behavior across all radii. We use the example in Figure 6 to illustrate what we have in mind. Here $\mathbb{X}$ is
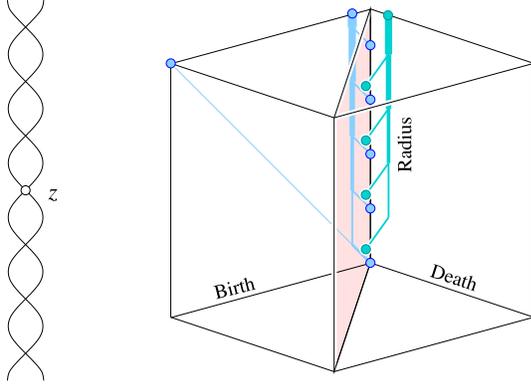

Figure 6: Left: the one-dimensional chain of loops, $\mathbb{X}$. Right: the dimension 1 $(\alpha|r)$-vineyard of $\mathbb{X}$ at z. The only significantly persistent vine runs roughly diagonally in the Birth-Radius plane and tracks a relative 1-cycle. All other vines run near the Radius-axis and track classes caused by the loops in the chain.

a one-dimensional space embedded in $\mathbb{R}^2$. It consists of a string of loops, each connected to the loop before and the loop after. Its dimension 1 vineyard at the point z contains a prominent vine that has high persistence across all values of r. This vine tracks a dimension 1 relative homology class and corresponds to the chain itself which, from a distance, may be seen as a single curve. It can be detected even for rather sparse samples. Furthermore, the vineyard contains two small vines per loop, one tracking a relative and the other an absolute homology class. The relative class emerges at the moment the ball B_r first intersects the loop. It attains its largest persistence when B_r reaches the maximum near the center of the loop after which time the corresponding point in the diagram stops moving and sweeps out a vertical vine. At the same moment the absolute class emerges and attains its largest persistence when B_r reaches the other end of the loop after which time the corresponding point stops moving and sweeps out a vertical vine, as before.

For the study of local homology we are primarily interested in small values of r, that is, the lower portion of the vineyard. Of course, what small means is in the eye of a beholder. On the other hand, the Local Homology Inference Theorem and its inverse can be used to make informed guesses. If the space $\mathbb{X}$ in Figure 6 is sampled sufficiently densely, then small values of r resolve it, and we are able to detect the three dimension 1 cycles in the local homology of z. Specifically, there are three vines emerging from the origin, each tracking a relative homology class. If the sampling is not sufficiently dense then we cannot distinguish $\mathbb{X}$ from a 1-manifold. Indeed, an arc passing through the vertices joining the loops could conceivably produce the same sample.

6 Algorithm

In this section, we describe an algorithm for constructing the series of $(\alpha|r)$-vineyards of a finite set of points as seen from a fixed point $z \in \mathbb{R}^n$. The algorithm is based on comparing various subcomplexes of the Delaunay triangulation of the finite set. We recall that a *simplicial complex* is a finite set of simplices that is closed under the face relation such that any two simplices are either disjoint or meet in a common face [24]. We will simultaneously use this geometric view as well as the more abstract interpretation in which a simplex is simply a finite set of points (its vertices in the geometric view) and a face is a subset.

Voronoi decompositions. Letting $u \in \mathbb{R}^n$ be a point with weight $w \in \mathbb{R}$, the *weighted square distance* of $x \in \mathbb{R}^n$ from u is $\pi_u(x) = \|x - u\|^2 - w$. For the common case in which the weight vanishes the weighted square distance is the squared Euclidean distance. Given a set of weighted points U, the *(weighted) Voronoi cell* of $u \in U$ is

$$V(u) \;=\; \{x \in \mathbb{R}^n \mid \pi_u(x) \leq \pi_v(x), v \in U\}.$$

For the time being we are interested in the case in which U is finite and all weights are zero. We index the points and use the shorter notation $V_i = V(u_i)$ for their Voronoi cells. Each V_i is the intersection of finitely many closed half-spaces and therefore a convex polyhedron. Collectively, the V_i cover the entire space thus forming the *Voronoi decomposition* of $\mathbb{R}^n$, which we denote as $\mathrm{Vor}(U|\mathbb{R}^n)$; see Figure 7. We are also interested in the Voronoi decompositions of the sublevel sets inside the ball and on the sphere, which we denote as $\mathrm{Vor}(U|U_\alpha \cap B_r)$ and $\mathrm{Vor}(U|U_\alpha \cap \partial B_r)$. The former consists of cells $V_i \cap U_\alpha \cap B_r$, which are convex and generically either empty or n-dimensional. The latter consists of cells $V_i \cap U_\alpha \cap \partial B_r$, which are intersections of spherical caps and generically either empty or $(n-1)$-dimensional but not necessarily topologically simple. For example, in Figure 7 we see a Voronoi edge that intersects ∂B_r twice so that one of the two incident Voronoi cells intersects ∂B_r in two components. To cope with the resulting difficulties, we introduce the set $Z(\alpha)$ of points $x \in \mathbb{R}^n$ that satisfy $\|x - z\|^2 - r^2 \leq \|x - u_i\|^2 - \alpha^2$ for all $u_i \in U$. This is the Voronoi cell of z in $\mathrm{Vor}(U \cup \{z\}|\mathbb{R}^n)$ in which every point has weight α^2 except for z which has weight r^2. To distinguish it from the other Voronoi cells we refer to $Z(\alpha)$ as the *power cell* of z. More important than $Z(\alpha)$ itself is the complement of its interior, $Z_0(\alpha) = B_r - \mathrm{int}\, Z(\alpha)$. We will see in Section 7 that $Z_0(\alpha)$ behaves topologically like
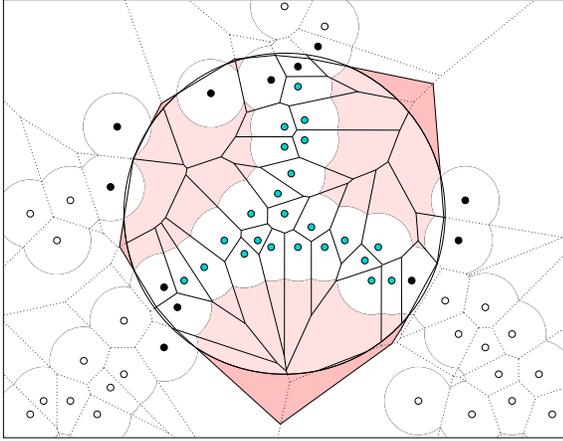
8

Figure 7: Voronoi decompositions of space, of the sublevel set of points at distance at most α from U, and of restricted versions of the same sublevel set. The Voronoi cells of the black dots contribute to the decompositions of $U_\alpha \cap \partial B_r$ and of $Z_0(\alpha)$, those of the gray and black dots contribute to the decomposition of $U_\alpha \cap B_r$, and the Voronoi cells of all dots contribute to the decompositions of U_α and of $\mathbb{R}^n$.
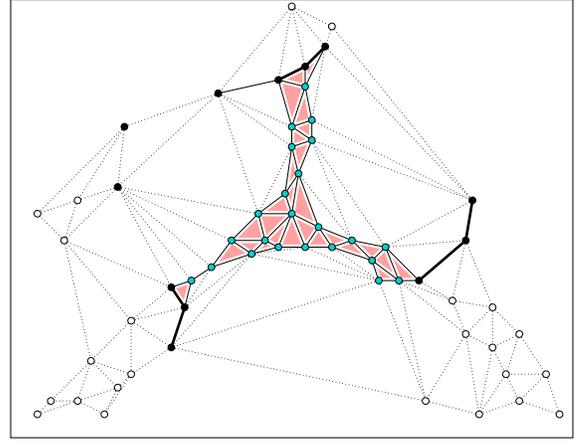

Figure 8: Delaunay triangulations dual to the Voronoi decompositions of the rectangular window in Figure 7 as well as of $U_\alpha \cap B_r$ and of $Z_0(\alpha)$. The drawing style identifies which vertices, edges, and triangles belong to which Delaunay triangulations.

$U_\alpha \cap \partial B_r$. A distinct advantage of the former set is that its intersections with the Voronoi cells are convex. Furthermore, it is not difficult to show that every point in $Z_0(\alpha)$ also belongs to U_α, that is, $Z_0(\alpha) \subseteq U_\alpha \cap B_r$.

Delaunay triangulations. Computationally more convenient than the Voronoi decompositions are their dual *Delaunay triangulations*. For a subset $X \subseteq \mathbb{R}^n$ this is the set $\mathrm{Del}(U|X)$ of simplices $\sigma \subseteq U$ for which X and $V_\sigma = \bigcap_{u_i \in \sigma} V_i$ have a non-empty intersection. In other words, $\mathrm{Del}(U|X)$ is the nerve of the collection of sets $X \cap V_i$. For $X = \mathbb{R}^n$ we get the usual notion of Delaunay triangulation and for $X \subset \mathbb{R}^n$ we get the restricted Delaunay triangulation as defined in [14]. Generically, $\mathrm{Del}(U|\mathbb{R}^n)$ is a simplicial complex geometrically realized in $\mathbb{R}^n$; see Figure 8. We are also interested in the restrictions to $U_\alpha \cap B_r$ and to $Z_0(\alpha)$. In these cases the Delaunay triangulations depend on α and we write $K(\alpha) = \mathrm{Del}(U|U_\alpha \cap B_r)$ and $K_0(\alpha) = \mathrm{Del}(U|Z_0(\alpha))$. Since the restricting domains are subsets of each other, the three Delaunay triangulations are subcomplexes of each other, namely $K_0(\alpha) \subseteq K(\alpha) \subseteq \mathrm{Del}(U|\mathbb{R}^n)$; see Figure 8.

Computing persistence. We now discuss the construction of the series of persistence diagrams of the distance function d_U restricted to the ball of fixed radius r around z. Specifically, we compute the diagram that describes the evolution of the homology classes in the sequence (3). Alternatively, we could compute the diagram of the sequence (4), which by the Isomorphism Lemma contains the same information as (3) but read backwards. We do neither and instead compute the diagrams from the respective first halves since this avoids the

need to subdivide the Delaunay triangulation and leads to a simpler and more efficient implementation. Indeed, we substitute the sequence of complexes $K(\alpha)$, for $0 \leq \alpha < \infty$, and the homomorphisms induced by inclusion for the first half of (3). A formal proof that this substitution does not affect the persistence diagrams will be given in Section 7. To do the actual computation, We use a *compatible ordering* of the simplices in $K(\infty) = \mathrm{Del}(U|B_r)$. This is an ordering $\sigma_1, \sigma_2, \ldots, \sigma_m$ such that

- the simplices in the complex $K(\alpha)$ precede all simplices in $K(\infty) - K(\alpha)$, for all α;
- all faces of a simplex precede that simplex.

Based on this ordering we set up the boundary matrix such that $D[i, j] = 1$ if σ_i is a codimension one face of σ_j and $D[i, j] = 0$ otherwise. Its rows and columns are ordered the same way as the simplices in the compatible ordering. The matrix is *reduced* if each column is either zero or its lowest 1 is in a unique row. In other words, the function that maps each non-zero column j to the row-index $i = low(j)$ of its lowest 1 is injective. Using left-to-right column operations we can reduce D in time $\mathrm{O}(m^3)$, and the pairs $i = low(j)$ give the points in the persistence diagrams; see the matrix version of the persistence algorithm [12] as described in [8].

For the restriction of d_U to $(B_r, \partial B_r)$ we substitute the sequence of pairs $(K(\alpha), K_0(\alpha))$, for $0 \leq \alpha < \infty$, and the homomorphisms induced by inclusion for the first half of (4). A formal proof that this substitution does not affect the persistence diagrams will again be given in Section 7. Each pair is represented by the first complex together with the cone from a dummy vertex, ω, over the second complex. In other words, we use all simplices $\sigma \in K(\alpha)$ and all simplices $\tau \cup \{\omega\}$ with $\tau \in K_0(\alpha)$. Strictly speaking this is not a complex because it does not contain ω, but this is intentional to get exactly the relative homology groups. As

before we use a compatible ordering of the simplices, we set up the boundary matrix accordingly, we reduce the matrix, and we finally derive the persistence diagram from the pairs $i = low(j)$.

Constructing vineyards. Recall that the series of vineyards, $\mathrm{Vnrd}(d_U | d_z)$, may be identified with the 1-parameter family of persistence diagrams, $\mathrm{Dgm}(d_U | B_r)$, for $0 \leq r < \infty$. This is also how we construct it, by maintaining the persistence diagrams while growing r from zero to infinity. Indeed, we just need to maintain the two compatible orderings of the simplices and recompute the reduced matrices whenever these orderings change. For each simplex $\sigma \in \mathrm{Del}(U | \mathbb{R}^n)$ there are functions in r^2 that characterize when σ belongs to $K(\alpha)$ and when to $K_0(\alpha)$. Importantly, these functions are mostly continuous so that the maintenance of the compatible orderings reduces to transpositions of contiguous simplices; see Appendices A and B. Furthermore, each function is piecewise algebraic, where the number of pieces and the degrees are bounded from above by some constant. It follows that the graphs of any two functions cross at most some constant number of times. The total number of transpositions is therefore in $\mathrm{O}(m^2)$. Finally, the reduced matrix can be maintained in time $\mathrm{O}(m)$ per transposition [8]. It follows that the entire algorithm takes time at most cubic in the number of simplices.

$(\alpha | r)$-VINEYARD THEOREM. Given the Delaunay triangulation of U in $\mathbb{R}^n$ and a point $z \in \mathbb{R}^n$, the series of $(\alpha | r)$-vineyards of the two distance functions, $\mathrm{Vnrd}(d_U | d_z)$, can be constructed in time $\mathrm{O}(m^3)$, where m is the number of simplices in $\mathrm{Del}(U | \mathbb{R}^n)$.

7   Correctness

In this section, we prove that substituting the complexes $K_0(\alpha) \subseteq K(\alpha)$ for the spaces $U_\alpha \cap \partial B_r \subseteq U_\alpha \cap B_r$ does not affect the persistence diagrams. We do this in two steps, first constructing homotopy equivalences between spaces and second embedding the induced isomorphisms between the corresponding homology groups in commuting diagrams.

Homotopy equivalence of pairs. Beyond homotopy equivalences between spaces we need them between pairs of spaces. Specifically, (X, X_0) is *homotopy equivalent* to (Y, Y_0), denoted as $(X, X_0) \simeq (Y, Y_0)$, if there exist maps of pairs in both directions whose compositions are homotopic to the respective identities [23, p. 27]. We begin by establishing a homotopy equivalence between the pairs $(U_\alpha \cap B_r, U_\alpha \cap \partial B_r)$ and $(U_\alpha \cap B_r, Z_0(\alpha))$.

POWER CELL LEMMA. Let U, α, z, r be such that $B_r - Z_0(\alpha) \neq \emptyset$. Then the identity on $U_\alpha \cap B_r$ is a homotopy equivalence of $(U_\alpha \cap B_r, U_\alpha \cap \partial B_r)$ and $(U_\alpha \cap B_r, Z_0(\alpha))$ as a map of pairs.

PROOF. It suffices to show that the restriction of the identity, $i : U_\alpha \cap \partial B_r \to Z_0(\alpha)$, is a homotopy equivalence. Let y be a point in $B_r - Z_0(\alpha)$. Every point x in $Z_0(\alpha)$ belongs to B_r but not to the interior of $Z(\alpha)$. The weighted square distance of x from z is therefore non-positive and not smaller than the smallest weighted square distance to a point in U. Hence $\|x - u_i\|^2 - \alpha^2 \leq 0$ for at least one $u_i \in U$ which implies $Z_0(\alpha) \subseteq U_\alpha$. Now draw the ray that starts at y and passes through x and let x' be the point where it crosses ∂B_r. We map x to x' and thus define a retraction $j : Z_0(\alpha) \to U_\alpha \cap \partial B_r$. The composition $i \circ j$ is the identity on $U_\alpha \cap \partial B_r$. The other composition, $j \circ i$, is homotopic to the identity of $Z_0(\alpha)$, as established by the straight-line homotopy $\lambda : Z_0(\alpha) \times [0, 1] \to Z_0(\alpha)$ defined by $\lambda(x, t) = (1 - t)x + tx'$. This implies that the identity is a homotopy equivalence as a map of pairs, as claimed. $\quad\boxdot$

We note that when $B_r \cap Z(\alpha) = \emptyset$ then there is no homotopy equivalence between the pairs. Indeed, we then have $B_r \subseteq U_\alpha$ so that $(U_\alpha \cap B_r) - (U_\alpha \cap \partial B_r) = B_r - \partial B_r$ is an open ball while $(U_\alpha \cap B_r) - Z_0(\alpha) = \emptyset$.

Nerves and barycentric subdivisions. To prepare the construction of our second homotopy equivalence, we consider maps between spheres and a union of convex sets. Specifically, let C be a finite collection of convex sets in $\mathbb{R}^n$, write N for the nerve of C, and let $\mathrm{Sd}\, N$ be the first barycentric subdivision of the nerve. We think of $\mathrm{Sd}\, N$ abstractly, as the simplicial complex that has a vertex $\hat{\sigma}$ for each simplex σ in N and a simplex τ for each increasing chain in the face relation of the simplices in N. We draw $\mathrm{Sd}\, N$ in $\mathbb{R}^n$ by mapping each vertex $\hat{\sigma}$ to a point $f(\hat{\sigma})$ in the intersection of the sets that correspond to the vertices of σ, as in Figure 9. Extending this map by piecewise linear interpolation to the simplices gives a map $f : |\mathrm{Sd}\, N| \to \mathbb{R}^n$. Note that each simplex in $\mathrm{Sd}\, N$ is contained in a single set in C which implies that the image of f is contained in $\bigcup C$. As
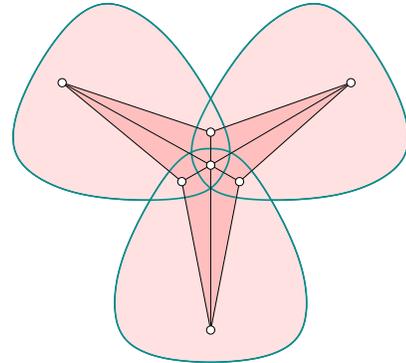

Figure 9: A collection of three convex sets in the plane and a piecewise linear drawing of the barycentric subdivision of the nerve in which each vertex maps to a point in the intersection of the corresponding sets.

suggested in Figure 9 such a drawing does not generally exist for the nerve, which is the reason we consider its barycentric subdivision.

Let now $\gamma : \mathbb{S}^p \to \bigcup C$ be a map of the p-dimensional sphere into the union of the convex sets. We are interested in constructing a map $\mu : \mathbb{S}^p \to |\operatorname{Sd} N|$ whose composition with f is homotopic to γ, $f \circ \mu \simeq \gamma$. For this purpose let L be a triangulation of $\mathbb{S}^p$ that is sufficiently fine such that the image of every simplex is contained in a single but not necessarily unique set in C. In other words, we map each simplex $\tau \in L$ to a set $C_\tau \in C$ such that $\gamma(\tau) \subseteq C_\tau$. Since τ corresponds to a vertex $\hat\tau$ in the barycentric subdivision of L and C_τ is also a vertex in the nerve this gives a map from the vertices of $\operatorname{Sd} L$ to the vertices of N. It is easy to see that this is a vertex map, that is, the vertices of a simplex in $\operatorname{Sd} L$ map to the vertices of a simplex in N, a necessary condition to extend the map by piecewise linear interpolation. However, the same is not true if we substitute $\operatorname{Sd} N$ for N. We therefore consider the second barycentric subdivision of L, $\operatorname{Sd}^2 L = \operatorname{Sd}(\operatorname{Sd} L)$, map the new vertices to the corresponding barycenters of simplices in N, and finally extend the thus obtained vertex map by piecewise linear interpolation to $\mu : |\operatorname{Sd}^2 L| \to |\operatorname{Sd} N|$.
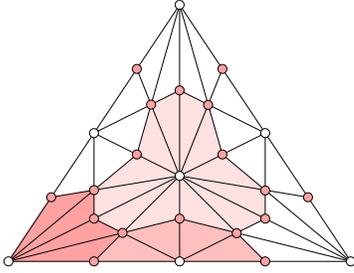

Figure 10: The second barycentric subdivision of a triangle in L. The shading indicates the stars of three vertices of the first barycentric subdivision.

We finally construct a homotopy $H : \mathbb{S}^p \times [0,1] \to \bigcup C$ that equals $f \circ \mu$ at $t = 0$ and γ at $t = 1$. It deforms the image under $f \circ \mu$ of the star of a vertex $\hat\tau$ of $\operatorname{Sd} L$ in $\operatorname{Sd}^2 L$ to the image under γ of the simplex τ of L. Such stars are illustrated in Figure 10. To create the deformation we move every new vertex $\hat v$ of $\operatorname{Sd}^2 L$ toward the vertex $\hat\tau$ of the simplex $v \in \operatorname{Sd} L$ whose corresponding simplex $\tau \in L$ has minimum dimension. To describe the resulting homotopy, we write $g(\hat v) = \hat\tau$, noting that $g(\hat v) = \hat v$ if $\hat v$ is a vertex that already belongs to the first barycentric subdivision. For the vertices of $\operatorname{Sd}^2 L$ the homotopy is defined by

$$H(\hat v, t) = (1-t)f(\mu(\hat v)) + t\gamma(g(\hat v))$$

and then extended by piecewise linear interpolation for every t. To see that this homotopy is well defined we note that the image of every simplex is contained in the same convex set during the entire homotopy. Let η be a simplex in the star of a vertex $\hat\tau$ of $\operatorname{Sd} L$ in $\operatorname{Sd}^2 L$. At the beginning, when $t = 0$,

the image of the vertices of η all belong to the set C_τ because they are all equal or adjacent to $\hat\tau$ in $\operatorname{Sd}^2 L$ and f maps barycenters to the intersection of the sets that correspond to the vertices of its simplex. By convexity, the image of the entire simplex η is contained in C_τ. At the end, when $t = 1$, the image of η is a subset of $\gamma(\tau)$, which is contained in C_τ by assumption of L being sufficiently fine. Finally, every point moves linearly so if the image of η is contained in C_τ both at the beginning and at the end then it is contained in C_τ during the entire homotopy.

Applying Whitehead's Theorem. We use a consequence of Whitehead's Theorem [19, p. 346] to construct our second homotopy equivalence, between the pairs $(U_\alpha \cap B_r, Z_0(\alpha))$ and $(K(\alpha), K_0(\alpha))$. To state the theorem, assume the convex sets in the collection C are the maximal cells in a CW complex. If $\bigcup C$ is connected and the homotopy groups of $\bigcup C$ and of the nerve N of C are isomorphic, for all dimensions p, then the map $f : |\operatorname{Sd} N| \to \bigcup C$ described above is a homotopy equivalence. Constructing a homotopy equivalence thus reduces to proving that the homotopy groups are isomorphic.

NERVE SUBDIVISION LEMMA. Let C be the collection of maximal cells of a CW complex, each a convex set in $\mathbb{R}^n$, N the nerve of C, and $f : |\operatorname{Sd} N| \to \bigcup C$ obtained by piecewise linear interpolation of its values at the vertices. If $f(\hat\sigma)$ is contained in the intersection of the cells that correspond to the vertices of σ, for each simplex $\sigma \in N$, then f is a homotopy equivalence.

PROOF. First we note that f induces a bijection between the connected components of $\operatorname{Sd} N$ and of $\bigcup C$. We then apply the following argument to each component separately. Equivalently, we assume without loss of generality that $\operatorname{Sd} N$ and $\bigcup C$ are both connected.

We want to show that the induced map on the p-dimensional homotopy group, $f_* : \pi_p(\operatorname{Sd} N) \to \pi_p(\bigcup C)$, is an isomorphism. We do this in two steps first showing that f_* is surjective and second that it is injective. To show that it is surjective we prove that for each map $\gamma : \mathbb{S}^p \to \bigcup C$ there is a map $\mu : \mathbb{S}^p \to |\operatorname{Sd} N|$ such that $f \circ \mu$ is homotopic to γ. But this we already did in the preceding paragraph when we discussed maps of the p-sphere to $\bigcup C$. It remains to show that f_* is injective. For this we consider a map $\gamma : \mathbb{S}^p \to \bigcup C$ that extends to a map on the ball bounded by the sphere, $\bar\gamma : \mathbb{B}^{p+1} \to \bigcup C$. We use the same construction as before to define a map $\bar\mu : \mathbb{B}^{p+1} \to |\operatorname{Sd} N|$ such that $f \circ \bar\mu \simeq \bar\gamma$. From the construction it is clear that the restriction μ of $\bar\mu$ to the sphere satisfies $f \circ \mu \simeq \gamma$. It follows that the preimage of zero under f_* is zero, in other words, the kernel of f_* is zero. Hence, f_* is injective and therefore bijective. Whitehead's Theorem applies showing that f is indeed a homotopy equivalence, as required. ⧠

To use the Nerve Subdivision Lemma in our context, we let C be the collection of cells $V_i \cap U_\alpha \cap B_r$ in the Voronoi

decomposition of the restricted sublevel set. Recall that $K(\alpha)$ is the nerve of this collection of sets. Next we construct a map $h_\alpha : |\operatorname{Sd} K(\alpha)| \to U_\alpha \cap B_r$ by specifying it at the vertices and extending it by piecewise linear interpolation. Recall that V_σ is the intersection of the Voronoi cells of all vertices of σ. To define the map we set

$$h_\alpha(\hat\sigma) \;=\; \arg \min_{x \in V_\sigma \cap U_\alpha \cap B_r} d_U^2(x) - d_z^2(x).$$

By construction, $h_\alpha(\hat\sigma)$ belongs to the intersection of the cells that correspond to the vertices of σ. We can therefore apply the Nerve Subdivision Lemma and conclude that h_α is a homotopy equivalence. We are also interested in the restriction of h_α to the barycentric subdivision of $K_0(\alpha)$. Recall that $\sigma \in K(\alpha)$ belongs to $K_0(\alpha)$ iff $V_\sigma \cap Z_0(\alpha)$ is non-empty. By construction, the point $h_\alpha(\hat\sigma)$ then lies in this intersection. The restriction of h_α is therefore a map $h'_\alpha : |\operatorname{Sd} K_0(\alpha)| \to Z_0(\alpha)$ that again satisfies the assumptions of the Nerve Subdivision Lemma. Hence, h_α is a homotopy equivalence as a map of pairs.

Commuting diagrams. The Power Cell and Nerve Subdivision Lemmas imply that the series $\mathsf{H}(U_\alpha \cap B_r, U_\alpha \cap \partial B_r)$, $\mathsf{H}(U_\alpha \cap B_r, Z_0(\alpha))$, and $\mathsf{H}(K(\alpha), K_0(\alpha))$ are isomorphic. However, to conclude that the corresponding sequences of homology groups give rise to the same persistence diagrams we need more, namely that the groups form a commuting diagram whose vertical maps are isomorphisms. We draw the diagram of spaces and maps between them from which the commuting diagram can be obtained by application of the homology functor:

$$
\begin{array}{ccc}
(U_\alpha \cap B_r, U_\alpha \cap \partial B_r) & \xrightarrow{j_\alpha^{\alpha'}} & (U_{\alpha'} \cap B_r, U_{\alpha'} \cap \partial B_r) \\
\downarrow i_\alpha & & \downarrow i_{\alpha'} \\
(U_\alpha \cap B_r, Z_0(\alpha)) & \xrightarrow{j_\alpha^{\alpha'}} & (U_{\alpha'} \cap B_r, Z_0(\alpha')) \\
\uparrow h_\alpha & & \uparrow h_{\alpha'} \\
(K(\alpha), K_0(\alpha)) & \xrightarrow{g_\alpha^{\alpha'}} & (K(\alpha'), K_0(\alpha')),
\end{array}
$$

where $\alpha \le \alpha'$. By the Power Cell and Nerve Subdivision Lemmas, the vertical maps induce isomorphisms between the homology groups of the spaces. The maps i and j are inclusions which implies that the upper square of the corresponding diagram of homology groups commutes. To prove the same for the lower square we consider the maps $e = j_\alpha^{\alpha'} \circ h_\alpha$ and $e' = h_{\alpha'} \circ g_\alpha^{\alpha'}$. Consider the map $H : |K(\alpha)| \times [0,1] \to U_\alpha \cap B_r$ defined by $H(x,t) = h_{\alpha_t} \circ g_\alpha^{\alpha_t}(x)$, where $\alpha_t = (1-t)\alpha + t\alpha'$. Since the maps g and j are inclusions and the maps h vary continuously with α, H is a homotopy between e and e'. This implies that the induced homomorphisms between the corresponding homology groups are the same, $e_* = e'_*$.

To summarize, we have isomorphisms connecting the groups in the columns of a commuting diagram. It follows that each of the three rows gives rise to the same series of

persistence diagrams. In other words, our algorithm which computes persistence diagrams using the complexes $K(\alpha)$ and $K_0(\alpha)$ is correct.

8 Discussion

The main contribution of this paper is the development of topological data analysis methods for the algorithmic study of sampled stratified spaces. Specifically, we show how to assess the local homology at a point and prove that a sufficiently dense sample implies the correctness of our assessment. While non-trivial, the described algorithm is readily implementable and runs in time at most cubic in the number of simplices in the Delaunay triangulation.

We expect that in practice the rate-limiting step of our algorithm will be the construction of the Delaunay triangulation. In most cases, only a small subcomplex of the Delaunay triangulation is relevant for the assessment. More generally, we may restrict the construction to simplices that connect points at distance at most some threshold r from each other. Examples of complexes that limit themselves to such simplices are the alpha, Čech, and Vietoris-Rips complexes; see [9, 13]. It would be interesting to develop fast output-sensitive algorithms for these complexes and to substitute them for the Delaunay triangulation of $\mathbb{R}^n$ on which the methods in this paper are currently based.

The ability to assess the local homology of a stratified space at a point from a finite sample is an important step in a more ambitious program. The larger goal is the construction of the stratified space or a description of the class of stratified spaces that possibly give rise to the observed sample. More desirable still would be a hierarchy of stratified spaces that describes the data set on a continuum of scale levels.

References

[1] N. AMENTA AND M. BERN. Surface reconstruction by Voronoi filtering. *Discrete Comput. Geom.* **22** (1999), 481–504.

[2] M. BELKIN AND P. NIYOGI. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* **15** (2003), 1373–1396.

[3] P. BENDICH, J. HARER AND H. KING. Persistent intersection homology for stratified spaces. Manuscript, Math. Dept., Duke Univ., Durham, North Carolina, 2007.

[4] C. BREGLER AND S. M. OMOHUNDRO. Nonlinear manifold learning for visual speech recognition. *In* "Proc. 5th Internat. Conf. Comput. Vision, 1995", IEEE Comput. Soc. Press, Los Alamitos, California, 494–499.

[5] F. CHAZAL, D. COHEN-STEINER AND A. LIEUTIER. A sampling theory for compact sets in Euclidean space. *In* "Proc. 22nd Ann. Sympos. Comput. Geom., 2006", 319–326.

[6] D. COHEN-STEINER, H. EDELSBRUNNER AND J. HARER. Stability of persistence diagrams. *Discrete Comput. Geom.* **37** (2007), 103–120.

[7] D. COHEN-STEINER, H. EDELSBRUNNER AND J. HARER. Extending persistence using Poincaré and Lefschetz duality. *Found. Comput. Math.*, to appear.

[8] D. COHEN-STEINER, H. EDELSBRUNNER AND D. MOROZOV. Vines and vineyards by updating persistence in linear time. *In* "Proc. 22nd Ann. Sympos. Comput. Geom., 2006", 119–126.

[9] V. DE SILVA AND G. CARLSSON. Topological estimation using witness complexes. *In* "Proc. Sympos. Point-Based Graphics, 2004", 157–166.

[10] T. K. DEY. *Curve and Surface Reconstruction.* Cambridge Univ. Press, England, 2007.

[11] H. EDELSBRUNNER. *Geometry and Topology for Mesh Generation.* Cambridge Univ. Press, England, 2001.

[12] H. EDELSBRUNNER, D. LETSCHER AND A. ZOMORODIAN. Topological persistence and simplification. *Discrete Comput. Geom.* **28** (2002), 511–533.

[13] H. EDELSBRUNNER AND E. P. MÜCKE. Three-dimensional alpha shapes. *ACM Trans. Comput. Graphics* **13** (1994), 43–72.

[14] H. EDELSBRUNNER AND N. R. SHAH. Triangulating topological spaces. *Internat. J. Comput. Geom. Appl.* **7** (1997), 365–378.

[15] M. GORESKY AND R. MACPHERSON. Intersection homology I. *Topology* **19** (1982), 135–162.

[16] M. GORESKY AND R. MACPHERSON. *Stratified Morse Theory.* Springer-Verlag, Heidelberg, Germany, 1988.

[17] N. HABEGGER AND L. SAPER. Intersection cohomology of cs-spaces and Zeeman's filtration. *Invent. Math.* **105** (1991), 247–272.

[18] G. HARO, G. RANDALL AND G. SAPIRO. Stratification learning: detecting mixed density and dimensionality in high dimensional point clouds. *In* "Proc. 19th Conf. Adv. Neural Inform. Process. Syst., 2006", to appear.

[19] A. HATCHER. *Algebraic Topology.* Cambridge Univ. Press, England, 2002.

[20] H. HOPPE, T. DE ROSE, T. DUCHAMP, J. MCDONALD AND W. STÜTZLE. Surface reconstruction from unorganized points. *Comput. Graphics* (1992), 71–78.

[21] B. HUGHES AND S. WEINBERGER. Surgery and stratified spaces. In *Surveys on Surgery Theory*, S. Cappell, A. Ranicki and J. Rosenberg (eds.), Princeton Univ. Press, Princeton, New Jersey, 2000, 311–342.

[22] J. LERAY. Sur la forme des espaces topologiques et sur les points fixes des représentations. *J. Math. Pure Appl.* **24** (1945), 95–167.

[23] C. R. F. MAUNDER. *Algebraic Topology.* Dover, New York, 1996.

[24] J. R. MUNKRES. *Elements of Algebraic Topology.* Addison-Wesley, Redwood City, California, 1984.

[25] P. NIYOGI, S. SMALE AND S. WEINBERGER. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geometry*, to appear.

[26] J. WANG, M. M. OLIVEIRA AND A. E. KAUFMAN. Reconstructing manifold and non-manifold surfaces from point clouds. *In* "IEEE Conf. Visualization, 2005", 53.

[27] S. WEINBERGER. *The Topological Classification of Stratified Spaces.* Univ. Chicago Press, Chicago, Illinois, 1994.

[28] A. ZOMORODIAN AND G. CARLSSON. Computing persistent homology. *Discrete Comput. Geom.* **33** (2005), 249–274.

Appendix A

Given a point $z \in \mathbb{R}^n$ and a radius $r \geq 0$, we use the restriction of the Delaunay triangulation to the ball B_r and the pair $(B_r, \partial B_r)$, both centered at z, to assess the local homology of the data at z. Specifically, we consider the complexes

$$
\begin{aligned}
K(\alpha) &= \mathrm{Del}(U|U_\alpha \cap B_r); \\
K_0(\alpha) &= \mathrm{Del}(U|Z_0(\alpha)),
\end{aligned}
$$

where we recall that $Z_0(\alpha) = B_r - \mathrm{int}\, Z(\alpha)$ is contained in U_α. In this appendix, we study under what conditions these complexes contain a simplex in $\mathrm{Del}(U|\mathbb{R}^n)$.

Threshold for A. A necessary condition for a simplex $\sigma \in \mathrm{Del}(U|\mathbb{R}^n)$ to belong to $K(\alpha)$ is it belongs to the Delaunay triangulation of the sublevel set,

$$
A(\alpha) = \mathrm{Del}(U|U_\alpha).
$$

This complex is also known as the alpha complex of U [13]. To characterize when σ belongs to $A(\alpha)$ we consider the smallest $(n-1)$-sphere that passes through the points of $\sigma \in U$ and encloses none of the points of U. We call this the *smallest empty circumsphere* of σ and let $c_0 \in \mathbb{R}^n$ be its center and ϱ_0 its radius. If $\sigma = \{u_i\}$ is a vertex then the $(n-1)$-sphere degenerates to the point $c_0 = u_i$ and we have $\varrho_0 = 0$. Generically, c_0 and ϱ_0 are unique and we set $\alpha_\sigma = \varrho_0$.

CONDITION A. Let σ be a simplex in $\mathrm{Del}(U|\mathbb{R}^n)$. Then

$$
\sigma \in A(\alpha) \quad \text{iff} \quad \alpha_\sigma^2 \leq \alpha^2.
$$

There is an alternative geometric interpretation of this condition in terms of the smallest but possibly non-empty circumsphere of σ. It passes through the points of σ but may enclose other points of U. Unless $\dim \sigma = n$, its center c and radius ϱ are not necessarily the same as c_0 and ϱ_0. The point c is also the center of the common intersection of the balls of radius α centered at the vertices of σ and the plane of dimension $n - \dim \sigma$ whose points are equidistant from these vertices. If this common intersection is non-empty then it is either a point or a ball of dimension $n - \dim \sigma$ and square radius $\alpha^2 - \varrho^2$. The interpretation of Condition A is now that α_σ^2 is the smallest value of α^2 for which this ball has a non-empty intersection with $V_\sigma = \bigcap_{u_i \in \sigma} V_i$. In other words, $\sigma \in A(\alpha)$ iff $U_\alpha \cap V_\sigma \neq \emptyset$ which is consistent with the definition of $A(\alpha)$.

Thresholds for E and E_0. We take an indirect approach to the restrictions of the Delaunay triangulation to $U_\alpha \cap B_r$ and to $Z_0(\alpha)$. We begin by considering the restrictions to the power cell and its boundary,

$$
\begin{aligned}
E(\alpha) &= \mathrm{Del}(U|Z(\alpha)); \\
E_0(\alpha) &= \mathrm{Del}(U|\partial Z(\alpha)).
\end{aligned}
$$

We note that $E(\alpha)$ subdivides the underlying space of the closed star of z in $W(\alpha)$, the Delaunay triangulation of $U \cup \{z\}$ in which z has weight r^2 and all other points have weight α^2. Furthermore $E_0(\alpha)$ is the link of z in $W(\alpha)$. To characterize when σ belongs to $E(\alpha)$ and $E_0(\alpha)$ we set $\eta_\sigma^2 = \varphi_\sigma^2 = r^2 + \varrho_0^2 - \|z - c_0\|^2$ if σ is an n-simplex. Otherwise, we set η_σ^2 equal to the maximum value η_τ^2 assigned to any n-simplex τ that has σ as a face. Symmetrically, we set φ_σ^2 equal to the minimum value φ_τ^2 assigned to any n-simplex τ that has σ as a face.

CONDITION E. Let σ be a simplex in $\mathrm{Del}(U|\mathbb{R}^n)$. Then

$$\sigma \in E(\alpha) \quad \text{iff} \quad \alpha^2 \leq \eta_\sigma^2;$$
$$\sigma \in E_0(\alpha) \quad \text{iff} \quad \varphi_\sigma^2 \leq \alpha^2 \leq \eta_\sigma^2.$$

There is again an alternative geometric interpretation of this condition. For a point $x \in V_\sigma$ let $\varrho(x)$ be the distance to the points $u_i \in \sigma$. Then x belongs to $Z(\alpha)$ iff $\|x - z\|^2 - r^2 \leq \varrho(x)^2 - \alpha^2$. The interpretation is now that $\varphi_\sigma^2 \leq \alpha^2 \leq \eta_\sigma^2$ iff V_σ has a non-empty intersection with $\partial Z(\alpha)$. For $\alpha^2 = \varphi_\sigma^2$ the rest of V_σ lies inside $Z(\alpha)$ while for $\alpha^2 = \eta_\sigma^2$ the rest of V_σ lies outside $Z(\alpha)$.

Thresholds for F and F_0. Next we consider the restrictions of the sublevel set to the power cell and its boundary,

$$F(\alpha) = \mathrm{Del}(U|U_\alpha \cap Z(\alpha));$$
$$F_0(\alpha) = \mathrm{Del}(U|U_\alpha \cap \partial Z(\alpha)).$$

By definition, $F(\alpha)$ is a subset of $A(\alpha) \cap E(\alpha)$ and, similarly, $F_0(\alpha)$ is a subset of $A(\alpha) \cap E_0(\alpha)$, but generally we do not have equality. By definition of the power cell we have $U_\alpha \cap V_\sigma \subseteq Z(\alpha)$ iff $U_\alpha \cap V_\sigma \subseteq B_r$ as well as $U_\alpha \cap V_\sigma \cap Z(\alpha) = \emptyset$ iff $U_\alpha \cap V_\sigma \cap B_r = \emptyset$. We use this to formulate a test for deciding when a simplex in $E(\alpha)$ belongs to $F(\alpha)$ and the same for $E_0(\alpha)$ and $F_0(\alpha)$. Specifically, we give a condition when $U_\alpha \cap V_\sigma$ intersects ∂B_r. Consider the affine hull of the simplex, $D = \mathrm{aff}\, \sigma$, and the affine hull of the corresponding intersection of Voronoi cells, $V = \mathrm{aff}\, V_\sigma$. By construction, D is a plane of dimension $\dim \sigma$, V is a plane of dimension $n - \dim \sigma$, and the two intersect orthogonally in the point c defined earlier; see Figure 11. Assuming $\dim \sigma < n$, the balls of radius α centered at the vertices of σ intersect V in a ball of dimension $n - \dim \sigma$ whose center is c and whose square radius is $\alpha^2 - \varrho^2$. Let d be the distance of z from D and v the distance from V. When $|d - (r^2 - v^2)^{1/2}| = (\alpha^2 - \varrho^2)^{1/2}$ then the $(n - \dim \sigma)$-ball centered at c touches the sphere of radius r around z, and depending on the sign of $d - (r^2 - v^2)^{1/2}$ it does this either from the inside or from the outside. Taking the square and solving for α^2 we get $w(r^2) = r^2 + \varrho^2 + d^2 - v^2 - 2d(r^2 - v^2)^{1/2}$. This function has a minimum at $r^2 = d^2 + v^2$ for which the
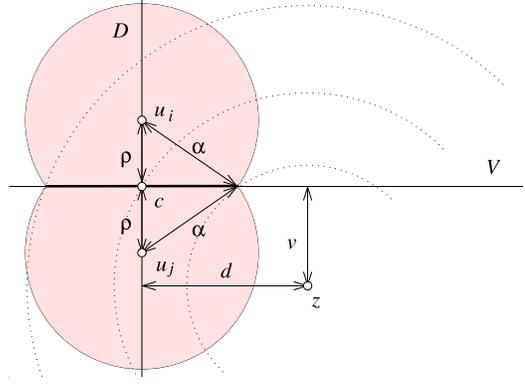

Figure 11: The point c is the center of the smallest circumsphere of the edge $\sigma = \{u_i, u_j\}$. The distance of z from c is $(d^2 + v^2)^{1/2}$.

value is $w(r^2) = \varrho^2$. We now set

$$\kappa_\sigma^2 = \begin{cases} w(r^2) & \text{if } r^2 \leq d^2 + v^2; \\ \varrho^2 & \text{if } d^2 + v^2 \leq r^2; \end{cases}$$
$$\lambda_\sigma^2 = \begin{cases} \varrho^2 & \text{if } r^2 \leq d^2 + v^2; \\ w(r^2) & \text{if } d^2 + v^2 \leq r^2. \end{cases}$$

They are both undefined for $r^2 < v^2$, the case when $w(r^2)$ is undefined. The above derivation does not make sense for an n-simplex σ but the following condition does, provided we set $\kappa_\sigma^2 = \lambda_\sigma^2 = \varrho^2$.

CONDITION F. Let σ be a simplex in $\mathrm{Del}(U|\mathbb{R}^n)$. Then

$$\sigma \in F(\alpha) \quad \text{iff} \quad \max\{\alpha_\sigma^2, \kappa_\sigma^2\} \leq \alpha^2 \leq \eta_\sigma^2;$$
$$\sigma \in F_0(\alpha) \quad \text{iff} \quad \max\{\alpha_\sigma^2, \varphi_\sigma^2, \kappa_\sigma^2, \lambda_\sigma^2\} \leq \alpha^2 \leq \eta_\sigma^2.$$

The geometric interpretation of Condition F should be clear. The only way for a simplex σ to belong to $A(\alpha)$ and $E(\alpha)$ but not to $F(\alpha)$ is that its circumcenter c lies outside the sphere around z, $r^2 < d^2 + v^2$, and the $(n - \dim \sigma)$-ball does not touch this sphere, $\alpha^2 < \kappa_\sigma^2$. Similarly, the only way for σ to belong to $A(\alpha)$, $E_0(\alpha)$, and $F(\alpha)$ but not to $F_0(\alpha)$ is that c lies inside the ball around z, $d^2 + v^2 < r^2$, and the $(n - \dim \sigma)$-ball does not touch the sphere, $\alpha^2 < \lambda_\sigma^2$.

Thresholds for K and K_0. We now derive conditions for the restrictions to $U_\alpha \cap B_r$ and $Z_0(\alpha)$. Specifically, we use the fact that both these two spaces are swept out by the boundary of the sublevel set restricted to the power cell.

SWEEP LEMMA. Let $\alpha \geq 0$. Then

$$U_\alpha \cap B_r = \bigcup_{0 \leq s \leq \alpha} \partial U_s \cap Z(s);$$
$$Z_0(\alpha) = \bigcup_{0 \leq s \leq \alpha} U_s \cap \partial Z(s).$$

14

PROOF. To establish the first equation we show that $x \in \partial U_s$ implies $x \in B_r$ iff $x \in Z(s)$. Indeed, $x \in \partial U_s$ implies $\|x - u_i\|^2 - s^2 = 0$, where $u_i \in U$ minimizes the distance to x. Thus $x \in Z(s)$ iff $\|x - z\|^2 - r^2 \leq 0$, as required. This implies that $\partial U_s \cap Z(s)$ sweeps out $U_\alpha \cap B_r$ as we increase s from 0 to α.

To prove the second equation we show that $x \in \partial Z(s)$ implies $x \in B_r$ iff $x \in U_\alpha$. Indeed, if $x \in \partial Z(s)$ then $\|x - z\|^2 - r^2 = \|x - u_i\|^2 - s^2$, where $u_i \in U$ again minimizes the distance to x. Both sides are non-positive at the same time, as required. This implies that $U_s \cap \partial Z(s)$ sweeps out $Z_0(\alpha) = B_r - \text{int } Z(\alpha)$ as we increase s from 0 to α. ⬚

The first equation in the Sweep Lemma implies that $U_\alpha \cap B_r$ is the union of the $U_s \cap Z(s)$, even without taking the boundary of U_s. Hence, $K(\alpha)$ is the union of the $F(s)$, for $0 \leq s \leq \alpha$. Similarly, the second equation implies that $K_0(\alpha)$ is the union of the $F_0(s)$, for $0 \leq s \leq \alpha$. We observe that if η_σ^2 is less than α_σ^2 or κ_σ^2 then $F(s)$ is empty, for all s, and so are $K(\alpha)$ and $K_0(\alpha)$. Otherwise, we get the conditions by dropping the upper bounds in Condition F.

CONDITION K. Let σ be a simplex in $\text{Del}(U | \mathbb{R}^n)$. If $\eta_\sigma^2 < \max\{\alpha_\sigma^2, \kappa_\sigma^2\}$ then $K(\alpha) = K_0(\alpha) = \emptyset$. Otherwise,

$$\begin{aligned} \sigma \in K(\alpha) \quad &\text{iff} \quad \max\{\alpha_\sigma^2, \kappa_\sigma^2\} \leq \alpha^2; \\ \sigma \in K_0(\alpha) \quad &\text{iff} \quad \max\{\alpha_\sigma^2, \varphi_\sigma^2, \kappa_\sigma^2, \lambda_\sigma^2\} \leq \alpha^2. \end{aligned}$$

Indeed, the upper bound just guarantees that $U_\alpha \cap V_\sigma$ has a non-empty intersection with $Z(\alpha)$. Since we increase s from 0 to α there is a non-empty intersection between $U_s \cap V_\sigma$ and $Z(s)$, for some s, iff there is a non-empty intersection between $U_\alpha \cap V_\sigma$ and B_r, which is captured by the remaining inequality.

Thresholds for A and A_0. We can further simplify the condition by restricting the Delaunay triangulation to the sublevel set outside the power cell,

$$A_0(\alpha) \quad = \quad \text{Del}(U | U_\alpha - \text{int } Z(\alpha)).$$

Equivalently, $A_0(\alpha) = (A(\alpha) - K(\alpha)) \cup K_0(\alpha)$. By construction, $A_0(\alpha)$ is a subcomplex of $A(\alpha)$ and the difference is $A(\alpha) - A_0(\alpha) = K(\alpha) - K_0(\alpha)$. We also consider the diagram of relative homology groups with induced homomorphisms,

$$\begin{array}{ccc} (K(\alpha), K_0(\alpha)) & \xrightarrow{g_\alpha^{\alpha'}} & (K(\alpha'), K_0(\alpha')), \\ \downarrow i_\alpha & & \downarrow i_{\alpha'} \\ (A(\alpha), A_0(\alpha)) & \xrightarrow{j_\alpha^{\alpha'}} & (A(\alpha'), A_0(\alpha')), \end{array}$$

Since all maps are inclusions the diagram commutes and so does the corresponding diagram of relative homology groups. Finally, excision implies that the vertical maps induce isomorphisms. It follows that the persistence diagrams

we get from the $(A(\alpha), A_0(\alpha))$ are indeed the same as the one of the $(K(\alpha), K_0(\alpha))$. The reason for making this final substitution is computational convenience.

CONDITION A'. Let σ be a simplex in $\text{Del}(U | \mathbb{R}^n)$. Then

$$\begin{aligned} \sigma \in A(\alpha) \quad &\text{iff} \quad \alpha_\sigma^2 \leq \alpha^2; \\ \sigma \in A_0(\alpha) \quad &\text{iff} \quad \max\{\alpha_\sigma^2, \varphi_\sigma^2, \lambda_\sigma^2\} \leq \alpha^2. \end{aligned}$$

Appendix B

In this appendix, we extend the results of Appendix A to the situation in which the radius of the restricting ball varies. It is convenient to add r to the notation, writing $K(\alpha, r) = K(\alpha)$ and similar for other complexes.

Absolute homology. Recall that the first half of the sequence (3) gives the same persistence diagrams as the sequence of the $K(\alpha, r)$, where r is fixed. Each simplex σ in the Delaunay triangulation undergoes the same kind of evolution as α goes from 0 to ∞:

Step 1. σ becomes a member of $K(\alpha, r)$;

a step that may also be skipped. The thresholds that determine membership in this complex vary continuously with the radius. Looking at the squares and parametrizing by r^2, we get three functions, $\alpha_\sigma^2, \kappa_\sigma^2, \eta_\sigma^2 : [0, \infty) \to [0, \infty)$. Rewriting the first half of Condition K we find how these functions control membership in $K(\alpha, r)$.

ABSOLUTE HOMOLOGY EVOLUTION LEMMA.
Let σ be a simplex in $\text{Del}(U | \mathbb{R}^n)$ and $r \geq 0$. If $\eta_\sigma^2(r^2) < \max\{\alpha_\sigma^2(r^2), \kappa_\sigma^2(r^2)\}$ then $K(\alpha, r) = \emptyset$. Otherwise,

$$\sigma \in K(\alpha, r) \quad \text{iff} \quad \max\{\alpha_\sigma^2(r^2), \kappa_\sigma^2(r^2)\} \leq \alpha^2.$$
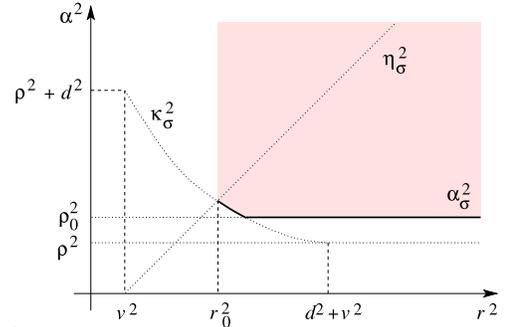

Figure 12: The simplex σ belongs to $K(\alpha, r)$ iff the point (α^2, r^2) lies in the shaded region above the function that discriminates between membership and non-membership.

The inequality is illustrated in Figure 12. The smallest value of r^2 at which σ belongs to any of the $K(\alpha, r)$ is

when $\max\{\alpha_\sigma^2(r^2), \kappa_\sigma^2(r^2)\} = \eta_\sigma^2(r^2)$. Letting this value be r_0^2, we have a continuous function from $[r_0^2, \infty)$ to $[0, \infty)$ that discriminates membership of σ in $K(\alpha, r)$ from non-membership. In general, this function consists of two portions, the one on the left contributed by κ_σ^2 and the one on the right contributed by α_σ^2. If $c = c_0$ then $r_0^2 \leq d^2 + v^2$ which guarantees that the left portion is non-empty. Otherwise, it is possible that the function is constant over the entire interval it is defined.

Relative homology. Recall that the first half of the sequence (4) gives the same persistence diagrams as the sequence of the $(A(\alpha, r), A_0(\alpha, r))$, where r is again fixed. Each simplex $\sigma \in \mathrm{Del}(U|\mathbb{R}^n)$ undergoes the same kind of evolution as α goes from 0 to ∞:

Step 1. σ becomes a member of $A(\alpha, r)$;

Step 2. σ becomes a member of $A_0(\alpha, r)$,

two steps that may also occur simultaneously. Similar to the absolute homology case we get three continuous functions, $\alpha_\sigma^2, \varphi_\sigma^2, \lambda_\sigma^2 : [0, \infty) \to [0, \infty)$. Rewriting Condition A' we find how these functions control membership in $(A(\alpha, r), A_0(\alpha, r))$.

RELATIVE HOMOLOGY EVOLUTION LEMMA. Let σ be a simplex in $\mathrm{Del}(U|\mathbb{R}^n)$ and $r \geq 0$. Then

$$\begin{aligned}
\sigma \in A(\alpha, r) \quad &\text{iff} \quad \alpha_\sigma^2(r^2) \leq \alpha^2; \\
\sigma \in A_0(\alpha, r) \quad &\text{iff} \quad \max\{\alpha_\sigma^2(r^2), \varphi_\sigma^2(r^2), \lambda_\sigma^2(r^2)\} \leq \alpha^2.
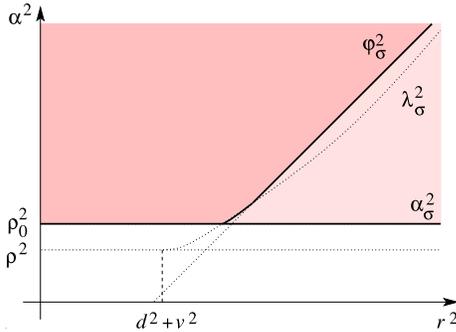\end{aligned}$$

Figure 13: The simplex σ belongs to $A(\alpha, r)$ and $A_0(\alpha, r)$ iff the point (α^2, r^2) lies in the shaded region above both membership discriminating functions, and σ belongs to $A(\alpha, r)$ but not to $A_0(\alpha, r)$ iff the point lies in the light shaded region between the two functions.

Figure 13 illustrates the inequalities for a simplex σ whose smallest circumsphere is different from its smallest empty circumsphere: $c \neq c_0$ and $\varrho < \varrho_0$. Often the picture is even simpler. For example, if $\dim \sigma = n$ then $\varrho = \varrho_0$ and the function that controls the membership of σ in $A_0(\alpha, r)$ simplifies to $r^2 + \varrho^2 - (d^2 + v^2)$. It follows that the region in which σ belongs to $A(\alpha, r)$ but not to $A_0(\alpha, r)$ is a wedge to the right of the corner point at $r^2 = d^2 + v^2$ and $\alpha^2 = \varrho^2$.

Crossings and transpositions. The two Evolution Lemmas introduce three *membership functions* per simplex,

$$\begin{aligned}
f_{\sigma, K}(r^2) &= \max\{\alpha_\sigma(r^2), \kappa_\sigma^2(r^2)\}; \\
f_{\sigma, A}(r^2) &= \alpha_\sigma(r^2); \\
f_{\sigma, A_0}(r^2) &= \max\{\alpha_\sigma(r^2), \varphi_\sigma^2(r^2), \lambda_\sigma^2(r^2)\}.
\end{aligned}$$

All three are continuous except for $f_{\sigma, K}$ which is undefined for $r^2 < r_0^2$. We set $f_{\sigma, K}(r^2) = \infty$ for $0 \leq r^2 < r_0^2$, and viewing infinity as just another value we thus get a function with a single discontinuity. We will see shortly that having just one discontinuity per function does not substantially affect the results we reap. Each membership function consists of pieces contributed by α_σ^2, φ_σ^2, κ_σ^2, and λ_σ^2. The α_σ^2 are constant functions and the φ_σ^2 are functions of degree one in the variable r^2. The κ_σ^2 and λ_σ^2 are also of degree one in r^2 except that they have an additional square root term. Each membership function consists of a constant number of pieces. The only case in which this is perhaps not entirely obvious is for φ_σ^2 which is a point-wise minimum of functions φ_τ^2 over all n-simplices τ in $\mathrm{Del}(U|\mathbb{R}^n)$ that contain σ as a face. There can be an arbitrary number of such n-simplices but since the corresponding functions are of the form $\varphi_\tau^2(r^2) = r^2 + \text{const}$, only one provides all the minima. Two functions $f, g : [0, \infty) \to [0, \infty)$ *cross* at r_1 if

$$[f(r_1 - \varepsilon) - g(r_1 - \varepsilon)][f(r_1 + \varepsilon) - g(r_1 + \varepsilon)] \quad < \quad 0$$

for all sufficiently small $\varepsilon > 0$. For continuous functions this happens at a point at which they agree, $f(f_1) = g(r_1)$. Similarly, we define when f and g cross at an interval along which they agree. Any two pieces of the membership functions have constant complexity and cross at most a constant number of times. Since each membership function consists only of a constant number of such pieces, this implies that any two membership functions cross at most some constant number of times. Drawing the graphs of the membership functions of all m simplices in $\mathrm{Del}(U|\mathbb{R}^n)$ we therefore get an arrangement of $3m$ curves with at most some constant times m^2 vertices, edges, and regions. When we sweep the arrangement with a vertical line from left to right we do at most some constant times m^2 transpositions at the vertices and another constant times m^2 transpositions because of the m discontinuities. The $(\alpha|r)$-Vineyard Theorem follows.

Appendix C

$Y_0 \subseteq Y \subseteq \mathbb{R}^n$	topological spaces
$z \in \mathbb{R}^n$	point
$d_Y, d_z : \mathbb{R}^n \to \mathbb{R}$	distance function
$Y_\alpha, Y^\alpha, B_r, B^r$	sublevel, superlevel sets
$\mathsf{H}(Y), \mathsf{H}(Y, Y_0)$	series of homology groups
$\mathrm{Dgm}(d_Y | B_r)$	series of persistence diagram
$\#^a(d_Y)$	series of #points in region
$\mathrm{Vnrd}(d_Y | d_z)$	series of $(\alpha | r)$-vineyards
$\mathbb{X}$	stratified space
$\mathbb{X}_i - \mathbb{X}_{i-1}$	union of i-strata
$\varepsilon > 0$	constant
$R_{\mathbb{X}}(\varepsilon), R'_U(\varepsilon), R''_U(\varepsilon)$	sets of radii
$\pi_u(x), \pi_z(x)$	weighted square distances
$u_i, u_j \in U$	sampled points
$\sigma \subseteq U$	(abstract) simplex
$V_i = V(u_i)$	Voronoi cells
$Z_0(\alpha) \subseteq Z(\alpha)$	complement, power cell
$\mathrm{Vor}(U | X)$	Voronoi decomposition
$\mathrm{Del}(U | X)$	Delaunay triangulation
$D[i, j]; m$	boundary matrix, #simplices
$K_0(\alpha) \subseteq K(\alpha)$	simplicial complexes
$C, N, \mathrm{Sd}\, N$	set of cells, nerve, subdivision
$L, \mathrm{Sd}\, L, \mathrm{Sd}^2 L$	triangulations of sphere
$\sigma, \tau, \upsilon, \eta; \hat{\sigma}, \hat{\tau}, \hat{\upsilon}$	simplices; barycenters
$\gamma, \mu, \bar{\gamma}, \bar{\mu}$	maps
$i_\alpha, j_\alpha^{\alpha'}, g_\alpha^{\alpha'}$	inclusions
$i, j, f, h_\alpha; H$	maps
$W(\alpha)$	weighted Delaunay tri.
$A_0(\alpha) \subseteq A(\alpha)$	restricted to sublevel set
$E_0(\alpha) \subseteq E(\alpha)$	restricted to (∂) power cell
$F_0(\alpha) \subseteq F(\alpha)$	further restricted to U_α
$c, c_0; \varrho, \varrho_0$	centers; radii
$D, V; d, v$	planes; distances
$\alpha_\sigma^2, \eta_\sigma^2, \varphi_\sigma^2, \kappa_\sigma^2, \lambda_\sigma^2$	thresholds or threshold functions
$f_{\sigma,K}, f_{\sigma,A}, f_{\sigma,A_0}$	membership functions

Table 2: List of important notation.