Reconstruction Using Witness Complexes

Leonidas J. Guibas* Steve Y. Oudot†

Abstract

We present a novel reconstruction algorithm that, given an input point set sampled from an object S, builds a one-parameter family of complexes that approximate S at different scales. At a high level, our method is very similar in spirit to Chew's surface meshing algorithm, with one notable difference: the restricted Delaunay triangulation is replaced by the *witness complex*, which makes our algorithm applicable in any metric space. To prove its correctness on curves and surfaces, we highlight the relationship between the witness complex and the restricted Delaunay triangulation in 2d and in 3d. Specifically, we prove that both complexes are equal in 2d and closely related in 3d, under some mild sampling assumptions.

1 Introduction

The problem of reconstructing a curve or a surface from scattered data points has received a lot of attention in the past. Although it is ill-posed in general, since infinitely many shapes with different topological types can interpolate a given point cloud, a number of provably good methods have been proposed. The common denominator of these methods is the assumption that the input point set is densely sampled from a sufficiently regular shape: this assumption makes the reconstruction problem well-posed, since all sufficiently regular shapes interpolating the point set have the same topological type and are close to one another geometrically. It suffices then to approximate any of these shapes to get the right answer. The notion of ε-sample, introduced by Amenta and Bern [1], provides a sound mathematical framework for this kind of approach, the corresponding set of reconstructible shapes being the class of manifolds with *positive reach* [21]. A number of provably-good algorithms are based on the ε-sampling theory – see [7] for a survey, and several extensions have been proposed to reconstruct manifolds in higher-dimensional spaces [12] or from noisy point cloud data [20]. The theory itself has been recently extended to a larger class of shapes, known as the class of Lipschitz manifolds [6]. In all these methods, the Delaunay triangulation of the input point set plays a prominent role since the final reconstruction is extracted from it.

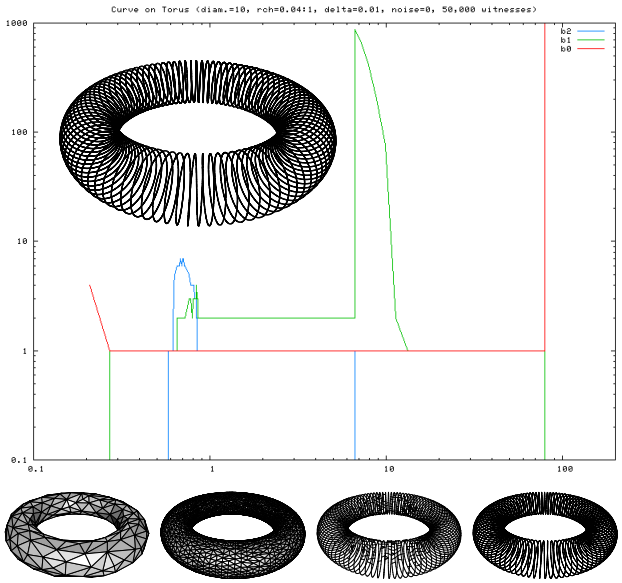

Figure 1: One-parameter family of complexes built by the algorithm, and their Betti numbers.

This approach to surface reconstruction is limited because it assumes implicitely that a point cloud should always represent a single class of shapes. Consider the example of a closed helical curve rolled around a torus in $\mathbb{R}^3$ – see Figure 1. Take a very dense uniform point sample of the curve: what does this point set represent, the curve or the torus? Although both objects are well-sampled according to Amenta and Bern's sampling theory, classical reconstruction methods always choose a single shape, here the curve or the torus, by restricting themselves either to a certain dimension or to a certain scale: for instance, the reconstruction method of [12] or the dimension detection algorithm of [19] will detect the curve but not the torus, since the point set is a sparse sample of the curve but not of the torus. Now, we claim that the result of the reconstruction should not be either the curve or the torus, but both of them. More generally, the result of the reconstruction should be a one-parameter family of complexes, whose

*Computer Science Department, Stanford University. guibas@cs.stanford.edu

†Computer Science Department, Stanford University. steve.oudot@stanford.edu

elements approximate the original shape at different scales, as illustrated in Figure 1. This point of view, inspired from recent results in Computational Topology [8, 11, 25], stands in sharp contrast with previous work on reconstruction and is echoed in the literature on non-linear dimensionality reduction [28, 30] and topological persistence [16, 31].

This paper presents a novel reconstruction algorithm that, given an input point set W sampled from an object S, builds a one-parameter family of complexes that approximate S at different scales. At a high level, the method is very similar to Chew's surface meshing algorithm [5, 14]: it constructs a subset L of W iteratively, while maintaining a subcomplex of the Delaunay triangulation of L. The one-parameter family of complexes obtained from this iterative process is the result of the algorithm. The difference with Chew's approach is that, instead of maintaining the restricted Delaunay triangulation of L, we maintain its witness complex $\mathcal{C}^W(L)$. The main advantages are that the underlying object S does not have to be known, and that the full-dimensional Delaunay triangulation $\mathcal{D}(L)$ does not have to be computed. Moreover, the algorithm can be used in any metric space, ultimately enabling new applications of the Delaunay-based reconstruction ideas.

The witness complex can be viewed as a weak version of the Delaunay triangulation, well-defined and computable in any metric space – see Section 2 below. As such, it has played an important role in the context of topological data analysis [18]. It was first introduced by de Silva [17], who proved that $\mathcal{C}^W(L)$ is a subcomplex of $\mathcal{D}(L)$ whenever the points of L lie in general position in a Euclidean space. Moreover, if the set W of witnesses spans the whole ambient space, then $\mathcal{C}^W(L)$ is equal to $\mathcal{D}(L)$. Now, the question is whether this property holds when the points of W are sampled from a subset S of the ambient space, such as for instance a submanifold: in [18], Carlsson and de Silva observed that $\mathcal{C}^W(L)$ is then closely related to the restricted Delaunay triangulation $\mathcal{D}_{|S}(L)$, and they conjectured that both objects should coincide under some sampling assumptions on W and L. We prove that this conjecture is valid for a curve in the plane, but not for a surface in 3d. In the latter case, we show how to relax the definition of the witness complex so that it contains $\mathcal{D}_{|S}(L)$, and then how to extract a subcomplex that approximates $\mathcal{D}_{|S}(L)$ (and hence S). This proves that our reconstruction algorithm is correct when applied to point samples of Lipschitz curves or surfaces.

We are only aware of one related result: in [3], Attali *et al.* show that $\mathcal{C}^W(L)$ and $\mathcal{D}_{|S}(L)$ coincide whenever the set W of witnesses spans an entire submanifold of $\mathbb{R}^n$ of dimension one or two. This result differs from ours in two ways: our set W can be finite, which makes our result more practical, yet in return our set L has to be sparse compared to W, for $\mathcal{C}^W(L)$ to contain $\mathcal{D}_{|S}(L)$. This sparseness condition is not an issue in practice, since the set L is constructed by the algorithm. Other noticeable differences are that our manifolds can have singularities, and that our point samples can be noisy. Our assumption on the input point set W is fairly mild, since it amounts to saying that the Hausdorff distance between W and S is sufficiently small. In particular, there is no sparseness condition on W, and the amplitude of the noise can be as large as the sampling density. This noise model, introduced in [10] and used in subsequent work on reconstruction [8, 11, 24, 25], is less restrictive than its predecessors [13, 20, 23], and it makes our algorithm more practical.

The paper is organized as follows. In Section 2, we recall several concepts that will be used later on. In Section 3, we present our structural results. Specifically, we prove that the restricted Delaunay triangulation and the witness complex are equal in 2d (Section 3.1) and closely related in 3d (Section 3.2), even when the data are noisy (Section 3.3). In Section 4, we introduce our reconstruction algorithm and present some experimental results.

2 Background and definitions

Let S be a subset of $\mathbb{R}^2$ (resp. $\mathbb{R}^3$), L a finite set of points in $\mathbb{R}^2$ (resp. $\mathbb{R}^3$), and ε a positive number.

Definition 2.1.
- *L is an ε-noisy sample of S if no point of L is farther than ε from S.*
- *L is an ε-sample of S if no point of S is farther than ε from L.*
- *L is ε-sparse if the pairwise distances between the points of L are at least ε.*

A 0-noisy sample is called a noise-free sample. When the first two conditions of the definition apply simultaneously, for a same ε, the Hausdorff distance between L and S is bounded by ε. We denote by $\mathcal{D}(L)$ the Delaunay triangulation of L.

Definition 2.2. *The Delaunay triangulation of L* restricted to *S, or $\mathcal{D}_{|S}(L)$ for short, is the subcomplex of $\mathcal{D}(L)$ made of the Delaunay faces whose dual Voronoi faces intersect S.*

Let W be another set of points in $\mathbb{R}^2$ (resp. $\mathbb{R}^3$), finite or infinite.

Definition 2.3.
- *Given a point $w \in W$ and a simplex $\sigma = [p_0, \cdots, p_l]$ with vertices in L, w* witnesses *σ if $p_0, \cdots, p_l$ belong to*

the $l+1$ nearest neighbors of w, that is, $\forall i \in \{0, \cdots, l\}$, $\forall q \in L \setminus \{p_0, \cdots, p_l\}$, $\mathrm{d}(w, p_i) \leq \mathrm{d}(w, q)$.
• *The witness complex of L relative to W, or $\mathcal{C}^W(L)$ for short, is the maximum abstract simplicial complex with vertices in L, whose faces are witnessed by points of W.*

The fact that $\mathcal{C}^W(L)$ is an abstract simplicial complex means that a simplex belongs to the complex only if all its faces do. By de Silva's result[1] [17], we have $\mathcal{C}^W(L) \subseteq \mathcal{D}(L)$ for any sets W and L such that the points of L lie in general position (which will be assumed implicitly in the rest of the paper). This implies that $\mathcal{C}^W(L)$ is always an embedded simplicial complex. In the sequel, L will be referred to as the set of landmarks, and W as the set of witnesses.

Lipschitz curves and surfaces. Boissonnat and Oudot [6] introduced a new framework for the analysis of Delaunay-based sampling algorithms. This framework relies on a quantity, called the *Lipschitz radius*, which plays a role equivalent to the *local feature size* of Amenta and Bern [1], on a broader class of shapes – the class of Lipschitz curves and surfaces.

Definition 2.4. *Let S be the boundary of a bounded open subset $\mathcal{O}$ of $\mathbb{R}^2$ (resp. $\mathbb{R}^3$). Given a point $p \in S$, the k-Lipschitz radius of S at p, or $\mathrm{lr}_k(p)$ for short, is the maximum radius r such that $\mathcal{O} \cap B(p, r)$ is the intersection of $B(p, r)$ with the hypograph of some k-Lipschitz univariate (resp. bivariate) function. We call $\mathrm{lr}_k(S)$ the infimum of lr_k over S.*

It is proved in [6] that $\mathrm{lr}_k(S) > 0$ whenever S is a k-Lipschitz curve in $\mathbb{R}^2$ or surface in $\mathbb{R}^3$. In such a case, one can attach to each point $p \in S$ a so-called *k-Lipschitz normal* $\mathbf{n}_k(p)$ and a so-called *k-Lipschitz support plane* $T_k(p)$, which play a role similar to the usual normal vector and tangent plane in the Lipschitz setting. The main result of [6] is the following:

Theorem 2.1. *Let S be a k-Lipschitz surface in $\mathbb{R}^3$ and $L \subset S$ a finite point set, such that:*
H1 *L is an ε-sample of S, with $\varepsilon < \frac{1}{7}\,\mathrm{lr}_k(S)$,*
H2 *the triangles of $\mathcal{D}_{|S}(L)$ have radius-edge ratios of at most ϱ, with $\varrho < \frac{\cos 2\theta}{2 \sin \theta}$, where $\theta = \arctan k$.*
Then, $\mathcal{D}_{|S}(L)$ is a 2-manifold isotopic to S, at Hausdorff distance at most ε from S, and whose oriented normals approximate the k-Lipschitz normals of S within an angle of $\arcsin(2\varrho \sin\theta)$.

Another useful result, proved in [26], is an equivalent of Proposition 13 of [4] for Lipschitz surfaces:

Lemma 2.1. *Let S be a k-Lipschitz surface in $\mathbb{R}^3$, with $k < 1$. Then, $\forall p \in S$, $\forall r \leq \mathrm{lr}_k(p)$, $S \cap B(p, r)$ is a topological disk.*

Similar results exist in the planar case, with similar yet simpler proofs (omitted here):

Lemma 2.2. *Let S be a k-Lipschitz curve in $\mathbb{R}^2$, with $k < 1$. Then, $\forall p \in S$, $\forall r \leq \mathrm{lr}_k(p)$, $S \cap B(p, r)$ is a topological arc. Moreover, the orthogonal projection of $S \cap B(p, r)$ onto $T_k(p)$ is a segment whose vertices are the orthogonal projections of the two endpoints of $S \cap B(p, r)$.*

Theorem 2.2. *If S is a k-Lipschitz curve in the plane, with $k < 1$, and if L is an ε-sample of S, with $\varepsilon < \mathrm{lr}_k(S)$, then $\mathcal{D}_{|S}(L)$ is a polygonal curve homeomorphic to S and at Hausdorff distance at most ε from S.*

3 Structural results

In this section, we highlight the relationship between the witness complex and the restricted Delaunay triangulation in 2d and in 3d. Let S be a k-Lipschitz manifold, *i.e.* either a k-Lipschitz curve in the plane (Section 3.1) or a k-Lipschitz surface in 3d (Section 3.2), for some constant $k \geq 0$. For convenience, we define $\theta = \arctan k \in [0, \pi/2[$. Let W be a δ-noisy δ-sample of S and $L \subset W$ an ε-sparse ε-sample of W. The constants δ and ε will be made explicit later on. Clearly, L is a $(\delta + \varepsilon)$-sample of S. We assume that no vertex of the Voronoi diagram of L lies on S, a condition that can always be satisfied by an infinitesimal perturbation of the points of L since S has non-zero codimension. In Sections 3.1 and 3.2, we assume further that L is noise-free ($L \subset S$). The case of a noisy set of landmarks is deferred to Section 3.3.

3.1 The planar case.

Theorem 3.1. *Assume that $\theta < \arcsin\frac{1}{8} \approx 7.2\,\mathrm{deg}$ and that $\delta < \min\{\frac{1-8\sin\theta}{12},\ \frac{3\cos\theta-2}{4(9\cos\theta+6)}\}\,\mathrm{lr}_k(S)$. If ε satisfies $\max\left\{\frac{12\sin\theta}{1-8\sin\theta},\ \frac{6}{3\cos\theta-2}\right\}\delta < \varepsilon < \frac{1}{8}\,\mathrm{lr}_k(S) - \frac{3}{2}\,\delta$, then $\mathcal{C}^W(L)$ coincides with $\mathcal{D}_{|S}(L)$.*

The lower bound on ε means that the set W of witnesses must be sufficiently dense[2] compared to the set L of landmarks, for the simplices of $\mathcal{D}_{|S}(L)$ to be witnessed. An illustration is given in Figure 2, which shows that $\mathcal{C}^W(L)$ contains $\mathcal{D}_{|S}(L)$ when L is sparse (left picture), whereas when $L = W$ ($\varepsilon = 0$), $\mathcal{C}^W(L)$ coincides with the nearest neighbor graph of L, which has nothing to do with $\mathcal{D}_{|S}(L)$ (right picture). The upper bound on ε

[1] In his paper, de Silva distinguishes between *weak witnesses* and *strong witnesses*. Here, all witnesses are weak.

[2] In particular, we have $\varepsilon > 6\delta$.

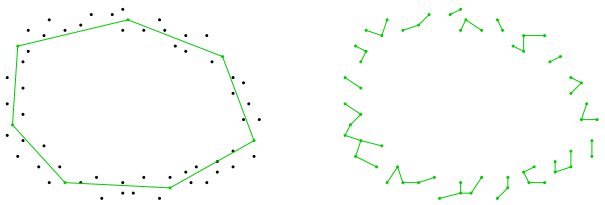

Figure 2: A set of witnesses sampling a smooth closed curve with noise, and two different subsets of landmarks (in green) together with their witness complexes.

ensures that the set L of landmarks is sufficiently dense, so that the nice properties of $\mathcal{D}_{|S}(L)$ stated in Section 2 hold, and that classical local arguments can be applied to show that $\mathcal{C}^W(L)$ is included in $\mathcal{D}_{|S}(L)$.

The proof of the theorem proceeds in two stages: first, we show that $\mathcal{D}_{|S}(L)$ is included in $\mathcal{C}^W(L)$ (Lemma 3.1), then we show that $\mathcal{C}^W(L)$ is included in $\mathcal{D}_{|S}(L)$ (Lemma 3.2).

LEMMA 3.1. *Assume that* $\theta < \frac{\pi}{6}$ *and that* $\delta < \min\{\frac{1-2\sin\theta}{2}, \frac{3\cos\theta-2}{2(3\cos\theta+4)}\}\ \mathrm{lr}_k(S)$. *If* ε *satisfies* $\max\{\frac{2\sin\theta}{1-2\sin\theta}, \frac{6}{3\cos\theta-2}\}\ \delta < \varepsilon < \frac{1}{2}\ \mathrm{lr}_k(S) - \delta$, *then* $\mathcal{D}_{|S}(L)$ *is included in* $\mathcal{C}^W(L)$.

Proof. Let $e = [u, v]$ be an edge of $\mathcal{D}_{|S}(L)$. By definition of $\mathcal{D}_{|S}(L)$, the dual Voronoi edge of e intersects S at some point c. Let $r = \mathrm{d}(c, u) = \mathrm{d}(c, v) = \mathrm{d}(c, L)$, which is at most $\varepsilon + \delta$ since L is a $(\delta + \varepsilon)$-sample of S. Since W is a δ-sample of S, there is some $w \in W$ at distance at most δ from c. Then, u and v are both included in $B(w, \varepsilon + 2\delta)$.

Let p be any point of $L \setminus \{u, v\}$. We will prove that $p \notin B(w, \varepsilon + 2\delta)$, which means that w witnesses e. Consider the portion of S that lies in $B(c, r)$. Since $r \leq \varepsilon + \delta < \mathrm{lr}_k(S)$, we know from Lemma 2.2 that $S \cap B(c, r)$ is a topological arc whose endpoints are u, v and whose orthogonal projection onto $T_k(c)$ is the line segment $[\bar{u}, \bar{v}]$, where $\bar{u}, \bar{v}$ are the orthogonal projections of u, v. If p does not belong to $B(c, \mathrm{lr}_k(S))$, then it does not belong to $B(w, \varepsilon + 2\delta)$ either, since $B(w, \varepsilon + 2\delta) \subseteq B(c, \varepsilon + 3\delta)$, which by hypothesis is included in $B(c, \mathrm{lr}_k(S))$. Otherwise, since $p \in L \setminus \{u, v\}$, $S \cap B(c, r)$ contains c but not p, therefore $[\bar{u}, \bar{v}]$ contains c but not the projection $\bar{p}$ of p, because the projection from $S \cap B(c, \mathrm{lr}_k(S))$ to $T_k(c)$ is one-to-one. As a consequence, $\mathrm{d}(c, p)$ is at least $\mathrm{d}(c, \bar{p}) \geq \mathrm{d}(c, \{\bar{u}, \bar{v}\}) + \mathrm{d}(\bar{p}, \{\bar{u}, \bar{v}\})$. Since L is ε-sparse, we have $\mathrm{d}(p, u) \geq \varepsilon$, $\mathrm{d}(p, v) \geq \varepsilon$, and $\mathrm{d}(c, u) = \mathrm{d}(c, v) \geq \frac{\mathrm{d}(u,v)}{2} \geq \frac{\varepsilon}{2}$. Moreover, since $S \cap B(c, \mathrm{lr}_k(S))$ is the graph of a k-Lipschitz univariate function defined over $T_k(c)$, we have $\mathrm{d}(\bar{p}, \bar{u}) \geq \mathrm{d}(p, u)\cos\theta$, $\mathrm{d}(\bar{p}, \bar{v}) \geq \mathrm{d}(p, v)\cos\theta$, $\mathrm{d}(c, \bar{u}) \geq \mathrm{d}(c, u)\cos\theta$, and $\mathrm{d}(c, \bar{v}) \geq \mathrm{d}(c, v)\cos\theta$. As a result,

$$\begin{aligned} \mathrm{d}(c, p) &\geq \mathrm{d}(c, \bar{p}) \geq \mathrm{d}(c, \{\bar{u}, \bar{v}\}) + \mathrm{d}(\bar{p}, \{\bar{u}, \bar{v}\}) \\ &\geq \tfrac{\varepsilon}{2}\cos\theta + \varepsilon\cos\theta = \tfrac{3}{2}\varepsilon\cos\theta. \end{aligned}$$

This expression is greater than $\varepsilon + 3\delta$ since $\varepsilon > \frac{6\delta}{3\cos\theta-2}$, by hypothesis. It follows that p is farther than $\varepsilon + 3\delta$ from c, and hence farther than $\varepsilon + 2\delta$ from w. Thus, w witnesses $[u, v]$. Similarly, every other edge of $\mathcal{D}_{|S}(L)$ is witnessed by some point of W. Since $L \subseteq W$, the vertices of $\mathcal{D}_{|S}(L)$ witness themselves, hence the 1-skeleton of $\mathcal{D}_{|S}(L)$ is included in $\mathcal{C}^W(L)$.

Finally, we assumed that no Voronoi vertex lies on S, which implies that $\mathcal{D}_{|S}(L)$ has no simplex of dimension two or more. Hence, $\mathcal{D}_{|S}(L)$ is equal to its 1-skeleton, which is included in $\mathcal{C}^W(L)$. This proves the lemma. □

The proof of the second lemma uses similar arguments and is therefore omitted.

LEMMA 3.2. *Assume that* $\theta < \arcsin\frac{1}{8}$ *and that* $\delta < \frac{1-8\sin\theta}{12}\ \mathrm{lr}_k(S)$. *If* ε *satisfies* $\frac{12\sin\theta}{1-8\sin\theta}\ \delta < \varepsilon < \frac{1}{8}\ \mathrm{lr}_k(S) - \frac{3}{2}\ \delta$, *then* $\mathcal{C}^W(L)$ *is included in* $\mathcal{D}_{|S}(L)$.

3.2 The 3d case. Unlike in the planar case, the witness complex and the restricted Delaunay triangulation of points sampled from a surface in 3d may not always coincide, even in situations where the sets of witnesses and landmarks satisfy strong sampling conditions. The reason is that, when a tetrahedron t of $\mathcal{D}(L)$ has almost cocircular vertices, the chance for any of the diagonal edges of t to be witnessed by a point of W is small – such a tetrahedron is called a *sliver* in the literature [29]. In order to give an intuition of this fact, let us assume for simplicity that the surface is flat and that the vertices of t are cocircular, as in Figure 3 (left). The order-2 Voronoi diagram of the vertices is then degenerate, the Voronoi cells of the diagonal edges being reduced to a single point p that lies at the intersection of the edges of the diagram. Therefore, any diagonal edge can be witnessed only by p, which means that the probability for any triangle of the quadrangle to be witnessed when W is finite is zero. As a result, holes appear with probability one in the witness complex, as illustrated in Figure 4 (left).

When the vertices of tetrahedron t are almost-cocircular, as in Figure 3 (right), the order-2 Voronoi cell of one diagonal edge is empty, while the cell of the other diagonal edge is arbitrarily small. Thus, the probability for any triangle of the quadrangle to be witnessed when W is finite is also arbitrarily small. Although it is always possible to perturbate the point set L so that the points are in general position, guaranteeing that the

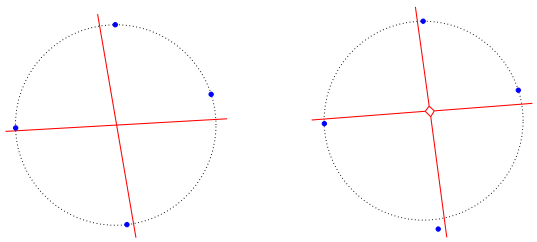

Figure 3: Order-2 Voronoi diagrams in the plane.

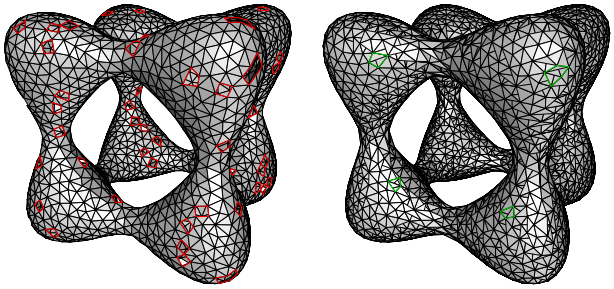

Figure 4: Witness complex and ν-witness complex.

order-2 Voronoi cells of the edges of $\mathcal{D}_{|S}(L)$ are sufficiently large requires large perturbations, which are not tractable in practice since the underlying surface S is unknown.

3.2.1 The ν-witness complex. Our approach for dealing with the above issue consists in relaxing the definition of the witness complex, so that the latter includes the restricted Delaunay triangulation. This requires to modify the concept of witness:

DEFINITION 3.1. *Given an integer m, a point $w \in W$ and a simplex $\sigma = [p_0, \cdots, p_l]$ with vertices in L, w m-witnesses σ if all the $\mathrm{d}(w, p_i)$ are among the m smallest values of the set $\{\mathrm{d}(w, q),\ q \in L\}$.*

Observe that, in the case where $m \leq l$, some vertices of σ must be equidistant to w for w to m-witness σ. In particular, if $m = 1$, then all the points of σ must be equidistant to w, which means that w is a strong witness of σ, or equivalently, that σ is a Delaunay simplex. If $m = 0$, then no point $w \in W$ can m-witness σ. In [18], the authors use m-witnesses only for edges. More generally, we use them for simplices of all dimensions:

DEFINITION 3.2. *Given a countable sequence ν of integers, the ν-witness complex of L relative to W, or $\mathcal{C}^W_\nu(L)$ for short, is the maximum abstract simplicial complex with vertices in L, such that each i-face is ν_i-witnessed by some point of W.*

Since the simplices of $\mathcal{C}^W_\nu(L)$ have their vertices in L, their dimension is at most $|L|-1$. Hence, in the sequence of integers, only ν_0 through $\nu_{|L|-1}$ are used. There is a natural relationship between $\mathcal{C}^W(L)$ and $\mathcal{C}^W_\nu(L)$: if $\nu_i \geq i+1\ \forall i$, then $\mathcal{C}^W_\nu(L)$ contains $\mathcal{C}^W(L)$; in contrast, if $\nu_i \leq i+1\ \forall i$, then $\mathcal{C}^W_\nu(L)$ is included in $\mathcal{C}^W(L)$; thus, $\mathcal{C}^W(L) = \mathcal{C}^W_\nu(L)$ whenever $\nu_i = i+1\ \forall i$. In addition, if $\nu_i = 0$ for some i, then the i-skeleton of $\mathcal{C}^W_\nu(L)$ is empty, since a simplex cannot be 0-witnessed, and therefore the dimension of $\mathcal{C}^W_\nu(L)$ is at most $i-1$.

THEOREM 3.2. *Assume that $\theta < \arccos(2\sin\pi/7) \approx 29.7\,\mathrm{deg}$ and that $\delta < \frac{\cos\theta - 2\sin\pi/7}{3\cos\theta + 2\sin\pi/7}\,\mathrm{lr}_k(S)$. If ε satisfies $\frac{8\sin\pi/7}{\cos\theta - 2\sin\pi/7}\,\delta < \varepsilon < \mathrm{lr}_k(S) - 3\delta$, then, for any sequence ν of integers such that $\nu_0 \geq 1$, $\nu_1 \geq 6$ and $\nu_2 \geq 6$, $\mathcal{D}_{|S}(L)$ is included in $\mathcal{C}^W_\nu(L)$.*

Proof. Since the vertices of $\mathcal{D}_{|S}(L)$ belong to L, which is included in W, they witness themselves and thus belong to $\mathcal{C}^W_\nu(L)$. In addition, since we assumed that no Voronoi vertex lies on S, $\mathcal{D}_{|S}(L)$ contains no simplex of dimension three or more.

Let σ be a simplex (edge or triangle) of $\mathcal{D}_{|S}(L)$, and let $B(c, r)$ be a Delaunay ball centered on S and circumscribing σ. Since W is a δ-sample of S, there is a point $w \in W$ at distance at most δ from c. Then, σ lies in the ball $B(w, r+\delta)$, which is included in $B(c, r+2\delta)$.

CLAIM 3.1. *The ball $B(c, r+2\delta)$ contains at most six points of L.*

Proof. Since L is a $(\delta+\varepsilon)$-sample of S, the radius r of the surface Delaunay ball is at most $\delta+\varepsilon$. Therefore, $r + 2\delta \leq \varepsilon + 3\delta$, which is less than $\mathrm{lr}_k(S)$ according to the hypothesis of Theorem 3.2. It follows that $S \cap B(c, r+2\delta)$ is the graph of a k-Lipschitz bivariate function defined over the plane $T_k(c)$.

Let $p_1, \cdots, p_l$ be the points of $L \cap B(c, r+2\delta)$. We call $\bar{p}_1, \cdots, \bar{p}_l$ their orthogonal projections onto $T_k(c)$. Since L is ε-sparse, the p_i are at least ε away from one another. And since they belong to $S \cap B(c, \mathrm{lr}_k(S))$, which is the graph of a k-Lipschitz bivariate function defined over the plane $T_k(c)$, their projections $\bar{p}_i$ are at least $\varepsilon\cos\theta$ away from one another. Moreover, since $B(c, r)$ is a Delaunay ball, the p_i are at least r away from c, hence the $\bar{p}_i$ are at least $r\cos\theta$ away from c.

The rest of the proof depends on whether $r \geq \varepsilon$ or $r < \varepsilon$. In fact, the overall ideas are the same, but some technical details differ.

• If $r \geq \varepsilon$, then, inside $T_k(c)$, c and the $\bar{p}_i$ are centers of pairwise-disjoint open disks of radius $\frac{\varepsilon}{2}\cos\theta$. Let $D_c, D_1, \cdots, D_l$ denote these disks. Since the p_i belong to $B(c, r+2\delta)$, the $\bar{p}_i$ belong to the disk $D(c, r+2\delta)$. Therefore, $D_c, D_1, \cdots, D_l$ form a congruent packing of the disk $D(c, r+2\delta+\frac{\varepsilon}{2}\cos\theta)$. Now, according to the

hypotheses of the claim, we have:

$$r+2\delta+\frac{\varepsilon}{2}\cos\theta \leq \varepsilon+3\delta+\frac{\varepsilon}{2}\cos\theta < \frac{\varepsilon}{2}\cos\theta\left(\frac{1}{\sin\pi/7}+1\right).$$

Hence, by a classical result on congruent packings of disks [22, 27], there are at most seven disks of radius $\frac{\varepsilon}{2}\cos\theta$ packed in $D(c, r+2\delta)$. The fact that D_c is one of them implies that $l \leq 6$, which proves the claim in the case where $r \geq \varepsilon$.

• If $r < \varepsilon$, then c and the $\bar{p}_i$ are centers of pairwise-disjoint open disks of radius $\frac{r}{2}\cos\theta$. Let $D_c, D_1, \cdots, D_l$ denote these disks. Since the p_i belong to $B(c, r+2\delta)$, $D_c, D_1, \cdots, D_l$ are included in $D(c, r+2\delta+\frac{r}{2}\cos\theta)$. Now, $B(c,r)$ is a Delaunay ball, hence its bounding sphere contains at least two points of L, which implies that $r \geq \varepsilon/2$. Moreover, the hypotheses of the claim state that $\delta < \frac{\varepsilon}{4}\left(\frac{\cos\theta}{2\sin\pi/7}-1\right)$, which is at most $\frac{r}{2}\left(\frac{\cos\theta}{2\sin\pi/7}-1\right)$. Therefore, $D_c, D_1, \cdots, D_l$ form a congruent packing of a disk of radius:

$$\begin{aligned} r+2\delta+\tfrac{r}{2}\cos\theta &< r+r\left(\tfrac{\cos\theta}{2\sin\pi/7}-1\right)+\tfrac{r}{2}\cos\theta \\ &= \tfrac{r}{2}\cos\theta\left(\tfrac{1}{\sin\pi/7}+1\right). \end{aligned}$$

It follows, by the same result as above on congruent packings of disks, that $l \leq 6$, which proves the claim in the case where $r < \varepsilon$. □

Claim 3.1 implies that the vertices of σ are among the six nearest neighbors of w. Since this is true for any edge or triangle of $\mathcal{D}_{|S}(L)$, and since the vertices of $\mathcal{D}_{|S}(L)$ belong to $\mathcal{C}^W_\nu(L)$, $\mathcal{C}^W_\nu(L)$ contains all the edges and triangles of $\mathcal{D}_{|S}(L)$. This ends the proof of the theorem. □

The next theorem guarantees that the simplices of $\mathcal{C}^W_\nu(L)$ are not too large as long as the ν_i remain bounded. It follows that the size of $\mathcal{C}^W_\nu(L)$ is linear in $|L|$, since L is sparse. This property can be generalized to higher dimensions, at the price of an exponential growth of the constant factor. This motivates the use of the witness complex instead of the Delaunay triangulation.

THEOREM 3.3. *Assume that δ, ε satisfy $\delta+\varepsilon < \frac{\cos\theta}{\sqrt{6}}\ \mathrm{lr}_k(S)$. Then, for any point $w \in W$, the distance between w and its sixth nearest neighbor among the points of L is at most $\delta+\left(\frac{\sqrt{6}}{\cos\theta}+1\right)(\delta+\varepsilon)$. As a consequence, for any sequence ν of integers such that $\nu_1 \leq 6$, the total number of simplices of $\mathcal{C}^W_\nu(L)$ is at most $2^{O\left((\delta+\varepsilon/\varepsilon\cos\theta)^3\right)}|L|$, which is linear with respect to $|L|$ as far as θ is fixed and δ is within a constant factor of ε.*

Proof. Let $w \in W$ and let $\bar{w}$ be a point of S closest to w. Since W is a δ-noisy sample of S, we have $\mathrm{d}(w,\bar{w}) \leq \delta$. We call $p_1, \cdots, p_l$ the points of L that lie in $B\left(\bar{w}, \left(\frac{\sqrt{6}}{\cos\theta}+1\right)(\delta+\varepsilon)\right)$, and $\bar{p}_1, \cdots, \bar{p}_l$ their orthogonal projections onto the plane $T_k(\bar{w})$. We will prove that $l \geq 6$.

Since L is a $(\delta+\varepsilon)$-sample of S, the balls $B_i = B(p_i, \delta+\varepsilon)$ cover $S \cap B(\bar{w}, {}^{(\delta+\varepsilon)\sqrt{6}}/_{\cos\theta})$ (observe that, among the balls of radius $\delta+\varepsilon$ centered at the points of L, only the B_i intersect $B(\bar{w}, {}^{(\delta+\varepsilon)\sqrt{6}}/_{\cos\theta})$). It follows that, inside $T_k(\bar{w})$, the disks $D_i = D(\bar{p}_i, \delta+\varepsilon)$ cover the orthogonal projection of $S \cap B(\bar{w}, {}^{(\delta+\varepsilon)\sqrt{6}}/_{\cos\theta})$. Now, according to the hypothesis of the lemma, we have ${}^{(\delta+\varepsilon)\sqrt{6}}/_{\cos\theta} < \mathrm{lr}_k(S)$. Thus, by Lemma 2.1, $S \cap B(\bar{w}, {}^{(\delta+\varepsilon)\sqrt{6}}/_{\cos\theta})$ is a topological disk whose orthogonal projection onto $T_k(\bar{w})$ contains the projection D of the intersection of $B(\bar{w}, {}^{(\delta+\varepsilon)\sqrt{6}}/_{\cos\theta})$ with the cone of apex $\bar{w}$, of axis aligned with $\mathbf{n}_k(\bar{w})$ and of half-angle $\frac{\pi}{2}-\theta$. Therefore, the D_i cover D, which is a disk of center $\bar{w}$ and radius $(\delta+\varepsilon)\sqrt{6}$. Thus, the number of disks D_i is at least

$$\frac{\mathrm{Area}(D)}{\mathrm{Area}(D_i)} = \frac{6\pi(\delta+\varepsilon)^2}{\pi(\delta+\varepsilon)^2} = 6.$$

It follows that the number of points of L that lie in $B\left(\bar{w}, \left(\frac{\sqrt{6}}{\cos\theta}+1\right)(\delta+\varepsilon)\right)$ is at least 6. As a result, the distance from w to its sixth nearest landmark is at most $\mathrm{d}(w,\bar{w}) + \left(\frac{\sqrt{6}}{\cos\theta}+1\right)(\delta+\varepsilon) \leq \delta + \left(\frac{\sqrt{6}}{\cos\theta}+1\right)(\delta+\varepsilon)$.

Let us now bound the size of $\mathcal{C}^W_\nu(L)$. Let p be a point of L. From the above paragraph we deduce that the edges of $\mathcal{C}^W_\nu(L)$ incident to p are included in balls of radii at most $\delta+\left(\frac{\sqrt{6}}{\cos\theta}+1\right)(\delta+\varepsilon)$. Hence, all edges belong to a common ball of center p and radius $r \leq 2\delta + 2\left(\frac{\sqrt{6}}{\cos\theta}+1\right)(\delta+\varepsilon)$, which is equal to $2\varepsilon\left(1+\frac{\sqrt{6}}{\cos\theta}+\left(\frac{\sqrt{6}}{\cos\theta}+2\right)\frac{\delta}{\varepsilon}\right)$. The neighboring vertices $q_1, \cdots, q_l$ of p in $\mathcal{C}^W_\nu(L)$ belong to $B(p,r)$ as well. Now, since the points of L are farther than ε from one another, the q_i are centers of pairwise-disjoint balls of radius ${}^{\varepsilon}/_{2}$, hence their number l is at most $\frac{\mathrm{Vol}(B(p,r))}{\mathrm{Vol}(B(p,\varepsilon/2))} \leq 64\left(1+\frac{\sqrt{6}}{\cos\theta}+\left(\frac{\sqrt{6}}{\cos\theta}+2\right)\frac{\delta}{\varepsilon}\right)^3 = O\left(\frac{1}{\cos^3\theta}\left(1+\frac{\delta}{\varepsilon}\right)^3\right)$. Since every simplex of $\mathcal{C}^W_\nu(L)$ incident to p is uniquely defined as a subset of $\{q_1, \cdots, q_l\}$, the number of simplices of $\mathcal{C}^W_\nu(L)$ incident to p is at most 2^l, which gives the result. □

3.2.2 Manifold extraction. It follows from Theorem 3.2 that $\mathcal{C}^W_\nu(L)$ contains $\mathcal{D}_{|S}(L)$, but Figure 4

(right)[3] shows that $\mathcal{C}_\nu^W(L)$ is not restricted to $\mathcal{D}_{|S}(L)$ and contains additional simplices that are small enough to be ν-witnessed. Nevertheless, it is possible to extract from $\mathcal{C}_\nu^W(L)$ a simplicial surface $\hat{S}$ isotopic to S and at Hausdorff distance $O(\varepsilon+\delta)$ of S. The extraction procedure takes a number ϱ as parameter and proceeds as follows:

1. Since the goal is to extract a 2-manifold, only the 2-skeleton of $\mathcal{C}_\nu^W(L)$ is considered. Since it may not be an embedded complex, we intersect it with $\mathcal{D}(L)$. The result is a pure 2-dimensional subcomplex $\mathcal{C}$ of $\mathcal{D}(L)$.
2. To guarantee that the output simplicial surface has no skinny triangle, we delete from $\mathcal{C}$ all the triangles of radius-edge ratio greater than ϱ.
3. We greedily remove from $\mathcal{C}$ all the triangles incident to *sharp edges*. An edge is sharp if all its incident triangles in $\mathcal{C}$ lie in a small wedge of angle at most $\pi/2$. This definition applies in particular to edges that are incident to one single triangle.
4. By a depth-first walk in the dual graph of the remaining part of $\mathcal{C}$, we extract the outer boundary of $\mathcal{C}$.

Observe that steps 3. and 4. correspond to the manifold extraction procedure of [1, 2]. As argued in these papers, the outcome is a simplicial complex $\hat{S}$ whose dihedral angles are greater than $\pi/2$. Moreover, thanks to step 2., the radius-edge ratios of the facets of $\hat{S}$ are at most ϱ. However, two issues arise: first, by greedily removing non-Delaunay triangles or triangles with sharp edges or large radius-edge ratios from $\mathcal{C}_\nu^W(L)$, steps 1. through 3. might end up with an empty complex $\mathcal{C}$. As a result, $\hat{S}$ may be empty. Second, the outer boundary of $\mathcal{C}$ might not be an embedded surface since it may contain multiple vertices or edges. By proceeding with a depth-first search on the dual graph of $\mathcal{C}$, step 4. duplicates multiple vertices and edges, so that the resulting complex $\hat{S}$ is a simplicial surface whose immersion in $\mathbb{R}^3$ coincides with the outer boundary of $\mathcal{C}$.

THEOREM 3.4. *Let* $\varrho = 1 + \frac{1-2\sin\pi/7}{8\sin\pi/7} \approx 1.038$. *Assume that* $\theta < \arctan\frac{\sqrt{3}}{1+4\varrho} \approx 18.6\deg$. *If* δ, ε *satisfy* $\frac{8\sin\pi/7}{\cos\theta-2\sin\pi/7}\,\delta < \varepsilon < \frac{\cos^3\theta}{(\cos\theta+\sqrt{6})(4+3\cos^2\theta)\varrho\sqrt{3}}\,\mathrm{lr}_k(S) - \frac{2\cos\theta+\sqrt{6}}{\cos\theta}\,\delta$, *then, for any sequence* ν *of integers such that* $\nu_0 \geq 1$ *and* $\nu_1 = \nu_2 = 6$, *the simplicial complex* $\hat{S}$ *extracted from* $\mathcal{C}_\nu^W(L)$ *with parameter* ϱ *is an embedded surface isotopic to* S *and at Hausdorff distance at most* $\left(\delta + \left(\frac{\sqrt{6}}{\cos\theta}+1\right)(\delta+\varepsilon)\right)\frac{\varrho\sqrt{3}}{\cos^2\theta}$ *from* S.

[3] Thanks to a bug in Geomview, we can see some hidden triangles in the vicinity of slivers, such as in green areas.

The proof of the theorem (omitted here) is roughly the same as in Section 5 of [6]. Here is a short overview:

- First, we show that the triangles of $\mathcal{D}_{|S}(L)$ are not skinny and make large dihedral angles. This fact, combined with Theorem 3.2, implies that complex $\mathcal{C}$ contains $\mathcal{D}_{|S}(L)$ after step 3. above. We deduce that $\hat{S}$ is not empty, since $\mathcal{D}_{|S}(L)$ is a manifold without boundary, by Theorem 2.1.
- Second, we use Theorem 3.3 (ii) of [6] to show that $\hat{S}$ is a Lipschitz surface, which implies in particular that it is an embedded surface.
- Third, we use Proposition 6.4 of [6] to bound the Hausdorff distance $\mathrm{d}_\mathcal{H}(\hat{S}, S)$ between $\hat{S}$ and S. We show that $\mathrm{d}_\mathcal{H}(\hat{S}, S)$ is small compared to the Lipschitz radii of $\hat{S}$ and S.
- Finally, we apply Theorem 6.2 of [9] to show that $\hat{S}$ and S are isotopic.

3.3 Dealing with noisy data. Our previous results hold provided that the set L of landmarks lies on the curve or surface S. Theorem 3.5 below shows that this condition is not mandatory, under some restrictions on the densities of W and L. Let $\lambda_0 \approx 0.078$ denote the smallest positive root of the polynomial $64\lambda^6 + 832\lambda^5 + 1008\lambda^4 - 160\lambda^3 - 4\lambda^2 - 12\lambda + 1$, and for any $\theta \in [0, \frac{\pi}{2}]$, let $\lambda(\theta)$ be the smallest positive root of $16(4\sin^2\theta - 1)\lambda^6 + 32\lambda^5 - 12(2+3\sin^2\theta)\lambda^4 + 8\lambda^3 + (4\sin^2\theta + 63)\lambda^2 + 64\lambda - 16$.

THEOREM 3.5. *Let* S *be a* k*-Lipschitz surface in* $\mathbb{R}^3$, *and let* L *be a* δ*-noisy* ε*-sparse* $(\delta+\varepsilon)$*-sample of* S. *Assume that* δ, ε *satisfy the following conditions, where* $\theta = \arctan k$:

$$\begin{cases} \delta < \min\left\{\frac{1}{4}, \frac{\sqrt{2}-4\sin\theta}{2(\sqrt{2}+4\sin\theta)}, \frac{\cos(2\theta)-2\sin\theta}{2(\cos(2\theta)+2\sin\theta)}, \lambda_0, \lambda(\theta)\right\}\,\varepsilon \\ \delta < \min\left\{\frac{1}{14} - \frac{\varepsilon}{2\mathrm{lr}_k(S)}, \frac{1}{6} - \frac{7\varepsilon}{12\mathrm{lr}_k(S)}\right\}\,\mathrm{lr}_k(S) \end{cases}$$

Then, there exists a k'*-Lipschitz surface* S', *passing through the points of* L, *isotopic to* S, *and at Hausdorff distance at most* $\varepsilon + 3\delta$ *from* S, *such that:*

$$\begin{cases} k' = \tan\left(\arcsin\left(\frac{4((1+2\sin\theta)\frac{\delta}{\varepsilon}+\sin\theta)}{2\sin\left(\arcsin\frac{1-2\frac{\delta}{\varepsilon}}{2(1+2\frac{\delta}{\varepsilon})} - 2\arcsin\frac{2\frac{\delta}{\varepsilon}}{1-2\frac{\delta}{\varepsilon}}\right)}\right)\right) \\ \mathrm{lr}_{k'}(S') \geq \mathrm{lr}_k(S) - (2\varepsilon+7\delta) > \frac{1}{2}\,\mathrm{lr}_k(S) \end{cases}$$

If the set W of witnesses is a δ-noisy δ-sample of a k-Lipschitz surface S, for some sufficiently small δ (as compared to $\mathrm{lr}_k(S)$), then Theorem 3.5 ensures that there exists an interval of values of ε such that any ε-sparse ε-sample L of W lies on a k'-Lipschitz surface S', with $k' = O(k + \delta/\varepsilon)$ and $\mathrm{lr}_k(S') = \Omega(\mathrm{lr}_k(S))$. The structural results of Section 3.2 apply then to S', W, L. And since S' is isotopic to S and close to it for the Hausdorff distance, these results hold for S, W, L as

well, with slightly worse constants. There exists also a version of Theorem 3.5 for Lipschitz curves, which can be combined with the structural results of Section 3.1.

The proof of the theorem (omitted here) consists in building an isotopy $\phi : [0,1] \times S \to \mathbb{R}^3$ such that $S' = \phi(1,S)$ is a k'-Lipschitz surface passing through the points of L, with $k' = O(k + \delta/\varepsilon)$ and $\mathrm{lr}_{k'}(S') = \Omega(\mathrm{lr}_k(S))$. Intuitively, since the points of L lie ε away from one another, with ε large compared to the amplitude δ of the noise, the surface S can be *snapped* onto the points of L without changing its normals too much. This can be easily seen on simple examples, such as for instance when S is the x-axis in $\mathbb{R}^2$ (in this case, the snapped curve is the polygonal chain connecting the points of L in the order of their abscissae).

4 Application to reconstruction

4.1 Algorithm. The algorithm works in any arbitrary metric space. It takes as input a finite point set W, identified as the set of witnesses, and an optional countable sequence ν of integers, whose default value is $\nu_i = i+1\ \forall i$ (which corresponds to $\mathcal{C}^W_\nu = \mathcal{C}^W$). The algorithm constructs a set $L \subseteq W$ of landmarks iteratively, starting with $L = \emptyset$, and in the meantime it maintains $\mathcal{C}^W_\nu(L)$. At each iteration, the witness lying furthest away from L is inserted in L, and $\mathcal{C}^W_\nu(L)$ is updated as described below. The process stops when $L = W$. The output of the algorithm is either the one-parameter family of complexes $\mathcal{C}^W_\nu(L)$ built throughout the process, or simply the diagram of their Betti numbers, computed on the fly using the persistence algorithm[4] of [31]. 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 landmarks (the points of W have been sorted according to their order of insertion in L) and to rebuild its witness complex.

4.2 Update of $\mathcal{C}^W_\nu(L)$. Our strategy to update $\mathcal{C}^W_\nu(L)$ relies on the following observation: when a witness p is inserted in L, every simplex that appears in $\mathcal{C}^W_\nu(L)$ is incident to p, whereas every simplex that disappears from $\mathcal{C}^W_\nu(L)$ has a face that is no longer ν-witnessed. It follows that the ν-witnesses of all these simplices belong to the reverse κ-nearest landmarks[5] of p, where $\kappa = \min\{|L|,\ \max_i \nu_i\}$. Hence, $\mathcal{C}^W_\nu(L)$ can be updated by performing a reverse κ-nearest landmarks search on p, and then, for each witness w in the outcome, a κ-nearest landmarks search on w, to determine which simplices to insert or delete from $\mathcal{C}^W_\nu(L)$. A number of dynamic data structures exist that can perform these queries efficiently – see [15] for a survey. Note however that κ can be as large as $|W|$, a case in which the above queries take linear time. Moreover, when $\nu_i \geq |L|$ $\forall i$, $\mathcal{C}^W_\nu(L)$ coincides with the complete hypergraph of L and hence has an exponential size. Nevertheless, in Euclidean space $\mathbb{R}^n$, κ is more likely to be a constant depending (exponentially) on n, which reduces the size of $\mathcal{C}^W_\nu(L)$ to $O(|L|)$, by Theorem 3.3. The total time spent to maintain $\mathcal{C}^W_\nu(L)$ is then $O(|W|^2)$, since any newly-created landmark has $\Theta(|W|)$ reverse κ-nearest landmarks (these can be detected naively by an exhaustive search on the set W), each of which witnesses a constant number of simplices (these can be found by maintaining the lists of κ-nearest landmarks of the witnesses). We conjecture that it should be possible to reduce the time complexity to $O(|W| \log |W|)$, under some sparseness condition on W.

4.3 2d and 3d cases. We take $\nu = (1,2,3)$ in 2d and $\nu = (1,6,6,4)$ in 3d, as prescribed by the theory. Moreover, we replace $\mathcal{C}^W_\nu(L)$ by its intersection with $\mathcal{D}(L)$. This makes sense because, $\mathcal{D}_{|S}(L)$ being a subset of $\mathcal{D}(L)$, Theorems 3.1 and 3.4 hold the same if $\mathcal{C}^W_\nu(L)$ is replaced by $\mathcal{C}^W_\nu(L) \cap \mathcal{D}(L)$. The advantage of the latter complex is that it can be stored as a subcomplex of $\mathcal{D}(L)$, which allows to speed-up the (reverse) κ-nearest landmarks queries in practice. Another thing in 3d is that we also maintain the subcomplex $\hat{S}$ extracted by the procedure of Section 3.2.2.

4.4 Theoretical guarantees. Let $L(i)$ denote the set L at the end of iteration i of the algorithm. Calling $\varepsilon(i)$ the minimum number such that $L(i)$ is an $\varepsilon(i)$-sample of W, we have the following:

LEMMA 4.1. *At any iteration i, $L(i)$ is an $\varepsilon(i)$-sparse $\varepsilon(i)$-sample of W.*

Proof. At each iteration $j \leq i$ of the algorithm, the witness $p(j)$ farthest from $L(j-1)$ is inserted in $L(j-1)$. Right before this insertion, $L(j-1)$ is an $\varepsilon(j-1)$-sample of W. This means that the distance from $p(j)$ to $L(j-1)$ is $\varepsilon(j-1)$. Since L keeps growing during the process, we have $\varepsilon(j) \geq \varepsilon(j+1)$, $\forall j \leq i$. Thus, each point inserted in L before or at iteration i is at least $\varepsilon(i-1)$ away from L at the time of its insertion. This implies that $L(i)$ is $\varepsilon(i-1)$-sparse, and therefore also $\varepsilon(i)$-sparse. □

From this lemma and from Theorems 3.1, 3.4 and 3.5, we deduce that, if W is a δ-noisy δ-sample of some k-Lipschitz manifold S in $\mathbb{R}^2$ (resp. $\mathbb{R}^3$), for sufficiently small values of δ and k, then there exists an interval of

[4] The filtration used in [31] is rebuilt at each iteration, since some simplices are deleted from our complex $\mathcal{C}^W_\nu(L)$.

[5] These are the witnesses that have p among their κ-nearest landmarks.

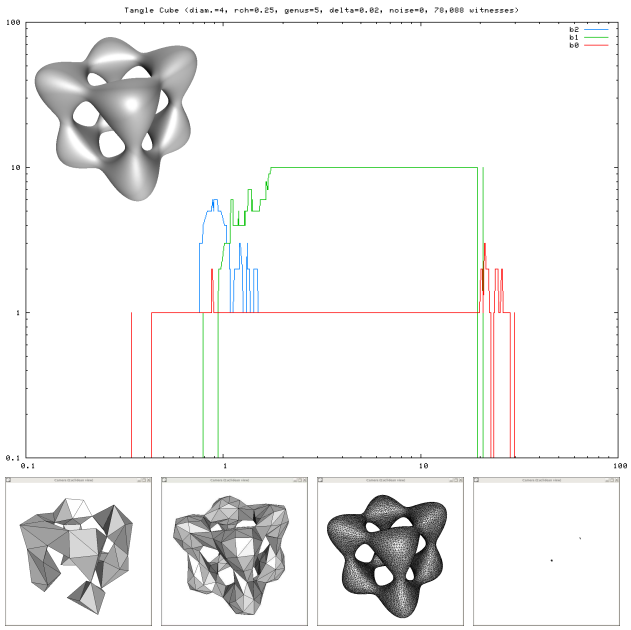

Figure 5: Diagram of Betti numbers of $\hat{S}$ for the Tanglecube point set.

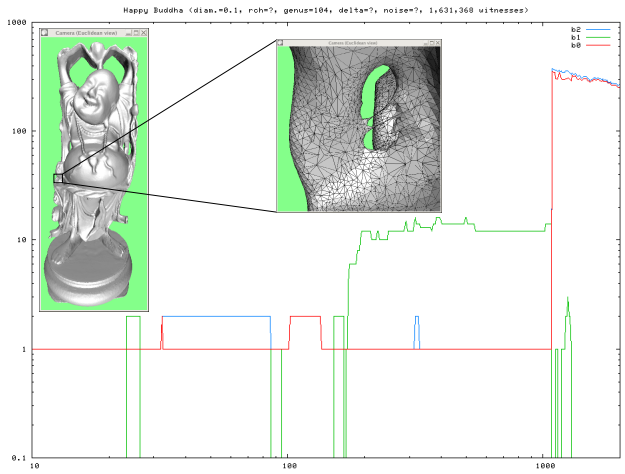

Figure 6: Diagram of Betti numbers of $\hat{S}$ for the Happy Buddha point set.

values of $\varepsilon(i)$ such that $\mathcal{C}^W(L(i))$ (resp. $\hat{S}$) is a correct approximation of S. Therefore, the topological type of $\mathcal{C}^W(L)$ (resp. $\hat{S}$) stabilizes for some time during the course of the algorithm, and so do topological invariants such as homology groups. The duration of the stabilized phase depends on the ratio $\delta/\mathrm{lr}_k(S)$.

These guarantees hold provided that the input point set W is a δ-noisy δ-sample of the underlying manifold S, for some sufficiently small value of δ. This is equivalent to saying that the Hausdorff distance between W and S is bounded by δ. In particular, there is no sparseness condition on W, and the amplitude of the noise can be as large as the sampling density.

4.5 Experimental results and discussion. Figure 5 shows three main phases in the evolution of the Betti numbers of $\hat{S}$ (the x-axis represents $1/\varepsilon(i)$ on a logarithmic scale): first, their behaviour is erratic and the topology of $\hat{S}$ keeps changing, because $\varepsilon(i)$ is too large compared to $\mathrm{lr}_k(S)$; then, the Betti numbers stabilize and a plateau appears in the diagram, as predicted by the theory; finally, $\varepsilon(i)$ becomes too small compared to δ, and holes appear in the complex, which doom the manifold extraction process. Some snapshots of $\hat{S}$ during the three phases are given at the bottom of Figure 5. The plateau in Figure 6 is smaller, due to the fairly high value of δ in the data set. Observe that the topological type of $\hat{S}$ on the plateau (genus = 6) coincides with the one of the original physical object, but not with the one of the model in the repository (genus = 104). The reason is that, due to noise and holes in the data, classical reconstruction techniques fail because they look at the point cloud at one scale only. In contrast, our method provides reconstructions at various scales and generates plateaus whenever the topological type is stable enough to be plausibly that of the underlying object. For instance, in Figure 6, a new plateau appears right before the data structure becomes unstable: this plateau indicates that a new handle (shown on the model on the left) has been detected. Another example is given in Figure 1, where the diagram has two well-separated plateaus corresponding to two plausible reconstructions: a torus, and a simple closed curve drawn on that torus. To handle the change in dimension (shown at the bottom row of the figure), we maintained $\mathcal{C}^W_\nu(L)$ for both $\nu = (1, 2, 3, 4)$ and $\nu = (1, 6, 6, 4)$ simultaneously, and determined at each step the complex to keep according to their Betti number β_2.

5 Conclusion

We have introduced a new reconstruction method, based on the witness complex. This method uses inter-sample distances alone and can therefore be applied in any metric space. Moreover, it stands in sharp contrast with previous work in the area, since it is multiscale and gives some insights on the various plausible topological types of the original object. We believe that this approach to manifold reconstruction is highly practical and has a number of potential applications, such as for instance topological noise removal or mesh compression. As a side product, in order to prove our algorithm correct on Lipschitz curves and surfaces, we have highlighted

the relationship between the witness complex and the restricted Delaunay triangulation in 2d and 3d.

Acknowledgements. The authors acknowledge the support of DARPA grant 32905, NSF grants FRG-0354543 and ITR 0205671, and NIH grant GM-072970. They wish to thank V. de Silva for helpful discussions, and A. Zomorodian for providing them with his implementation of the persistence algorithm.

References

[1] N. Amenta and M. Bern. Surface reconstruction by Voronoi filtering. *Discrete Comput. Geom.*, 22(4):481–504, 1999.

[2] 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.

[3] D. Attali, H. Edelsbrunner, and Y. Mileyko. Weak witnesses for Delaunay triangulations of submanifolds. Manuscript, in preparation.

[4] J.-D. Boissonnat and F. Cazals. Natural neighbor coordinates of points on a surface. *Computational Geometry: Theory and Applications*, 19:155–173, 2001.

[5] J.-D. Boissonnat and S. Oudot. Provably good sampling and meshing of surfaces. *Graphical Models*, 67(5):405–451, September 2005.

[6] 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.

[7] 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.

[8] 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.

[9] F. Chazal and A. Lieutier. Weak feature size and persistent homology: Computing homology of solids in $\mathbb{R}^n$ from noisy data samples. Technical Report 378, Institut de Mathématiques de Bourgogne, 2004. Partially published in *Proc. 21st Annu. ACM Sympos. on Comput. Geom.*, pages 255–262, 2005.

[10] F. Chazal and A. Lieutier. The λ-medial axis. *Graphical Models*, 67(4):304–331, July 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] 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] S.-W. Cheng, S. Funke, M. Golin, P. Kumar, S.-H. Poon, and E. Ramos. Curve reconstruction from noisy samples. *Computational Geometry: Theory and Applications*, 31(1-2):63–100, 2005.

[14] L. P. Chew. Guaranteed-quality mesh generation for curved surfaces. In *Proc. 9th Annu. ACM Sympos. Comput. Geom.*, pages 274–280, 1993.

[15] K. L. Clarkson. Nearest-neighbor searching and metric space dimensions. In G. Shakhnarovich, T. Darrell, and P. Indyk, editors, *Nearest-Neighbor Methods for Learning and Vision: Theory and Practice*, pages 15–59. MIT Press, 2006.

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

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

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

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

[20] T. K. Dey and S. Goswami. Provable surface reconstruction from noisy samples. *Computational Geometry: Theory and Application*, 35(1-2):124–141, 2006.

[21] H. Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 93:418–491, 1959.

[22] R. L. Graham. Sets of points with given minimum separation (solution to Problem E1921). *Amer. Math. Monthly*, 75:192–193, 1968.

[23] R. Kolluri. Provably good moving least squares. In *Proc. 16th ACM-SIAM Symposium on Discrete Algorithms*, pages 1008–1017, 2005.

[24] B. Mederos, N. Amenta, L. Velho, and L. H. de Figueiredo. Surface reconstruction for noisy point clouds. In *Proc. 3rd Sympos. on Geometry Processing*, pages 53–62, 2005.

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

[26] S. Oudot, L. Rineau, and M. Yvinec. Meshing volumes with curved boundaries. Invited to the special issue of *Engineering With Computers* on IMR 2005.

[27] U. Pirl. Der Mindestabstand von n in der Einheitskreisscheibe gelegenen Punkten. *Math. Nachr.*, 40:111–124, 1969.

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

[29] J. R. Shewchuk. What is a good linear element? Interpolation, conditioning, and quality measures. In *Proc. 11th Internat. Meshing Roundtable*, pages 115–126, 2002.

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

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