Manifold Reconstruction in Arbitrary Dimensions using Witness Complexes

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

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

Steve Y. Oudot
Computer Science Dept.
Stanford University
Stanford, CA 94305
steve.oudot@stanford.edu

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 assumptions. Unfortunately, these results do not extend to higher-dimensional manifolds, even under stronger sampling conditions. In this paper, we show how the sets of witnesses and landmarks can be enriched, so that the nice relations that exist between both complexes still hold on higher-dimensional manifolds. We also use our structural results to devise 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 does not require the input point sample W to be sparse. Its time complexity is bounded by $c(d)|W|^2$, where $c(d)$ is a constant depending solely on the dimension d of the ambient space.

Categories and Subject Descriptors: I.3.5 [Computer Graphics]: Curve, surface, solid, and object representations

General Terms: Algorithms, Theory.

Keywords: Witness complex, restricted Delaunay triangulation, manifold reconstruction, sampling conditions.

1. INTRODUCTION

A number of areas of Science and Engineering deal with point clouds lying on 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* problem, which is at the core of non-linear dimensionality reduction techniques [25 27], finds applications in many areas, including machine learning [4], pattern recognition [26], scientific visualization [28], image or signal processing [22], and neural computation [16]. 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 [8 13 29] offer a nice mathematical framework. However, in many situations it is desirable to construct a simplicial complex with the same topological type as the manifold, and close to it geometrically.

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 [6] for a survey. In these methods, the output complex is extracted from $\mathcal{D}(W)$, and it 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 smooth or Lipschitz curves in the plane and surfaces in 3-space, under mild sampling conditions [1 2 5].

Generalizing these ideas to arbitrary dimensions and co-dimensions is difficult however, because a given point set W may sample well various manifolds with different homotopy types, as illustrated in Figure 1. To overcome this issue, 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 [12 17 20]. 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 [12 17 20] contain a wealth of relevant ideas and results, some of which are used in this paper.

A different and very promising approach [7 9 21 23], 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

SCG'07, June 6–8, 2007, Gyeongju, South Korea.

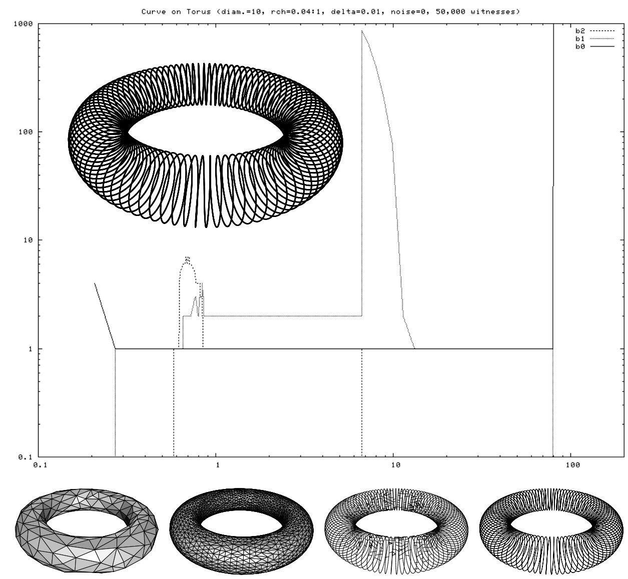

Figure 1: A helical curve drawn on a torus (top). Any dense sampling of the curve is also a dense sampling of the torus. To deal with this ambiguity, the algorithm of [21] builds a sequence of complexes (bottom) approximating the input at various scales, and maintains their Betti numbers (top).

one-parameter family of complexes. This approach to reconstruction stands in sharp contrast with previous work in the area. In [7 9 23], the family of complexes is derived from the α-offsets of the input W, where α ranges from zero to infinity. The theoretical guarantees of [7] hold in a very wide setting, since W is assumed to be a sufficiently dense sampling of a general compact set. In [21], the family is given by the *witness complex* of L relative to W, or $\mathcal{C}^W(L)$ for short, where L is a subset of W constructed iteratively, starting with $L = \emptyset$, and inserting each time the point of W lying furthest away from L. 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 [3 21]. The assumptions on L in [21] are more stringent than the ones on W, but this is not an issue since L is generated by the algorithm. Unfortunately, the structural results of [3 21] do not hold in higher dimensions, which makes it very unlikely that the algorithm works on any manifold S of dimension greater than two. The reason is that the normals of $\mathcal{D}^S(L)$ can be arbitrarily wrong if $\mathcal{D}^S(L)$ contains badly-shaped simplices, called *slivers* [12]. As a consequence, $\mathcal{C}^W(L)$ may not be included in $\mathcal{D}^S(L)$, which may not be homotopy equivalent to S [24]. This is true even if W and L satisfy strong sampling conditions, such as being arbitrarily dense uniform samples of S.

In this paper, we show how to enrich W and L in order to make the structural results of [3 21] hold for higher-dimensional manifolds. To each point $p \in L$ we assign a non-negative weight $\omega(p)$, so that $\mathcal{D}^S(L)$ and $\mathcal{C}^W(L)$ are now replaced by their weighted versions, $\mathcal{D}^S_\omega(L)$ and $\mathcal{C}^W_\omega(L)$. The idea of assigning weights to the vertices comes from [10], 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.* [12], who showed that, under sufficient sampling conditions on L, there exists a distribution of weights ω such that $\mathcal{D}^S_\omega(L)$ is homeomorphic to S. Our main result is that, for the same distribution of weights and under similar conditions on L, $\mathcal{C}^W_\omega(L)$ is included in $\mathcal{D}^S_\omega(L)$, for all $W \subseteq S$. Combined with the fact that $\mathcal{C}^S_\omega(L)$ contains $\mathcal{D}^S_\omega(L)$, we get that $\mathcal{C}^S_\omega(L) = \mathcal{D}^S_\omega(L)$, which is homeomorphic to S. This is a generalization of the result of [3] 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 radius ζ. For sufficiently large ζ, this enlarged set W^ζ contains S, and hence $\mathcal{D}^S_\omega(L) \subseteq \mathcal{C}^{W^\zeta}_\omega(L)$. And if ζ is not too large, then $\mathcal{C}^{W^\zeta}_\omega(L)$ is still included in $\mathcal{D}^S_\omega(L)$. Thus, we obtain $\mathcal{C}^{W^\zeta}_\omega(L) = \mathcal{D}^S_\omega(L)$ under sufficient conditions on W, L, ζ. Here again, the condition on L is stringent, but the one on W is mild.

These structural results suggest combining the method of [21] with the sliver exudation technique of [10], in order to make it work on higher-dimensional manifolds. 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 $\mathcal{C}^{W^\zeta}_\omega(L)$. This algorithm is very simple conceptually, and to some extent it can be viewed as a dynamic version of the algorithm of [12], where the Delaunay triangulation of W would be constructed progressively as the weights of the points of W are computed. This raises a number of questions, such as whether the weight assignment process removes all the slivers from the vicinity of $\mathcal{D}^S_\omega(L)$, since some of the weights are assigned in early stages of the course of the algorithm. We prove that all slivers are indeed removed at some point, and that consequently $\mathcal{C}^{W^\zeta}_\omega(L)$ is homeomorphic to S, provided that W is a dense enough sampling of S.

Our assumption on W is much less stringent than the one used in [12], which makes our algorithm more practical. On the other hand, our assumption that S is a smooth manifold is stronger than the one of [7], but we get stronger guarantees, namely $\mathcal{C}^{W^\zeta}_\omega(L)$ is homeomorphic to S, whereas the complex of [7] is only homotopy equivalent to S. Observe also that the witness complex is always embedded in the ambient space $\mathbb{R}^d$, which is not the case of the nerve of an offset. This is important if one wants to do further processing on the reconstruction. By considering the α-shape of W instead of the nerve of its α-offset, the authors of [7] get rid of this issue. Nevertheless, they cannot get rid of the so-called *curse of dimensionality* because, as α varies from zero to infinity, the α-complex spans the entire d-dimensional Delaunay triangulation of W, whose size is $\Theta\left(|W|^{\lceil d/2 \rceil}\right)$. In contrast, the witness complex can be maintained by considering only the $\kappa(d)$ nearest neighbors in L of each point of W, which reduces the time complexity of our algorithm to $O\left(c(d)|W|^2\right)$, where $\kappa(d)$ and $c(d)$ are constants depending solely on d. This is worse than the bound of [12] ($c(d)|W| \log |W|$) since we cannot benefit from the sparseness assumption on W, which is essential for the correctness of the algorithm of [12].

The outline of the paper is as follows: after recalling the necessary background in Section 2, we present our main structural result (Theorem 3.1) in Section 3, and we introduce our reconstruction algorithm in Section 4.

2. BACKGROUND AND DEFINITIONS

2.1 Manifolds and samples

The ambient space is $\mathbb{R}^d$, $d \geq 3$, equipped with the usual Euclidean norm $\|p\| = \sqrt{\sum_{i=1}^d p_i^2}$. All manifolds considered in this paper are compact closed submanifolds of $\mathbb{R}^d$, of dimension two or more. The case of curves has already been addressed[1] in [21]. 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.

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, and L is *ε-sparse* if the pairwise Euclidean distances between the points of L are at least ε. Note that an ε-sparse sample of a compact set is always finite. Parameter ε is sometimes made adaptative in the literature [1], its value depending on the distance to the medial axis of the manifold.

2.2 Simplex shape

Given $k+1$ points $p_0, \cdots, p_k \in \mathbb{R}^d$, $[p_0, \cdots, p_k]$ denotes the k-simplex of vertices $p_0, \cdots, p_k$. The geometric realization of this simplex is the convex hull of $p_0, \cdots, p_k$, which has dimension k if the vertices are affinely independent. Following [11], we call *sliver measure* of $[p_0, \cdots, p_k]$ the ratio[2]:

$$\varrho([p_0, \cdots, p_k]) = \frac{\text{vol}([p_0, \cdots, p_k])}{\min\{\|p_i - p_j\|,\ 0 \leq i < j \leq k\}^k}, \qquad (1)$$

where $\text{vol}([p_0, \cdots, p_k])$ denotes the volume of the convex hull of $p_0, \cdots, p_k$ 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 to the case $k = 0$ by imposing $\varrho([p_0]) = 1$ for any vertex p_0. Given $\bar{\varrho} \geq 0$, simplex $[p_0, \cdots, p_k]$ is said to be a *$\bar{\varrho}$-sliver* if $\varrho([p_0, \cdots, p_k]) < \bar{\varrho}^k/k!$. 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 large sliver measure does not always mean being well-shaped. Take an isoceles triangle t of base b and height $h \geq \frac{b\sqrt{3}}{2}$. Its sliver measure is $\varrho(t) = \frac{h \cdot b}{2b^2} = \frac{h}{2b}$. Reducing b while keeping h constant makes t arbitrarily skinny but $\varrho(t)$ arbitrarily large – t is called a *dagger* in the literature [10]. 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 $\omega : L \to [0, \infty)$. The quantity $\max_{u \in L, v \in L \setminus \{u\}} \frac{\omega(u)}{\|u - v\|}$ is called the *relative amplitude* of ω. Given $p \in \mathbb{R}^d$, the *weighted distance* from p to some weighted point $v \in L$ is $\|p - v\|^2 - \omega(v)^2$. This is actually not a metric, since it is not symmetric. 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 [10].

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 [19]. For any simplex σ of $\mathcal{D}_\omega(L)$, $\text{V}_\omega(\sigma)$ denotes the face of the weighted Voronoi diagram of L that is dual to σ. Moreover, given any subset X of $\mathbb{R}^d$, we call $\mathcal{D}_\omega^X(L)$ the weighted Delaunay triangulation of L restricted to X. 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, $\text{V}_\omega(\sigma)$ becomes $\text{V}(\sigma)$, and $\mathcal{D}_\omega^X(L)$ becomes $\mathcal{D}^X(L)$.

DEFINITION 2.1. *Let $W, L \subseteq \mathbb{R}^d$ be such that L is finite, and let ω be a distribution of weights on L.*

- *Given a point $w \in W$ and a simplex $\sigma = [p_0, \cdots, p_k]$ with vertices in L, w ω-witnesses σ if $p_0, \cdots, p_k$ are among the $k+1$ nearest neighbors of w in the weighted metric, that is, $\forall i \in \{0, \cdots, k\}$, $\forall q \in L \setminus \{p_0, \cdots, p_k\}$, $\|w - p_i\|^2 - \omega(p_i)^2 \leq \|w - q\|^2 - \omega(q)^2$.*
- *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 [14 15]. 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 for the standard Euclidean norm, and is therefore noted $\mathcal{C}^W(L)$.

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

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

$\mathcal{D}_\omega^W(L)$ is sometimes called the *strong witness complex of L relative to W* in the literature [15]. Given 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 metric, 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.

THEOREM 2.3 (THM. 2 AND §3 OF [14]). *Let $W, L \subseteq \mathbb{R}^d$ be such that L is finite. Let also ω be any distribution of weights on L, such that the weighted point set L lies in general position. Let σ be a simplex with vertices in L. If σ and all its subsimplices are ω-witnessed by points of W, then σ belongs to $\mathcal{D}_\omega(L)$. In other words, $\mathcal{C}_\omega^W(L)$ is a subcomplex of $\mathcal{D}_\omega(L)$. Moreover, the dual weighted Voronoi face of σ intersects the convex hull of the ω-witnesses of σ and its subsimplices.*

It is always possible to perturb the points of L or their weights infinitesimally, so as to make them lie in general position. Therefore, in the rest of the paper we assume implicitely that the weighted point set L is in general position.

On 1- and 2-dimensional manifolds, the unweighted witness complex is closely related to the unweighted restricted Delaunay triangulation [3 21], which provides good topological and geometric approximations [1 2]. However, as proved

[1]The results of [21] apply to Lipschitz curves in the plane, but they can be extended to smooth curves in higher dimensions in a straightforward manner.

[2]This definition generalizes the one of [10] to higher dimensions. It departs from the definition of [12], which introduced a bug in the proof of Lemma 10 of the same paper.

in [24], these properties do not extend to higher-dimensional manifolds, even under stronger sampling conditions:

THEOREM 2.4 (SEE [24]). *For any positive constant* $\mu < 1/3$, *there exists a closed and compact hypersurface* S *of positive reach in* $\mathbb{R}^4$, *and an* $\Omega(\varepsilon)$*-sparse* $O(\varepsilon)$*-sample* L *of* S, *with* $\varepsilon = \mu\,\mathrm{rch}(S)$, *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 exists a* δ*-sample* W *of* S *such that* $\mathcal{C}^W(L)$ *neither contains nor is contained in* $\mathcal{D}^S(L)$. *Furthermore,* W *can be made indifferently finite or infinite.*

The proof of this theorem builds on an example of [12, §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 change arbitrarily under 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 [21], 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

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 $\mathrm{K}^\theta(p)$ for short, the cone of semi-aperture θ around the tangent space of S at p: $\mathrm{K}^\theta(p) = \{q \in \mathbb{R}^d \mid \angle(pq, T(p)) \leq \theta\}$. The name *cocone* refers to the fact that $\mathrm{K}^\theta(p)$ 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 $\omega : L \to [0, \infty)$, the *weighted* θ*-cocone complex* of L, noted $\mathrm{K}^\theta_\omega(L)$, 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 $[p_0, \cdots, p_k]$ of $\mathcal{D}_\omega(L)$ belongs to $\mathrm{K}^\theta_\omega(L)$ if, and only if, its dual weighted Voronoi face intersects $\mathrm{K}^\theta(p_0) \cup \cdots \cup \mathrm{K}^\theta(p_k)$. Note that the cones in [12] are defined around approximations of the tangent spaces of S at the points of L. However, the results of [12] 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 [12]). *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* $\varepsilon \leq c_{\bar\varrho}\,\mathrm{rch}(S)$, *and for any* $\omega : L \to [0, \infty)$ *of relative amplitude less than* $\frac{1}{2}$, *if* $\mathrm{K}^{\pi/32}_\omega(L)$ *has no* $\bar\varrho$*-sliver, then* $\mathrm{K}^{\pi/32}_\omega(L)$ *coincides with* $\mathcal{D}^S_\omega(L)$, *which is homeomorphic to* S.

It is also proved in [12] 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 $\mathrm{K}^{\pi/32}_\omega(L)$. This is true provided that $\bar\varrho$ is sufficiently small compared to $\bar\omega$, and that $\varepsilon \leq c_{\bar\varrho}\,\mathrm{rch}(S)$ (as in Theorem 2.5, the choice of $\bar\varrho$ influences the bound on ε). The results of [12] 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 [12, §13].

2.5 Useful results

The following results, adapted from [12 20], will be used in the sequel:

LEMMA 2.6 (LEMMA 6 OF [20]). *Let* S *be a manifold, and let* $p, q \in S$ *be such that* $\|p - q\| < \mathrm{rch}(S)$. *Then, the Euclidean distance from* q *to* $T(p)$ *is at most* $\frac{\|p-q\|^2}{2\,\mathrm{rch}(S)}$.

LEMMA 2.7 (LEMMA 2 OF [12]). *Let* S *be a manifold and* $\theta \in [0, \pi/2]$ *an angle value. Let* $v \in S$ *and* $p \in \mathrm{K}^\theta(v)$ *be such that* $\|p - v\| < \mathrm{rch}(S)$. *Then, the Euclidean distance from* p *to* S *is at most* $\left(\sin\theta + \frac{\|p-v\|}{2\,\mathrm{rch}(S)}\right)\|p - v\|$.

LEMMA 2.8 (LEMMA 3 OF [12]). *Let* S *be a manifold and* $\theta \in [0, \pi/2)$ *an angle value. Let* L *be an* ε*-sample of* S, *with* $\varepsilon < \frac{4}{9}(1 - \sin\theta)^2\,\mathrm{rch}(S)$. *For any* $\omega : L \to [0, \infty)$ *of relative amplitude less than* $\frac{1}{2}$, *for any* $v \in L$ *and any* $p \in \mathrm{K}^\theta(v) \cap V_\omega(v)$, *we have* $\|p - v\| \leq \frac{3\varepsilon}{1-\sin\theta}$.

3. STRUCTURAL RESULTS

In this section, $\bar\omega \in (0, 1/2)$ and $\bar\varrho > 0$ are fixed constants. For any $k \geq 0$, we define $c_1(k) = 4\,(1 + 2\bar\omega + k(1 + 3\bar\omega))$ and $c_2(k) = 4\,(3 + 6\bar\omega + 2k(1 + 3\bar\omega))$. Let S be a manifold in $\mathbb{R}^d$, W a (finite or infinite) δ-sample of S in $\mathbb{R}^d$, 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 $\mathcal{D}^S_\omega(L) \subseteq \mathcal{D}^{W^\zeta}_\omega(L) \subseteq \mathcal{C}^{W^\zeta}_\omega(L)$.
- 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 $\mathcal{C}^W(L)$ are now replaced by their weighted versions, $\mathcal{D}^S_\omega(L)$ and $\mathcal{C}^W_\omega(L)$. Given any angle value $\theta \in (0, \pi/2)$, our main structural result (Theorem 3.1) states that, under sufficient conditions on $\delta, \varepsilon, \zeta$, $\mathrm{K}^\theta_\omega(L)$ contains $\mathcal{C}^{W^\zeta}_\omega(L)$ for any $\omega : L \to [0, \infty)$ of relative amplitude at most $\bar\omega$. Therefore, $\mathcal{C}^{W^\zeta}_\omega(L) \subseteq \mathcal{D}^S_\omega(L)$ whenever $\theta \leq \frac{\pi}{32}$ and ω removes all $\bar\varrho$-slivers from $\mathrm{K}^{\pi/32}_\omega(L)$, by Theorem 2.5.

THEOREM 3.1. *Given any angle value* $\theta \in (0, \pi/2)$, *if* $\delta, \varepsilon, \zeta$ *satisfy the following conditions:*

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

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

then $\mathcal{D}^S_\omega(L) \subseteq \mathcal{D}^{W^\zeta}_\omega(L) \subseteq \mathcal{C}^{W^\zeta}_\omega(L) \subseteq \mathrm{K}^\theta_\omega(L)$ *for any distribution of weights* ω *on* L *of relative amplitude at most* $\bar\omega$. *If in addition* $\theta \leq \pi/32$, $\varepsilon \leq c_{\bar\varrho}\,\mathrm{rch}(S)$, *and* ω *is such that* $\mathrm{K}^{\pi/32}_\omega(L)$ *contains no* $\bar\varrho$*-sliver*[3], *then Theorem 2.5 implies that* $\mathcal{D}^{W^\zeta}_\omega(L) = \mathcal{C}^{W^\zeta}_\omega(L) = \mathrm{K}^{\pi/32}_\omega(L) = \mathcal{D}^S_\omega(L)$, *which is homeomorphic to* S.

[3] As mentioned after Theorem 2.5, such distributions of weights exist, provided that $\bar\varrho$ is sufficiently small.

H1 requires that W is dense compared to L, which must be dense compared to $\mathrm{rch}(S)$. This is very similar in spirit to the condition of [21]. H2 bounds the dilation parameter ζ. The smaller the angle θ of semi-aperture of the cocones, the smaller ε must be for $\mathrm{K}^\theta_\omega(L)$ to contain $\mathcal{D}^S_\omega(L)$, and the smaller ζ must be for $\mathrm{K}^\theta_\omega(L)$ to contain $\mathcal{C}^{W^\zeta}_\omega(L)$. Condition $\zeta \geq \delta$ ensures that $\mathcal{D}^S_\omega(L) \subseteq \mathcal{C}^{W^\zeta}_\omega(L)$.

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 $\mathcal{C}^S_\omega(L)$ coincides with $\mathcal{D}^S_\omega(L)$ for a suitable distribution of weights $\omega : L \to [0,\infty)$, 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.* [3] to higher dimensions, with an additional sparseness condition on L. This condition is mandatory for the existence of a suitable ω on manifolds of dimension three or higher [10 12]. The general case of Theorem 3.1 allows to have $W \subsetneq S$, which is more practical.

Another useful property, stated as Theorem 3.2 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[4] with H2: as a consequence, two different values of ζ have to be used in order to bound $\mathrm{K}^\theta_\omega(L)$, as emphasized in our algorithm – see Section 4.

THEOREM 3.2. *If $\theta \leq \frac{\pi}{32}$, and if δ, ε satisfy Condition H1 of Theorem 3.1, then, for any $\zeta \geq \frac{7\sin\theta}{2(1-\sin\theta)}\,\varepsilon$, for any $\omega : L \to [0,\infty)$ of relative amplitude at most $\bar{\omega}$, $\mathrm{K}^\theta_\omega(L) \subseteq \mathcal{D}^{W^\zeta}_\omega(L) \subseteq \mathcal{C}^{W^\zeta}_\omega(L)$.*

The rest of Section 3 is devoted to the proofs of Theorems 3.1 and 3.2, and we give an overview below.

Overview of the proofs

The proof of Theorem 3.1 is given in Section 3.1. Showing that $\mathcal{D}^S_\omega(L) \subseteq \mathcal{D}^{W^\zeta}_\omega(L) \subseteq \mathcal{C}^{W^\zeta}_\omega(L)$ 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 $\mathcal{D}^S_\omega(L) \subseteq \mathcal{D}^{W^\zeta}_\omega(L)$, which by Corollary 2.2 is included in $\mathcal{C}^{W^\zeta}_\omega(L)$ for any $\omega : L \to [0,\infty)$. Showing that $\mathcal{C}^{W^\zeta}_\omega(L) \subseteq \mathrm{K}^\theta_\omega(L)$ 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 metric are bounded (Lemma 3.4). From this fact we derive some bounds on the Euclidean distances between the simplices of $\mathcal{C}^{W^\zeta}_\omega(L)$ and their ω-witnesses (Lemma 3.5). 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 $\mathcal{C}^{W^\zeta}_\omega(L) \subseteq \mathrm{K}^\theta_\omega(L)$ (Lemma 3.6).

The proof of Theorem 3.2 is given in Section 3.2. Intuitively, if L is a sufficiently dense sampling of S, then, for any simplex $\sigma \in \mathrm{K}^\theta_\omega(L)$ and any vertex v of σ such that $\mathrm{V}_\omega(\sigma) \cap \mathrm{K}^\theta(v) \neq \emptyset$, the points of $\mathrm{V}_\omega(\sigma) \cap \mathrm{K}^\theta(v)$ lie close to $T(v)$ and hence close to S. Therefore, they belong to W^ζ whenever ζ is large enough, which implies that $\sigma \in \mathcal{D}^{W^\zeta}_\omega(L) \subseteq \mathcal{C}^{W^\zeta}_\omega(L)$.

[4]Indeed, the upper bound on ζ in H2 is always smaller than the lower bound on ζ in Theorem 3.2.

3.1 Proof of Theorem 3.1

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

LEMMA 3.3. *Under Condition H1 of Theorem 3.1, for any $v \in L$ and any $\omega : L \to [0,\infty)$ 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 $\mathrm{V}(v)$ of v in the unweighted Voronoi diagram of L. Since L is an $(\varepsilon+\delta)$-sample of S, $S_v \cap \mathrm{V}(v)$ is contained in the ball $B(v)$ of center v and radius $\varepsilon+\delta$. By H1, the radius of this ball is at most $\mathrm{rch}(S)$, hence we have $S_v \cap B(v) \subsetneq S_v$. This implies that the boundary of $\mathrm{V}(v)$ 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. □

Here is now the core argument of the proof of Theorem 3.1:

LEMMA 3.4. *Under Condition H1 of Theorem 3.1, for any $\omega : L \to [0,\infty)$ of relative amplitude at most $\bar{\omega}$, for any $p \in S$, and for any $k \leq d$, the Euclidean distance between p and its $(k+1)$th nearest landmark in the weighted metric 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 metric, 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 3.3. 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 $v_1, \cdots, v_k$ denote the k nearest landmarks of p in the weighted metric. By induction, we have $\{v_1, \cdots, v_k\} \subset B(p, r_{k-1})$. Moreover, by the case $k=0$ above, for any $i \leq k$ we have $S \cap \mathrm{V}_\omega(v_i) \subseteq B(v_i, (1+2\bar{\omega})(\varepsilon+\delta))$, 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 $r_{k-1} + (1+2\bar{\omega})(\varepsilon+\delta) \leq \mathrm{rch}(S)$, hence we have $S \cap B(p, r_{k-1} + (1+2\bar{\omega})(\varepsilon+\delta)) \subsetneq S_p$, which implies two things: first, $v_1, \cdots, v_k$ 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 $\bigcup_{i=1}^k (S \cap \mathrm{V}_\omega(v_i))$. Let q be a point of intersection: q lies on the bisector hyperplane between some v_i and some point $v \in L \setminus \{v_1, \cdots, v_k\}$, *i.e.* $q \in \mathrm{V}_\omega(v_i) \cap \mathrm{V}_\omega(v)$. 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 $v \in L \setminus \{v_1, \cdots, v_k\}$, 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 metric. 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 3.3. □

Using Lemma 3.4, we can bound the Euclidean distances between the simplices of $\mathcal{C}_\omega^{W^\zeta}(L)$ and their ω-witnesses. Our bounds depend on quantities $c_1(k), c_2(k)$, defined at the top of Section 3:

Lemma 3.5. *Under Conditions H1–H2 of Theorem 3.1, for any $\omega : L \to [0, \infty)$ 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, 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', $\|c' - v\| \leq c_2(k)\, \varepsilon$.

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 metric. Unfortunately, we cannot apply Lemma 3.4 directly to bound the Euclidean distance between c and its $(k+1)$ nearest landmarks in the weighted metric, 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 3.4, 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 metric. 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 lth nearest landmark of c in the weighted metric. 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 3.3 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 3.4, the Euclidean distance between w' and its nearest landmark u' in the weighted metric 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 metric. 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 3.3. 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). $\square$

We can now show that $\mathcal{C}_\omega^{W^\zeta}(L) \subseteq \mathrm{K}_\omega^\theta(L)$, which concludes the proof of Theorem 3.1:

Lemma 3.6. *Under Conditions H1–H2 of Theorem 3.1, for any distribution of weights $\omega : L \to [0, \infty)$ 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, $\mathrm{V}_\omega(\sigma) \cap \mathrm{K}^\theta(v) \neq \emptyset$. As a consequence, $\mathcal{C}_\omega^{W^\zeta}(L) \subseteq \mathrm{K}_\omega^\theta(L)$.*

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 $\mathrm{V}_\omega(v)$. And since v belongs to $\mathrm{K}^\theta(v)$, we have $\mathrm{V}_\omega(v) \cap \mathrm{K}^\theta(v) \neq \emptyset$, which proves the lemma in the case $k = 0$.

Assume now that $1 \leq k \leq d$. By Lemma 3.5 (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 $\mathrm{rch}(S)$, by H1–H2, hence the Euclidean distance from w' to $T(v)$ is at most $\frac{\varepsilon^2}{2\,\mathrm{rch}(S)}\,(c_2(d) + \zeta/\varepsilon)^2$, by Lemma 2.6. It follows that the Euclidean distance from c' to $T(v)$ is at most $\zeta + \frac{\varepsilon^2}{2\,\mathrm{rch}(S)}\,(c_2(d) + \zeta/\varepsilon)^2$. 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 of the subsimplices of σ do not lie farther than $\zeta + \frac{\varepsilon^2}{2\,\mathrm{rch}(S)}\,(c_2(d) + \zeta/\varepsilon)^2$ 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 $\zeta + \frac{\varepsilon^2}{2\,\mathrm{rch}(S)}\,(c_2(d) + \zeta/\varepsilon)^2$ away from $T(v)$, and at least $\frac{1}{2}(1 - \bar{\omega}^2)\varepsilon$ away from v, in the Euclidean metric. Therefore, $\sin\angle(vp, T(v)) \leq \frac{2\zeta}{(1-\bar{\omega}^2)\varepsilon} + \frac{\varepsilon}{(1-\bar{\omega}^2)\mathrm{rch}(S)}\,(c_2(d) + \zeta/\varepsilon)^2$, which is at most $\sin\theta$, by H1–H2. As a consequence, p belongs to $\mathrm{K}^\theta(v)$. Now, recall that p is a point of $\mathrm{V}_\omega(\sigma)$. Hence, $\mathrm{V}_\omega(\sigma) \cap \mathrm{K}^\theta(v) \neq \emptyset$, which means that σ is a simplex of $\mathrm{K}_\omega^\theta(L)$. 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 $\mathrm{K}_\omega^\theta(L)$. $\square$

3.2 Proof of Theorem 3.2

Let σ be a simplex of $\mathrm{K}_\omega^\theta(L)$. 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 $\varepsilon + \delta < \frac{4}{9}(1 - \sin\theta)^2\ \mathrm{rch}(S)$, Lemma 2.8 states that $\|c - v\| \leq \frac{3(\varepsilon+\delta)}{1-\sin\theta}$, which is less than $\mathrm{rch}(S)$ since by assumption we have $\varepsilon + \delta < \frac{1-\sin\theta}{3}\ \mathrm{rch}(S)$. Therefore, by Lemma 2.7, we have $\|q' - c\| \leq \frac{3(\varepsilon+\delta)}{1-\sin\theta}\left(\sin\theta + \frac{3(\varepsilon+\delta)}{2(1-\sin\theta)\ \mathrm{rch}(S)}\right)$, 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, $\|c - w\| \leq \delta + \frac{3(\varepsilon+\delta)}{1-\sin\theta}\left(\sin\theta + \frac{3(\varepsilon+\delta)}{2(1-\sin\theta)\ \mathrm{rch}(S)}\right)$, which is at most ζ by hypothesis. It follows that $c \in W^\zeta$, and hence that σ belongs to $\mathcal{D}_\omega^{W^\zeta}(L)$. This concludes the proof of Theorem 3.2. $\square$

4. MANIFOLD RECONSTRUCTION

In this section, we use our structural results in the context of manifold reconstruction. From now on, we fix $\bar{\omega} = \frac{1}{4}$ and $\theta = \frac{\pi}{32}$, and we set $\bar{\varrho}$ to an arbitrarily small positive value.

We give an overview of the approach in Section 4.1, then we present the algorithm in Section 4.2, and we prove its correctness in Section 4.3. Our theoretical guarantees are discussed in Section 4.4. In Section 4.5, we give some details on how the various components of the algorithm can be implemented. These details are then used in Section 4.6 to bound the space and time complexities of the algorithm.

4.1 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 $\varepsilon \leq c_{\bar{\varrho}}$ rch(S) satisfying Condition H1 of Theorem 3.1. Then, Theorems 3.1 and 3.2 would guarantee that $\mathcal{D}_\omega^S(L) \subseteq \mathcal{C}_\omega^{W^{\zeta_1}}(L) \subseteq \mathrm{K}_\omega^{\pi/32}(L) \subseteq \mathcal{C}_\omega^{W^{\zeta_2}}(L)$ for any distribution of weights 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 $\zeta_2 = \frac{7\sin\pi/32}{2(1-\sin\pi/32)}\ \varepsilon$. We could then apply the pumping strategy of [10] 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 $\mathrm{K}_\omega^{\pi/32}(L)$) 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 rch(S) are unknown, finding an $\varepsilon \leq c_{\bar{\varrho}}$ rch(S) that provably satisfies H1 can be difficult, if not impossible. This is why we combine the above approach with the multi-scale reconstruction scheme of [21], 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 homology type of S – some examples can be found in [21].

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 $S_1, \cdots, S_l$, as it is the case for instance in the example of Figure 1. In such a situation, the values $\delta_1, \cdots, \delta_l$ 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 $\mathcal{C}_\omega^{W^{\zeta_1}}(L)$ – see Figure 1.

4.2 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[5] from L in the Euclidean metric 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 Section 4.1. 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 increased to ω_p, and $\mathcal{C}_\omega^{W^{\zeta_1}}(L), \mathcal{C}_\omega^{W^{\zeta_2}}(L)$ are updated.

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. It is in fact sufficient to output the diagram of their Betti numbers, computed on the fly using the persistence algorithm of [29]. 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}.

[5]At the first iteration, since L is empty, p is chosen arbitrarily in W.

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 $p := \mathrm{argmax}_{w\in W}\min_{v\in L}\|w-v\|$;
  // p is chosen arbitrarily in W if $L = \emptyset$
  $L := L \cup \{p\}$; $\omega(p) := 0$;
  $\varepsilon := \max_{w\in W}\min_{v\in L}\|w-v\|$;
  $\zeta_1 := \frac{3}{8}\sin\frac{\pi}{32}\ \varepsilon$; $\zeta_2 := \frac{7\sin\pi/32}{2(1-\sin\pi/32)}\ \varepsilon$;
  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.

Figure 2: Pseudo-code of the algorithm.

The pseudo-code of the algorithm is given in Figure 2. The pumping procedure is the same as in [12], with $\mathrm{K}_\omega^{\pi/32}(L)$ replaced by $\mathcal{C}_\omega^{W^{\zeta_2}}(L)$. Given a point p just inserted in L, of weight $\omega(p) = 0$, Pump(p) progressively increases the weight of p up 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 ω_p in the range of weights that maximized the minimum sliver measure in the star of p.

4.3 Theoretical guarantees

Assume that the input point set W is a δ-sample of some unknown manifold S. For any $i > 0$, let $L(i)$, $\omega(i)$, $\varepsilon(i)$, $\zeta_1(i)$, $\zeta_2(i)$ denote respectively L, ω, ε, ζ_1, ζ_2 at the end of the ith iteration of the main loop of the algorithm. Since $L(i)$ grows with i, $\varepsilon(i)$ is a decreasing function of i. Moreover, since the algorithm always inserts in L the point of W lying furthest away from L, it is easily seen that $L(i)$ is an $\varepsilon(i)$-sparse $\varepsilon(i)$-sample of W [21, Lemma 4.1].

THEOREM 4.1. *$\mathcal{C}_{\omega(i)}^{W^{\zeta_1(i)}}(L(i))$ coincides with $\mathcal{D}_{\omega(i)}^S(L(i))$ and is homeomorphic to S whenever $\varepsilon(i) \leq c_{\bar{\varrho}}$ rch(S) satisfies H1, which eventually happens during the course of the algorithm if δ is small enough.*

The proof of the theorem uses the following intermediate result, whose proof (omitted here) is roughly the same as in Section 8 of [12]:

LEMMA 4.2. *Whenever* $\delta \leq \varepsilon(i) \leq \frac{2\mathrm{rch}(S)}{7d-1} - \delta$, *the star of* $p(i)$ *in* $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ *contains no* $\bar{\varrho}$*-sliver.*

PROOF OF THEOREM 4.1. Let $i > 0$ be an iteration such that $\varepsilon(i) \leq c_{\bar{\varrho}}\ \mathrm{rch}(S)$ satisfies H1. Then, $\zeta_1(i) = \frac{2(1-\bar{\omega}^2)\sin\theta}{5}\ \varepsilon(i)$ satisfies H2, and $\zeta_2(i) = \frac{7\sin\pi/32}{2(1-\sin\pi/32)}\ \varepsilon(i)$ satisfies the hypothesis of Theorem 3.2. Since W is a δ-sample of S by assumption, and since $L(i)$ is an $\varepsilon(i)$-sparse $\varepsilon(i)$-sample of W, Theorems 3.1 and 3.2 imply that $\mathcal{D}_{\omega(i)}^{S}(L(i)) \subseteq \mathcal{C}_{\omega(i)}^{W^{\zeta_1(i)}}(L(i)) \subseteq \mathrm{K}_{\omega(i)}^{\pi/32}(L(i)) \subseteq \mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$. Let us show that $\mathrm{K}_{\omega(i)}^{\pi/32}(L(i))$ contains no $\bar{\varrho}$-sliver, which by Theorem 2.5 will give 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 because $\mathrm{K}_{\omega(i)}^{\pi/32}(L(i)) \subseteq \mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$. Since $\varepsilon(i)$ satisfies H1, we have $\delta \leq \varepsilon(i) \leq \frac{2\mathrm{rch}(S)}{7d-1} - \delta$. Therefore, Lemma 4.2 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 Lemma 4.2 was not yet satisfied.

Let $i_0 \leq i$ be the first iteration such that $\varepsilon(i_0) \leq \frac{2\mathrm{rch}(S)}{7d-1} - \delta$. For any iteration j between i_0 and i, we have $\varepsilon(j) \leq \varepsilon(i_0) \leq \frac{2\mathrm{rch}(S)}{7d-1} - \delta$ and $\varepsilon(j) \geq \varepsilon(i) \geq \delta$, hence Lemma 4.2 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 $\bar{\varrho}$-sliver, by Lemma 4.2. Therefore, the star of $p(j)$ in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ still contains no $\bar{\varrho}$-sliver.

Consider now an iteration $j \leq i_0 - 1$. Since $\varepsilon(j)$ is greater than $\frac{2\mathrm{rch}(S)}{7d-1} - \delta$, we cannot ensure that the star of $p(j)$ in $\mathcal{C}_{\omega(j)}^{W^{\zeta_2(j)}}(L(j))$ contains no $\bar{\varrho}$-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, for any such neighbor q, there is a point $c \in W^{\zeta_2(i)}$ that $\omega(i)$-witnesses edge $[p(j), q]$. According to Lemma 3.5 (i), $\|c - p(j)\|$ and $\|c - q\|$ are at most $c_1(1)\ \varepsilon(i)$. Hence, we have $\|p(j) - q\| \leq \|p(j) - c\| + \|c - q\| \leq 2c_1(1)\ \varepsilon(i)$. Now, recall that $L(i_0-1)$ is $\varepsilon(i_0-1)$-sparse, where $\varepsilon(i_0-1) > \frac{2\mathrm{rch}(S)}{7d-1} - \delta$, which by H1 is greater than $2c_1(1)\ \varepsilon(i)$. It follows that q cannot belong to $L(i_0-1)$, since $p(j)$ does. Hence, the star of q in $\mathcal{C}_{\omega(i)}^{W^{\zeta_2(i)}}(L(i))$ contains no $\bar{\varrho}$-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 $\bar{\varrho}$-sliver either. This concludes the proof of Theorem 4.1. □

4.4 Discussion

Theorem 4.1 guarantees the existence of a plateau in the diagram of the Betti numbers of $\mathcal{C}_{\omega}^{W^{\zeta_1}}(L)$, but it does not tell exactly where this plateau is located in the diagram, because δ and $\mathrm{rch}(S)$ are unknown. Nevertheless, it does give a guarantee on the length of the plateau, which, in view of H1, is of the order of $(a\ \mathrm{rch}(S) - b\ \delta)$, where a, b are two constants. Hence, for sufficiently small δ, the plateau is long enough to be detected by the user or some statistical method applied 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.

Observe that the structural results of Section 3 still hold[6] if we relax the assumption on W and allow its points to lie slightly off the manifold, at a Hausdorff distance of δ. This gives a deeper meaning to Theorem 4.1, 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 [21], enlarges greatly the range of applications of our method.

Note finally that our theoretical guarantees still hold if $\mathcal{C}_{\omega}^{W^{\zeta_1}}(L)$ 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 or the strong witness complex, with similar theoretical guarantees.

4.5 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.

4.5.1 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, which is a costly operation. Fortunately, as shown below, we can restrict ourselves to constructing 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,

[6] W^{ζ} still covers S when $\zeta \geq \delta$, and the proof that $\mathcal{C}_{\omega}^{W^{\zeta}}(L) \subseteq \mathrm{K}_{\omega}^{\theta}(L)$ holds the same, with an additional term δ in the bounds of Lemmas 3.3 through 3.6.

Lemma 4.3. *$\mathcal{C}(i)$ 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 $k \leq d$, Lemma 3.5 (i) ensures that the $k+1$ nearest landmarks of c in the weighted metric, $p_0, \cdots, p_k$, belong to $B(c, c_1(d)\varepsilon(i))$. Since $\|w - c\| \leq \zeta(i)$, $p_0, \cdots, p_k$ belong to $B(w, r(i))$, where $r(i) = \varepsilon(i)\left(c_1(d) + {}^{\zeta(i)}/_{\varepsilon(i)}\right)$. 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 $B\left(w, r(i) + {}^{\varepsilon(i)}/_2\right)$, thus their number is at most $\left(1 + {}^{2r(i)}/_{\varepsilon(i)}\right)^d = \left(1 + 2c_1(d) + {}^{2\zeta(i)}/_{\varepsilon(i)}\right)^d$, which is bounded by $\kappa(d)$ since $\zeta(i) \leq \zeta_2(i) < {}^{\varepsilon(i)}/_2$. Therefore, $p_0, \cdots, p_k$ belong to $\Lambda(w)$. As a result, in the weighted metric, the $k+1$ nearest neighbors of c in $\Lambda(w)$ are the same as its $k+1$ 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 $\mathcal{C}_{\omega(i)}^{W^{\zeta(i)}}(L(i))$. □

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 $O\left(\kappa(d)^{d+1}\right) = d^{O(d^2)}$ cells [18], 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 $O(\kappa(d)) = d^{O(d^2)}$.

4.5.2 Pumping procedure

As mentioned in Section 4.2, 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 $|W|d^{O(d^2)}$.

For the sake of efficiency, we need to reduce the size of the sequence of events to $d^{O(d^2)}$ before any further processing. To this end, we prune out most of the events, keeping only those involving simplices whose vertices belong to the $\kappa(d)$ nearest neighbors of p among L in the Euclidean metric. By the same argument as above, there are at most $d^{O(d^2)}$ such simplices. Nevertheless, the sequence of events may still be much larger, because a simplex may appear in or disappear from the star of p several times[7]. To bound the total number of events, 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 and simplified, 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 4.5.1) 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 precisely the number of events, bounded by $d^{O(d^2)}$. At each iteration, the update of $\mathcal{C}$ takes time at most $|W|d^{O(d^2)}$, 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 bounded by $|W|d^{O(d^2)}$.

The downside is that the pumping procedure works with a wrong sequence of events, since some events have been discarded during the precomputation. As a result, the pumping process may not remove all the $\bar{\varrho}$-slivers from the star of p. Nevertheless, whenever $\varepsilon(i) \leq c_{\bar{\varrho}}\ \mathrm{rch}(S)$ satisfies H1, no simplex is discarded during the precomputation phase, by an argument similar to the one used in the proof of Lemma 4.3. Furthermore, Lemma 4.2 still holds, with exactly the same proof – described in Section 8 of [12] and omitted here. Hence, Theorem 4.1 still applies.

4.5.3 Computation of p and ε

At each iteration of the main loop of the algorithm, the computation of $p = \mathrm{argmax}_{w\in W} \min_{v\in L} \|w - v\|$ 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.

4.6 Complexity of the algorithm

Theorem 4.4. *The space and time complexities of the algorithm are respectively $|W|d^{O(d^2)}$ and $|W|^2 d^{O(d^2)}$.*

Proof. As reported in Section 4.5.1, the size of the complex maintained by the algorithm never exceeds $|W|d^{O(d^2)}$, 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 4.5.3, and the pumping of p takes $|W|d^{O(d^2)}$ time, by Section 4.5.2. 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 4.5.1. Therefore, the time spent during 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 $|W|^2 d^{O(d^2)}$, since the number of iterations of the main loop of the algorithm is $|W|$. □

Note that the bounds of Theorem 4.4 do not take into account the (optional) computation of the Betti numbers of

[7]The cell of a k-simplex σ in the weighted Voronoi diagram of L of order $k+1$ evolves as the weight of p increases, and therefore it may intersect W^{ζ} several times during the process, since W^{ζ} is not convex.

$\mathcal{C}_{\omega(i)}^{W^{\zeta_1(i)}}(L(i))$ at each iteration of the algorithm, which can take a time cubic in the size of the complex [29].

5. CONCLUSION

We have proved that the structural properties of the witness complex on low-dimensional manifolds can be extended to manifolds of arbitrary dimensions and co-dimensions. To avoid pathological cases generated by slivers, we have assigned weights to the landmarks, so that, for carefully-chosen distributions of weights, the witness complex is included in the weighted restricted Delaunay triangulation, which is homeomorphic to the underlying manifold. Moreover, both complexes coincide if the set of witnesses is dilated by a ball of suitable radius. We have also introduced a conceptually simple reconstruction algorithm, based on the witness complex, which combines a farthest-point refinement scheme with a vertex pumping strategy. Using our structural results, we have proved the correctness of the algorithm when applied to sufficiently dense samplings of smooth manifolds. These mild assumptions on the input make the approach relevant on practical data, even though constructing arrangements of hyperplanes is not desirable in practice. We also believe the approach is generic enough to be applicable to related problems, such as manifold sampling and meshing.

Note that our structural results still hold in the somewhat more general setting where W is an *adaptive* δ-sample of S, in the sense of [1], and where L is an (ε, κ)-sample of W, in the sense of [12]. The proofs are roughly the same, with an additional twist which slightly degrades the constants.

6. ACKNOWLEDGEMENTS

The authors thank the anonymous referees for their insightful 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.

7. 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] D. Attali, H. Edelsbrunner, and Y. Mileyko. Weak witnesses for Delaunay triangulations of submanifolds. In *Proc. ACM Sympos. on Solid and Physical Modeling*, 2007. To appear.

[4] M. Belkin and P. Niyogi. Semi-supervised learning on Riemannian manifolds. *Machine Learning*, 56(1-3):209–239, 2004.

[5] 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.

[6] 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.

[7] F. Chazal, D. Cohen-Steiner, and A. Lieutier. A sampling theory for compact sets in Euclidean space. In *Proc. 22nd Annu. ACM Sympos. Comput. Geom.*, pages 319–326, 2006.

[8] 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.

[9] 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.

[10] 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.

[11] S.-W. Cheng, T. K. Dey, and E. A. Ramos. Manifold reconstruction from point samples. Journal version, in preparation.

[12] 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.

[13] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. In *Proc. 21st Annu. ACM Sympos. Comput. Geom.*, pages 263–271, 2005.

[14] V. de Silva. A weak definition of Delaunay triangulation. Manuscript. Preprint available at `http://math.stanford.edu/comptop/preprints/weak.pdf`, October 2003.

[15] V. de Silva and G. Carlsson. Topological estimation using witness complexes. In *Proc. Sympos. Point-Based Graphics*, pages 157–166, 2004.

[16] D. DeMers and G. Cottrell. Nonlinear dimensionality reduction. *Advances in Neural Information Processing Systems 5*, pages 580–587, 1993.

[17] T. K. Dey, J. Giesen, S. Goswami, and W. Zhao. Shape dimension and approximation from samples. *Discrete and Computational Geometry*, 29:419–434, 2003.

[18] H. Edelsbrunner. *Algorithms in Combinatorial Geometry*, volume 10 of *EATCS Monographs on Theoretical Computer Science*. Springer-Verlag, 1987.

[19] H. Edelsbrunner. *Geometry and Topology for Mesh Generation*. Cambridge, 2001.

[20] 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.

[21] L. G. Guibas and S. Y. Oudot. Reconstruction using witness complexes. In *Proc. 18th ACM-SIAM Sympos. on Discrete Algorithms*, pages 1076–1085, 2007.

[22] 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.

[23] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, to appear.

[24] S. Y. Oudot. On the topology of the restricted Delaunay triangulation and witness complex in higher dimensions. Manuscript. Preprint available at `http://geometry.stanford.edu/member/oudot/drafts/Delaunay_hd.pdf`, November 2006.

[25] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[26] 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.

[27] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[28] 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.

[29] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.