



Manifold Reconstruction in Arbitrary Dimensions using Witness Complexes

Jean-Daniel Boissonnat, Leonidas J. Guibas, Steve Y. Oudot

► To cite this version:

Jean-Daniel Boissonnat, Leonidas J. Guibas, Steve Y. Oudot. Manifold Reconstruction in Arbitrary Dimensions using Witness Complexes. Discrete and Computational Geometry, 2009, pp.37. hal-00488434

HAL Id: hal-00488434

<https://hal.science/hal-00488434>

Submitted on 2 Jun 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Manifold Reconstruction in Arbitrary Dimensions using Witness Complexes

Jean-Daniel Boissonnat
INRIA, Géométrie Team
2004 route des lucioles
06902 Sophia-Antipolis, France
boissonn@sophia.inria.fr

Leonidas J. Guibas
Dept. Computer Science
Stanford University
Stanford, CA 94305
guibas@cs.stanford.edu

Steve Y. Oudot
Dept. Computer Science
Stanford University
Stanford, CA 94305
steve.oudot@inria.fr*

Abstract

It is a well-established fact that the witness complex is closely related to the restricted Delaunay triangulation in low dimensions. Specifically, it has been proved that the witness complex coincides with the restricted Delaunay triangulation on curves, and is still a subset of it on surfaces, under mild sampling conditions. In this paper, we prove that these results do not extend to higher-dimensional manifolds, even under strong sampling conditions such as uniform point density. On the positive side, we show how the sets of witnesses and landmarks can be enriched, so that the nice relations that exist between restricted Delaunay triangulation and witness complex hold on higher-dimensional manifolds as well. We derive from our structural results an algorithm that reconstructs manifolds of any arbitrary dimension or co-dimension at different scales. The algorithm combines a farthest-point refinement scheme with a vertex pumping strategy. It is very simple conceptually, and it does not require the input point sample to be sparse. Its running time is bounded by $c(d)n^2$, where n is the size of the input point cloud, and $c(d)$ is a constant depending solely (yet exponentially) on the dimension d of the ambient space. Although this running time makes our reconstruction algorithm rather theoretical, recent work has shown that a variant of our approach can be made tractable in arbitrary dimensions, by building upon the results of this paper.

1 Introduction

A number of areas of Science and Engineering deal with point clouds lying on or near submanifolds of Euclidean spaces. Such data can be either collected through measurements of natural phenomena, or generated by simulations. Given a finite set of sample points W , presumably drawn from an unknown manifold S , the goal is to retrieve some information about S from W . This *manifold learning* process, which is at the core of non-linear dimensionality reduction techniques [31, 33], finds applications in many areas, including machine learning [6], pattern recognition [32], scientific visualization [34], image or signal processing [28], and neural computation [19]. The nature of the sought-for information is very application-dependent, and sometimes it is enough to inquire about the topological invariants of the manifold, a case in which techniques such as topological persistence

*This work was done while this author was a post-doctoral fellow at Stanford University. His email there is no longer valid.

[10, 16, 23] offer a nice mathematical framework. However, in some situations it is desirable to build faithful approximations to the manifold, both in terms of topology and geometry.

This problem has received a lot of attention from the computational geometry community, which proposed elegant solutions in low dimensions, based on the use of the Delaunay triangulation $\mathcal{D}(W)$ of the input point set W – see [8] for a survey. In these methods, the approximation takes the form of a simplicial complex extracted from $\mathcal{D}(W)$. This complex is equal or close to $\mathcal{D}^S(W)$, the Delaunay triangulation of W restricted to the manifold S . What makes the Delaunay-based approach attractive is that, not only does it behave well on practical examples, but its performance is guaranteed by a sound theoretical framework. Indeed, the restricted Delaunay triangulation is known to provide good topological and geometric approximations of sampled planar curves or surfaces in 3-space, under mild sampling conditions [1, 2, 7]. The analysis on curves extends in fact beyond the planar case, to Euclidean spaces of arbitrary dimensions [21]; yet, the corresponding reconstruction algorithms use the prior knowledge that the object underlying the input data is 1-dimensional, and they are therefore unable to deal with higher-dimensional sampled objects.

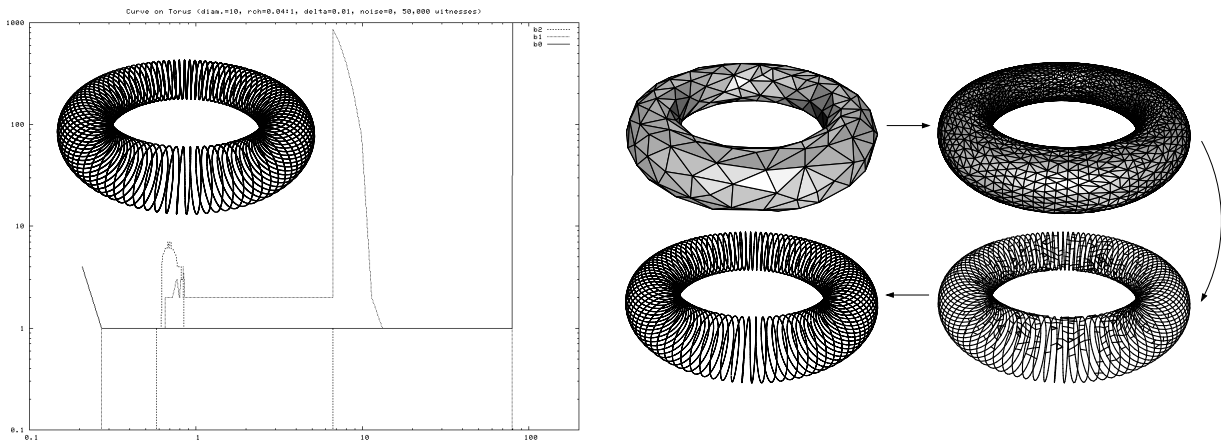


Figure 1: Helical curve drawn on a torus (left). From a uniform sampling of the curve, the method of [27] builds a one-parameter family of complexes (right) and maintains their Betti numbers (left).

Generalizing the above ideas to tackle the reconstruction of manifolds of (unknown) arbitrary dimensions and co-dimensions seems difficult, mainly because a given point set W may well-sample various objects with different topological types. For instance, in the scenario depicted in Figure 1, any dense point sample of the helical curve is also a dense point sample of the torus on which the curve is drawn, thus making the dimension of the object underlying the data uncertain. Another source of uncertainty is noise, which makes it difficult to distinguish between sampling artefacts and small topological features of the underlying object, and which becomes all the more problematic when the intrinsic dimensionality of the data remains unknown. To overcome these issues, it has been suggested to strengthen the sampling conditions, so that they can be satisfied by only one class of manifolds sharing the same topological invariants [15, 20, 26]. In some sense, this is like choosing arbitrarily between the possible reconstructions, and ignoring the rest of the information carried by the input. Moreover, the new sets of conditions on the input are so strict that they are hardly satisfiable in practice, thereby making the contributions rather theoretical. Yet [15, 20, 26] contain a wealth of relevant ideas and results, some of which are exploited in this paper.

A different and very promising approach [9, 11, 27, 30], reminiscent of topological persistence,

builds a one-parameter family of complexes that approximate S at various scales. The claim is that, for sufficiently dense W , the family contains a long sequence of complexes carrying the same homotopy type as S . In fact, there can be several such sequences, each one corresponding to a plausible reconstruction – see Figure 1. Therefore, performing a reconstruction on W boils down to finding the long *stable* sequences in the one-parameter family of complexes. This approach to reconstruction stands in sharp contrast with previous work in the area. In [9, 11, 30], the family of complexes is derived from the α -offsets (or, equivalently, from the α -shapes) of the input W , where α ranges from zero to infinity. The theoretical guarantees of [9] hold in a very wide setting, since W is assumed to be a sufficiently dense sampling of a general compact set. In [27], the family is given by the *witness complex* of L relative to W , or $\mathcal{C}^W(L)$ for short, where $L \subseteq W$ is a subset of landmarks constructed iteratively, starting with $L = \emptyset$, and inserting each time the point of W lying furthest away from L . Here, the points of $W \setminus L$ do not belong to the simplicial complex, but they are used as *witnesses*, to drive the construction of the complex. It is known that $\mathcal{C}^W(L)$ coincides with $\mathcal{D}^S(L)$ on smooth curves, and is still included in it on smooth surfaces [5, 27]. The assumptions on L in [27] are more stringent than the ones on W , but this is not an issue since L is generated by the algorithm.

Pursuing the work initiated in [5, 15, 27], this paper studies the behavior of the restricted Delaunay triangulations and witness complexes of points sampled from manifolds of arbitrary dimensions (in particular, manifolds of dimensions 3 and higher). Our first structural results (Section 3) are negative, namely: the structural properties of restricted Delaunay triangulations and witness complexes on curves and surfaces do not hold on manifolds of dimensions 3 or higher. The underlying intuition is that the normals of $\mathcal{D}^S(L)$ can be arbitrarily wrong if $\mathcal{D}^S(L)$ contains badly-shaped simplices, called *slivers* [15]. As a consequence, $\mathcal{C}^W(L)$ may not be included in $\mathcal{D}^S(L)$, which may not be homeomorphic nor even homotopy equivalent to S (see Theorem 2.4 below). This is true even if W and L satisfy strong sampling conditions, such as being arbitrarily dense uniform samples of S . A direct consequence of these negative results is that the algorithm of [27] cannot work on manifolds of dimensions 3 or more.

On the positive side, we show how W and L can be enriched, so as to make the structural results of [5, 27] hold on higher-dimensional manifolds (Section 4). To this end, we assign a non-negative weight $\omega(p)$ to each point $p \in L$, such that $\mathcal{D}^S(L)$ and $\mathcal{C}^W(L)$ are now replaced by their weighted versions, $\mathcal{D}_\omega^S(L)$ and $\mathcal{C}_\omega^W(L)$. The idea of assigning weights to the vertices comes from [13], where it was used to remove slivers from 3-dimensional Delaunay triangulations. This sliver removal technique was extended to higher dimensions by Cheng *et al.* [15], who showed that, under sufficient sampling conditions on L , there exists a distribution of weights ω such that $\mathcal{D}_\omega^S(L)$ is homeomorphic to S . Our main result is that, for the same distribution of weights and under similar conditions on L , $\mathcal{C}_\omega^W(L)$ is included in $\mathcal{D}_\omega^S(L)$, for all $W \subseteq S$. Since this is true in particular for $W = S$, combined with the fact that $\mathcal{C}_\omega^S(L)$ contains $\mathcal{D}_\omega^S(L)$, we get that $\mathcal{C}_\omega^S(L) = \mathcal{D}_\omega^S(L)$, which is homeomorphic to S . This is a generalization of the result of [5] to higher-dimensional manifolds. It is not quite practical since W has to be equal to S . In the more realistic case where W is a finite subset of S , we enlarge it by replacing its points by balls of same radius ζ : for sufficiently large values of ζ , this enlarged set W^ζ contains S , and hence $\mathcal{D}_\omega^S(L) \subseteq \mathcal{C}_\omega^{W^\zeta}(L)$. And if ζ is not too large, then $\mathcal{C}_\omega^{W^\zeta}(L)$ is still included in $\mathcal{D}_\omega^S(L)$. Thus, we obtain $\mathcal{C}_\omega^{W^\zeta}(L) = \mathcal{D}_\omega^S(L)$ under sufficient conditions on W, L, ζ . Here again, the condition on L is stringent, but the one on W is mild.

Our positive structural results suggest combining the method of [27] with the sliver exudation technique of [13], in order to make the approach work on higher-dimensional manifolds (Section 5).

Our combination is the simplest possible: at each iteration of the algorithm, we insert a new point p in L , compute its best possible weight, and update $\mathcal{C}_\omega^{W^\zeta}(L)$. This algorithm is very simple conceptually, and to some extent it can be viewed as a dynamic version of the algorithm of [15], where the Delaunay triangulation of W would be constructed progressively as the weights of the points of W are computed. This raises the question of whether the weight assignment process removes all the slivers from the vicinity of $\mathcal{D}_\omega^S(L)$, since some of the weights are assigned in early stages of the course of the algorithm. We prove that all slivers are eventually removed at some point (Theorem 5.2), and that consequently $\mathcal{C}_\omega^{W^\zeta}(L)$ is homeomorphic to S , provided that W is a dense enough sampling of S . The resulting algorithm has a running time bounded by $c(d)|W|^2$, where $c(d)$ is some constant depending solely (yet exponentially) on d . This bound is comparable to the ones obtained in previous work [15], as will be discussed in Section 6.

2 Background and definitions

The ambient space is \mathbb{R}^d , $d \geq 3$, equipped with the usual Euclidean norm $\|p\| = \sqrt{\sum_{i=1}^d p_i^2}$.

2.1 Manifolds and samples

All manifolds considered in this paper are compact closed submanifolds of \mathbb{R}^d , of dimension two or higher. The case of curves has already been addressed¹ in [27]. The *reach* of a manifold S , or $\text{rch}(S)$ for short, is the minimum distance of a point on S to the medial axis of S . All manifolds in this paper are assumed to have a positive reach. This is equivalent to saying that they are C^1 -continuous, and that their normal vector field satisfies a Lipschitz condition [24].

Given a (finite or infinite) subset L of a manifold S , and a positive parameter ε , L is an ε -*sample* of S if every point of S is at Euclidean distance at most ε to L . This sampling condition amounts to saying that L achieves some minimal local density everywhere on the manifold S . Note however that L does not have to be uniformly sampled, although the bound ε is uniform. To make the density of L uniform, one can add the constraint that L is ε -*sparse*, that is, the pairwise Euclidean distances between the points of L are all at least ε . Note that an ε -sparse sample of a compact set is always finite.

2.2 Simplex shape

Given $k + 1$ points $p_0, \dots, p_k \in \mathbb{R}^d$, $[p_0, \dots, p_k]$ denotes the k -simplex of vertices p_0, \dots, p_k . The geometric realization of this simplex is the convex hull of p_0, \dots, p_k , which has dimension k if the vertices are affinely independent. Following [14, 29], we call *sliver measure* of $[p_0, \dots, p_k]$ the ratio²:

$$\varrho([p_0, \dots, p_k]) = \frac{\text{vol}([p_0, \dots, p_k])}{\min\{\|p_i - p_j\|, 0 \leq i < j \leq k\}^k},$$

where $\text{vol}([p_0, \dots, p_k])$ denotes the volume of the convex hull of p_0, \dots, p_k in \mathbb{R}^k . In the special case where $k = 1$, we have $\varrho([p_0, p_1]) = 1$ for any edge $[p_0, p_1]$. We extend the notion of sliver measure

¹The results of [27] apply to Lipschitz curves in the plane, but they can be extended to smooth curves in higher dimensions in a straightforward manner.

²The definition of [14, 29] generalizes the one of [13] to higher-dimensional simplices. It departs from the definition of [15], which introduced a bug in the proof of Lemma 10 of the same paper.

to the case $k = 0$ by imposing $\varrho([p_0]) = 1$ for any vertex p_0 . Given $\bar{\varrho} \geq 0$, simplex $[p_0, \dots, p_k]$ is said to be a $\bar{\varrho}$ -*sliver* if $\varrho([p_0, \dots, p_k]) < \bar{\varrho}^k/k!$. For sufficiently small $\bar{\varrho}$, this means that the volume of the simplex is small compared to the volume of the diametral k -ball of its shortest edge. As a result, the simplex is badly-shaped. Note however that having a good sliver measure does not always mean being well-shaped [13]. Parameter $\bar{\varrho}$ is called the *sliver bound* in the sequel.

2.3 Weighted points, Delaunay triangulation, and witness complex

Given a finite point set $L \subset \mathbb{R}^d$, a *distribution of weights on L* is a non-negative real-valued function $\omega : L \rightarrow [0, \infty)$. The quantity $\max_{u \in L, v \in L \setminus \{u\}} \frac{\omega(u)}{\|u-v\|}$ is called the *relative amplitude* of ω . Given an unweighted point $u \in \mathbb{R}^d$, the *weighted distance* of u to some weighted point $v \in L$ is $\|u-v\|^2 - \omega(v)^2$. Similarly, the weighted distance between two points $u, v \in L$ is $\|u-v\|^2 - \omega(u)^2 - \omega(v)^2$. This is actually not a metric, since it satisfies neither the non-negativity condition nor the triangle inequality.

Given a finite point set $L \subset \mathbb{R}^d$, and a distribution of weights ω on L , we denote by $\mathcal{D}_\omega(L)$ the *weighted Delaunay triangulation* of L , which is the dual complex of the Voronoi diagram of L in the weighted metric. For any simplex σ of $\mathcal{D}_\omega(L)$, we denote by $V_\omega(\sigma)$ the face of the weighted Voronoi diagram of L that is dual to σ . Whenever the relative amplitude of ω is less than $\frac{1}{2}$, the points of L have non-empty cells in the weighted Voronoi diagram of L , and in fact each point of L belongs to its own cell [13]. As a consequence, every point of L is a vertex of $\mathcal{D}_\omega(L)$.

Given any subset W of \mathbb{R}^d , we call $\mathcal{D}_\omega^W(L)$ the weighted Delaunay triangulation of L restricted to W . In the special case where all the weights are equal, $\mathcal{D}_\omega(L)$ coincides with the standard Euclidean Delaunay triangulation, and is therefore noted $\mathcal{D}(L)$. Similarly, $V_\omega(\sigma)$ becomes $V(\sigma)$, and $\mathcal{D}_\omega^W(L)$ becomes $\mathcal{D}^W(L)$.

Definition 2.1 *Let W, L be two arbitrary subsets of \mathbb{R}^d such that L is finite, and let $\omega : L \rightarrow [0, \infty)$ be a distribution of weights on L .*

- *Given a point $w \in W$ and a simplex $\sigma = [p_0, \dots, p_k]$ with vertices in L , w is a ω -witness of σ (or simply w ω -witnesses σ) if p_0, \dots, p_k are among the $k+1$ nearest neighbors of w in the weighted distance, that is, $\forall i \in \{0, \dots, k\}, \forall q \in L \setminus \{p_0, \dots, p_k\}, \|w - p_i\|^2 - \omega(p_i)^2 \leq \|w - q\|^2 - \omega(q)^2$.*
- *The ω -witness complex of L relative to W , or $\mathcal{C}_\omega^W(L)$ for short, is the maximum abstract simplicial complex with vertices in L , whose faces are ω -witnessed by points of W .*

This definition comes from [17, 18]. From now on, W will be referred to as the set of witnesses, and L as the set of landmarks. In the special case where all the weights are equal, $\mathcal{C}_\omega^W(L)$ coincides with the witness complex in the standard Euclidean norm, and is therefore noted $\mathcal{C}^W(L)$.

Given a distribution of weights ω on L , for any simplex $\sigma = [p_0, \dots, p_k]$ of $\mathcal{D}_\omega^W(L)$ and any point $w \in W$ lying on the weighted Voronoi face dual to σ , we have $\|w - p_i\|^2 - \omega(p_i)^2 = \|w - p_j\|^2 - \omega(p_j)^2 \leq \|w - q\|^2 - \omega(q)^2$ for all $i, j \in \{0, \dots, k\}$ and all $q \in L \setminus \{p_0, \dots, p_k\}$. Hence, w is a ω -witness of all the subsimplices of σ , which therefore belong to $\mathcal{C}_\omega^W(L)$. As a result,

Corollary 2.2 *For any subsets $W, L \subseteq \mathbb{R}^d$ with L finite, for any $\omega : L \rightarrow [0, \infty)$, $\mathcal{D}_\omega^W(L) \subseteq \mathcal{C}_\omega^W(L)$.*

$\mathcal{D}_\omega^W(L)$ is sometimes called the *strong witness complex* of L relative to W in the literature [18]. The following result, due to de Silva [17], relates the weighted witness complex $\mathcal{C}_\omega^W(L)$ to the full weighted Delaunay triangulation $\mathcal{D}_\omega(L)$:

Theorem 2.3 (Corollary 7.6 of [17]) *For any subsets $W, L \subseteq \mathbb{R}^d$ with L finite, for any $\omega : L \rightarrow [0, \infty)$, we have $\mathcal{C}_\omega^W(L) \subseteq \mathcal{D}_\omega(L)$. Moreover, for any simplex σ of $\mathcal{C}_\omega^W(L)$, the weighted Voronoi face dual to σ intersects the convex hull of the ω -witnesses (among the points of W) of σ and of its subsimplices.*

On 1- and 2-dimensional manifolds, the unweighted witness complex is closely related to the unweighted restricted Delaunay triangulation [5, 27], which provides good topological and geometric approximations [1, 2]. Unfortunately, these properties do not extend to higher-dimensional manifolds, even under stronger sampling conditions:

Theorem 2.4 *For any positive constant $\mu < 1/3$, there is a closed compact hypersurface S of positive reach in \mathbb{R}^4 , and an $\Omega(\varepsilon)$ -sparse $O(\varepsilon)$ -sample L of S , with $\varepsilon = \mu \text{rch}(S)$, such that $\mathcal{D}^S(L)$ is not homotopy equivalent to S . The constants hidden in the Ω and O notations are absolute and do not depend on μ . In addition, for any $\delta > 0$, there is a δ -sample W of S such that $\mathcal{C}^W(L)$ neither contains nor is contained in $\mathcal{D}^S(L)$. Here, W can be made indifferently finite or infinite.*

The proof of this theorem, detailed in Section 3 of the paper, builds on an example of [15, §11]. The intuition is that, when $\mathcal{D}^S(L)$ contains slivers, it is possible to make its normals turn by a large angle (say $\pi/2$) by perturbing the points of L infinitesimally. Then, the combinatorial structure of $\mathcal{D}^S(L)$ can be changed arbitrarily by small perturbations of S . The consequence is that $\mathcal{D}^S(L)$ may not be homotopy equivalent to S , and it may not contain all the simplices of $\mathcal{C}^W(L)$ either. In addition, as emphasized in [27], for any $k \geq 2$, the k -simplices of $\mathcal{D}^S(L)$ may have arbitrarily small cells in the restricted Voronoi diagram of L of order $k + 1$, which implies that they may not be witnessed in W if ever $W \subsetneq S$.

2.4 Weighted cocone complex

Our proofs make use of the so-called *cocone complex*, which was first introduced by Amenta, Choi, Dey, and Leekha [3] in the context of 3d surface reconstruction from point samples, and which was later on adapted by Cheng, Dey, and Ramos [15] to the case of weighted point sets lying on manifolds of arbitrary dimensions and co-dimensions.

Given a finite subset L of \mathbb{R}^d , and a distribution of weights ω on L , we say that the weighted points of L lie in *general position* if no point of \mathbb{R}^d is equidistant to $d + 2$ points of L in the weighted distance, and if no $d + 1$ points on the convex hull of L are coplanar. Under this assumption, every simplex of $\mathcal{D}_\omega(L)$ has dimension at most d . It is always possible to perturb the points of L or their weights infinitesimally, so that the weighted point set L lies in general position. Therefore, in the rest of the paper we will assume implicitly that the weighted points of L lie in general position.

At any point p on a manifold S , there exist a tangent space $T(p)$ and a normal space $N(p)$. These two subspaces of \mathbb{R}^d are orthogonal, and their direct sum is \mathbb{R}^d . For any angle value $\theta \in [0, \pi/2]$, we call θ -*cocone* of S at p , or $K^\theta(p)$ for short, the cone of semi-aperture θ around the tangent space of S at p : $K^\theta(p) = \{q \in \mathbb{R}^d \mid \angle(pq, T(p)) \leq \theta\}$. The name *cocone* refers to the fact that $K^\theta(p)$ is the complement of a cone of semi-aperture $\frac{\pi}{2} - \theta$ around the normal space of S at p .

Given an angle $\theta \in [0, \pi/2]$, a manifold S , a finite point set $L \subset S$, and a distribution of weights $\omega : L \rightarrow [0, \infty)$, the *weighted θ -cocone complex* of L , noted $K_\omega^\theta(L)$, is the subcomplex of $\mathcal{D}_\omega(L)$ made of the simplices whose dual weighted Voronoi faces intersect the θ -cocone of at least one of their vertices. This means that a simplex $[p_0, \dots, p_k]$ of $\mathcal{D}_\omega(L)$ belongs to $K_\omega^\theta(L)$ if, and only if, its dual weighted Voronoi face intersects $K^\theta(p_0) \cup \dots \cup K^\theta(p_k)$. Note that the cones in [15] are defined

around approximations of the tangent spaces of S at the points of L . However, the results of [15] hold *a fortiori* when the approximations of the tangent spaces are error-free, which is the case here.

Theorem 2.5 (Lemmas 13, 14, 18 of [15]) *For any sliver bound $\bar{\varrho} > 0$, there exists a constant $c_{\bar{\varrho}} > 0$ such that, for any manifold S , for any ε -sparse 2ε -sample L of S with $\varepsilon \leq c_{\bar{\varrho}} \text{rch}(S)$, for any distribution of weights $\omega : L \rightarrow [0, \infty)$ of relative amplitude less than $\frac{1}{2}$, if $K_{\omega}^{\pi/32}(L)$ has no $\bar{\varrho}$ -sliver, then $K_{\omega}^{\pi/32}(L)$ coincides with $\mathcal{D}_{\omega}^S(L)$, which is homeomorphic to S .*

It is also proved in [15] that, for any $\bar{\omega} \in (0, 1/2)$, there exists a distribution of weights ω on L , of relative amplitude at most $\bar{\omega}$, that removes all $\bar{\varrho}$ -slivers from $K_{\omega}^{\pi/32}(L)$. This is true provided that $\bar{\varrho}$ is sufficiently small compared to $\bar{\omega}$, and that $\varepsilon \leq c_{\bar{\varrho}} \text{rch}(S)$ (as in Theorem 2.5, the choice of $\bar{\varrho}$ influences the bound on ε). Note that the results of [15] hold in fact in the slightly more general setting where ε is a (non-uniform) 1-Lipschitz function, everywhere bounded by a fraction of the distance to the medial axis of S – see [15, §13].

2.5 Useful results

We will make use of the following Lemma from [26], whose proof holds the same in our context:

Lemma 2.6 (Lemma 6 of [26]) *Let S be a manifold, and let $p, q \in S$ be such that $\|p - q\| < \text{rch}(S)$. Then, $\inf\{\|q - q'\|, q' \in T(p)\} \leq \frac{\|p - q\|^2}{2 \text{rch}(S)}$, and $\angle(pq, T(p)) \leq \arcsin \frac{\|p - q\|}{2 \text{rch}(S)}$.*

We will also use the following results from [15], whose proofs have been adapted to our context in Appendix A:

Lemma 2.7 (Lemma 2 (ii) of [15]) *Let S be a manifold and $\theta \in [0, \pi/2]$ an angle value. Let $v \in S$ and $p \in K^{\theta}(v)$ be such that $\|p - v\| < \frac{1}{4} \text{rch}(S)$. Let q be the orthogonal projection of p onto $T(v)$, and q' the point of S closest to q . Then, $\|q' - p\| \leq \left(\sin \theta + \frac{2\|p - v\|}{\text{rch}(S)}\right) \|p - v\|$.*

Lemma 2.8 (Lemma 3 of [15]) *Let S be a manifold and $\theta \in [0, \pi/2)$ an angle value. Let L be an ε -sample of S , with $\varepsilon < \frac{1}{9}(1 - \sin \theta)^2 \text{rch}(S)$. For any distribution of weights $\omega : L \rightarrow [0, \infty)$ of relative amplitude less than $\frac{1}{2}$, for any $v \in L$ and any $p \in K^{\theta}(v) \cap V_{\omega}(v)$, we have $\|p - v\| \leq \frac{3\varepsilon}{1 - \sin \theta}$.*

3 Negative structural results

This section provides a series of counter-examples that prove the statement of Theorem 2.4 correct. Our first example shows that $\mathcal{D}^S(L)$ may not always be homeomorphic to S , even though L is an $\Omega(\varepsilon)$ -sparse $O(\varepsilon)$ -sample of S , for arbitrarily small ε (Lemma 3.1). Our construction builds on an example of [15, §11], which deals with hypersurfaces in \mathbb{R}^4 . The intuitive idea is that, when $\mathcal{D}^S(L)$ contains badly-shaped tetrahedra, it is possible to make its normals turn by a large angle (say $\frac{\pi}{2}$) by perturbing the points of L infinitesimally. It follows that the combinatorial structure of $\mathcal{D}^S(L)$ can be modified by small perturbations of S . We then extend our counter-example to show that $\mathcal{D}^S(L)$ may even not be homotopy equivalent to S (Lemma 3.2). Finally, we show that $\mathcal{C}^W(L)$ may not be included in $\mathcal{D}^S(L)$, even for arbitrarily dense sets $W \subseteq S$ (Lemma 3.3). The fact that $\mathcal{C}^W(L)$ may not contain $\mathcal{D}^S(L)$ either if $W \subsetneq S$ has already been proved in [27].

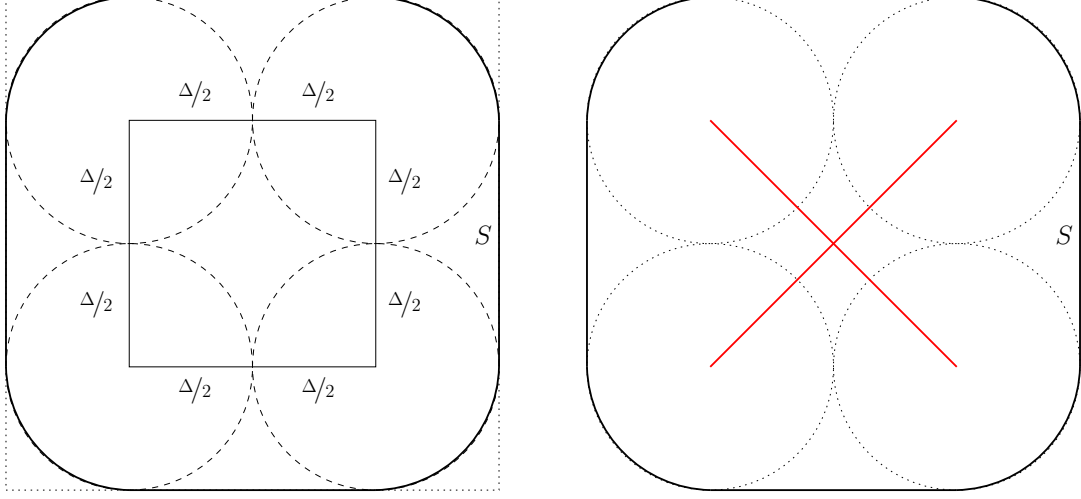


Figure 2: Left: 2-d version of hypersurface S (bold), defined as the boundary of the Minkowski sum of hypercube $[-\frac{\Delta}{2}, \frac{\Delta}{2}]^2$ (solid) with the ball of radius $\frac{\Delta}{2}$ centered at the origin (copies of this ball are dashed). Hypercube $[-\Delta, \Delta]^2$ is marked by dotted lines. Right: S and its medial axis.

Lemma 3.1 *For any positive constant $\mu < \frac{1}{3}$, there exist a compact closed hypersurface S in \mathbb{R}^4 and an $\Omega(\varepsilon)$ -sparse $O(\varepsilon)$ -sample L of S , with $\varepsilon = \mu \text{rch}(S)$, such that $\mathcal{D}^S(L)$ is not homeomorphic to S . The constants hidden in the Ω and O notations are absolute and do not depend on μ .*

Proof. Let $\Delta = \frac{2}{\mu}$. In \mathbb{R}^4 , endowed with an orthonormal frame (x, y, z, t) , we construct a hypersurface S of reach $\frac{\Delta}{2} = \frac{1}{\mu}$. Consider the Minkowski sum of hypercube $[-\frac{\Delta}{2}, \frac{\Delta}{2}]^4$ with the ball of radius $\frac{\Delta}{2}$ centered at the origin. The result is a smoothed-out version of hypercube $[-\Delta, \Delta]^4$, as illustrated in Figure 2 (left). Let S be its boundary. The reach of S is $\frac{\Delta}{2}$, as shown in Figure 2 (right). Let $\varepsilon = \mu \text{rch}(S) = 1$, and let $\delta > 0$ be an arbitrarily small parameter. Consider points $u = (1, 0, 0, \Delta)$, $v = (1, 1, 0, \Delta)$, $w = (0, 1, 0, \Delta)$, and $p_0 = (0, 0, \delta, \Delta)$. Let $c_0 = (\frac{1}{2}, \frac{1}{2}, \frac{\delta}{2}, \Delta)$. It is easily seen that c_0 is the circumcenter of $[u, v, w, p_0]$. Moreover, all these points belong to S , which coincides with hyperplane $t = \Delta$ in their vicinity. Let $r_0 = \|c_0 - u\| = \|c_0 - v\| = \|c_0 - w\| = \|c_0 - p_0\|$. We generate an ε -sparse 2ε -sample L_0 of S by an iterative process, starting with $L_0 = \{u, v, w, p_0\}$, and inserting at each iteration the point of S lying furthest away from the current point set L_0 , until the farthest point of S is no farther than 2ε from L_0 . Since S is compact, the process terminates, and the outcome is a 2ε -sample of S . Moreover, since u, v, w, p_0 lie at least ε away from one another, and since every point inserted in L_0 lies at least 2ε away from L_0 at the time of its insertion, L_0 is ε -sparse. Finally, no point of ball $B(c_0, r_0)$ lies farther from $\{u, v, w, p_0\}$ than $2r_0 = 2\sqrt{\frac{1}{2} + \frac{\delta^2}{4}}$, which is less than 2ε since δ is arbitrarily small. It follows that the interior of $B(c_0, r_0)$ contains no point of L_0 , which implies that $[u, v, w, p_0]$ belongs to $\mathcal{D}^S(L_0)$, its dual Voronoi edge intersecting S at c_0 . Observe also that, since u, v, w, p_0 belong to hyperplane $t = \Delta$, the normal of $[u, v, w, p_0]$ is aligned with vector $(0, 0, 0, 1)$, as shown in Figure 3 (left).

We now deform S slightly and create a small bump at c_0 , such that the top of the bump is moved by $\frac{\delta}{2}$ into the t -dimension, outward the hypercube. This bump changes the local feature size of S . However, since δ is arbitrarily small, the radius of curvature of the bump can be forced to be at least $\frac{\Delta}{2}$, which implies that the reach of S remains equal to $\frac{\Delta}{2} = \frac{1}{\mu}$. Furthermore, since

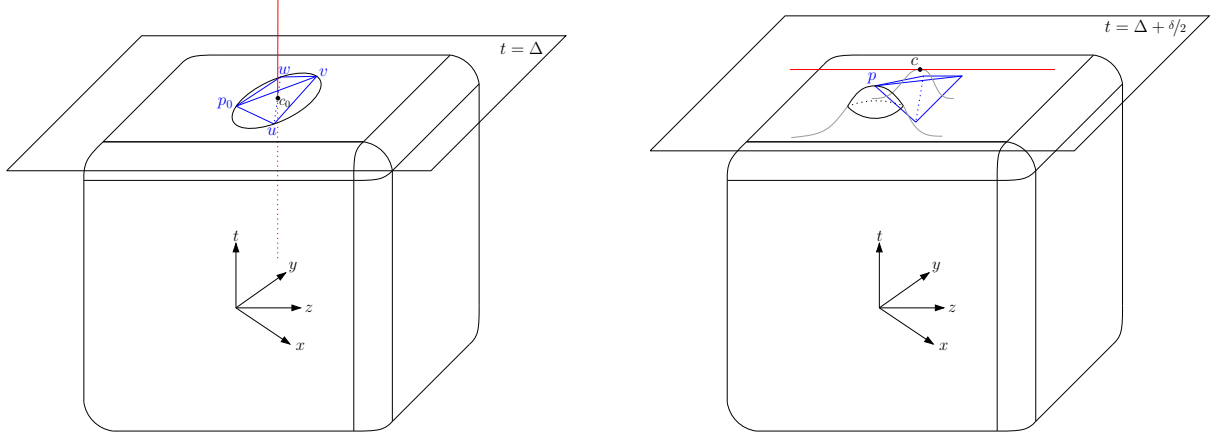


Figure 3: Left: tetrahedron $[u, v, w, p_0]$ and its dual Voronoi edge. Right: after perturbation of S .

c_0 is the center of a Delaunay ball of radius greater than 1, we can assume that δ is small enough for the points of L_0 to remain on S . Let $c = (\frac{1}{2}, \frac{1}{2}, \frac{\delta}{2}, \Delta + \frac{\delta}{2})$ be at the top of the bump. Since the points of L_0 are located in hyperplane $t = \Delta$ in the vicinity of $[u, v, w, p_0]$, c is equidistant to u, v, w, p_0 , and closer to these points than to any other point of L_0 . This implies that the open ball $B_c = B(c, \|c - u\|)$ contains no point of L_0 and has u, v, w, p_0 on its bounding sphere. Hence, B_c is a Delaunay ball circumscribing $[u, v, w, p_0]$, and c belongs to the Voronoi edge dual to $[u, v, w, p_0]$. Moreover, since u, v, w and $(0, 0, 0, \Delta)$ are cocircular, ∂B_c passes also through $(0, 0, 0, \Delta)$.

We deform S further by creating another small bump, at point $(0, 0, 0, \Delta)$ this time, so as to move this point by δ into the t -dimension, outward the hypercube. Let $p = (0, 0, 0, \Delta + \delta)$ be the top of the bump — see Figure 3 (right). A quick computation shows that $\|c - p\| = \|c - u\|$, which implies that $p \in \partial B_c$. Here again, by choosing δ sufficiently small, we can make sure that the radius of curvature of the bump is at least $\frac{\Delta}{2}$, which means that the reach of the deformed hypersurface is still $\frac{\Delta}{2} = \frac{1}{\mu}$. We can also make sure that the bump of p is disjoint from the bump of c since $\|c - p\| > \frac{1}{\sqrt{2}}$, and that the points of $L_0 \setminus \{p_0\}$ remain³ on S . It follows that B_c is empty of points of L , where L is defined by $L = L_0 \cup \{p\} \setminus \{p_0\}$. Since ∂B_c contains u, v, w, p , B_c is a Delaunay ball circumscribing $[u, v, w, p]$. Equivalently, c belongs to the Voronoi edge e dual to $[u, v, w, p]$. Note also that L is an $(\varepsilon - \delta)$ -sparse $(2\varepsilon + \delta)$ -sample of S .

Since $[u, v, w, p]$ is included in hyperplane $z = 0$, its dual Voronoi edge e is aligned with $(0, 0, 1, 0)$, as illustrated in Figure 3 (right). This edge is incident to four Voronoi 2-faces, which are dual to the four facets of $[u, v, w, p]$. These 2-faces can be seen as extrusions, into the z -dimension $(0, 0, 1, 0)$, of the edges of the Voronoi diagram of $\{u, v, w, p\}$ inside hyperplane $z = 0$. Among these Voronoi edges, two lie above the plane $t = \Delta + \frac{\delta}{2}$, and two lie below. As a result, in \mathbb{R}^4 , two Voronoi 2-faces incident to e lie above hyperplane $t = \Delta + \frac{\delta}{2}$. These two Voronoi 2-faces do not intersect S , except at c and possibly at the bump of p . Now, the circumradii of the facets of $[u, v, w, p]$ are at most $\|c - u\| = \frac{\sqrt{1+\delta^2}}{\sqrt{2}} < \mu \operatorname{rch}(S)$, thus, inside hyperplane $z = 0$, Amenta and Bern's normal lemma [1, Lemma 7] states that the edges of the Voronoi diagram of $\{u, v, w, p\}$ make angles of at most $\arcsin \frac{\mu\sqrt{3}}{1-\mu} < \frac{\pi}{3}$ with vector $(0, 0, 0, 1)$. As a consequence, any Voronoi 2-face f incident to e

³They lie at least ε away from p_0 , and hence at least $\varepsilon - \delta$ away from $(0, 0, 0, \Delta)$.

in \mathbb{R}^4 makes an angle of at most $\frac{\pi}{3}$ with the plane passing through c , of directions $(0, 0, 1, 0)$ and $(0, 0, 0, 1)$ (note that the affine hull $\text{aff}(f)$ intersects this plane along the line $\text{aff}(e)$). Since p lies $\frac{1}{\sqrt{2}}$ away from this plane and only $\frac{\delta}{2}$ above c , for sufficiently small δ the Voronoi 2-faces incident to e lying above hyperplane $t = \Delta + \frac{\delta}{2}$ do not intersect the bump of p . As a consequence, they intersect S only at c , and therefore their dual Delaunay triangles are incident to exactly one tetrahedron of $\mathcal{D}^S(L)$, namely $[u, v, w, p]$. Hence, $\mathcal{D}^S(L)$ is not a closed hypersurface, and for this reason it cannot be homeomorphic to S . \square

Observe that the construction performed in the above proof leads to a degenerate case, where the Voronoi edge e dual to tetrahedron $[u, v, w, p]$ intersects S tangentially at c . This degeneracy can be removed by inflating the bump of c infinitesimally, so that it intersects e twice and transversally, but still no other Voronoi edge.

Note also that tetrahedron $[u, v, w, p]$ is a sliver, since vertex p lies close to the affine hull of $[u, v, w]$. The original counter-example of [15] was designed to highlight the fact that the normals of slivers in the restricted Delaunay triangulation may differ significantly from the normals of the underlying manifold. This is not true for non-sliver simplices, as shown in Lemma 15 of [15]. In this respect, the fact that $[u, v, w, p]$ is a sliver is crucial for our counter-example to work.

Lemma 3.2 *For any positive constant $\mu < \frac{1}{3}$, there exist a compact closed hypersurface S in \mathbb{R}^4 and an $\Omega(\varepsilon)$ -sparse $O(\varepsilon)$ -sample L of S , with $\varepsilon = \mu \text{rch}(S)$, such that $\mathcal{D}^S(L)$ is not homotopy equivalent to S . The constants hidden in the Ω and O notations are absolute and do not depend on μ .*

Proof. Let $\Delta = \frac{2}{\mu}$, and let $\delta > 0$ be an arbitrarily small parameter. We begin our analysis with the example built in the proof of Lemma 3.1. We will modify S and L in such a way that tetrahedron $[p, u, v, w]$ will no longer belong to $\mathcal{D}^S(L)$ while its four facets will still. This will prevent $\mathcal{D}^S(L)$ from being homotopy equivalent to S .

Consider point $q = (\frac{1+\sqrt{2}}{2}, \frac{1}{2}, \delta^2, \Delta + \delta)$. The distance of q to hyperplane $t = \Delta$ is δ , which is arbitrarily small. Hence, as explained in the proof of Lemma 3.1, it is possible to deform S slightly by creating a small bump of radius of curvature at least $\frac{\Delta}{2}$ at point $(\frac{1+\sqrt{2}}{2}, \frac{1}{2}, \delta^2, \Delta)$, so that S now passes through q while its reach remains $\frac{\Delta}{2} = \frac{1}{\mu}$. Moreover, since q lies farther than $\frac{1}{2}$ from $\{p, u, v, w\}$, we can assume without loss of generality that its bump does not affect the positions of p, u, v, w .

The circumcenter of pentahedron $[p, u, v, w, q]$ is $c' = (\frac{1}{2}, \frac{1}{2}, \frac{\delta^2}{2}, \Delta + \frac{\delta}{2})$, and its circumradius r' is less than $\varepsilon = 1$ (for sufficiently small δ). It follows that every point of S lying in the ball $B(c', r')$ is at distance $O(\varepsilon)$ of $\{p, u, v, w, q\}$. In addition, q is farther than $\frac{\varepsilon}{2}$ from $\{p, u, v, w\}$. Therefore, if we modify L by inserting q and deleting all the points that lie in the interior of $B(c', r')$, L remains an $\Omega(\varepsilon)$ -sparse $O(\varepsilon)$ -sample of S . Moreover, $[p, u, v, w, q]$ is now a Delaunay pentahedron, whose dual Voronoi vertex is c' .

Note that line (c, c') is the affine hull of the Voronoi edge e dual to $[p, u, v, w]$, and that c' is an endpoint of e — see Figure 4 (top). Recall that, among the four 2-faces incident to e , two lie above hyperplane $t = \Delta + \frac{\delta}{2}$. Let f_{puv} and f_{pvw} denote these two 2-faces. They are dual to triangles $[p, u, v]$ and $[p, v, w]$ respectively, since p lies above hyperplane $t = \Delta$, which contains $[u, v, w]$. Moreover, f_{puv} and f_{pvw} are convex polygons, whose boundaries are two cycles of Voronoi edges that intersect each other along e . In the cycle of ∂f_{puv} , one edge adjacent to e , noted e_{puvq} , is dual to tetrahedron $[p, u, v, q]$. Similarly, in the cycle of ∂f_{pvw} , one edge adjacent to e , noted e_{pvwq} , is dual to $[p, v, w, q]$.

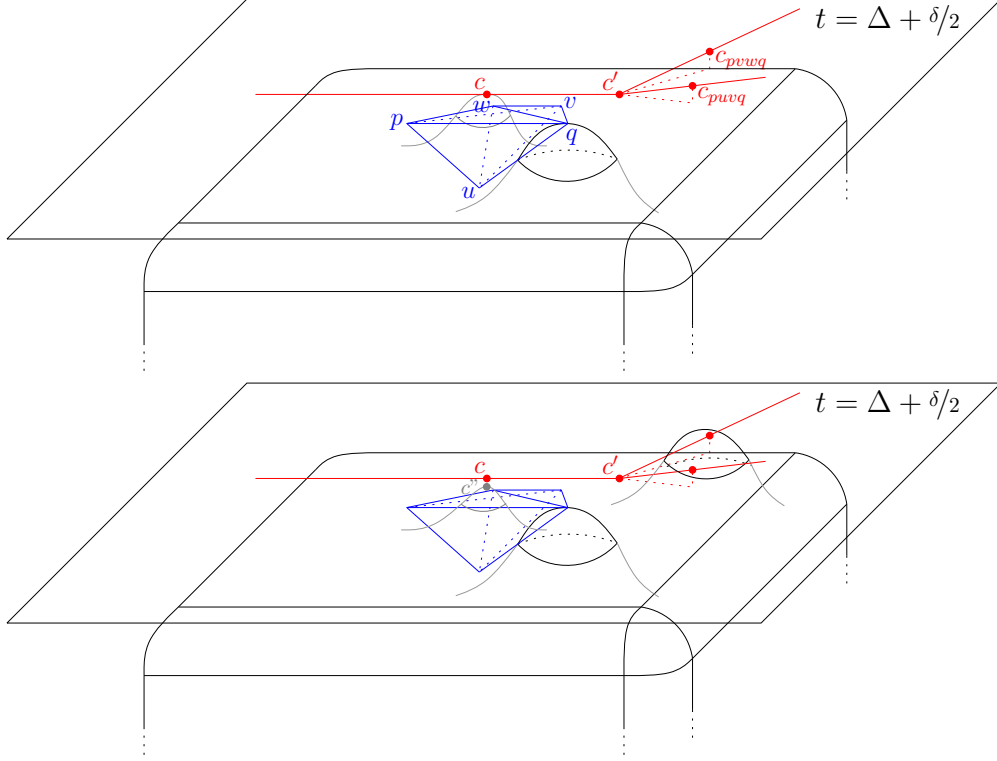


Figure 4: Top: pentahedron $[p, u, v, w, q]$ and the duals of $[p, u, v, w]$, $[p, u, v, q]$, and $[p, v, w, q]$. Bottom: the bump of c_{puvq} and c_{pvwq} (the bump of c has been slightly deflated).

Note that c' is an endpoint of both e_{puvq} and e_{pvwq} . Moreover, it can be easily checked that the line $\text{aff}(e_{puvq})$ passes also through point $c_{puvq} = \left(\frac{1}{2} + \frac{\delta^2(\delta^2+1)}{1+\sqrt{2}}, \frac{1}{2}, -\frac{1}{2}, \Delta + \frac{\delta}{2} + \frac{\delta(\delta^2+1)}{1+\sqrt{2}}\right)$, while the line $\text{aff}(e_{pvwq})$ passes through $c_{pvwq} = \left(\frac{1}{2}, \frac{1}{2} + \delta^2(\delta^2+1), -\frac{1}{2}, \Delta + \frac{\delta}{2} + \delta(\delta^2+1)\right)$. This implies that e_{puvq} and e_{pvwq} make angles of $O(\delta)$ with hyperplane $t = \Delta + \frac{\delta}{2}$. So, we are in a situation where tetrahedron $[p, u, v, w]$ has a horizontal dual edge, while two of its adjacent tetrahedra, namely $[p, u, v, q]$ and $[p, v, w, q]$, have almost horizontal dual edges, as illustrated at the top of Figure 4.

Since $\|c_{puvq} - p\| = \|c_{puvq} - u\| = \|c_{puvq} - v\| = \|c_{puvq} - q\| < \|c_{puvq} - w\|$, which is less than $\varepsilon = 1$ for sufficiently small δ , we can modify⁴ L such that the ball $B(c_{puvq}, \|c_{puvq} - q\|)$ contains no point of L in its interior, while L still remains an $\Omega(\varepsilon)$ -sparse $O(\varepsilon)$ -sample of S . Similarly, we can assume without loss of generality that $B(c_{pvwq}, \|c_{pvwq} - q\|)$ is a Delaunay ball. It follows that $c_{puvq} \in e_{puvq}$ and $c_{pvwq} \in e_{pvwq}$. Since c_{puvq} and c_{pvwq} lie $O(\delta)$ away from each other, $O(\delta)$ above hyperplane $t = \Delta$, and $\Omega(\varepsilon)$ away from L , we can deform S by creating a bump passing through c_{puvq} and c_{pvwq} , of height $O(\delta)$ and radius of curvature at least $\frac{\Delta}{2}$, while maintaining the points of L on S — see Figure 4 (bottom). Moreover, since c_{puvq} and c_{pvwq} also lie $\Omega(\varepsilon)$ away from c , we can assume without loss of generality that their bump does not touch the bump of c . It follows that tetrahedra $[p, u, v, w]$, $[p, u, v, q]$, and $[p, v, w, q]$ belong to $\mathcal{D}^S(L)$, while S is still tangent at c to the Voronoi edge e dual to tetrahedron $[p, u, v, w]$. We call S^+ the current version of hypersurface S , and $\mathcal{D}^{S^+}(L)$ the Delaunay triangulation of L restricted to S^+ .

⁴For instance, we can simply delete the points of L that lie in the interior of the ball $B(c_{puvq}, \|c_{puvq} - q\|)$.

Our last operation consists in deflating slightly the bump of c , such that S no longer intersects e , and thus $[p, u, v, w]$ no longer belongs to $\mathcal{D}^S(L)$, as illustrated at the bottom of Figure 4. Note however that $[p, u, v, q]$ and $[p, v, w, q]$ (and thus triangles $[p, u, v]$ and $[p, v, w]$) are still in $\mathcal{D}^S(L)$, since the bump of c_{puvq} and c_{pvwq} is disjoint from the bump of c . Recall also that the Voronoi 2-faces dual to $[p, u, w]$ and $[u, v, w]$ lie below hyperplane $t = \Delta + \frac{\delta}{2}$, and that they make angles of at most $\frac{\pi}{3}$ with vector $(0, 0, 0, 1)$. Since the deflation of the bump of c is arbitrarily small, the Voronoi 2-faces dual to $[p, u, w]$ and $[u, v, w]$ still intersect S . As a consequence, the two triangles remain in $\mathcal{D}^S(L)$. We call S^- the current version of hypersurface S , and $\mathcal{D}^{S^-}(L)$ the Delaunay triangulation of L restricted to S^- .

The result of these operations is that, although L is an $\Omega(\varepsilon)$ -sparse $O(\varepsilon)$ -sample of both hypersurfaces S^+ and S^- , whose homotopy types and reaches are the same as the ones of S , $\mathcal{D}^{S^+}(L)$ and $\mathcal{D}^{S^-}(L)$ are different. Specifically, tetrahedron $[p, u, v, w]$ is contained in $\mathcal{D}^{S^+}(L)$ but not in $\mathcal{D}^{S^-}(L)$, whereas its facets belong to both complexes. It follows that the Euler characteristics of $\mathcal{D}^{S^+}(L)$ and $\mathcal{D}^{S^-}(L)$ differ⁵, which implies that the complexes have different homotopy types. Therefore, at least one of them is not homotopy equivalent to the 3-sphere S . \square

Lemma 3.2 proves the first part of the statement of Theorem 2.4. We now turn our focus to witness complexes and show the following result, which, combined with the negative results of [27], proves the second part of the statement of Theorem 2.4:

Lemma 3.3 *For any positive constants $\mu, \nu < \frac{1}{3}$, there exist a compact closed hypersurface S in \mathbb{R}^4 , an $\Omega(\varepsilon)$ -sparse $O(\varepsilon)$ -sample L of S , and a δ -sample W of S , with $\varepsilon = \mu \text{rch}(S)$ and $\delta = \nu \text{rch}(S)$, such that $\mathcal{C}^W(L)$ is not included in $\mathcal{D}^S(L)$. The constants hidden in the Ω and O notations are absolute and do not depend on μ nor ν . Moreover, W can be made indifferently finite or infinite, and arbitrarily dense.*

Proof. Let S^- , L , e , and c be defined as in the proof of Lemma 3.2. Recall that tetrahedron $[p, u, v, w]$ does not belong to $\mathcal{D}^{S^-}(L)$, whereas its facets do. We assume without loss of generality that c is not an endpoint of the Voronoi edge e , which means that the bounding sphere of the Delaunay ball $B(c, \|c - p\|)$ contains no point of L other than p, u, v, w . This condition can be ensured by an infinitesimal perturbation of the points of $L \setminus \{p, u, v, w\}$. Let $d_c = \min_{p' \in L \setminus \{p, u, v, w\}} \|c - p'\|$. This quantity is greater than $\|c - p\|$ since $B(c, \|c - p\|)$ contains no point of $L \setminus \{p, u, v, w\}$.

Consider any (finite or infinite) set of witnesses $W \subseteq S^-$ such that, for each facet σ of $[p, u, v, w]$, W contains at least one point of $S^- \cap V(\sigma)$ (every such point witnesses σ and its subsimplices). Assume further that W contains the top point of the bump of c (call this point c''). In the last stage of the perturbation of S described in the proof of Lemma 3.2, we slightly deflated the bump of c , such that c'' lies strictly below c . Note that p is the vertex of $[p, u, v, w]$ lying furthest away from c'' . Since the deflation is arbitrarily small, we can assume without loss of generality that $\|c - c''\| < \frac{1}{2}(d_c - \|c - p\|)$. This implies that the ball $B(c'', \|c'' - p\|) \subseteq B(c, \|c - p\| + 2\|c - c''\|)$ is included in the interior of $B(c, d_c)$. As a result, $B(c'', \|c'' - p\|)$ contains no point of $L \setminus \{p, u, v, w\}$. Since p, u, v, w belong to $B(c'', \|c'' - p\|)$, tetrahedron $[p, u, v, w]$ is witnessed by c'' . And since the facets of $[p, u, v, w]$ and their subsimplices are witnessed by points of W , $[p, u, v, w]$ belongs to the witness complex $\mathcal{C}^W(L)$. However, we saw in the proof of Lemma 3.2 that $[p, u, v, w]$ does not belong to $\mathcal{D}^{S^-}(L)$. \square

⁵Specifically, $\chi(\mathcal{D}^{S^+}(L)) = \chi(\mathcal{D}^{S^-}(L)) - 1$.

4 Positive structural results

In this section, $\bar{\omega} \in (0, 1/2)$ and $\bar{\varrho} > 0$ are fixed constants. Moreover, for convenience, we introduce the following quantities depending on integer $k \geq 0$:

$$c_1(k) = 4(1 + 2\bar{\omega} + k(1 + 3\bar{\omega})), \quad (1)$$

$$c_2(k) = 4(3 + 6\bar{\omega} + 2k(1 + 3\bar{\omega})). \quad (2)$$

These quantities will be instrumental in our proofs, where k will stand for the dimension of a simplex, or a manifold, or the ambient space.

Let S be a manifold in \mathbb{R}^d , W a (finite or infinite) δ -sample of S , and L a finite ε -sparse ε -sample of W , for two parameters δ, ε to be specified later on. Note that L is an $(\varepsilon + \delta)$ -sample of S . According to Theorem 2.4, $\mathcal{C}^W(L)$ may not coincide with $\mathcal{D}^S(L)$, even under strong assumptions on δ, ε . Specifically:

- some simplices of $\mathcal{D}^S(L)$ may not belong to $\mathcal{C}^W(L)$ if W does not span S entirely. Our solution to this problem is to enlarge the set of witnesses, in order to make it cover S . More precisely, we dilate W by a ball of radius ζ centered at the origin, so that the set of witnesses is now $W^\zeta = \bigcup_{w \in W} B(w, \zeta)$. For $\zeta \geq \delta$, this set contains S , hence $\mathcal{D}^S(L) \subseteq \mathcal{D}^{W^\zeta}(L) \subseteq \mathcal{C}^{W^\zeta}(L)$.
- some simplices of $\mathcal{C}^W(L)$ may not belong to $\mathcal{D}^S(L)$ if the latter contains $\bar{\varrho}$ -slivers. To remedy this problem, we assign non-negative weights to the landmarks, so that $\mathcal{D}^S(L)$ and $\mathcal{C}^{W^\zeta}(L)$ are now replaced by their weighted versions, $\mathcal{D}_\omega^S(L)$ and $\mathcal{C}_\omega^{W^\zeta}(L)$. Given an angle value $\theta \in (0, \pi/2)$, our main structural result (Theorem 4.1) states that, under sufficient conditions on $\delta, \varepsilon, \zeta$, we have $\mathcal{C}_\omega^{W^\zeta}(L) \subseteq K_\omega^\theta(L)$ for any $\omega : L \rightarrow [0, \infty)$ of relative amplitude at most $\bar{\omega}$. Therefore, $\mathcal{C}_\omega^{W^\zeta}(L) \subseteq \mathcal{D}_\omega^S(L)$ whenever $\theta \leq \frac{\pi}{32}$ and ω removes all $\bar{\varrho}$ -slivers from $K_\omega^{\pi/32}(L)$, by Theorem 2.5.

Theorem 4.1 *Let S be a manifold in \mathbb{R}^d , W a (finite or infinite) δ -sample of S , and L a finite ε -sparse ε -sample of W . Given three constants: an amplitude bound $\bar{\omega} \in (0, 1/2)$, a sliver bound $\bar{\varrho} > 0$, and an angle parameter $\theta \in (0, \pi/2)$, if $\delta, \varepsilon, \zeta$ satisfy the following conditions:*

$$\text{H1} \quad \frac{5}{2(1-\bar{\omega}^2)\sin\theta} \delta \leq \varepsilon \leq \frac{5(1-\bar{\omega}^2)\sin\theta}{(2(1-\bar{\omega}^2)\sin\theta + 5c_2(d))^2} \text{rch}(S),$$

$$\text{H2} \quad \zeta \in \left[\delta, \frac{2(1-\bar{\omega}^2)\sin\theta}{5} \varepsilon \right],$$

then $\mathcal{D}_\omega^S(L) \subseteq \mathcal{D}_\omega^{W^\zeta}(L) \subseteq \mathcal{C}_\omega^{W^\zeta}(L) \subseteq K_\omega^\theta(L)$ for any $\omega : L \rightarrow [0, \infty)$ of relative amplitude at most $\bar{\omega}$. If in addition $\theta \leq \pi/32$, $\varepsilon \leq c_{\bar{\varrho}} \text{rch}(S)$, and ω is such that $K_\omega^{\pi/32}(L)$ contains no $\bar{\varrho}$ -sliver⁶, then Theorem 2.5 implies that $\mathcal{D}_\omega^{W^\zeta}(L) = \mathcal{C}_\omega^{W^\zeta}(L) = K_\omega^{\pi/32}(L) = \mathcal{D}_\omega^S(L)$, which is homeomorphic to S .

H1 requires that W is dense compared to L , which must be dense compared to $\text{rch}(S)$. This is very similar in spirit to the condition of [27]. H2 bounds the dilation parameter ζ . The smaller the angle θ of semi-aperture of the cocones, the smaller ε must be for $K_\omega^\theta(L)$ to contain $\mathcal{D}_\omega^S(L)$, and the smaller ζ must be for $K_\omega^\theta(L)$ to contain $\mathcal{C}_\omega^{W^\zeta}(L)$. Condition $\zeta \geq \delta$ ensures that $\mathcal{D}_\omega^S(L) \subseteq \mathcal{C}_\omega^{W^\zeta}(L)$.

In the special case where $W = S$ (which implies $\delta = 0$) and $\zeta = 0$, H2 and the left-hand side of H1 become void, and the theorem states that $\mathcal{C}_\omega^S(L)$ coincides with $\mathcal{D}_\omega^S(L)$ for a suitable distribution of weights $\omega : L \rightarrow [0, \infty)$, provided that L is an ε -sparse ε -sample of S , for a small enough ε . This is an extension of the result of Attali *et al.* [5] to higher dimensions, with an additional sparseness

⁶As mentioned after Theorem 2.5, such distributions of weights exist, provided that $\bar{\varrho}$ is sufficiently small.

condition⁷ on L . The general case of Theorem 4.1 allows to have $W \subsetneq S$, which is of practical interest.

Another useful property, stated as Theorem 4.2 below, is that the weighted witness complex actually includes the weighted cocone complex for any distribution of weights of relative amplitude at most $\bar{\omega}$, provided that ζ is sufficiently large compared to ε . Note that this condition on ζ is incompatible with H2: as a consequence, two different values of ζ must be used to bound $K_\omega^\theta(L)$, as emphasized in our algorithm – see Section 5.

Theorem 4.2 *If $\theta \leq \frac{\pi}{32}$, and if δ, ε satisfy Condition H1 of Theorem 4.1, then, for any $\zeta \geq 6 \sin \theta \varepsilon$, for any $\omega : L \rightarrow [0, \infty)$ of relative amplitude at most $\bar{\omega}$, $K_\omega^\theta(L) \subseteq \mathcal{D}_\omega^{W^\zeta}(L) \subseteq \mathcal{C}_\omega^{W^\zeta}(L)$.*

Overview of the proofs. The rest of Section 4 is devoted to the proofs of Theorems 4.1 and 4.2.

The proof of Theorem 4.1 is given in Section 4.1 below. Showing that $\mathcal{D}_\omega^S(L) \subseteq \mathcal{D}_\omega^{W^\zeta}(L) \subseteq \mathcal{C}_\omega^{W^\zeta}(L)$ is in fact easy: since W is a δ -sample of S , W^ζ contains S whenever $\zeta \geq \delta$, which is guaranteed by H2; it follows immediately that $\mathcal{D}_\omega^S(L) \subseteq \mathcal{D}_\omega^{W^\zeta}(L)$, which by Corollary 2.2 is included in $\mathcal{C}_\omega^{W^\zeta}(L)$ for any $\omega : L \rightarrow [0, \infty)$. Showing that $\mathcal{C}_\omega^{W^\zeta}(L) \subseteq K_\omega^\theta(L)$ requires more work, but the core argument is very simple, namely: the Euclidean distances between a point of S and its k nearest landmarks in the weighted distance are bounded (Lemma 4.4). From this fact we derive some bounds on the Euclidean distances between the simplices of $\mathcal{C}_\omega^{W^\zeta}(L)$ and their ω -witnesses (Lemma 4.5). These bounds are then used to show that the ω -witnesses of a simplex σ lie in the θ -cocones of the vertices of σ , from which we deduce that $\mathcal{C}_\omega^{W^\zeta}(L) \subseteq K_\omega^\theta(L)$ (Lemma 4.6).

The proof of Theorem 4.2 is given in Section 4.2. Intuitively, if L is a sufficiently dense sampling of S , then, for any simplex $\sigma \in K_\omega^\theta(L)$ and any vertex v of σ such that $V_\omega(\sigma) \cap K^\theta(v) \neq \emptyset$, the points of $V_\omega(\sigma) \cap K^\theta(v)$ lie close to $T(v)$ and hence close to S . Therefore, they belong to W^ζ whenever ζ is large enough, which implies that $\sigma \in \mathcal{D}_\omega^{W^\zeta}(L) \subseteq \mathcal{C}_\omega^{W^\zeta}(L)$.

4.1 Proof of Theorem 4.1

As explained above, all we have to do is to show that, whenever Conditions H1–H2 are satisfied, $\mathcal{C}_\omega^{W^\zeta}(L)$ is included in $K_\omega^\theta(L)$ for any $\omega : L \rightarrow [0, \infty)$ of relative amplitude at most $\bar{\omega}$. We will use the following bound on the weights:

Lemma 4.3 *Under Condition H1 of Theorem 4.1, for any $v \in L$ and any $\omega : L \rightarrow [0, \infty)$ of relative amplitude at most $\bar{\omega}$, $\omega(v)$ is at most $2\bar{\omega}(\varepsilon + \delta)$.*

Proof. Let $v \in L$, and let S_v be the connected component of S containing v . Consider the cell $V(v)$ of v in the unweighted Voronoi diagram of L . Since L is an $(\varepsilon + \delta)$ -sample of S , $S_v \cap V(v)$ is contained in the ball $B(v)$ of center v and radius $\varepsilon + \delta$. By H1, the radius of this ball is at most $\text{rch}(S)$, hence we have $S_v \cap B(v) \subsetneq S_v$. This implies that the boundary of $V(v)$ intersects S_v . Let p be a point of intersection. There is a point $u \in L$ such that $\|p - u\| = \|p - v\|$, which is at most $\varepsilon + \delta$. Hence, $\|v - u\| \leq 2(\varepsilon + \delta)$. We deduce that $\omega(v) \leq \bar{\omega}\|v - u\| \leq 2\bar{\omega}(\varepsilon + \delta)$, since $\bar{\omega}$ bounds the relative amplitude of ω . \square

Here is now the core argument of the proof of Theorem 4.1:

⁷This condition is mandatory for the existence of a suitable ω on manifolds of dimension three or higher [13, 15].

Lemma 4.4 *Under Condition H1 of Theorem 4.1, for any $p \in S$, for any $\omega : L \rightarrow [0, \infty)$ of relative amplitude at most $\bar{\omega}$, for any non-negative integer $k \leq d$, the Euclidean distance between p and its $(k+1)$ th nearest landmark in the weighted distance is at most $r_k = (1 + 2\bar{\omega} + 2k(1 + 3\bar{\omega}))(\varepsilon + \delta)$.*

Proof. The proof is by induction on k . Assume first that $k = 0$. Let $v_1 \in L$ be the nearest neighbor of p in the weighted distance, and $u \in L$ the nearest neighbor of p in the Euclidean metric. Observe that u may or may not be equal to v_1 . Since L is an $(\varepsilon + \delta)$ -sample of S , $\|p - u\|$ is at most $\varepsilon + \delta$. Moreover, we have $\|p - v_1\|^2 - \omega(v_1)^2 \leq \|p - u\|^2 - \omega(u)^2$, which gives $\|p - v_1\|^2 \leq \|p - u\|^2 + \omega(v_1)^2 - \omega(u)^2 \leq (\varepsilon + \delta)^2 + \omega(v_1)^2 - \omega(u)^2$. Note that $-\omega(u)^2$ is non-positive, and that $\omega(v_1)^2$ is at most $4\bar{\omega}^2(\varepsilon + \delta)^2$, by Lemma 4.3. Therefore, $\|p - v_1\|^2 \leq (1 + 4\bar{\omega}^2)(\varepsilon + \delta)^2 \leq (1 + 2\bar{\omega})^2(\varepsilon + \delta)^2$, which proves the lemma in the case $k = 0$.

Assume now that $k \geq 1$, and that the result holds up to $k - 1$. Let v_1, \dots, v_k denote the k nearest landmarks of p in the weighted distance. By induction, we have $\{v_1, \dots, v_k\} \subset B(p, r_{k-1})$. Moreover, by the case $k = 0$ above, for any $i \leq k$ we have $S \cap V_\omega(v_i) \subseteq B(v_i, (1 + 2\bar{\omega})(\varepsilon + \delta))$, which is included in $B(p, r_{k-1} + (1 + 2\bar{\omega})(\varepsilon + \delta))$. Let S_p be the connected component of S that contains p . It follows from H1 that $r_{k-1} + (1 + 2\bar{\omega})(\varepsilon + \delta) \leq \text{rch}(S)$, hence we have $S \cap B(p, r_{k-1} + (1 + 2\bar{\omega})(\varepsilon + \delta)) \subsetneq S_p$, which implies two things: first, v_1, \dots, v_k belong to S_p ; second, their weighted Voronoi cells do not cover S_p entirely. As a consequence, S_p must intersect the boundary of $\bigcup_{i=1}^k (S \cap V_\omega(v_i))$. Let q be a point of intersection: q lies on the bisector hyperplane between some v_i and some point $v \in L \setminus \{v_1, \dots, v_k\}$, i.e. $q \in V_\omega(v_i) \cap V_\omega(v)$. By the case $k = 0$ above, $\|q - v_i\|$ and $\|q - v\|$ are at most $(1 + 2\bar{\omega})(\varepsilon + \delta)$. Moreover, we have $\|p - v_i\| \leq r_{k-1}$, by induction. Therefore, $\|p - v\| \leq \|p - v_i\| + \|v_i - q\| + \|q - v\| \leq r_{k-1} + 2(1 + 2\bar{\omega})(\varepsilon + \delta)$. Since $v \in L \setminus \{v_1, \dots, v_k\}$, we have $\|p - v_{k+1}\|^2 - \omega(v_{k+1})^2 \leq \|p - v\|^2 - \omega(v)^2$, where v_{k+1} is the $(k+1)$ th nearest landmark of p in the weighted distance. Hence, $\|p - v_{k+1}\|^2 \leq \|p - v\|^2 + \omega(v_{k+1})^2 \leq (\|p - v\| + \omega(v_{k+1}))^2$, which is at most $(r_{k-1} + 2(1 + 2\bar{\omega})(\varepsilon + \delta) + 2\bar{\omega}(\varepsilon + \delta))^2 = r_k^2$, by Lemma 4.3. \square

Using Lemma 4.4, we can bound the Euclidean distances between the k -simplices of $\mathcal{C}_\omega^{W^\zeta}(L)$ and their ω -witnesses. Our bounds depend on quantities $c_1(k)$ and $c_2(k)$, defined respectively in Eqs. (1) and (2):

Lemma 4.5 *Under Conditions H1–H2 of Theorem 4.1, for any $\omega : L \rightarrow [0, \infty)$ of relative amplitude at most $\bar{\omega}$, for any k -simplex σ of $\mathcal{C}_\omega^{W^\zeta}(L)$ ($k \leq n$), and for any vertex v of σ , we have:*

- (i) *for any ω -witness $c \in W^\zeta$ of σ , $\|c - v\| \leq c_1(k) \varepsilon$,*
- (ii) *for any subsimplex $\sigma' \subseteq \sigma$ and any ω -witness $c' \in W^\zeta$ of σ' , $\|c' - v\| \leq c_2(k) \varepsilon$.*
- (iii) *there is a point p on the dual weighted Voronoi face of σ , such that, for any vertex v of σ , $\|p - v\| \leq c_2(k) \varepsilon$.*

Proof. Let σ be a k -simplex of $\mathcal{C}_\omega^{W^\zeta}(L)$. By definition, there is a point $c \in W^\zeta$ that ω -witnesses σ . This means that the vertices of σ are the $(k+1)$ nearest landmarks of c in the weighted distance. Unfortunately, we cannot apply Lemma 4.4 directly to bound the Euclidean distance between c and its $(k+1)$ nearest landmarks in the weighted distance, because c may not belong to S . However, since $c \in W^\zeta$, there is some point $w \in W \subseteq S$ such that $c \in B(w, \zeta)$. By Lemma 4.4, the ball $B(w, r_k)$ contains at least $(k+1)$ landmarks. Thus, at least one landmark u in $B(w, r_k)$ does not belong to the k nearest landmarks of c in the weighted distance. This means that, for any $l \leq k+1$, $\|c - v_l\|^2 - \omega(v_l)^2 \leq \|c - u\|^2 - \omega(u)^2$, where v_l denotes the l th nearest landmark of c in the weighted distance. It follows that $\|c - v_l\|^2 \leq \|c - u\|^2 + \omega(v_l)^2 - \omega(u)^2 \leq (\|c - u\| + \omega(v_l))^2$, which by Lemma 4.3 is at most $(\|c - u\| + 2\bar{\omega}(\varepsilon + \delta))^2$. Since $\|c - u\| \leq \|c - w\| + \|w - u\| \leq \zeta + r_k$,

we get $\|c - v_l\| \leq \zeta + r_k + 2\bar{\omega}(\varepsilon + \delta)$, which by H1–H2 is bounded by $c_1(k) \varepsilon$. Since this is true for any $l \leq k + 1$, the Euclidean distance between c and any vertex of σ is at most $c_1(k) \varepsilon$. This proves (i).

Let σ' be any subsimplex of σ . Since σ belongs to $\mathcal{C}_\omega^{W^\zeta}(L)$, σ' is ω -witnessed by some point $c' \in W^\zeta$. Let $w' \in W \subseteq S$ be such that $c' \in B(w', \zeta)$. According to Lemma 4.4, the Euclidean distance between w' and its nearest landmark u' in the weighted distance is at most $r_0 = (1 + 2\bar{\omega})(\varepsilon + \delta)$. Hence, $\|c' - u'\| \leq \zeta + r_0$. Let v'_1 be the nearest landmark of c' in the weighted distance. We have $\|c' - v'_1\|^2 \leq \|c' - u'\|^2 + \omega(v'_1)^2 - \omega(u')^2 \leq (\|c' - u'\| + \omega(v'_1))^2$, which is at most $(\zeta + r_0 + 2\bar{\omega}(\varepsilon + \delta))^2$, by Lemma 4.3. Since c' ω -witnesses σ' , v'_1 is a vertex of σ' and hence also a vertex of σ . Thus, for any vertex v of σ , we have $\|c' - v\| \leq \|c' - v'_1\| + \|v'_1 - v\|$, which by (i) is at most $\zeta + r_0 + 2\bar{\omega}(\varepsilon + \delta) + 2c_1(k) \varepsilon$. This quantity is bounded by $c_2(k) \varepsilon$, since under H1–H2 δ and ζ are at most ε . This proves (ii).

By Theorem 2.3, we know that σ is a weighted Delaunay simplex, and that its dual weighted Voronoi face intersects the convex hull of the ω -witnesses of the subsimplices of σ . Let p be a point of intersection. According to (ii), for any vertex v of σ , the ω -witnesses (in W^ζ) of σ and of its subsimplices belong to $B(v, c_2(k) \varepsilon)$, which therefore also contains their convex hull, and hence also p . Thus, we have $\|p - v\| \leq c_2(k) \varepsilon$, which proves (iii). \square

Now we can show that $\mathcal{C}_\omega^{W^\zeta}(L) \subseteq K_\omega^\theta(L)$, which concludes the proof of Theorem 4.1:

Lemma 4.6 *Under Conditions H1–H2 of Theorem 4.1, for any distribution of weights $\omega : L \rightarrow [0, \infty)$ of relative amplitude at most $\bar{\omega}$, for any k -simplex σ of $\mathcal{C}_\omega^{W^\zeta}(L)$ ($k \leq d$), and for any vertex v of σ , $V_\omega(\sigma) \cap K^\theta(v) \neq \emptyset$. As a consequence, $\mathcal{C}_\omega^{W^\zeta}(L) \subseteq K_\omega^\theta(L)$.*

Proof. Let σ be a k -simplex of $\mathcal{C}_\omega^{W^\zeta}(L)$, and let $v \in L$ be any vertex of σ . If $k = 0$, then $\sigma = [v]$. Since the relative amplitude of ω is at most $\bar{\omega} < \frac{1}{2}$, v belongs to its weighted Voronoi cell $V_\omega(v)$. And since v belongs to $K^\theta(v)$, we have $V_\omega(v) \cap K^\theta(v) \neq \emptyset$, which proves the lemma in the case $k = 0$.

Assume now that $1 \leq k \leq d$. By Lemma 4.5 (ii), for any simplex $\sigma' \subseteq \sigma$, for any ω -witness c' of σ' , we have $\|c' - v\| \leq c_2(k) \varepsilon \leq c_2(d) \varepsilon$. Since $c' \in W^\zeta$, there is a point $w' \in W \subseteq S$ such that $\|w' - c'\| \leq \zeta$, which implies that $\|w' - v\| \leq \varepsilon(c_2(d) + \zeta/\varepsilon)$. This quantity is at most $\text{rch}(S)$, by H1–H2, hence the Euclidean distance between w' and $T(v)$ is at most $\frac{\varepsilon^2}{2 \text{rch}(S)} (c_2(d) + \zeta/\varepsilon)^2$, by Lemma 2.6. It follows that the Euclidean distance from c' to $T(v)$ is at most $\zeta + \frac{\varepsilon^2}{2 \text{rch}(S)} (c_2(d) + \zeta/\varepsilon)^2$. This holds for any ω -witness $c' \in W^\zeta$ of any simplex $\sigma' \subseteq \sigma$. Since σ is a simplex of $\mathcal{C}_\omega^{W^\zeta}(L)$, Theorem 2.3 tells us that σ belongs to $\mathcal{D}_\omega(L)$, and that the dual weighted Voronoi face of σ intersects the convex hull of the ω -witnesses of the subsimplices of σ . Let p be a point of intersection. Since the ω -witnesses (in $\mathcal{C}_\omega^{W^\zeta}(L)$) of the subsimplices of σ do not lie farther than $\zeta + \frac{\varepsilon^2}{2 \text{rch}(S)} (c_2(d) + \zeta/\varepsilon)^2$ from $T(v)$, neither does p , which belongs to their convex hull.

Let us now give a lower bound on $\|p - v\|$, which will allow us to conclude afterwards. Since the dimension k of σ is at least one, σ has at least two vertices. Hence, there exists at least one point $u \in L \setminus \{v\}$ such that $\|p - v\|^2 - \omega(v)^2 = \|p - u\|^2 - \omega(u)^2$, which gives $\|p - v\|^2 = \|p - u\|^2 + \omega(v)^2 - \omega(u)^2$. Since $\bar{\omega}$ bounds the relative amplitude of ω , we have $\omega(u)^2 \leq \bar{\omega}^2 \|u - v\|^2$. Moreover, $\omega(v)^2$ is non-negative. Hence, $\|p - v\|^2 \geq \|p - u\|^2 - \bar{\omega}^2 \|u - v\|^2$. By the triangle inequality, we obtain $\|p - v\|^2 \geq (\|p - v\| - \|u - v\|)^2 - \bar{\omega}^2 \|u - v\|^2$, which gives $\|p - v\| \geq \frac{1}{2}(1 - \bar{\omega}^2) \|u - v\|$. This implies that $\|p - v\| \geq \frac{1}{2}(1 - \bar{\omega}^2)\varepsilon$, since L is ε -sparse.

To conclude, p is at most $\zeta + \frac{\varepsilon^2}{2 \text{rch}(S)} (c_2(d) + \zeta/\varepsilon)^2$ away from $T(v)$, and at least $\frac{1}{2}(1 - \bar{\omega}^2)\varepsilon$ away from v , in the Euclidean metric. Therefore, $\sin \angle(vp, T(v)) \leq \frac{2\zeta}{(1 - \bar{\omega}^2)\varepsilon} + \frac{\varepsilon}{(1 - \bar{\omega}^2)\text{rch}(S)} (c_2(d) + \zeta/\varepsilon)^2$,

which is at most $\sin \theta$, by H1–H2. As a consequence, p belongs to $K^\theta(v)$. Now, recall that p is a point of $V_\omega(\sigma)$. Hence, $V_\omega(\sigma) \cap K^\theta(v) \neq \emptyset$, which means that σ is a simplex of $K_\omega^\theta(L)$. Since this is true for any k -simplex σ of $\mathcal{C}_\omega^{W^\zeta}(L)$ ($k \leq d$), and since, by Theorem 2.3, $\mathcal{C}_\omega^{W^\zeta}(L)$ is included in $\mathcal{D}_\omega(L)$, which has no simplex of dimension greater than d because the weighted point set L lies in general position, $\mathcal{C}_\omega^{W^\zeta}(L)$ is included in $K_\omega^\theta(L)$. \square

4.2 Proof of Theorem 4.2

Let σ be a simplex of $K_\omega^\theta(L)$. By definition, the dual weighted Voronoi face of σ intersects the θ -cocone of at least one vertex v of σ . Let c be a point of intersection. Since L is an $(\varepsilon + \delta)$ -sample of S , with $\varepsilon + \delta < \frac{1}{9}(1 - \sin \theta)^2 \text{rch}(S)$, Lemma 2.8 states that $\|c - v\| \leq \frac{3(\varepsilon + \delta)}{1 - \sin \theta}$, which is less than $\frac{1}{4} \text{rch}(S)$ since by assumption we have $\varepsilon + \delta < \frac{1 - \sin \theta}{12} \text{rch}(S)$. Therefore, by Lemma 2.7, we have $\|q' - c\| \leq \frac{3(\varepsilon + \delta)}{1 - \sin \theta} \left(\sin \theta + \frac{6(\varepsilon + \delta)}{(1 - \sin \theta) \text{rch}(S)} \right)$, where q' is a point of S closest to c . Since W is a δ -sample of S , there exists some point $w \in W$ such that $\|q' - w\| \leq \delta$. Hence, $\|c - w\| \leq \delta + \frac{3(\varepsilon + \delta)}{1 - \sin \theta} \left(\sin \theta + \frac{6(\varepsilon + \delta)}{(1 - \sin \theta) \text{rch}(S)} \right)$, which by H1 is at most $6 \sin \theta \varepsilon \leq \zeta$. It follows that $c \in W^\zeta$, and hence that σ belongs to $\mathcal{D}_\omega^{W^\zeta}(L)$, which by Corollary 2.2 is included in $\mathcal{C}_\omega^{W^\zeta}(L)$. This concludes the proof of Theorem 4.2. \square

5 Application to multiscale manifold reconstruction

In this section, we use our structural results in the context of manifold reconstruction. From now on, we fix $\bar{\omega} = \frac{1}{4}$, $\theta = \frac{\pi}{32}$, and $\bar{\varrho} = \frac{2^{-7-(1+4c_1(1))^d}}{c_1(1)c_3(d)}$, where $c_1(k), c_2(k)$ are defined as in Eqs. (1) and (2), and where $c_3(k) = c_2(k)(1 + 2c_1(1)) + 4\bar{\omega}$. We give an overview of the approach in Section 5.1, then we present the algorithm in Section 5.2, and we prove its correctness in Section 5.3. Our theoretical guarantees are discussed in Section 6. In Section 5.4, we give some details on how the various components of the algorithm can be implemented. These details are then used in Section 5.5 to bound the space and time complexity of the algorithm.

5.1 Overview of the approach

Let W be a finite input point set drawn from some unknown manifold S , such that W is a δ -sample of S , for some unknown $\delta > 0$. Imagine we were able to construct an ε -sparse ε -sample L of W , for some $\varepsilon \leq c_{\bar{\varrho}} \text{rch}(S)$ satisfying Condition H1 of Theorem 4.1. Then, Theorems 4.1 and 4.2 would guarantee that $\mathcal{D}_\omega^S(L) \subseteq \mathcal{C}_\omega^{W^{\zeta_1}}(L) \subseteq K_\omega^{\pi/32}(L) \subseteq \mathcal{C}_\omega^{W^{\zeta_2}}(L)$ for any distribution of weights ω of relative amplitude at most $\bar{\omega}$, where $\zeta_1 = \frac{2(1-\bar{\omega}^2)\sin \theta}{5} \varepsilon = \frac{3}{8} \sin \frac{\pi}{32} \varepsilon$ and $\zeta_2 = 6 \sin \frac{\pi}{32} \varepsilon$. We could then apply the pumping strategy of [13] to $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$, which would remove all $\bar{\varrho}$ -slivers from $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$ (and hence from $K_\omega^{\pi/32}(L)$), provided that $\bar{\varrho}$ is sufficiently small. As a result, $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ would coincide with $\mathcal{D}_\omega^S(L)$ and be homeomorphic to S , by Theorem 2.5.

Unfortunately, since δ and $\text{rch}(S)$ are unknown, finding an $\varepsilon \leq c_{\bar{\varrho}} \text{rch}(S)$ that provably satisfies H1 can be difficult, if not impossible. This is why we combine the above approach with the multiscale reconstruction scheme of [27], which generates a monotonic sequence of samples $L \subseteq W$. Since L keeps growing during the process, ε keeps decreasing, eventually satisfying H1 if δ is small

enough. At that stage, $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ becomes homeomorphic to S , and thus a plateau appears in the diagram of the Betti numbers of $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$, showing the values of S – see [27] for some examples.

Note that the only assumption made on the input point set W is that it is a δ -sample of some manifold, for some small enough δ . In particular, W is not assumed to be a sparse δ -sample. As a result, W may well-sample several manifolds S_1, \dots, S_l , as it is the case for instance in the example of Figure 1. In such a situation, the values $\delta_1, \dots, \delta_l$ for which W is a δ_i -sample of S_i differ, and they generate distinct plateaus in the diagram of the Betti numbers of $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$.

Another noticeable aspect of our approach, compared to some previous algorithms [15], is that it does not require any preprocessing of the input data, like estimating the dimension or tangent spaces of the underlying manifold.

5.2 The algorithm

The input is a finite point set $W \subset \mathbb{R}^d$. Initially, $L = \emptyset$ and $\varepsilon = \zeta_1 = \zeta_2 = +\infty$.

At each iteration, the point $p \in W$ lying furthest away from L in the Euclidean metric⁸ is inserted in L and pumped. Specifically, once p has been appended to L , its weight $\omega(p)$ is set to zero, and ε is set to $\max_{w \in W} \min_{v \in L} \|w - v\|$. $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ and $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$ are updated accordingly, ζ_1 and ζ_2 being defined as in the overview section. Then, p is pumped: the pumping procedure $Pump(p)$ determines the weight ω_p of p that maximizes the minimum sliver measure among the simplices of the star of p in $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$. After the pumping, $\omega(p)$ is set to ω_p , and $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$, $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$ are updated. The pseudo-code is given opposite.

Input: $W \subset \mathbb{R}^d$ finite.

Init: Let $L := \emptyset$; $\varepsilon := +\infty$; $\zeta_1 := +\infty$; $\zeta_2 := +\infty$;

While $L \subsetneq W$ do

Let $p := \operatorname{argmax}_{w \in W} \min_{v \in L} \|w - v\|$;

// p is chosen arbitrarily in W if $L = \emptyset$

$L := L \cup \{p\}$; $\omega(p) := 0$;

$\varepsilon := \max_{w \in W} \min_{v \in L} \|w - v\|$;

$\zeta_1 := \frac{3}{8} \sin \frac{\pi}{32} \varepsilon$; $\zeta_2 := 6 \sin \frac{\pi}{32} \varepsilon$;

Update $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ and $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$;

$\omega(p) := Pump(p)$;

Update $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ and $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$;

End-while

Output: the sequence of complexes $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ obtained after every iteration of the *While* loop.

The algorithm terminates when $L = W$. The output is the one-parameter family of complexes $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ built throughout the process, or simply the diagram of their Betti numbers, computed on the fly using the persistence algorithm of [23]. With this diagram, the user can determine the scale at which to process the data: it is then easy to generate the corresponding subset of weighted landmarks (the points of W have been sorted according to their order of insertion in L , and their weights have been stored) and to rebuild its weighted witness complex relative to W^{ζ_1} .

The pumping procedure is the same as in [15], with $K_\omega^{\pi/32}(L)$ replaced by $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$. Given a point p just inserted in L , $Pump(p)$ makes the weight of p vary from zero to $\bar{\omega} \min\{\|p - v\|, v \in L \setminus \{p\}\}$, while maintaining the star of p in $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$. The combinatorial structure of the star changes only at a finite set of event times. Between consecutive event times, the minimum sliver measure among the simplices of the star is constant and therefore computed once. In the end, the procedure returns an arbitrary value in the range of weights that maximize the minimum sliver measure in the star of p .

⁸At the first iteration, since L is empty, p is chosen arbitrarily in W .

5.3 Guarantees on the output

Assume that the input point set W is a δ -sample of some unknown manifold S . For any $i > 0$, let $p(i)$ be the point inserted in L at the i th iteration of the algorithm, and let $L(i), \omega(i), \varepsilon(i), \zeta_1(i), \zeta_2(i)$ denote respectively $L, \omega, \varepsilon, \zeta_1, \zeta_2$ at the end of that iteration. We have $L(i) = L(i-1) \cup \{p(i)\}$, therefore $L(i)$ keeps growing with i . As consequence, $\varepsilon(i)$ is a decreasing function of i . Moreover:

Lemma 5.1 *For any iteration $i > 0$, $L(i)$ is an $\varepsilon(i)$ -sparse $\varepsilon(i)$ -sample of W .*

Proof. The fact that $L(i)$ is an $\varepsilon(i)$ -sample of W follows directly from the definition of $\varepsilon(i)$. Now, at each iteration $j \leq i$ of the algorithm, the witness $p(j)$ lying furthest away from $L(j-1)$ is appended to $L(j-1)$. Right before this insertion, $L(j-1)$ is an $\varepsilon(j-1)$ -sample of W . This means that the Euclidean distance from $p(j)$ to $L(j-1)$ is $\varepsilon(j-1)$. Since $j \leq i$, we have $\varepsilon(j-1) \geq \varepsilon(j) \geq \varepsilon(i)$, hence each point inserted in L before or at iteration i is at least $\varepsilon(i)$ away from L at the time of its insertion. This implies that $L(i)$ is $\varepsilon(i)$ -sparse. \square

Combining Lemma 5.1 with Theorems 2.5, 4.1, and 4.2, we can prove our main result:

Theorem 5.2 $\mathcal{C}_{\omega(i)}^{W^{\zeta_1(i)}}(L(i))$ coincides with $\mathcal{D}_{\omega(i)}^S(L(i))$ and is homeomorphic to S whenever $\varepsilon(i) \leq c_{\bar{\varrho}} \text{rch}(S)$ satisfies H1, which eventually happens in the course of the algorithm if δ is small enough.

The proof of the theorem consists mainly in showing that all slivers are eventually removed from $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$, even though some of the weights are assigned in early stages of the course of the algorithm. Our line of argument relies on an intermediate result, stated as Claim 5.2.1 below, whose proof is deferred to the end of Section 5.3:

Proof. Let $i > 0$ be an iteration such that $\varepsilon(i) \leq c_{\bar{\varrho}} \text{rch}(S)$ satisfies H1. Then, $\zeta_1(i) = \frac{3}{8} \sin \frac{\pi}{32} \varepsilon(i) = \frac{2(1-\bar{\omega}^2) \sin \theta}{5} \varepsilon(i)$ satisfies H2, and $\zeta_2(i) = 6 \sin \frac{\pi}{32} \varepsilon(i)$ satisfies the hypothesis of Theorem 4.2. Since W is a δ -sample of S by assumption, and since $L(i)$ is an $\varepsilon(i)$ -sparse $\varepsilon(i)$ -sample of W by Lemma 5.1, Theorems 4.1 and 4.2 imply that $\mathcal{D}_{\omega(i)}^S(L(i)) \subseteq \mathcal{C}_{\omega(i)}^{W^{\zeta_1(i)}}(L(i)) \subseteq \mathcal{K}_{\omega(i)}^{\pi/32}(L(i)) \subseteq \mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$. Let us show that $\mathcal{K}_{\omega(i)}^{\pi/32}(L(i))$ contains no $\bar{\varrho}$ -sliver, which by Theorem 2.5 gives the result. In fact, we will prove that $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ contains no $\bar{\varrho}$ -sliver, which is sufficient since $\mathcal{K}_{\omega(i)}^{\pi/32}(L(i)) \subseteq \mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$.

Claim 5.2.1 *For any iteration $j > 0$ such that $\delta \leq \varepsilon(j) \leq \frac{2\text{rch}(S)}{7d-1} - \delta$, the star of $p(j)$ in $\mathcal{C}_{\omega(j)}^{W^{\zeta_2(j)}}(L(j))$ contains no $\bar{\varrho}$ -sliver.*

Since $\varepsilon(i)$ satisfies H1, we have $\delta \leq \varepsilon(i) \leq \frac{2\text{rch}(S)}{7d-1} - \delta$. Therefore, Claim 5.2.1 guarantees that the star of $p(i)$ in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ contains no $\bar{\varrho}$ -sliver. However, this does not mean that $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ itself contains no $\bar{\varrho}$ -sliver, because some of the points of $L(i)$ were pumped at early stages of the course of the algorithm, when the assumption of Claim 5.2.1 was not yet satisfied.

Let $i_0 \leq i$ be the first iteration such that $\varepsilon(i_0) \leq \frac{2\text{rch}(S)}{7d-1} - \delta$. For any iteration j between i_0 and i , we have $\varepsilon(j) \leq \varepsilon(i_0) \leq \frac{2\text{rch}(S)}{7d-1} - \delta$ and $\varepsilon(j) \geq \varepsilon(i) \geq \delta$, hence Claim 5.2.1 guarantees that the pumping procedure removes all $\bar{\varrho}$ -slivers from the star of $p(j)$ in $\mathcal{C}_{\omega(j)}^{W^{\zeta_2(j)}}(L(j))$ at iteration j . For any k between j and i , the update of $\mathcal{C}_{\omega}^{W^{\zeta_2}}(L)$ after the pumping of $p(k)$ may modify the star of $p(j)$. However, since the pumping of $p(k)$ only increases its weight, the new simplices in the star

of $p(j)$ belong also to the star of $p(k)$, which contains no $\bar{\rho}$ -sliver, by Claim 5.2.1. Therefore, the star of $p(j)$ in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ still contains no $\bar{\rho}$ -sliver. Consider now an iteration $j \leq i_0 - 1$. Since $\varepsilon(j)$ is greater than $\frac{2\text{rch}(S)}{7d-1} - \delta$, we cannot ensure that the star of $p(j)$ in $\mathcal{C}_{\omega(j)}^{W^{\zeta_2(j)}}(L(j))$ contains no $\bar{\rho}$ -sliver. However, we claim that the points of $L(i)$ that are neighbors of $p(j)$ in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ were inserted in L on or after iteration i_0 . Indeed, since $\varepsilon(i)$ satisfies H1, we have $\delta \leq \varepsilon(i) \leq \frac{\text{rch}(S)}{3} - \delta$. Moreover, $\zeta_2(i) = 6 \sin \frac{\pi}{32} \varepsilon(i) \leq \varepsilon(i)$. Hence, Lemma A.1 ensures that every neighbor of $p(j)$ in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ is at Euclidean distance at most $2c_1(1) \varepsilon(i)$ from $p(j)$. Now, recall that $L(i_0 - 1)$ is $\varepsilon(i_0 - 1)$ -sparse, where $\varepsilon(i_0 - 1) > \frac{2\text{rch}(S)}{7d-1} - \delta$, which by H1 is greater than $2c_1(1) \varepsilon(i)$. It follows that no point of $L(i_0 - 1)$ is a neighbor of $p(j)$ in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$. This means that, for any neighbor q of $p(j)$ in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$, q was inserted in L on or after iteration i_0 . Hence, the star of q in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ contains no $\bar{\rho}$ -sliver. Since this is true for any neighbor q of $p(j)$ in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$, the star of $p(j)$ does not contain any $\bar{\rho}$ -sliver either. This concludes the proof of Theorem 5.2. \square

We end Section 5.3 by giving the proof of Claim 5.2.1:

Proof of Claim 5.2.1. The argument is the same as in Section 8 of [15], and it relies on Lemmas 4 through 10 of [15], which have been adapted to our context in Appendix A. Note that we assumed $\delta \leq \varepsilon(j) \leq \frac{2\text{rch}(S)}{7d-1} - \delta$, and that we have $\zeta_2(j) = 6 \sin \frac{\pi}{32} \varepsilon(j) \leq \varepsilon(j)$, which means that $\delta, \zeta_2(j), \varepsilon(j)$ satisfy the hypotheses of Lemmas A.1 through A.6. We will show that there exists a value of $\omega(p(j))$ such that the star of $p(j)$ in $\mathcal{C}_{\omega}^{W^{\zeta_2(j)}}(L(j))$ contains no $\bar{\rho}$ -sliver. The result will follow, since the pumping procedure selects a value of $\omega(p(j))$ that maximizes the minimum sliver measure in the star of $p(j)$.

Let $l(j) = \min\{\|p(j) - v\|, v \in L(j) \setminus \{p(j)\}\}$. Let $\Sigma(j)$ be the set of all the simplices that appear in the star of $p(j)$ in $\mathcal{C}_{\omega}^{W^{\zeta_2(j)}}(L(j))$ during the course of the pumping procedure. Since $\bar{\rho} < 1$, the 0- and 1-simplices of $\Sigma(j)$ are not $\bar{\rho}$ -slivers. Now, for any simplex $\sigma \in \Sigma(j)$ of dimension at least two that is a $\bar{\rho}$ -sliver, if the facets of σ are not $\bar{\rho}$ -slivers, then Lemma A.6 ensures that σ appears in $\mathcal{C}_{\omega}^{W^{\zeta_2(j)}}(L(j))$ only when the squared weight of $p(j)$ belongs to an interval of length less than $8\bar{\rho}c_1(1)c_3(d) \varepsilon(j)^2$. Moreover, by Lemma A.3, the total number of neighbors that $p(j)$ has in $\mathcal{C}_{\omega}^{W^{\zeta_2(j)}}(L(j))$ as its weight ranges from 0 to $\bar{\omega}l(j)$ is bounded by $(1 + 4c_1(1))^d$. Therefore, the overall number of simplices that can appear in the star of $p(j)$ as its weight evolves is at most $2^{(1+4c_1(1))^d}$. It follows that the set of values of the squared weight of $p(j)$ for which the star of $p(j)$ contains $\bar{\rho}$ -slivers whose facets are not $\bar{\rho}$ -slivers, is contained in a union of intervals of total length less than $2^{3+(1+4c_1(1))^d} \bar{\rho}c_1(1)c_3(d) \varepsilon(j)^2 = 2^{-4} \varepsilon(j)^2 = \bar{\omega}^2 \varepsilon(j)^2$, which is at most $\bar{\omega}^2 l(j)^2$ since $L(j)$ is $\varepsilon(j)$ -sparse. Therefore, there are values of $\omega(p(j))$ within $[0, \bar{\omega}l(j)]$ for which the star of $p(j)$ in $\mathcal{C}_{\omega}^{W^{\zeta_2(j)}}(L(j))$ contains no $\bar{\rho}$ -sliver whose facets are not $\bar{\rho}$ -slivers. In fact, the star of p contains no $\bar{\rho}$ -sliver at all, since otherwise it would contain a $\bar{\rho}$ -sliver whose facets are not $\bar{\rho}$ -slivers (take for instance a $\bar{\rho}$ -sliver of smallest dimension in the star of p). This ends the proof of Claim 5.2.1. \square

5.4 Details of the implementation

Before we can analyze the complexity of the algorithm, we need to give some details on how its various components can be implemented.

5.4.1 Update of the witness complex

Although the algorithm is conceptually simple, its implementation requires to be able to update $\mathcal{C}_\omega^{W^\zeta}(L)$, where $\zeta \in \{\zeta_1, \zeta_2\}$. This task is significantly more difficult than updating $\mathcal{C}_\omega^W(L)$, mainly because W^ζ is not finite, which makes it impossible to perform the ω -witness test of Definition 2.1 on every single point of W^ζ individually. However, since we are working in Euclidean space \mathbb{R}^d , for any $w \in W$ it is actually possible to perform the ω -witness test on the whole ball $B(w, \zeta)$ at once. Each time a point is inserted in L or its weight is increased, $\mathcal{C}_\omega^{W^\zeta}(L)$ is updated by iterating over the points $w \in W$ and performing the ω -witness test on $B(w, \zeta)$. This test boils down to intersecting $B(w, \zeta)$ with the weighted Voronoi diagrams of L of orders 1 through $d + 1$. In the space of spheres \mathbb{R}^{d+1} , this is equivalent to intersecting a vertical cylinder of base $B(w, \zeta)$ with the cells of an arrangement of $|L|$ hyperplanes. This operation is very costly, hence we only construct the arrangement of the hyperplanes of the $\kappa(d)$ nearest landmarks of w in the Euclidean metric, where $\kappa(d) = (2 + 2c_1(d))^d$. This means that we do not actually maintain $\mathcal{C}_\omega^{W^\zeta}(L)$, but another complex \mathcal{C} , which might not contain $\mathcal{C}_\omega^{W^\zeta}(L)$ nor be contained in it. Nevertheless,

Lemma 5.3 $\mathcal{C}(i)$ coincides with $\mathcal{C}_{\omega(i)}^{W^{\zeta(i)}}(L(i))$ whenever $\varepsilon(i)$ satisfies H1.

Proof. Let w be a point of W , and let $\Lambda(w) \subseteq L(i)$ denote the set of its $\kappa(d)$ nearest landmarks in the Euclidean metric. For any point $c \in B(w, \zeta(i))$ and any positive integer $k \leq d + 1$, Lemma 4.5 (i) ensures that the k nearest landmarks of c in the weighted distance, p_1, \dots, p_k , belong to $B(c, c_1(d)\varepsilon(i))$. Since $\|w - c\| \leq \zeta(i)$, p_1, \dots, p_k belong to $B(w, r(i))$, where $r(i) = \varepsilon(i)(c_1(d) + \zeta(i)/\varepsilon(i))$. Now, $L(i)$ is $\varepsilon(i)$ -sparse, hence the points of $L(i) \cap B(w, r(i))$ are centers of pairwise-disjoint balls of radius $\varepsilon(i)/2$. These balls are included in $B(w, r(i) + \varepsilon(i)/2)$, thus their number is at most $(1 + 2r(i)/\varepsilon(i))^d = (1 + 2c_1(d) + 2\zeta(i)/\varepsilon(i))^d$, which is bounded by $\kappa(d)$ since $\zeta(i) \leq \zeta_2(i) < \varepsilon(i)/2$. Therefore, p_1, \dots, p_k belong to $\Lambda(w)$. As a result, in the weighted distance, the k nearest neighbors of c in $\Lambda(w)$ are the same as its k nearest neighbors in $L(i)$. Since this is true for any $w \in W$ and any $c \in B(w, \zeta(i))$, and since all the balls of radius $\zeta(i)$ centered at the points of W are tested at each update of the complex, $\mathcal{C}(i)$ coincides with $\mathcal{C}_{\omega(i)}^{W^{\zeta(i)}}(L(i))$. \square

It follows from this lemma that our guarantees on the output of the algorithm still hold if $\mathcal{C}_\omega^{W^\zeta}(L)$ is replaced by \mathcal{C} . Note also that, since any arrangement of $\kappa(d)$ hyperplanes in \mathbb{R}^{d+1} has $O(\kappa(d)^{d+1}) = d^{O(d^2)}$ cells [22], each point of W generates at most $d^{O(d^2)}$ simplices in \mathcal{C} . It follows that the size of our complex is bounded by $|W|d^{O(d^2)}$. The same bound clearly holds for the time spent updating \mathcal{C} after a point insertion or a weight increase, provided that the $\kappa(d)$ nearest landmarks of a witness can be computed in $d^{O(d^2)}$ time. In practice, we maintain the lists of $\kappa(d)$ nearest landmarks of the witnesses in parallel to \mathcal{C} . Each time a new point is inserted in L , the list of each witness is updated in $O(\kappa(d))$ time as we iterate over all the witnesses to update \mathcal{C} . Using these lists, we can retrieve the $\kappa(d)$ nearest landmarks of any witness in time $O(\kappa(d)) \leq d^{O(d^2)}$.

5.4.2 Pumping procedure

As mentioned in Section 5.2, only a finite number of events occur while a point $p \in L$ is being pumped. The sequence of events can be precomputed before the beginning of the pumping process, by iterating over the points of W that have p among their $\kappa(d)$ nearest landmarks. For each such point w , we detect the sequence of simplices incident to p that start or stop being ω -witnessed by points of $B(w, \zeta_2)$ as the weight of p increases. In the space of spheres \mathbb{R}^{d+1} , this is equivalent

to looking at how the cells of the arrangement of $\kappa(d)$ hyperplanes evolve as the hyperplane of p translates vertically. The number of events that are generated by a point of W is of the order of the size of the arrangement of $\kappa(d)$ hyperplanes in \mathbb{R}^{d+1} , hence the total number of events is at most $|W|d^{O(d^2)}$, and the time spent computing them is also bounded by $|W|d^{O(d^2)}$.

To reduce the total number of events, we discard the ones that involve simplices whose vertices do not belong to the $\kappa(d)$ nearest neighbors of p among L in the Euclidean metric. As a result, only the simplices whose vertices belong to the $\kappa(d)$ nearest neighbors of p are considered in the sequel. By the same argument as above, the number of such simplices is at most $d^{O(d^2)}$. Note however that a k -simplex σ may appear in or disappear from the star of p several times: indeed, its cell in the weighted Voronoi diagram of L of order $k+1$ evolves as the weight of p increases, and it may intersect W^ζ several times during the process, since W^ζ is not convex. Therefore, for each simplex σ , we report only the first time where σ appears in the star of p , and the last time where it disappears from the star of p . Thus, the number of events reported per simplex is at most two, which implies that the total number of events reported is bounded by $d^{O(d^2)}$.

Once the sequence of events has been computed, the pumping procedure iterates over the events. At each iteration, the weight of p is increased, and $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$ (or rather the complex \mathcal{C} of Section 5.4.1) is updated. The minimum sliver measure in the star of p is computed on the fly, during the update of \mathcal{C} . The number of iterations of the pumping procedure is bounded by the number of events, which is at most $d^{O(d^2)}$. At each iteration, the update of \mathcal{C} takes $|W|d^{O(d^2)}$ time, which also includes the computation of the minimum sliver measure in the star of p . All in all, the time complexity of the pumping procedure is at most $|W|d^{O(d^2)}$.

The downside is that some events have been discarded during the precomputation, which means that the pumping procedure may work with a wrong sequence of events, and therefore may not be able to remove the \bar{q} -slivers from the star of p . Nevertheless,

Lemma 5.4 *Whenever $\varepsilon(i)$ satisfies H1, the neighbors of $p(i)$ in $\mathcal{C}_\omega^{W^{\zeta_2(i)}}(L(i))$ belong to the $\kappa(d)$ nearest neighbors of $p(i)$ among $L(i)$ in the Euclidean metric, for any $\omega : L(i) \rightarrow [0, \infty)$ of relative amplitude at most $\bar{\omega}$.*

Proof. Since $\varepsilon(i)$ satisfies H1, Lemma A.1 guarantees that the neighbors of $p(i)$ in $\mathcal{C}_\omega^{W^{\zeta_2(i)}}(L(i))$ belong to the ball $B(p(i), 2c_1(1)\varepsilon(i))$. Since $L(i)$ is $\varepsilon(i)$ -sparse, the points of $L(i) \cap B(p(i), 2c_1(1)\varepsilon(i))$ are centers of pairwise-disjoint balls of radius $\varepsilon(i)/2$, included in $B(p(i), (2c_1(1) + 1/2)\varepsilon(i))$. The number of such balls is therefore bounded by $(4c_1(1) + 1)^d$, which is at most $(2 + 2c_1(d))^d = \kappa(d)$ since $d \geq 3$. It follows that the neighbors of $p(i)$ in $\mathcal{C}_\omega^{W^{\zeta_2(i)}}(L(i))$ belong to its $\kappa(d)$ nearest neighbors among $L(i)$ in the Euclidean metric. \square

It follows from Lemma 5.4 that, whenever $\varepsilon(i)$ satisfies H1, no simplex is discarded during the precomputation. Note that some events may still be discarded, because we keep only the first and last event of each simplex. Intuitively, this means that the pumping procedure deals with more simplices than it should. However, since by Lemma A.6 the first and last events of a \bar{q} -sliver occur at very close moments, Claim 5.2.1 still holds, with exactly the same proof. Therefore, the theoretical guarantees of Section 5.3 still apply.

5.4.3 Computation of p and ε

At each iteration of the main loop of the algorithm, the computation of $p = \operatorname{argmax}_{w \in W} \min_{v \in L} \|w - v\|$ and $\varepsilon = \max_{w \in W} \min_{v \in L} \|w - v\|$ is done naively by iterating over the points of W and computing

their Euclidean distance to L . This takes $O(|W|)$ time once the sets of $\kappa(d)$ nearest landmarks have been updated.

5.5 Time and space complexities of the algorithm

Using the results of Section 5.4, we can bound the complexity of the algorithm:

Theorem 5.5 *The space complexity of the algorithm is at most $|W|d^{O(d^2)}$, and its time complexity is at most $|W|^2d^{O(d^2)}$.*

Proof. As reported in Section 5.4.1, the size of the complex maintained by the algorithm remains bounded by $|W|d^{O(d^2)}$, therefore the space complexity of the algorithm is at most $|W|d^{O(d^2)}$. At each iteration of the main loop of the algorithm, the computation of p and ε takes $O(|W|)$ time, according to Section 5.4.3, and the pumping of p takes $|W|d^{O(d^2)}$ time, by Section 5.4.2. Moreover, each update of $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ and $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$ takes $d^{O(d^2)}$ time, by Section 5.4.1. Therefore, the time spent in each iteration of the main loop of the algorithm is at most $|W|d^{O(d^2)}$. It follows that the time complexity of the algorithm is bounded by $|W|^2d^{O(d^2)}$, since the number of iterations of the main loop of the algorithm is $|W|$. \square

6 Discussion and perspectives

Running time. In view of Theorem 5.5, the worst-case running time of our reconstruction algorithm is bounded by $c(d)n^2$, where n is the size of the input point cloud and $c(d)$ is a constant depending exponentially on the ambient dimension d . This bound is comparable to the ones obtained in previous work [15], and it makes the corresponding approaches (including ours) rather theoretical. The reduction in the exponent from dependence on the ambient dimension to dependence on the intrinsic dimension of the data is of fundamental theoretical importance and may lead to practical algorithms down the road. Recent work by Chazal and Oudot [12] illustrates the potential of the witness complex as a data structure in this context, by showing that a variant of our approach can solve the related problem of homology inference from point cloud data within a time that depends only on the size and intrinsic dimensionality of the input data. This achievement was made possible by the structural results introduced in the present paper, upon which the ones of [12] are built.

The exponential dependence of the running time of our algorithm on the ambient dimension d is easily explained by the fact that the input point cloud is replaced by a union of balls in the calculations, which forces the algorithm to compute arrangements of hyperplanes in \mathbb{R}^{d+1} . The resort to a union of balls is made necessary by the fact that using only the data points to drive the construction of the witness complex may result in the appearance of holes in the complex, as pointed out in Sections 3 and 4 of the paper. Several other attempts at *filling in* these holes have been made, starting with the work of Carlsson and de Silva [18], in which the notion of witness is relaxed so that the resulting complex contains a larger number of simplices. A variant of this relaxation, proposed by Attali *et al.* [4], enables to fill in the holes while limiting the overhead in terms of size. A noticeable advantage of using relaxed witness complexes instead of standard witness complexes over unions of balls, is that the construction of the data structure only requires to compare distances, thus removing the need for costly geometric predicates and reducing the

algorithm's complexity. Nevertheless, there remains to provide theoretical guarantees for these various relaxations.

Interpretation of the output. Theorem 5.2 guarantees the existence of a plateau in the diagram of the Betti numbers of $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$, but in practice it does not tell where this plateau is located in the diagram, because δ and $\text{rch}(S)$ remain unknown. Nevertheless, it does give a guarantee on the length of the plateau, which, in view of H1, is of the order of $(a \text{rch}(S) - b \delta)$, where a, b are two constants. Hence, for sufficiently small δ , the plateau is long enough to be detected by the user or any statistical method used in a post-processing step. If W samples several manifolds, then several plateaus may appear in the diagram: each one of them shows a plausible reconstruction, depending on the scale at which the data set W is processed – see Figure 1. In-between the plateaus, the algorithm goes through transition phases where the behavior of the reconstruction cannot be controlled. This is the case in particular at the end of the refinement process, where ε becomes too small compared to δ (as per Hypothesis H1) for the guarantees stated in Theorem 4.1 to hold.

A variant of the algorithm uses the existence of a plateau in the diagram of Betti numbers as an early termination criterion. Specifically, the process stops as soon as a sufficiently large plateau appears in the diagram, corresponding to a plausible reconstruction. The minimum plateau length required for early termination can be specified by the user through some input parameter. This variant of the algorithm prevents the landmark set L from becoming too large, especially in cases where the input point cloud W oversamples the underlying manifold. This idea can be further exploited through an interactive process, where the user or a software agent queries our algorithm for the first significant plateau in the diagram, and then decides whether the corresponding reconstruction is suitable for the application considered or not; in the negative, our algorithm is queried for the next significant plateau in the diagram, and so on. It should be noted however that the presence of a plateau in the diagram of Betti numbers only provides partial evidence that a suitable reconstruction exists, since stability of the Betti numbers does not mean stability of the corresponding homology generators, which itself does not mean stability of the homeomorphism type, as pointed out in [27].

Handling noise in the data. It turns out that the structural results of Section 4 still hold⁹ if we relax the assumption on W and allow its points to lie slightly off the manifold S , at a Hausdorff distance of δ . This gives a deeper meaning to Theorem 5.2, which now guarantees that the algorithm generates a plateau whenever there exists a manifold S that is well-sampled by L and that passes close to the points of W . This holds in particular for small deformations S of the manifold S_0 from which the points of W have been drawn: the consequence is that the topological features of S_0 (connected components, handles, *etc.*) are captured progressively by the algorithm, by decreasing order of size in the ambient space. For instance, if S_0 is a double torus whose handles have significantly different sizes, then the algorithm first detects the larger handle, generating a plateau with $\beta_0 = \beta_2 = 1$ and $\beta_1 = 2$, then later on it detects also the smaller handle, generating a new plateau with $\beta_0 = \beta_2 = 1$ and $\beta_1 = 4$. This property, illustrated in the experimental section of [27], enlarges greatly the range of applications of our method.

Our theoretical results also hold when $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ and $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$ are replaced by $\mathcal{D}_\omega^{W^{\zeta_1}}(L)$ and $\mathcal{D}_\omega^{W^{\zeta_2}}(L)$ respectively. This means that the algorithm can use indifferently the weak witness complex

⁹ W^ζ still covers S when $\zeta \geq \delta$, and the proof that $\mathcal{C}_\omega^{W^\zeta}(L) \subseteq K_\omega^\theta(L)$ holds the same, with an additional term δ in the bounds of Lemmas 4.3 through 4.6.

or the strong witness complex, with similar theoretical guarantees. In Section 5, we focus mainly on the weak witness complex, which is potentially more practical, yet the definitive choice between both complexes should depend on the application considered.

Intrinsic versus extrinsic. Our approach to manifold reconstruction is purely *extrinsic*, that is, based on the pairwise distances between the data points in the ambient space \mathbb{R}^d . What if intrinsic distances along the manifold S are used instead? Is there any hope to include the witness complex into the (intrinsically defined) Delaunay triangulation? De Silva recently gave a partial answer to this question, showing that the Weak Witness Theorem 2.3 holds on Riemannian manifolds of arbitrary dimensions that have constant sectional curvature [17]. Gao *et al.* independently proved that a class of 2-dimensional length spaces also have this property [25]. What is the answer in the general case, where dimensions and sectional curvatures can be arbitrary?

Acknowledgements

The authors thank the anonymous referees for their detailed comments. The first author was supported in part by the Program of the EU as Shared-cost RTD (FET Open) Project under Contract No IST-006413 (ACS Algorithms for Complex Shapes). The second and third authors were supported by NSF grants FRG-0354543 and ITR-0205671, by NIH grant GM-072970, and by DARPA grant 32905.

References

- [1] N. Amenta and M. Bern. Surface reconstruction by Voronoi filtering. *Discrete Comput. Geom.*, 22(4):481–504, 1999.
- [2] N. Amenta, M. Bern, and D. Eppstein. The crust and the β -skeleton: Combinatorial curve reconstruction. *Graphical Models and Image Processing*, 60:125–135, 1998.
- [3] N. Amenta, S. Choi, T. K. Dey, and N. Leekha. A simple algorithm for homeomorphic surface reconstruction. *Internat. Journal of Comput. Geom. and Applications*, 12:125–141, 2002.
- [4] D. Attali, H. Edelsbrunner, J. Harer, and Y. Mileyko. Alpha-beta witness complexes. In *Proc. 10th Workshop on Algorithms and Data Structures*, pages 386–397, 2007.
- [5] D. Attali, H. Edelsbrunner, and Y. Mileyko. Weak witnesses for Delaunay triangulations of submanifolds. In *Proc. ACM Sympos. on Solid and Physical Modeling*, pages 143–150, 2007.
- [6] M. Belkin and P. Niyogi. Semi-supervised learning on Riemannian manifolds. *Machine Learning*, 56(1-3):209–239, 2004.
- [7] J.-D. Boissonnat and S. Oudot. Provably good sampling and meshing of Lipschitz surfaces. In *Proc. 22nd Annu. Sympos. Comput. Geom.*, pages 337–346, 2006.
- [8] F. Cazals and J. Giesen. Delaunay triangulation based surface reconstruction. In J.D. Boissonnat and M. Teillaud, editors, *Effective Computational Geometry for Curves and Surfaces*, pages 231–273. Springer, 2006.

- [9] F. Chazal, D. Cohen-Steiner, and A. Lieutier. A sampling theory for compact sets in Euclidean space. In *Proc. 22nd Annu. Sympos. Comput. Geom.*, pages 319–326, 2006.
- [10] F. Chazal and A. Lieutier. Weak feature size and persistent homology: Computing homology of solids in \mathbb{R}^n from noisy data samples. In *Proc. 21st Annual ACM Symposium on Computational Geometry*, pages 255–262, 2005.
- [11] F. Chazal and A. Lieutier. Topology guaranteeing manifold reconstruction using distance function to noisy data. In *Proc. 22nd Annu. Sympos. on Comput. Geom.*, pages 112–118, 2006.
- [12] F. Chazal and S. Y. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. Comput. Geom.*, pages 232–241, 2008.
- [13] S.-W. Cheng, T. K. Dey, H. Edelsbrunner, M. A. Facello, and S.-H. Teng. Sliver exudation. *Journal of the ACM*, 47(5):883–904, 2000.
- [14] S.-W. Cheng, T. K. Dey, and E. A. Ramos. Manifold reconstruction from point samples. Journal version, in preparation.
- [15] S.-W. Cheng, T. K. Dey, and E. A. Ramos. Manifold reconstruction from point samples. In *Proc. 16th Sympos. Discrete Algorithms*, pages 1018–1027, 2005.
- [16] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. In *Proc. 21st ACM Sympos. Comput. Geom.*, pages 263–271, 2005.
- [17] V. de Silva. A weak characterisation of the Delaunay triangulation. *Geometriae Dedicata*, 135(1):39–64, August 2008.
- [18] V. de Silva and G. Carlsson. Topological estimation using witness complexes. In *Proc. Sympos. Point-Based Graphics*, pages 157–166, 2004.
- [19] D. DeMers and G. Cottrell. Nonlinear dimensionality reduction. *Advances in Neural Information Processing Systems 5*, pages 580–587, 1993.
- [20] T. K. Dey, J. Giesen, S. Goswami, and W. Zhao. Shape dimension and approximation from samples. *Discrete and Computational Geometry*, 29:419–434, 2003.
- [21] T. K. Dey and P. Kumar. A simple provable algorithm for curve reconstruction. In *Proc. 10th ACM-SIAM Sympos. Discrete Algorithms*, pages 893–894, January 1999.
- [22] H. Edelsbrunner. *Algorithms in Combinatorial Geometry*, volume 10 of *EATCS Monographs on Theoretical Computer Science*. Springer-Verlag, 1987.
- [23] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28:511–533, 2002.
- [24] H. Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.
- [25] J. Gao, L. J. Guibas, S. Y. Oudot, and Y. Wang. Geodesic delaunay triangulation and witness complex in the plane. In *Proc. 18th ACM-SIAM Sympos. on Discrete Algorithms*, pages 571–580, 2008.

- [26] J. Giesen and U. Wagner. Shape dimension and intrinsic metric from samples of manifolds with high co-dimension. *Discrete and Computational Geometry*, 32:245–267, 2004.
- [27] L. J. Guibas and S. Y. Oudot. Reconstruction using witness complexes. *Discrete and Computational Geometry*, 40:325–356, 2008.
- [28] A. B. Lee, K. S. Pederson, and D. Mumford. The nonlinear statistics of high-contrast patches in natural images. *Internat. J. of Computer Vision*, 54(1-3):83–103, 2003.
- [29] X.-Y. Li. Generating well-shaped d-dimensional Delaunay meshes. *Theoretical Computer Science*, 296(1):145–165, 2003.
- [30] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, March 2008.
- [31] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [32] M.-J. Seow, R. C. Tompkins, and V. K. Asari. A new nonlinear dimensionality reduction technique for pose and lighting invariant face recognition. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition - Workshops*, page 160, 2005.
- [33] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [34] M. Vlachos, C. Domeniconi, D. Gunopulos, G. Kollios, and N. Koudas. Non-linear dimensionality reduction techniques for classification and visualization. In *Proc. 8th ACM SIGKDD Internat. Conf. on Knowledge discovery and data mining*, pages 645–651, 2002.

A Appendix – Proofs of Lemmas 2 through 10 of [15]

Proof of Lemma 2.7 (Lemma 2 (ii) of [15]). Since $p \in K^\theta(v)$, we have $\angle pvq \leq \theta$. Hence, $\|p - q\| = \|p - v\| \sin \angle pvq \leq \|p - v\| \sin \theta$. Moreover, $\|q - v\| \leq \|p - v\| < \frac{1}{4} \text{rch}(S)$, hence, by Lemma 7 of [26] (whose proof holds the same in our context), $\|q - q'\| \leq \frac{2\|q-v\|^2}{\text{rch}(S)} \leq \frac{2\|p-v\|^2}{\text{rch}(S)}$. Now, by the triangle inequality, we have $\|p - q'\| \leq \|p - q\| + \|q - q'\|$, which, combined with the above inequalities, gives: $\|p - q'\| \leq \|p - v\| \sin \theta + \frac{2\|p-v\|^2}{\text{rch}(S)}$. \square

Proof of Lemma 2.8 (Lemma 3 of [15]). Let $\varepsilon_1 = \frac{3\varepsilon}{1-\sin\theta}$. The hypothesis of the lemma implies that $\varepsilon_1 < \frac{1}{3}(1 - \sin\theta) \text{rch}(S)$. Assume for a contradiction that there exist $v \in L$ and $p \in K^\theta(v) \cap V_\omega(v)$ such that $\|p - v\| > \varepsilon_1$. Let $y \in [p, v]$ be such that $\|y - v\| = \varepsilon_1/2$. Since $p \in K^\theta(v)$, we have $[p, v] \subseteq K^\theta(v)$, and hence $y \in K^\theta(v)$. Let q be the orthogonal projection of y onto $T(v)$, and q' the point of S closest to q .

Since $y \in K^\theta(v)$ and $\|y - v\| = \frac{\varepsilon_1}{2} < \frac{1}{4} \text{rch}(S)$, Lemma 2.7 states that $\|y - q'\| \leq \frac{\varepsilon_1}{2} \left(\sin \theta + \frac{\varepsilon_1}{\text{rch}(S)} \right)$. Let B' be the closed ball of center q' and radius ε . By the triangle inequality, the distance between y and any point of B' is at most $\varepsilon + \|y - q'\|$, which is bounded by $\frac{\varepsilon_1}{3}(1 - \sin\theta) + \frac{\varepsilon_1}{2} \left(\sin \theta + \frac{\varepsilon_1}{\text{rch}(S)} \right) = \frac{\varepsilon_1}{2} \left(\frac{2}{3} + \frac{\sin\theta}{3} + \frac{\varepsilon_1}{\text{rch}(S)} \right)$. This quantity is less than $\frac{\varepsilon_1}{2}$ since by hypothesis we

have $\varepsilon_1 < \frac{1}{3}(1 - \sin \theta) \text{rch}(S)$. Therefore, B' is included in the open ball of center y and radius $\|y - v\| = \varepsilon_1/2$. In particular, B' does not contain v . Now, $q' \in S$, B' has radius ε , and L is an ε -sample of S , thus B' contains some point $u \in L$. This point is different from v since $v \notin B'$.

We have $(\|p - v\|^2 - \omega(v)^2) - (\|p - u\|^2 - \omega(u)^2) = \|p - v\|^2 - \|p - u\|^2 + \omega(u)^2 - \omega(v)^2$. Note that $\omega(u)^2$ is non-negative. Moreover, since $\bar{\omega}$ bounds the relative amplitude of ω , we have $\omega(v)^2 \leq \bar{\omega}^2 \|u - v\|^2$. Hence, $(\|p - v\|^2 - \omega(v)^2) - (\|p - u\|^2 - \omega(u)^2) \geq \|p - v\|^2 - \|p - u\|^2 - \bar{\omega}^2 \|u - v\|^2$. Now, recall that u belongs to B' , which is included in the ball of center y and radius $\|y - v\| = \varepsilon_1/2$. The latter is included in the diametral ball of $[p, v]$ since $y \in [p, v]$ and $\|p - v\| > \varepsilon_1$. Therefore, $\angle puv$ is obtuse, and we have $\|p - v\|^2 - \|p - u\|^2 \geq \|u - v\|^2$, which is greater than $\bar{\omega}^2 \|u - v\|^2$. It follows that $(\|p - v\|^2 - \omega(v)^2) - (\|p - u\|^2 - \omega(u)^2)$ is positive, which means that p does not belong to $V_\omega(v)$. This contradicts the hypothesis of the lemma. \square

Let now W be a (finite or infinite) δ -sample of some manifold S in \mathbb{R}^d , and let $L \subseteq W$ be a (finite) ε -sparse ε -sample of W . Note that L is an $(\varepsilon + \delta)$ -sample of S . Let $\bar{\omega}$ be an arbitrary value within $(0, 1/2)$. We define $c_1(k), c_2(k)$ as at the top of Section 4, and $c_3(k)$ as at the top of Section 5.

Lemma A.1 (Lemma 4 of [15]) *Let $\omega : L \rightarrow [0, \infty[$ be a distribution of weights of relative amplitude at most $\bar{\omega}$. If $\varepsilon + \delta \leq \frac{\text{rch}(S)}{2+4\bar{\omega}}$ and $\max\{\delta, \zeta\} \leq \varepsilon$, then, for any points $u, v \in L$ such that $[u, v] \in \mathcal{C}_\omega^{W^\zeta}(L)$, $\|u - v\| \leq 2c_1(1) \varepsilon$.*

Proof. Since $[u, v] \in \mathcal{C}_\omega^{W^\zeta}(L)$, there is a point $c \in W^\zeta$ that ω -witnesses $[u, v]$. By Lemma¹⁰ 4.5 (i), $\|c - u\|$ and $\|c - v\|$ are at most $c_1(1) \varepsilon$. Hence, $\|u - v\| \leq \|u - c\| + \|c - v\| \leq 2c_1(1) \varepsilon$. \square

Lemma A.2 (Lemma 5 (i) of [15]) *Let $\omega, \omega' : L \rightarrow [0, \infty[$ be two distributions of weights of relative amplitudes bounded from above by $\bar{\omega}$. If $\varepsilon + \delta \leq \frac{\text{rch}(S)}{2+4\bar{\omega}}$ and $\max\{\delta, \zeta\} \leq \varepsilon$, then, for any points $u, u', v \in L$ such that $[u, v] \in \mathcal{C}_\omega^{W^\zeta}(L)$ and $[u', v] \in \mathcal{C}_{\omega'}^{W^\zeta}(L)$, $\|u - v\| \leq 2c_1(1) \|u' - v\|$.*

Proof. By Lemma A.1, we have $\|u - v\| \leq 2c_1(1) \varepsilon$. Moreover, since L is ε -sparse, we have $\|u' - v\| \geq \varepsilon$. The result follows. \square

Given $\zeta \geq 0$ and $v \in L$, let $G(v) \subset L$ be the set of all $u \in L$ such that u is a neighbor of v in $\mathcal{C}_\omega^{W^\zeta}(L)$ for some weight distribution ω of relative amplitude at most $\bar{\omega}$.

Lemma A.3 (Lemma 6 of [15]) *If $\varepsilon + \delta \leq \frac{\text{rch}(S)}{2+4\bar{\omega}}$ and $\max\{\delta, \zeta\} \leq \varepsilon$, then, for any $v \in L$, the size of $G(v)$ is at most $(1 + 4c_1(1))^d$.*

Proof. By Lemma A.1, the points of $G(v)$ lie in the ball $B(v, 2c_1(1) \varepsilon)$. Moreover, they lie at least ε away from one another, since L is ε -sparse. Hence, they are centers of pairwise-disjoint balls of radius $\varepsilon/2$, included in $B(v, (2c_1(1) + 1/2) \varepsilon)$. It follows that the size of $G(v)$ is at most $\text{vol}(B(v, (2c_1(1) + 1/2) \varepsilon)) / \text{vol}(B(v, \varepsilon/2)) = (4c_1(1) + 1)^d$. \square

Using the fact that the points of $G(v)$ lie near the tangent space of S at v , it is possible to make the bound of Lemma A.3 depend solely on the intrinsic dimension of the manifold S . However, this result cannot be exploited in our reconstruction algorithm, where the dimension of S is unknown.

¹⁰The statement and proof of Lemma 4.5 hold in fact for $\max\{\delta, \zeta\} \leq \varepsilon$ and $\varepsilon + \delta \leq \frac{\text{rch}(S)}{2k(1+3\bar{\omega})-2\bar{\omega}}$.

Lemma A.4 (Lemma 7 of [15]) *If $\varepsilon + \delta \leq \frac{\text{rch}(S)}{2d(1+3\bar{\omega})-2\bar{\omega}}$ and $\max\{\delta, \zeta\} \leq \varepsilon$, then, for any distribution of weights $\omega : L \rightarrow [0, \infty[$ of relative amplitude at most $\bar{\omega}$, for any simplex $\sigma \in \mathcal{C}_\omega^{W^\zeta}$, the orthoradius-edge ratio of σ is at most $c_2(d)$.*

Proof. Let σ be a simplex of $\mathcal{C}_\omega^{W^\zeta}(L)$, and let c, r be respectively its orthocenter and orthoradius: we have $\{c\} = \text{aff}(\sigma) \cap \text{aff}(V_\omega(\sigma))$ and $r = \sqrt{\|c - v\|^2 - \omega(v)^2}$, where v is any vertex of σ . Recall that $\mathcal{D}_\omega(L)$ has no simplex of dimension greater than d , since the weighted point set L lies in general position. Hence, the dimension of σ is at most d , and by Lemma 4.5 (iii) there is a point $p \in V_\omega(\sigma)$ such that $\|p - v\| \leq c_2(d) \varepsilon$. Therefore,

$$r = \sqrt{\|c - v\|^2 - \omega(v)^2} \leq \sqrt{\|p - v\|^2 - \omega(v)^2} \leq \|p - v\| \leq c_2(d) \varepsilon.$$

Moreover, since L is ε -sparse, the shortest edge length of σ is at least ε . Hence, the orthoradius-edge ratio of σ is at most $c_2(d)$. \square

Lemma A.5 (Lemma 8 of [15]) *If $\varepsilon + \delta \leq \frac{\text{rch}(S)}{2d(1+3\bar{\omega})-2\bar{\omega}}$ and $\max\{\delta, \zeta\} \leq \varepsilon$, then, for any distribution of weights $\omega : L \rightarrow [0, \infty[$ of relative amplitude at most $\bar{\omega}$, for any simplex $\sigma \in \mathcal{C}_\omega^{W^\zeta}$ and any facet σ' of σ , the Euclidean distance between the orthocenter of σ and $\text{aff}(\sigma')$ is at most $c_3(d) \varepsilon$.*

Proof. Let c and c' be the orthocenters of σ and σ' respectively. Note that $c \in \text{aff}(V_\omega(\sigma)) \subset \text{aff}(V_\omega(\sigma'))$, and that $\{c'\} = \text{aff}(\sigma') \cap \text{aff}(V_\omega(\sigma'))$. Hence, c' is the orthogonal projection of c onto $\text{aff}(\sigma')$, and the Euclidean distance between c and $\text{aff}(\sigma')$ is precisely $\|c - c'\|$.

Let v' be any vertex of σ' . Note that v' is also a vertex of σ , since $\sigma' \subset \sigma$. Recall that $\mathcal{D}_\omega(L)$ has no simplex of dimension greater than d , since the weighted point set L lies in general position. Hence, the dimension of σ is at most d , and by Lemma 4.5 (iii), there is a point $p \in V_\omega(\sigma)$ such that $\|p - v'\| \leq c_2(d) \varepsilon$. Therefore, $\|c - v'\| \leq \|p - v'\| \leq c_2(d) \varepsilon$, since $\{c\} = \text{aff}(\sigma) \cap \text{aff}(V_\omega(\sigma))$.

If $\sigma' = [v']$, then we have $c' = v'$, which implies that $\|c - c'\| = \|c - v'\| \leq c_2(d) \varepsilon$, which is at most $c_3(d) \varepsilon$.

Else, σ' has at least one vertex u' different from v' . Let r' denote the orthoradius of σ' . By Lemma A.4 (applied to σ'), we have $r' \leq c_2(d) \|v' - u'\|$, which by Lemma A.1 is at most $2c_2(d)c_1(1) \varepsilon$. Using Lemma 4.3 and the triangle inequality, we get:

$$\begin{aligned} \|c - c'\| &\leq \|c - v'\| + \|v' - c'\| = \|c - v'\| + \sqrt{r'^2 + \omega(v')^2} \\ &\leq \|c - v'\| + r' + \omega(v') \leq \|c - v'\| + r' + 2\bar{\omega}(\varepsilon + \delta) \\ &\leq (c_2(d)(1 + 2c_1(1)) + 2\bar{\omega}(1 + \delta/\varepsilon)) \varepsilon, \end{aligned}$$

which is at most $c_3(d) \varepsilon$ since $\delta \leq \varepsilon$. \square

Lemma A.6 (Lemma 10 of [15]) *Assume that $\varepsilon + \delta \leq \frac{\text{rch}(S)}{2d(1+3\bar{\omega})-2\bar{\omega}}$ and that $\max\{\delta, \zeta\} \leq \varepsilon$. Let $\bar{\varrho} > 0$ be a constant. Let σ be a k -dimensional $\bar{\varrho}$ -sliver ($k > 0$) with vertices in L , such that the facets of σ are not $\bar{\varrho}$ -slivers. Let v be any vertex of σ . For any distribution of weights $\omega' : L \setminus \{v\} \rightarrow [0, \infty[$ of relative amplitude at most $\bar{\omega}$, there exists an interval $I(v, \sigma, \omega') \subseteq [0, \bar{\omega}^2 \min\{\|v - u\|^2 \mid u \in L \setminus \{v\}\}]$, of length less than $8\bar{\varrho}c_1(1)c_3(d) \varepsilon^2$, such that, for any distribution of weights $\omega : L \rightarrow [0, \infty[$ of relative amplitude at most $\bar{\omega}$ that coincides with ω' on $L \setminus \{v\}$, σ belongs to $\mathcal{C}_\omega^{W^\zeta}(L)$ only if $\omega(v)^2 \in I(v, \sigma, \omega')$.*

Proof. Let l be the shortest edge length of σ . Let σ_v denote the facet of σ opposite to v , and let d_v be the Euclidean distance of v to $\text{aff}(\sigma_v)$. We have $\frac{d_v \text{vol}(\sigma_v)}{k} = \text{vol}(\sigma)$, which is less than $\frac{\bar{\varrho}^k l^k}{k!}$ since σ is a $\bar{\varrho}$ -sliver. By hypothesis, σ_v is not a $\bar{\varrho}$ -sliver, hence $\text{vol}(\sigma_v) \geq \frac{\bar{\varrho}^{k-1} l_v^{k-1}}{(k-1)!}$, where l_v is the shortest edge length of σ_v . Note that we have $l_v \geq l$ since $\sigma_v \subset \sigma$. Combining these relations, we get $d_v < \frac{k(k-1)! \bar{\varrho}^k l^k}{k! \bar{\varrho}^{k-1} l^{k-1}} = \bar{\varrho} l$.

Let $\omega' : L \setminus \{v\} \rightarrow [0, \infty[$ be a distribution of weights of relative amplitude at most $\bar{\omega}$. For any possible weight $\omega_v \in [0, \bar{\omega} \min\{\|v - u\| \mid u \in L \setminus \{v\}\}]$ of v , we call $d_c(\omega_v)$ the signed Euclidean distance between the orthocenter c of σ and $\text{aff}(\sigma_v)$, the distance being positive if, inside $\text{aff}(\sigma)$, c and v lie on the same side of $\text{aff}(\sigma_v)$. By Lemma A.5, if σ belongs to $\mathcal{C}_\omega^{W^\zeta}(L)$, then $|d_c(\omega_v)| \leq c_3(d) \varepsilon$. Moreover, it has been proved in [13, Claim 13] that $d_c(\omega_v) = d_c(0) - \frac{\omega_v^2}{2d_v}$, $d_c(0)$ being the signed Euclidean distance of c to $\text{aff}(\sigma_v)$ when $\omega_v = 0$. Therefore, σ belongs to $\mathcal{C}_\omega^{W^\zeta}(L)$ only if:

$$\max\{0, 2d_v (d_c(0) - c_3(d) \varepsilon)\} \leq \omega_v^2 \leq \min\{2d_v (d_c(0) + c_3(d) \varepsilon), \bar{\omega}^2 \min\{\|v - u\|^2 \mid u \in L \setminus \{v\}\}\}.$$

The length of the admissible interval for ω_v^2 is at most $4d_v c_3(d) \varepsilon$, which is less than $4\bar{\varrho} l c_3(d) \varepsilon$, which by Lemma A.1 is bounded by $8\bar{\varrho} c_1(1) c_3(d) \varepsilon^2$. \square