Towards Persistence-Based Reconstruction in Euclidean Spaces

Frédéric Chazal
INRIA, Geometrica group
4, rue Jacques Monod
91893 ORSAY Cedex, France
frederic.chazal@inria.fr

Steve Y. Oudot
INRIA, Geometrica group
4, rue Jacques Monod
91893 ORSAY Cedex, France
steve.oudot@inria.fr

ABSTRACT

Manifold reconstruction has been extensively studied for the last decade or so, especially in two and three dimensions. Recent advances in higher dimensions have led to new methods to reconstruct large classes of compact subsets of $\mathbb{R}^d$. However, the complexities of these methods scale up exponentially with d, making them impractical in medium or high dimensions, even on data sets of low intrinsic dimensionality.

In this paper, we introduce a novel approach that stands in-between classical reconstruction and topological estimation, and whose complexity scales up with the intrinsic dimension of the data. Our algorithm combines two paradigms: greedy refinement, and topological persistence. Given a point cloud in $\mathbb{R}^d$, we build a set of landmarks iteratively, while maintaining a nested pair of abstract complexes, whose images in $\mathbb{R}^d$ lie close to the data, and whose persistent homology eventually coincides with the homology of the underlying shape. When the data points are densely sampled from a smooth m-submanifold X of $\mathbb{R}^d$, our method retrieves the homology of X in time at most $c(m)n^5$, where n is the size of the input and $c(m)$ is a constant depending solely on m.

To prove the correctness of our algorithm, we investigate on Čech, Rips, and witness complex filtrations in Euclidean spaces. More precisely, we show how previous results on unions of balls can be transposed to Čech filtrations, and from there to Rips and witness complex filtrations. Finally, investigating further on witness complexes, we quantify a conjecture of Carlsson and de Silva, which states that witness complex filtrations should have cleaner persistence barcodes than Čech or Rips filtrations, at least on smooth submanifolds of Euclidean spaces.

Categories and Subject Descriptors: F.2.2 [Analysis of Algorithms and Problem Complexity]: Non-numerical Algorithms and Problems — Geometrical problems and computations, Computations on discrete structures.

General Terms: Algorithms, Theory.

Keywords: Čech complex, Rips complex, witness complex, filtration, persistent homology, manifold reconstruction.

SCG'08, June 9–11, 2008, College Park, Maryland, USA.

1. INTRODUCTION

The problem of reconstructing unknown structures from finite collections of data samples is ubiquitous in the Sciences, where it has many different variants, depending on the nature of the data and on the targeted application. In the last decade or so, the computational geometry community has gained a lot of interest in manifold reconstruction, where the goal is to reconstruct submanifolds of Euclidean spaces from point clouds. Efficient solutions have been proposed in dimensions two and three, based on the use of the Delaunay triangulation — see [8] for a survey. Recently, significant steps were made towards a full understanding of the potential and limitations of the Delaunay-based approach in arbitrary dimensions [14, 30]. In parallel, new sampling theories were developped, such as the critical point theory for distance functions [9], which provides sufficient conditions for the topology of a shape $X \subset \mathbb{R}^d$ to be captured by the offsets of a point cloud L lying at small Hausdorff distance. These advances lay the foundations of a new theoretical framework for the reconstruction of smooth submanifolds [11, 29], and more generally of large classes of compact subsets of $\mathbb{R}^d$ [9, 10, 12]. Combined with the introduction of more lightweight data structures, such as the *witness complex* [16], they have led to new provably-good algorithms [6] whose complexities can be orders of magnitude below the one of the classical Delaunay-based approach. For instance, on a data set with n points in $\mathbb{R}^d$, the algorithm of [6] runs in time $2^{O(d^2)}n^2$, whereas the size of the Delaunay triangulation can be of the order of $n^{\lceil \frac{d}{2} \rceil}$. Unfortunately, $2^{O(d^2)}n^2$ remains too large for these methods to be practical, even when the data points lie on a low-dimensional submanifold.

A weaker yet similarly difficult version of the reconstruction paradigm is topological estimation, where the goal is to infer the topological invariants of X from an input point cloud L. This problem has received a lot of attention in the recent years, and it finds applications in a number of areas of Science, *e.g.* sensor networks [18], statistical analysis [7], or dynamical systems [28, 31]. A classical approach consists in building a nested sequence of spaces $\mathcal{K}^0 \subseteq \mathcal{K}^1 \subseteq \cdots \subseteq \mathcal{K}^m$, and in studying the persistence of homology classes throughout this sequence. It has been independently proved in [12] and [15] that the persistent homology of the sequence defined by the α-offsets of a point cloud L coincides with the homology of the underlying shape X under mild sampling conditions. Specifically, if the Hausdorff distance between L and X is less than ε, for some small enough ε, then, for all sufficiently small $\alpha \geq \varepsilon$, the canonical inclusion map $L^\alpha \hookrightarrow L^{\alpha+2\varepsilon}$ induces homomorphisms between homology

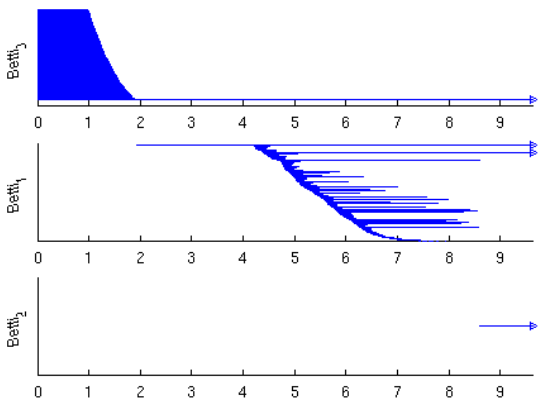

Figure 1: Topological analysis of a synthetic 1000-dimensional data set. The 50,000 data points have been sampled uniformly at random from a helical curve drawn on the 2d Clifford torus, embedded into $\mathbb{R}^{1000}$ via a quadratic mapping. The image shows the persistence barcode of the Rips filtration built over a carefully-chosen subset of 2000 landmarks.

groups, whose images are isomorphic to the homology groups of X. Combined with the structure theorem of [33], which states that the persistent homology of the sequence $\{L^\alpha\}_{\alpha\geq 0}$ is fully described by a finite set of intervals, called a *persistence barcode* or a *persistence diagram* — see Figure 1, the above result means that the homology of X can be deduced from this barcode, simply by removing the intervals of length less than 2ε, which are therefore viewed as topological noise.

From an algorithmic point of view, the persistent homology of a nested sequence of simplicial complexes (called a *filtration*) can be efficiently computed using the persistence algorithm [21, 33]. Among the many filtrations that can be built on top of a point set L, the α-shape enables to reliably recover the homology of the underlying space X, since it is known to be a deformation retract of L^α [20]. However, this property is useless in high dimensions, since computing the α-shape requires to build the full-dimensional Delaunay triangulation. It is therefore appealing to consider other filtrations that are easy to compute in arbitrary dimensions, such as the Rips and witness complex filtrations. In this paper, we produce an equivalent of the result of [12, 15] for these filtrations, and more generally for any filtration that is intertwined with the Čech filtration. Recall that, for all $\alpha > 0$, the Čech complex $\mathcal{C}^\alpha(L)$ is the nerve of the union of the open balls of same radius α about the points of L. It is known to be homotopy equivalent to L^α. However, combining this fact with the result of [12, 15] is not enough to prove that the persistent homology of $\mathcal{C}^\alpha(L) \hookrightarrow \mathcal{C}^{\alpha+2\varepsilon}(L)$ coincides with the homology of X, because it is unclear whether the homotopy equivalences $\mathcal{C}^\alpha(L) \to L^\alpha$ and $\mathcal{C}^{\alpha+2\varepsilon}(L) \to L^{\alpha+2\varepsilon}$ commute with the canonical inclusions $\mathcal{C}^\alpha(L) \hookrightarrow \mathcal{C}^{\alpha+2\varepsilon}(L)$ and $L^\alpha \hookrightarrow L^{\alpha+2\varepsilon}$. In the paper, we show that there exist homotopy equivalences that commute with canonical inclusions, at least at homology and homotopy levels. This enables us to extend the result of [12, 15] to the Čech filtration, and from there to the Rips and witness complex filtrations.

Another common concern in topological data analysis is the size of the vertex set on top of which the filtration is built. Indeed, in practical situations where the input data set W samples the underlying shape very finely, it makes sense to build the filtration on top of a small subset L of *landmarks* to avoid a waste of computational resources. However, downsampling the vertex set may result in a significant degradation in the quality of the persistence barcode. This is true in particular with the Čech and Rips filtrations, whose barcodes can have topological noise of amplitude depending directly on the density of the landmark set L. The introduction of the witness complex filtration appeared as an elengant way of solving this issue [17]. The witness complex of L relative to W, or $\mathcal{C}_W(L)$ for short, can be viewed as a relaxed version of the Delaunay triangulation of L, in which the points of $W \setminus L$ are used to drive the construction of the complex [16]. Due to its special nature, which takes advantage of the points of $W \setminus L$, the witness complex filtration is likely to give persistence barcodes whose topological noise depends on the density of W rather than on the one of L, as conjectured in [17]. We prove that this statement is only true to some extent, namely: whenever the points of W are sufficiently densely sampled from some smooth submanifold of $\mathbb{R}^d$, the topological noise in the barcode can be arbitrarily small compared to the density of L. Nevertheless, it cannot depend solely on the density of W. This shows that the witness complex filtration does provide cleaner persistence barcodes, though maybe not as clean as expected.

Taking advantage of the above results, we propose a novel approach to reconstruction that stands in-between the classical reconstruction and topological estimation paradigms. Our algorithm is a variant of the method of [6, 26] that combines greedy refinement and topological persistence. Given an input point cloud W, the algorithm builds a subset L of landmarks iteratively, and in the meantime it maintains a nested pair of simplicial complexes (Rips or witness complexes) and computes its persistent Betti numbers. The outcome of the algorithm is the diagram showing the evolution of these persistent Betti numbers. Using this diagram, a user or software agent can determine a relevant scale at which to process the data. It is then easy to rebuild the corresponding set of landmarks, as well as its nested pair of complexes. Although our method does not really compute an embedded complex that is close to X topologically and geometrically, it comes with theoretical guarantees, it is easily implementable, and it has reasonable complexity. Indeed, in the case where the input point cloud W is densely sampled from a smooth submanifold X of $\mathbb{R}^d$, we show that the complexity of our algorithm is bounded by $c(m)n^5$, where $c(m)$ is a quantity depending solely on the intrinsic dimension m of X, and n is the size of W. To the best of our knowledge, this is the first provably-good topological estimation or reconstruction method whose complexity scales up with the intrinsic dimension of the data. When X is a more general compact set in $\mathbb{R}^d$, our complexity bound becomes $c(d)n^5$.

The paper is organized as follows: after introducing the Čech, Rips, and witness complex filtrations in Section 2, we prove our structural results in Sections 3 and 4, focusing on compact subsets of $\mathbb{R}^d$ in Section 3, and on the particular case of smooth submanifolds in Section 4. Finally, we present our algorithm and its analysis in Section 5.

2. VARIOUS RELATED FILTRATIONS

The definitions, results and proofs of this section hold in any arbitrary metric space. However, for the sake of consistency with the rest of the paper, we state them in the

particular case of $\mathbb{R}^d$, endowed with the Euclidean norm $\|.\|$. Although our bounds can be proved to be tight in the general metric case, it is possible to work out somewhat tighter bounds in the Euclidean case, at the price of a loss of simplicity in the statements.

For any compact set $X \subset \mathbb{R}^d$, we call $\mathrm{diam}(X)$ the diameter of X, and $\mathrm{diam}_{\mathrm{CC}}(X)$ the *component-wise diameter* of X, defined by: $\mathrm{diam}_{\mathrm{CC}}(X) = \inf_i \mathrm{diam}(X_i)$, where the X_i are the path-connected components of X. Finally, given two compact sets X, Y in $\mathbb{R}^d$, we call $\mathrm{d}_{\mathcal{H}}(X,Y)$ their Hausdorff distance. Given a finite set L of points of $\mathbb{R}^d$ and a positive number α, we call L^α the union of the open balls of radius α centered at the points of L: $L^\alpha = \bigcup_{x\in L} B(x,\alpha)$. We also denote by $\{L^\alpha\}$ the open cover of L^α formed by the open balls of radius α centered at the points of L. The *Čech complex* of L of parameter α, or $\mathcal{C}^\alpha(L)$ for short, is the *nerve* of this cover, *i.e.* it is the abstract simplicial complex whose vertex set is L, and such that, for all $k \in \mathbb{N}$ and all $x_0, \cdots, x_k \in L$, $[x_0, \cdots, x_k]$ is a k-simplex of $\mathcal{C}^\alpha(L)$ if and only if $B(x_0,\alpha) \cap \cdots \cap B(x_k,\alpha)$ is non-empty. The *(Vietoris-)Rips complex* of L of parameter α, or $\mathcal{R}^\alpha(L)$ for short, is the abstract simplicial complex whose k-simplices correspond to unordered $(k+1)$-tuples of points of L which are pairwise within distance α of one another. The Rips complex is closely related to the Čech complex, according to the following standard result of computational topology:

LEMMA 2.1. *For all finite set $L \subset \mathbb{R}^d$ and all $\alpha > 0$, we have: $\mathcal{C}^{\frac{\alpha}{2}}(L) \subseteq \mathcal{R}^\alpha(L) \subseteq \mathcal{C}^\alpha(L)$.*

From now on, L is referred to as the landmark set. Let W be another (possibly infinite) subset of $\mathbb{R}^d$, referred to as the witness set. Let also $\alpha \in [0,\infty)$. Given a point $w \in W$ and a k-simplex σ with vertices in L, w is an *α-witness* of σ (or, equivalently, *w α-witnesses σ*) if the vertices of σ lie within distance $(\mathrm{d}_k(w)+\alpha)$ of w, where $\mathrm{d}_k(w)$ denotes the distance between w and its $(k+1)$th nearest landmark. The *α-witness complex* of L relative to W, or $\mathcal{C}_W^\alpha(L)$ for short, is the maximum abstract simplicial complex, with vertices in L, whose faces are α-witnessed by points of W.

When $\alpha = 0$, the α-witness complex coincides with the standard witness complex $\mathcal{C}_W(L)$, introduced in [16]. The α-witness complex is also closely related to the Čech complex, though the relationship is a bit more subtle than in the case of the Rips complex:

LEMMA 2.2. *Let $L, W \subseteq \mathbb{R}^d$ be such that L is finite. If every point of L lies within distance l of W, then for all $\alpha > l$ we have: $\mathcal{C}^{\frac{\alpha-l}{2}}(L) \subseteq \mathcal{C}_W^\alpha(L)$. In addition, if the distance from any point of W to its second nearest neighbor in L is at most l', then for all $\alpha > 0$ we have: $\mathcal{C}_W^\alpha(L) \subseteq \mathcal{C}^{2(\alpha+l')}(L)$.*

PROOF. Let $[x_0, \cdots, x_k]$ be a k-simplex of $\mathcal{C}^{\frac{\alpha-l}{2}}(L)$. This means that $\bigcap_{i=0}^k B(x_i, \frac{\alpha-l}{2}) \neq \emptyset$, and as a result, that $\|x_0 - x_i\| \leq \alpha - l$ for all $i = 0, \cdots, k$. Let w be a point of W closest to x_0. We have $\|w - x_0\| \leq l$, therefore $x_0, \cdots, x_k$ lie within distance α of w. Since the distances from w to its nearest points of L are non-negative, w is an α-witness of $[x_0, \cdots, x_k]$ and of all its faces. As a result, $[x_0, \cdots, x_k]$ is a simplex of $\mathcal{C}_W^\alpha(L)$. Consider now a k-simplex $[x_0, \cdots, x_k]$ of $\mathcal{C}_W^\alpha(L)$. If $k = 0$, then the simplex is a vertex $[x_0]$, and therefore it belongs to $\mathcal{C}^{\alpha'}(L)$ for all $\alpha' > 0$. Assume now that $k \geq 1$. Edges $[x_0, x_1], \cdots, [x_0, x_k]$ belong also to $\mathcal{C}_W^\alpha(L)$, hence they are α-witnessed by points of W. Let $w_i \in W$ be an α-witness of $[x_0, x_i]$. Distances $\|w_i - x_0\|$ and $\|w_i - x_i\|$ are bounded from above by $\mathrm{d}_2(w_i) + \alpha$, where $\mathrm{d}_2(w_i)$ is the distance from w_i to its second nearest point of L, which by assumption is at most l'. It follows that $\|x_0 - x_i\| \leq \|x_0 - w_i\| + \|w_i - x_i\| \leq 2\alpha + 2l'$. Since this is true for all $i = 0, \cdots, k$, we conclude that x_0 belongs to the intersection $\bigcap_{i=0}^k B(x_i, 2(\alpha + l'))$. As a result, $[x_0, \cdots, x_k]$ is a simplex of $\mathcal{C}^{2(\alpha+l')}(L)$. □

COROLLARY 2.3. *Let X be a compact subset of $\mathbb{R}^d$, and let $L \subseteq W \subseteq \mathbb{R}^d$ be such that L is finite. Assume that $\mathrm{d}_{\mathcal{H}}(X,W) \leq \delta$ and $\mathrm{d}_{\mathcal{H}}(W,L) \leq \varepsilon$, with $\varepsilon + \delta < \frac{1}{4}\mathrm{diam}_{\mathrm{CC}}(X)$. Then, for all $\alpha > \varepsilon$, $\mathcal{C}^{\frac{\alpha-\varepsilon}{2}}(L) \subseteq \mathcal{C}_W^\alpha(L) \subseteq \mathcal{C}^{2\alpha+6(\varepsilon+\delta)}(L)$. In particular, if $\delta \leq \varepsilon < \frac{1}{8}\mathrm{diam}_{\mathrm{CC}}(X)$ then, for all $\alpha \geq 2\varepsilon$ we have: $\mathcal{C}^{\frac{\alpha}{4}}(L) \subseteq \mathcal{C}_W^\alpha(L) \subseteq \mathcal{C}^{8\alpha}(L)$.*

PROOF. Since $\mathrm{d}_{\mathcal{H}}(W,L) \leq \varepsilon$, every point of L lies within distance ε of W. As a result, the first inclusion of Lemma 2.2 holds with $l = \varepsilon$, that is: $\mathcal{C}^{\frac{\alpha-\varepsilon}{2}}(L) \subseteq \mathcal{C}_W^\alpha(L)$. Now, for every point $w \in W$, there is a point $p \in L$ such that $\|w - p\| \leq \varepsilon$. Moreover, there is a point $x \in X$ such that $\|w - x\| \leq \delta$, since we assumed that $\mathrm{d}_{\mathcal{H}}(X,W) \leq \delta$. Let X_x be the path-connected component of X that contains x. Take an arbitrary value $\lambda \in \big(0, \frac{1}{2}\mathrm{diam}_{\mathrm{CC}}(X) - 2(\varepsilon+\delta)\big)$, and consider the open ball $B(w, 2(\varepsilon+\delta)+\lambda)$. This ball clearly intersects X_x, since it contains x. Furthermore, X_x is not contained entirely in the ball, since otherwise we would have: $\mathrm{diam}_{\mathrm{CC}}(X) \leq \mathrm{diam}(X_x) \leq 4(\varepsilon+\delta) + 2\lambda$, hereby contradicting the fact that $\lambda < \frac{1}{2}\mathrm{diam}_{\mathrm{CC}}(X) - 2(\varepsilon+\delta)$. Hence, there is a point $y \in X$ lying on the bounding sphere of $B(w, 2(\varepsilon+\delta)+\lambda)$. Let $q \in L$ be closest to y. We have $\|y - q\| \leq \varepsilon + \delta$, since our hypothesis implies that $\mathrm{d}_{\mathcal{H}}(X,L) \leq \mathrm{d}_{\mathcal{H}}(X,W) + \mathrm{d}_{\mathcal{H}}(W,L) \leq \delta + \varepsilon$. It follows then from the triangle inequality that $\|p - q\| \geq \|w - y\| - \|w - p\| - \|y - q\| \geq 2(\varepsilon+\delta) + \lambda - (\varepsilon+\delta) - (\varepsilon+\delta) = \lambda > 0$. Thus, q is different from p, and therefore the ball $B(w, 3(\varepsilon+\delta)+\lambda)$ contains at least two points of L. Since this is true for arbitrarily small values of λ, the distance from w to its second nearest neighbor in L is at most $3(\varepsilon+\delta)$. It follows that the second inclusion of Lemma 2.2 holds with $l' = 3(\varepsilon+\delta)$, that is: $\mathcal{C}_W^\alpha(L) \subseteq \mathcal{C}^{2(\alpha+3(\varepsilon+\delta))}(L)$. □

As mentioned at the head of the section, slightly tighter bounds can be worked out using specific properties of Euclidean spaces. For the case of the Rips complex, this was done by de Silva and Ghrist [18, 24]. Their approach can be combined with ours in the case of the witness complex.

3. PROPERTIES OF FILTRATIONS IN $\mathbb{R}^D$

This section uses classical concepts of algebraic topology: homotopy equivalences, deformation retractions, homology groups, homotopy groups, etc. We refer the reader to [27] for a good introduction to these concepts. Throughout the paper, we use singular homology with coefficients in an arbitrary field – omitted in the notations. Our results also hold at homotopy level, as detailed in Section 3.2.2.

Given a compact set $X \subset \mathbb{R}^d$, we denote by d_X the *distance function* defined by $\mathrm{d}_X(x) = \inf\{\|x - y\| : y \in X\}$. Although d_X is not differentiable, it is possible to define a notion of critical point for distance functions and we denote by $\mathrm{wfs}(X)$ the *weak feature size* of X, defined as the smallest positive critical value of d_X [10]. We do not explicitly use the notion of critical value in the following, but only its

relationship with the topology of the *offsets* $X^\alpha = \{x \in \mathbb{R}^d : \mathrm{d}_X(x) \leq \alpha\}$, stressed in the following result from [25]:

LEMMA 3.1. *If $0 < \alpha < \alpha'$ are such that there is no critical value of d_X in the closed interval $[\alpha, \alpha']$, then $X^{\alpha'}$ deformation retracts onto X^α.*

In particular, the hypothesis of the lemma is satisfied when $0 < \alpha_1 < \alpha_2 < \mathrm{wfs}(X)$. Therefore, all the offsets of X have the same homotopy type in the interval $(0, \mathrm{wfs}(X))$. In the sequel, we repeatedly make use of the following standard result of linear algebra:

LEMMA 3.2. *Given a sequence $A \to B \to C \to D \to E \to F$ of homomorphisms between finite-dimensional vector spaces, if* $\mathrm{rank}\,(A \to F) = \mathrm{rank}\,(C \to D)$*, then this quantity also equals the rank of $B \to E$. Similarly, if $A \to B \to C \to E \to F$ is a sequence of homomorphisms such that* $\mathrm{rank}\,(A \to F) = \dim C$*, then* $\mathrm{rank}\,(B \to E) = \dim C$.

3.1 Čech filtration

Since the Čech complex is the nerve of a union of balls, its homotopy type is closely related to the one of this union. We will use the following extension of Theorem 4.7 of [12]:

LEMMA 3.3. *Let X be a compact set and L a finite set in $\mathbb{R}^d$, such that $\mathrm{d}_{\mathcal{H}}(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{4}\,\mathrm{wfs}(X)$. Then, for all $\alpha, \alpha' \in [\varepsilon, \mathrm{wfs}(X) - \varepsilon]$ such that $\alpha' - \alpha \geq 2\varepsilon$, and for all $\lambda \in (0, \mathrm{wfs}(X))$, we have: $\forall k \in \mathbb{N}$, $H_k(X^\lambda) \cong \mathrm{im}\; i_*$, where $i_* : H_k(L^\alpha) \to H_k(L^{\alpha'})$ is the homomorphism between homology groups induced by the canonical inclusion $i : L^\alpha \hookrightarrow L^{\alpha'}$. Given an arbitrary point $x_0 \in X$, the same conclusion holds for homotopy groups with base-point x_0.*

PROOF. We can assume without loss of generality that $\varepsilon < \alpha < \alpha' - 2\varepsilon < \mathrm{wfs}(X) - 3\varepsilon$, since otherwise we can replace ε by any $\varepsilon' \in (d_H(X, L), \varepsilon)$. From the hypothesis we deduce the following sequence of inclusions:

$$X^{\alpha-\varepsilon} \hookrightarrow L^\alpha \hookrightarrow X^{\alpha+\varepsilon} \hookrightarrow L^{\alpha'} \hookrightarrow X^{\alpha'+\varepsilon} \tag{1}$$

By the Isotopy Lemma 3.1, for all $0 < \beta < \beta' < \mathrm{wfs}(X)$, the canonical inclusion $X^\beta \hookrightarrow X^{\beta'}$ is a homotopy equivalence. As a consequence, Eq. (1) induces a sequence of homomorphisms between homology groups, such that all homomorphisms between homology groups of $X^{\alpha-\varepsilon}, X^{\alpha+\varepsilon}, X^{\alpha'+\varepsilon}$ are isomorphisms. It follows then from Lemma 3.2 that $i_* : H_k(L^\alpha) \to H_k(L^{\alpha'})$ has same rank as these isomorphisms. Now, this rank is equal to the dimension of $H_k(X^\lambda)$, since the X^β are homotopy equivalent to X^λ for all $0 < \beta < \mathrm{wfs}(X)$. It follows that $\mathrm{im}\; i_* \cong H_k(X^\lambda)$, since our ring of coefficients is a field.

The case of homotopy groups is a little trickier, since the above rank argument cannot be used. However, we can use the same proof as in Theorem 4.7 of [12] to conclude. □

Observe that Lemma 3.3 does not guarantee the retrieval of the homology of X. Instead, it deals with sufficiently small offsets of X, which are homotopy equivalent to one another but possibly not to X itself [12, Fig. 4]. In the special case where X is a smooth submanifold of $\mathbb{R}^d$ however, X^λ and X are homotopy equivalent, and therefore the theorem guarantees the retrieval of the homology of X.

Consider now the Čech complex $\mathcal{C}^\alpha(L)$, for any value $\alpha > 0$. By definition, $\mathcal{C}^\alpha(L)$ is the nerve of the open cover $\{L^\alpha\}$ of L^α. Since the elements of $\{L^\alpha\}$ are convex, they form a *good* open cover of L^α, *i.e.* their intersections are either empty or contractible. It follows from the *nerve theorem* [27, Corollary 4G.3] that L^α and its nerve $\mathcal{C}^\alpha(L)$ are homotopy equivalent. We thus get the following diagram, where horizontal arrows are canonical inclusions and vertical arrows are homotopy equivalences provided by the nerve theorem:

$$\begin{array}{ccc} L^\alpha & \hookrightarrow & L^{\alpha'} \\ \uparrow & & \uparrow \\ \mathcal{C}^\alpha(L) & \hookrightarrow & \mathcal{C}^{\alpha'}(L) \end{array} \tag{2}$$

Unfortunately, the nerve theorem does not guarantee that this diagram commutes. However, standard arguments of algebraic topology imply the following result, where $\mathcal{NU}$ (resp. $\mathcal{NU}'$) stands for the nerve of the open cover $\mathcal{U}$ (resp. $\mathcal{U}'$):

LEMMA 3.4. *Let $X \subseteq X'$ be two paracompact spaces, and let $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ and $\mathcal{U}' = \{U'_\alpha\}_{\alpha \in A}$ be good open covers of X and X' respectively, based on a same finite parameter set A, such that $U_\alpha \subseteq U'_\alpha$ for all $\alpha \in A$. Then, there exist homotopy equivalences $\mathcal{NU} \to X$ and $\mathcal{NU}' \to X'$ that commute with the canonical inclusions $X \hookrightarrow X'$ and $\mathcal{NU} \hookrightarrow \mathcal{NU}'$ at homology and homotopy levels.*

Letting $X = L^\alpha$, $X' = L^{\alpha'}$, $\mathcal{U} = \{L^\alpha\}$, and $\mathcal{U}' = \{L^{\alpha'}\}$, we get from Lemma 3.4 that there exist homotopy equivalences $\mathcal{C}^\alpha(L) \to L^\alpha$ and $\mathcal{C}^{\alpha'}(L) \to L^{\alpha'}$ that make the diagram of Eq. (2) commute at homology and homotopy levels. Combined with Lemma 3.3, this fact implies the following result:

THEOREM 3.5. *Let X be a compact set and L a finite set in $\mathbb{R}^d$, such that $\mathrm{d}_{\mathcal{H}}(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{4}\,\mathrm{wfs}(X)$. Then, for all $\alpha, \alpha' \in [\varepsilon, \mathrm{wfs}(X) - \varepsilon]$ such that $\alpha' - \alpha > 2\varepsilon$, and for all $\lambda \in (0, \mathrm{wfs}(X))$, we have: $\forall k \in \mathbb{N}$, $H_k(X^\lambda) \cong \mathrm{im}\; j_*$, where $j_* : H_k(\mathcal{C}^\alpha(L)) \to H_k(\mathcal{C}^{\alpha'}(L))$ is the homomorphism between homology groups induced by the canonical inclusion $j : \mathcal{C}^\alpha(L) \hookrightarrow \mathcal{C}^{\alpha'}(L)$. Given a point $x_0 \in X$, the same result holds for homotopy groups with base-point x_0.*

Using the terminology of [33], this theorem guarantees that the homology of X^λ is obtained from the persistence barcode of the filtration $\{\mathcal{C}^\alpha(L)\}_{\alpha \geq 0}$ by removing the intervals of persistence less than 2ε.

We now give our proof[1] of Lemma 3.4, which consists in a generalization to our context of the main arguments of the proof of the nerve theorem provided in Section 4G of [27]:

Proof of Lemma 3.4. Recall that $\mathcal{U}$ is a good open cover of X, namely: $\forall k \in \mathbb{N}$, $\forall \alpha_0, \cdots, \alpha_k \in L$, $\bigcap_{l=0}^k U_\alpha$ is either empty or contractible. From this cover we construct a topological space ΔX as follows: let Δ^n denote the standard n-simplex, where $n = \#A - 1$. To each non-empty subset S of A we associate the face $[S]$ of Δ^n spanned by the elements of S, as well as the subspace $U_S = \bigcap_{s \in S} U_s$ of X. ΔX is then the subspace of $X \times \Delta^n$ defined by:

$$\Delta X = \bigcup_{\emptyset \neq S \subseteq A} U_S \times [S].$$

The subspace $\Delta X' \subseteq X' \times \Delta^n$ is built similarly. Note that we have $\Delta X \subseteq \Delta X'$, since the hypothesis of the lemma implies $U_S \subseteq U'_S$ for all $S \subseteq A$. Furthermore, the product

[1] Another proof of Lemma 3.4 is provided in [5], for the special case where $X = L^\alpha$, $X' = L^{\alpha'}$, $\mathcal{U} = \{L^\alpha\}$, and $\mathcal{U}' = \{L^{\alpha'}\}$ lie in $\mathbb{R}^d$. Our proof is simpler, and it holds in a more general setting.

structures of ΔX and $\Delta X'$ imply the existence of canonical projections $p : \Delta X \to X$ and $p' : \Delta X' \to X'$. These projections commute with the canonical inclusions $\Delta X \hookrightarrow \Delta X'$ and $X \hookrightarrow X'$, therefore the following diagram:

$$\begin{array}{ccc} X & \hookrightarrow & X' \\ \uparrow_p & & \uparrow_{p'} \\ \Delta X & \hookrightarrow & \Delta X' \end{array} \tag{3}$$

induces commutative diagrams at homology and homotopy levels. Moreover, since $\mathcal{U}$ is an open cover of X, which is paracompact, p is a homotopy equivalence [27, Prop. 4G.2]. The same holds for p', and therefore p and p' induce isomorphisms at homology and homotopy levels.

We now show that, similarly, there exist homotopy equivalences $\Delta X \to \mathcal{NU}$ and $\Delta X' \to \mathcal{NU}'$ that commute with the canonical inclusions $\Delta X \hookrightarrow \Delta X'$ and $\mathcal{NU} \hookrightarrow \mathcal{NU}'$. This follows in fact from the proof of Corollary 4G.3 of [27]. Indeed, using the notion of *complex of spaces* introduced in [27, Section 4G], it can be shown that ΔX is the realization of the complex of spaces associated with the cover $\mathcal{U}$ — see the proof of [27, Prop. 4G.2]. Its base is the barycentric subdivision Γ of $\mathcal{NU}$, where each vertex corresponds to a non-empty finite intersection U_S for some set $S \subseteq A$, and where each edge connecting two vertices $S \subset S'$ corresponds to the canonical inclusion $U_{S'} \hookrightarrow U_S$. In the same way, $\Delta X'$ is the realization of a complex of spaces built over the barycentric subdivision Γ' of $\mathcal{NU}'$. Now, since the non-empty finite intersections U_S (resp. U'_S) are contractible, the map $q : \Delta X \to \Gamma$ (resp. $q' : \Delta X' \to \Gamma'$) induced by sending each open set U_S (resp. U'_S) to a point is a homotopy equivalence [27, Prop. 4G.1 and Corol. 4G.3]. Furthermore, by construction, q is the restriction of q' to ΔX. Therefore,

$$\begin{array}{ccc} \Delta X & \hookrightarrow & \Delta X' \\ \downarrow_q & & \downarrow_{q'} \\ \Gamma & \hookrightarrow & \Gamma' \end{array} \tag{4}$$

is a commutative diagram where vertical arrows are homotopy equivalences. Now, it is well-known that Γ and Γ' are homeomorphic to $\mathcal{NU}$ and $\mathcal{NU}'$ respectively, and that the homeomorphisms commute with the inclusion. Combined with (3) and (4), this fact proves Lemma 3.4. □

3.2 Intertwined filtrations

3.2.1 Results on homology

Using Lemma 2.1 and Theorem 3.5, we get the following guarantees on the Rips filtration:

THEOREM 3.6. *Let $X \subset \mathbb{R}^d$ be a compact set and $L \subset \mathbb{R}^d$ a finite point set such that $\mathrm{d}_{\mathcal{H}}(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{9}\,\mathrm{wfs}(X)$. Then, for all $\alpha \in \left[2\varepsilon,\ \frac{1}{4}(\mathrm{wfs}(X) - \varepsilon)\right]$ and all $\lambda \in (0, \mathrm{wfs}(X))$, we have: $\forall k \in \mathbb{N}$, $H_k(X^\lambda) \cong \mathrm{im}\ j_*$, where j_* is the homomorphism between homology groups induced by the canonical inclusion $j : \mathcal{R}^\alpha(L) \hookrightarrow \mathcal{R}^{4\alpha}(L)$.*

PROOF. Lemma 2.1 provides the following sequence:

$$\mathcal{C}^{\frac{\alpha}{2}}(L) \hookrightarrow \mathcal{R}^\alpha(L) \hookrightarrow \mathcal{C}^\alpha(L) \hookrightarrow \mathcal{C}^{2\alpha}(L) \hookrightarrow \mathcal{R}^{4\alpha}(L) \hookrightarrow \mathcal{C}^{4\alpha}(L)$$

Since $\alpha \geq 2\varepsilon$, Theorem 3.5 implies that this sequence of inclusions induces a sequence of homomorphisms between homology groups, such that $H_k(\mathcal{C}^{\frac{\alpha}{2}}(L)) \to H_k(\mathcal{C}^{4\alpha}(L))$ and $H_k(\mathcal{C}^\alpha(L)) \to H_k(\mathcal{C}^{2\alpha}(L))$ have ranks equal to $\dim H_k(X^\lambda)$. Hence, by Lemma 3.2, rank j_* is also equal to $\dim H_k(X^\lambda)$. It follows that $\mathrm{im}\ j_* \cong H_k(X^\lambda)$. □

Similarly, Corollary 2.3 provides the following sequence:

$$\mathcal{C}^{\frac{\alpha}{4}}(L) \hookrightarrow \mathcal{C}^\alpha_W(L) \hookrightarrow \mathcal{C}^{8\alpha}(L) \hookrightarrow \mathcal{C}^{9\alpha}(L) \hookrightarrow \mathcal{C}^{36\alpha}_W(L) \hookrightarrow \mathcal{C}^{288\alpha}(L),$$

from which follows a result similar to Theorem 3.6 on the witness complex, by the same proof:

THEOREM 3.7. *Let X be a compact set in $\mathbb{R}^d$, and let $L \subseteq W \subseteq \mathbb{R}^d$ be such that L is finite. Assume that $\mathrm{d}_{\mathcal{H}}(X, W) \leq \delta$ and that $\mathrm{d}_{\mathcal{H}}(W, L) \leq \varepsilon$, with $\delta \leq \varepsilon < \min\{\frac{1}{8}\,\mathrm{diam}_{\mathrm{CC}}(X), \frac{1}{1153}\,\mathrm{wfs}(X)\}$. Then, for all $\alpha \in \left[4\varepsilon,\ \frac{1}{288}(\mathrm{wfs}(X) - \varepsilon)\right]$ and all $\lambda \in (0, \mathrm{wfs}(X))$, we have: $\forall k \in \mathbb{N}$, $H_k(X^\lambda) \cong \mathrm{im}\ j_*$, where j_* is the homomorphism between homology groups induced by the canonical inclusion $j : \mathcal{C}^\alpha_W(L) \hookrightarrow \mathcal{C}^{36\alpha}_W(L)$.*

More generally, the above arguments show that the homology of X^λ can be recovered from the persistence barcode of any filtration $\{F_\alpha\}_{\alpha \geq 0}$ that is intertwined with the Čech filtration in the sense of Lemmas 2.1 and 2.2. Note however that Theorems 3.6 and 3.7 suggest a different behavior of the barcode in this case, since its topological noise might scale up with α (specifically, it might be up to linear in α), whereas it is uniformly bounded by a constant in the case of the Čech filtration. This difference of behavior is easily explained by the way $\{F_\alpha\}_{\alpha \geq 0}$ is intertwined with the Čech filtration. A trick to get a uniformly-bounded noise is to represent the barcode of $\{F_\alpha\}_{\alpha \geq 0}$ on a logarithmic scale, that is, with $\log_2 \alpha$ instead of α in abcissa.

3.2.2 Results on homotopy

The results on homology obtained in Section 3.2.1 follow from simple algebraic arguments. Using a more geometric approach, we can get similar results on homotopy. From now on, $x_0 \in X$ is a fixed point and all the homotopy groups $\pi_k(X) = \pi_k(X, x_0)$ are assumed to be with base-point x_0. Theorems 3.6 and 3.7 extend to homotopy as follows:

THEOREM 3.8.

- *Under the hypotheses of Theorem 3.6, we have: $\forall k \in \mathbb{N}$, $\pi_k(X^\lambda) \cong \mathrm{im}\ j_*$, where $j_* : \pi_k(\mathcal{R}^\alpha(L)) \to \pi_k(\mathcal{R}^{4\alpha}(L))$ is the homomorphism between homotopy groups induced by the inclusion $\mathcal{R}^\alpha(L) \hookrightarrow \mathcal{R}^{4\alpha}(L)$.*
- *Under the hypotheses of Theorem 3.7, we have: $\forall k \in \mathbb{N}$, $\pi_k(X^\lambda) \cong \mathrm{im}\ j_*$, where $j_* : \pi_k(\mathcal{C}^\alpha_W(L)) \to \pi_k(\mathcal{C}^{36\alpha}_W(L))$ is the homomorphism between homotopy groups induced by the inclusion $\mathcal{C}^\alpha_W(L) \hookrightarrow \mathcal{C}^{36\alpha}_W(L)$.*

The proof of the theorem relies on the following result, which is an immediate generalization of Proposition 4.1 of [12]:

LEMMA 3.9. *Let X be a compact set and L a finite set in $\mathbb{R}^d$, such that $\mathrm{d}_{\mathcal{H}}(X, L) < \varepsilon$ for some $\varepsilon < \frac{1}{4}\,\mathrm{wfs}(X)$. Let $\alpha, \alpha' \in [\varepsilon, \mathrm{wfs}(X) - \varepsilon]$ be such that $\alpha' - \alpha \geq 2\varepsilon$. Given $k \in \mathbb{N}$, two* k-loops $\sigma_1, \sigma_2 : \mathbb{S}^k \to (L^\alpha, x_0)$ *in L^α are homotopic in $X^{\alpha'+\varepsilon}$ if and only if they are homotopic in $L^{\alpha'}$.*

Proof of Theorem 3.8. As mentioned at the beginning of the proof of Lemma 3.3, we can assume without loss of generality that $2\varepsilon < \alpha < \frac{1}{4}(\mathrm{wfs}(X) - \varepsilon)$. Consider the sequence of inclusions introduced in the proof of Theorem 3.6. We use the homotopy equivalences $h_\beta : L^\beta \to \mathcal{C}^\beta(L)$ provided by Lemma 3.4 for all values $\beta > 0$, which commute with inclusions at homotopy level. Note that, for any element σ of $\pi_k(\mathcal{C}^\beta(L))$, there exists a k-loop in L^β that is mapped through h_β to a k-loop representing the homotopy

class σ. In the following, we denote by σ_g such a k-loop. Let E, F and G be the images of $\pi_k(\mathcal{C}^{\frac{\alpha}{2}}(L))$ in $\pi_k(\mathcal{C}^{\alpha}(L))$, $\pi_k(\mathcal{C}^{2\alpha}(L))$ and $\pi_k(\mathcal{C}^{4\alpha}(L))$ respectively, through the homomorphisms induced by inclusion. We thus have a sequence of surjective homomorphisms: $\pi_k(\mathcal{C}^{\frac{\alpha}{2}}(L)) \to E \to F \to G$.

Note that, by Theorem 3.5, F and G are isomorphic to $\pi_k(X^{\lambda})$. Let $\sigma \in F$ be a homotopy class. Since F is the image of $\pi_k(\mathcal{C}^{\frac{\alpha}{2}}(L))$, we can assume without loss of generality that $\sigma_g \subset L^{\frac{\alpha}{2}}$. Assume that the image of σ in G is zero. Then, σ_g is null-homotopic in $L^{4\alpha}$ and, since $L^{4\alpha} \subset X^{4\alpha+\varepsilon}$, σ_g is also null-homotopic in $X^{4\alpha+\varepsilon}$. But $\sigma_g \subset L^{\frac{\alpha}{2}} \subset X^{\frac{\alpha}{2}+\varepsilon}$, and $X^{2\alpha+\varepsilon}$ deformation retracts onto $X^{\frac{\alpha}{2}+\varepsilon}$, by the Isotopy Lemma 3.1. Therefore, σ_g is null-homotopic in $X^{\frac{\alpha}{2}+\varepsilon}$, which is contained in $L^{2\alpha}$ since $\frac{\alpha}{2} + 2\varepsilon < 2\alpha$. Hence, σ_g is null-homotopic in $L^{2\alpha}$, *i.e.* $\sigma = 0$ in F. So, the homomorphism $F \to G$ is injective, and therefore it is an isomorphism. Thus, $F \to \pi_k(\mathcal{R}^{4\alpha}(L))$ is injective, and it is now enough to prove that the image of the homomorphism $\phi_* : \pi_k(\mathcal{R}^{\alpha}(L)) \to \pi_k(\mathcal{C}^{2\alpha}(L))$ induced by inclusion is F.

Obviously, F is contained in the image of ϕ_*. Now, let $\sigma \in \pi_k(\mathcal{R}^{\alpha}(L))$ and let $\phi_*(\sigma)_g$ be a k-loop in $L^{2\alpha}$ that is mapped through $h_{2\alpha}$ to a k-loop representing the homotopy class $\phi_*(\sigma)$. Since $\phi_*(\sigma)$ is in the image of ϕ_*, and since $\mathcal{R}^{\alpha}(L) \subset \mathcal{C}^{\alpha}(L)$, we can assume that $\phi_*(\sigma)_g$ belongs to L^{α}. Let $\tilde{\sigma}_g$ be the image of $\phi_*(\sigma)_g$ through a deformation retraction of $X^{2\alpha+\varepsilon}$ onto X^{α_0}, where $0 < \alpha_0 < \frac{\alpha}{2}$ is such that $\frac{\alpha}{2} - \alpha_0 > \varepsilon$. Obviously, $\tilde{\sigma}_g$ and $\phi_*(\sigma)_g$ are homotopic in $X^{2\alpha+\varepsilon}$, and it follows from Lemma 3.9 that $\tilde{\sigma}_g$ and $\phi_*(\sigma)_g$ are also homotopic in $L^{2\alpha}$. And since $\tilde{\sigma}_g$ is contained in $X^{\alpha_0} \subset L^{\frac{\alpha}{2}}$, the equivalence class of $h_{\frac{\alpha}{2}}(\tilde{\sigma}_g)$ in $\pi_k(\mathcal{C}^{\frac{\alpha}{2}}(L))$ is mapped to $\phi_*(\sigma) \in \pi_k(\mathcal{C}^{2\alpha}(L))$ through the homomorphism induced by $\mathcal{C}^{\frac{\alpha}{2}}(L) \hookrightarrow \mathcal{C}^{2\alpha}(L)$, which commutes with the homotopy equivalences. As a result, $\phi_*(\sigma)$ belongs to F, which is thus equal to im ϕ_*. This proves the first part of the theorem. The proof of the second part is mostly the same. □

4. SMOOTH SUBMANIFOLDS OF $\mathbb{R}^D$

In this section, we consider the case of submanifolds X of $\mathbb{R}^d$ that have positive *reach*. Recall that the reach of X, or rch(X) for short, is the minimum distance between the points of X and the points of its medial axis [1]. A point cloud $L \subset X$ is an *ε-sample* of X if every point of X lies within distance ε of L. In addition, L is *ε-sparse* if its points lie at least ε away from one another.

Theorem 4.1 below is a first attempt at quantifying a conjecture of Carlsson and de Silva [17], according to which the witness complex filtration should have *cleaner* persistence barcodes than the Čech and Rips filtrations, at least on smooth submanifolds of $\mathbb{R}^d$. By *cleaner* is meant that the amplitude of the topological noise in the barcodes should be smaller, and also that the long intervals should appear earlier. We prove this latter statement correct to some extent:

THEOREM 4.1. *There exist a constant $\varrho > 0$ and a continuous, non-decreasing map $\bar{\omega} : [0, \varrho) \to [0, \frac{1}{2})$, with $\bar{\omega}(0) = 0$, such that, for any submanifold X of $\mathbb{R}^d$, for all ε, δ satisfying $0 < \delta \leq \varepsilon < \varrho$ rch(X), for any δ-sample W of X and any ε-sparse ε-sample L of W, $\mathcal{C}^{\alpha}_W(L)$ contains a subcomplex $\mathcal{D}$ homeomorphic to X and such that the canonical inclusion $\mathcal{D} \hookrightarrow \mathcal{C}^{\alpha}_W(L)$ induces an injective homomorhism between homology groups, provided that α satisfies: $\frac{8}{3}(\delta + \bar{\omega}(\frac{\varepsilon}{\text{rch}(X)})^2\varepsilon) \leq \alpha < \frac{1}{2}$ rch$(X) - (3 + \frac{\sqrt{2}}{2})(\varepsilon + \delta)$.*

This theorem guarantees that, for values of α ranging from $O(\delta + \bar{\omega}(\frac{\varepsilon}{\text{rch}(X)})^2\varepsilon)$ to $\Omega(\text{rch}(X))$, the topology of X is captured by a subcomplex $\mathcal{D}$ that injects itself suitably in $\mathcal{C}^{\alpha}_W(L)$. As a result, long intervals showing the homology of X appear around $\alpha = O(\delta + \bar{\omega}(\frac{\varepsilon}{\text{rch}(X)})^2\varepsilon)$ in the persistence barcode of the witness complex filtration. This can be much sooner than the time $\alpha = 2\varepsilon$ prescribed by Theorem 3.7, since $\bar{\omega}(\frac{\varepsilon}{\text{rch}(X)})$ can be arbitrarily small. Specifically, the denser the landmark set L, the smaller the ratio $\frac{\varepsilon}{\text{rch}(X)}$, and therefore the smaller $\frac{8}{3}(\delta + \bar{\omega}(\frac{\varepsilon}{\text{rch}(X)})^2\varepsilon)$ compared to 2ε.

Our proof of Theorem 4.1 stresses the close relationship that exists between the α-witness complex and the so-called *weighted restricted Delaunay triangulation* $\mathcal{D}^X_{\omega}(L)$. Given a submanifold X of $\mathbb{R}^d$, a finite landmark set $L \subset \mathbb{R}^d$, and an assignment of non-negative weights to the landmarks, specified through a map $\omega : L \to [0, \infty)$, $\mathcal{D}^X_{\omega}(L)$ is the nerve of the restriction to X of the *power diagram*[2] of the weighted set L. By a result of Cheng *et al.* (see Theorem 4.2 below), $\mathcal{D}^X_{\omega}(L)$ is homeomorphic to X under a sufficient landmark density and under a suitable choice of weights – bounded from above by $\bar{\omega}\left(\frac{\varepsilon}{\text{rch}(X)}\right)$. The main point of our proof is then to show that $\mathcal{C}^{\alpha}_W(L)$ contains $\mathcal{D}^X_{\omega}(L)$ and that the latter injects itself *nicely* into the former.

In the special case where X is a smooth curve or surface, all weights can be taken to be zero, since the unweighted restricted Delaunay triangulation is known to be homeomorphic to X [1, 2]. As a result, function $\bar{\omega}$ is zero, and the long intervals showing the homology of X in the barcode of the witness complex filtration appear already at time $\alpha = O(\delta)$.

In the general case however, the upper bound on the appearance time of long bars cannot depend solely on δ, since otherwise, in the limit case where $\delta = 0$ (*i.e.* $W = X$), we would get that the homology groups of X can be injected into the ones of $\mathcal{C}_X(L)$, which is known to be true for curves and surfaces [3], but not for 3-manifolds [30]. Now, whether $O(\delta + \bar{\omega}(\frac{\varepsilon}{\text{rch}(X)})^2\varepsilon)$ is a tight upper bound or not is open.

The rest of Section 4 is devoted to the proof of Theorem 4.1. After introducing the weighted restricted Delaunay triangulation formally in Section 4.1, we stress its relationship with the α-witness complex in Section 4.2, and then we detail the proof of Theorem 4.1 in Section 4.3.

4.1 Weighted restricted Delaunay triangulation

Given a finite point set $L \subset \mathbb{R}^d$, an *assignment of weights over 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. Given a finite point set L and an assignment of weights ω over L, we denote by $\mathcal{V}_{\omega}(L)$ the power diagram of the weighted set L, and by $\mathcal{D}_{\omega}(L)$ its nerve, also known as the weighted Delaunay triangulation. If the relative amplitude of ω is at most $\frac{1}{2}$, then the points of L have non-empty cells in $\mathcal{V}_{\omega}(L)$, and in fact each point of L belongs to its own cell [13]. For any simplex σ of $\mathcal{D}_{\omega}(L)$, $\text{V}_{\omega}(\sigma)$ denotes its dual face in $\mathcal{V}_{\omega}(L)$.

Given a subset X of $\mathbb{R}^d$, we call $\mathcal{V}^X_{\omega}(L)$ the restriction of $\mathcal{V}_{\omega}(L)$ to X, and we denote by $\mathcal{D}^X_{\omega}(L)$ its nerve, also known as the weighted Delaunay triangulation of L restricted to

[2] More on power diagrams and on restricted Delaunay triangulations can be found in [4] and [22] respectively.

X. Observe that $\mathcal{D}^X_\omega(L)$ is a subcomplex of $\mathcal{D}_\omega(L)$. In the special case where all the weights are equal, $\mathcal{V}_\omega(L)$ and $\mathcal{D}_\omega(L)$ coincide with their standard Euclidean versions, $\mathcal{V}(L)$ and $\mathcal{D}(L)$. Similarly, $\mathrm{V}_\omega(\sigma)$ becomes $\mathrm{V}(\sigma)$, and $\mathcal{V}^X_\omega(L)$ and $\mathcal{D}^X_\omega(L)$ become respectively $\mathcal{V}^X(L)$ and $\mathcal{D}^X(L)$.

THEOREM 4.2 (LEMMAS 13, 14, 18 OF [14]).
There exist[3] *a constant* $\varrho > 0$ *and a non-decreasing continuous map* $\bar{\omega} : [0, \varrho) \to [0, \frac{1}{2})$*, such that, for any manifold* X *and any* ε*-sparse* 2ε*-sample* L *of* X*, with* $\varepsilon < \varrho\ \mathrm{rch}(X)$*, there is an assignment of weights* ω *of relative amplitude at most* $\bar{\omega}\left(\frac{\varepsilon}{\mathrm{rch}(X)}\right)$ *such that* $\mathcal{D}^X_\omega(L)$ *is homeomorphic to* X*.*

This theorem guarantees that the topology of X is captured by $\mathcal{D}^X_\omega(L)$ provided that the landmarks are sufficiently densely sampled on X, and that they are assigned suitable weights. Observe that the denser the landmark set, the smaller the weights are required to be, as specified by the map $\bar{\omega}$. In the particular case where X is a curve or a surface, $\bar{\omega}$ can be taken to be the constant zero map, since $\mathcal{D}^X(L)$ is homeomorphic to X [1, 2]. On higher-dimensional manifolds though, positive weights are required, since $\mathcal{D}^X(L)$ may fail to capture the topological invariants of X [30].

The proof of the theorem given in [14] shows that $\mathcal{V}^X_\omega(L)$ satisfies the so-called *closed ball property*, which states that every face of the weighted Voronoi diagram $\mathcal{V}_\omega(L)$ intersects the manifold X along a topological ball of proper dimension, if at all. Under this condition, there exists a homeomorphism h_0 between the nerve $\mathcal{D}^X_\omega(L)$ and X, as proved by Edelsbrunner and Shah [22]. Furthermore, h_0 sends every simplex of $\mathcal{D}^X_\omega(L)$ to a subset of the union of the restricted Voronoi cells of its vertices, that is: $\forall \sigma \in \mathcal{D}^X_\omega(L)$, $h_0(\sigma) \subseteq \bigcup_{v \text{ vertex of } \sigma} \mathrm{V}_\omega(v) \cap X$. This fact will be instrumental in the proof of Theorem 4.1.

4.2 Relationship between $\mathcal{D}^X_\omega(L)$ and $\mathcal{C}^\alpha_W(L)$

As mentioned in introduction, the use of the witness complex for topological data analysis is motivated by its relationship with the weighted restricted Delaunay triangulation:

LEMMA 4.3. *Let* X *be a compact set in* $\mathbb{R}^d$*,* $W \subseteq X$ *a* δ*-sample of* X*, and* $L \subseteq W$ *an* ε*-sparse* ε*-sample of* W*. Then, for all assignment of weights* ω *of relative amplitude* $\bar{\omega} \leq \frac{1}{2}$*,* $\mathcal{D}^X_\omega(L)$ *is included in* $\mathcal{C}^\alpha_W(L)$ *whenever* $\alpha \geq \frac{2}{1-\bar{\omega}^2}\left(\delta + \bar{\omega}^2\varepsilon\right)$*.*

This result implies in particular that $\mathcal{D}^X(L)$ is included in $\mathcal{C}^\alpha_W(L)$ whenever $\alpha \geq 2\delta$, since $\mathcal{D}^X(L)$ is nothing but $\mathcal{D}^X_\omega(L)$ for an assignment of weights of relative amplitude zero.

PROOF. Let σ be a simplex of $\mathcal{D}^X_\omega(L)$. If σ is a vertex, then it clearly belongs to $\mathcal{C}^\alpha_W(L)$ for all $\alpha \geq 0$, since $L \subseteq W$. Assume now that σ has positive dimension, and consider a point $c \in \mathrm{V}_\omega(\sigma) \cap X$. For any vertex v of σ and any point p of L (possibly equal to v), we have: $\|v-c\|^2 - \omega(v)^2 \leq \|p-c\|^2 - \omega(p)^2$, which yields: $\|v-c\|^2 \leq \|p-c\|^2 + \omega(v)^2 - \omega(p)^2$. Now, $\omega(p)^2$ is non-negative, while $\omega(v)^2$ is at most $\bar{\omega}^2\|v-p\|^2$, which gives: $\|v-c\|^2 \leq \|p-c\|^2 + \bar{\omega}^2\|v-p\|^2$. Replacing $\|v-p\|$ by $\|v-c\|+\|p-c\|$, we get a semi-algebraic expression of degree 2 in $\|v-c\|$, namely: $(1-\bar{\omega}^2)\|v-c\|^2 - 2\bar{\omega}^2\|p-c\|\|v-c\| - (1+\bar{\omega}^2)\|p-c\|^2 \leq 0$. It follows that $\|v-c\| \leq \frac{1+\bar{\omega}^2}{1-\bar{\omega}^2}\|p-c\|$. Let now w be a point of W closest to c in the Euclidean metric. Using the triangle inequality and the fact that $\|w-c\| \leq \delta$, we get: $\|v-w\| \leq \|v-c\| + \|w-c\| \leq \frac{1+\bar{\omega}^2}{1-\bar{\omega}^2}\|p-c\| + \delta$. This holds for any point $p \in L$, and in particular for the nearest neighbor p_w of w in L. Therefore, we have $\|v-w\| \leq \frac{1+\bar{\omega}^2}{1-\bar{\omega}^2}\|p_w-c\| + \delta$, which is at most $\frac{1+\bar{\omega}^2}{1-\bar{\omega}^2}(\|p_w-w\|+\delta)+\delta \leq \|p_w-w\| + \frac{2}{1-\bar{\omega}^2}\left(\delta+\bar{\omega}^2\varepsilon\right)$ because $\|w-c\| \leq \delta$ and $\|w-p_w\| \leq \varepsilon$. Since this inequality holds for any vertex v of σ, and since the Euclidean distances from w to all the landmarks are at least $\|p_w-w\|$, w is an α-witness of σ and of all its faces as soon as $\alpha \geq \frac{2}{1-\bar{\omega}^2}\left(\delta+\bar{\omega}^2\varepsilon\right)$. Since this holds for all simplex $\sigma \in \mathcal{D}^X_\omega(L)$, the lemma follows. □

4.3 Proof of Theorem 4.1

The proof relies on two technical results. The first one is Dugundji's extension theorem [19], which states that, given an abstract simplex σ and a continuous map $f : \partial\sigma \to \mathbb{R}^d$, f can be extended to a continuous map $f : \sigma \to \mathbb{R}^d$ such that $f(\sigma)$ is included in the Euclidean convex hull of $f(\partial\sigma)$, noted $\mathrm{CH}(f(\partial\sigma))$. This convexity property of f is used in the proof of the second technical result, stated as Lemma 4.5 below and proved at the end of the section.

Proof of Theorem 4.1. Since $\delta \leq \varepsilon$, L is an ε-sparse 2ε-sample of X, with $\varepsilon < \varrho\ \mathrm{rch}(X)$. Therefore, by Theorem 4.2, there exists an assignment of weights ω over L, of relative amplitude at most $\bar{\omega}\left(\frac{\varepsilon}{\mathrm{rch}(X)}\right)$, such that $\mathcal{D}^X_\omega(L)$ is homeomorphic to X. Taking $\mathcal{D} = \mathcal{D}^X_\omega(L)$, we then have: $\forall k \in \mathbb{N}$, $H_k(X) \cong H_k(\mathcal{D})$. Moreover, by Lemma 4.3, we know that $\mathcal{D} = \mathcal{D}^X_\omega(L)$ is included in $\mathcal{C}^\alpha_W(L)$, since $\alpha \geq \frac{8}{3}\left(\bar{\omega}\left(\frac{\varepsilon}{\mathrm{rch}(X)}\right)^2\varepsilon + \delta\right) \geq \frac{2}{1-\bar{\omega}\left(\frac{\varepsilon}{\mathrm{rch}(X)}\right)^2}\left(\bar{\omega}\left(\frac{\varepsilon}{\mathrm{rch}(X)}\right)^2\varepsilon + \delta\right)$.

There remains to prove that $j : \mathcal{D}^X_\omega(L) \hookrightarrow \mathcal{C}^\alpha_W(L)$ induces injective homomorphisms j_* between the homology groups of $\mathcal{D}^X_\omega(L)$ and $\mathcal{C}^\alpha_W(L)$. To do so, we will build a retraction $h : \mathcal{C}^\alpha_W(L) \to \mathcal{D}^X_\omega(L)$, *i.e.* a continuous map whose restriction to $\mathcal{D}^X_\omega(L)$, $h \circ j$, is the identity. This will imply that $h_* \circ j_* : H_k(\mathcal{D}^X_\omega(L)) \to H_k(\mathcal{D}^X_\omega(L))$ is an isomorphism (in fact, the identity), and thus that j_* is injective.

We begin our construction with the homeomorphism $h_0 : \mathcal{D}^X_\omega(L) \to X$ provided by the theorem of Edelsbrunner and Shah [22]. Taking h_0 as a map $\mathcal{D}^X_\omega(L) \to \mathbb{R}^d$, we extend it to a continuous map $\tilde{h}_0 : \mathcal{C}^\alpha_W(L) \to \mathbb{R}^d$ by the following iterative process: while there exists a simplex $\sigma \in \mathcal{C}^\alpha_W(L)$ such that $\tilde{h}_0$ is defined over the boundary of σ but not over its interior, we apply Dugundji's extension theorem, which extends $\tilde{h}_0$ to the entire simplex σ.

LEMMA 4.4. *The above iterative process extends* h_0 *to a map* $\tilde{h}_0 : \mathcal{C}^\alpha_W(L) \to \mathbb{R}^d$*.*

PROOF. We only need to prove that the process visits every simplex of $\mathcal{C}^\alpha_W(L)$. Assume for a contradiction that the process terminates while there still remain some unvisited simplices of $\mathcal{C}^\alpha_W(L)$. Consider one such simplex σ of minimal dimension. Either σ is a vertex, or there is at least one proper face of σ that has not yet been visited – since otherwise the process could visit σ. In the former case, σ is a point of L, and as such it is a vertex[4] of $\mathcal{D}^X_\omega(L)$, which means that h_0 is already defined over σ (contradiction). In the latter case, we get a contradiction with the fact that σ is of minimal dimension. □

[3]Here, quantities ϱ and $\bar{\omega}$ are the same as in Theorem 4.1. In fact, these quantities come from the lemmas of [14].

[4]Indeed, every point $p \in L$ lies on X and belongs to its own cell, since ω has relative amplitude less than $\frac{1}{2}$. Therefore, $\mathrm{V}_\omega(p) \cap X \neq \emptyset$, which means that p is a vertex of $\mathcal{D}^X_\omega(L)$.

Now that we have built a map $\tilde{h}_0 : \mathcal{C}^\alpha_W(L) \to \mathbb{R}^d$, our next step is to turn it into a map $\mathcal{C}^\alpha_W(L) \to X$. To do so, we compose it with the projection p_X that maps every point of $\mathbb{R}^d$ to its nearest neighbor on X, if the latter is unique. This projection is known to be well-defined and continuous over $\mathbb{R}^d \setminus \mathrm{M}$, where M denotes the medial axis of X [23].

LEMMA 4.5. *Let $X, W, L, \delta, \varepsilon$ satisfy the hypotheses of Theorem 4.1. Then, $\tilde{h}_0(\mathcal{C}^\alpha_W(L)) \cap \mathrm{M} = \emptyset$ provided that $\alpha < \frac{1}{2}\,\mathrm{rch}(X) - \left(3 + \frac{\sqrt{2}}{2}\right)(\varepsilon + \delta)$.*

Since by Lemma 4.5 we have $\tilde{h}_0(\mathcal{C}^\alpha_W(L)) \cap \mathrm{M} = \emptyset$, the map $p_X \circ \tilde{h}_0 : \mathcal{C}^\alpha_W(L) \to X$ is well-defined and continuous. Our final step is to compose it with h_0^{-1}, to get a continuous map $h = h_0^{-1} \circ p_X \circ \tilde{h}_0 : \mathcal{C}^\alpha_W(L) \to \mathcal{D}^X_\omega(L)$. The restriction of h to $\mathcal{D}^X_\omega(L)$ is simply $h_0^{-1} \circ p_X \circ h_0$, which coincides with $h_0^{-1} \circ h_0 = \mathrm{id}$ since $h_0(\mathcal{D}^X_\omega(L)) = X$. It follows that $h \circ j$ is the identity in $\mathcal{D}^X_\omega(L)$, and therefore that the induced map $h_* \circ j_*$ is also the identity. This implies that $j_* : H_k(\mathcal{D}^X_\omega(L)) \to H_k(\mathcal{C}^\alpha_W(L))$ is injective, which concludes the proof of Theorem 4.1. □

We end the section by providing the proof of Lemma 4.5:

Proof of Lemma 4.5. First, we claim that the image through $\tilde{h}_0$ of any simplex of $\mathcal{C}^\alpha_W(L)$ is included in the Euclidean convex hull of the restricted Voronoi cells of its simplices, that is:

$$\forall \sigma \in \mathcal{C}^\alpha_W(L),\ \tilde{h}_0(\sigma) \subseteq \mathrm{CH}\left(\bigcup_{v \text{ vertex of } \sigma} \mathrm{V}_\omega(v) \cap X\right).$$

This is clearly true if σ belongs to $\mathcal{D}^X_\omega(L)$, since in this case we have $\tilde{h}_0(\sigma) = h_0(\sigma) \subseteq \bigcup_{v \text{ vertex of } \sigma} \mathrm{V}_\omega(v) \cap X$, as mentioned after Theorem 4.2. Now, if the property holds for all the proper faces of a simplex $\sigma \in \mathcal{C}^\alpha_W(L)$, then by induction it also holds for the simplex itself. Indeed, for each proper face $\tau \subset \sigma$, we have $\tilde{h}_0(\tau) \subseteq \mathrm{CH}\left(\bigcup_{v \text{ vertex of } \tau} \mathrm{V}_\omega(v) \cap X\right)$, which is included in $\mathrm{CH}\left(\bigcup_{v \text{ vertex of } \sigma} \mathrm{V}_\omega(v) \cap X\right)$. Therefore, $\mathrm{CH}\left(\bigcup_{v \text{ vertex of } \sigma} \mathrm{V}_\omega(v) \cap X\right)$ contains $\mathrm{CH}\left(\tilde{h}_0(\partial\sigma)\right)$, which, by Dugundji's extension theorem, contains $\tilde{h}_0(\sigma)$. Thus, the property holds for every simplex of $\mathcal{C}^\alpha_W(L)$.

We can now prove that the image through $\tilde{h}_0$ of any arbitrary simplex σ of $\mathcal{C}^\alpha_W(L)$ does not intersect the medial axis of X. This is clearly true if σ is a simplex of $\mathcal{D}^X_\omega(L)$, since in this case $\tilde{h}_0(\sigma) = h_0(\sigma)$ is included in X. Assume now that $\sigma \notin \mathcal{D}^X_\omega(L)$. In particular, σ is not a vertex. Let v be an arbitrary vertex of σ. Consider any other vertex u of σ. Edge $[u, v]$ is α-witnessed by some point $w_{uv} \in W$. We then have $\|v - u\| \leq \|v - w_{uv}\| + \|w_{uv} - u\| \leq 2\mathrm{d}_2(w_{uv}) + 2\alpha$, where $\mathrm{d}_2(w_{uv})$ stands for the Euclidean distance from w_{uv} to its second nearest landmark. According to Lemma 3.4 of [6], we have $\mathrm{d}_2(w) \leq 3(\varepsilon + \delta)$, since L is an $(\varepsilon + \delta)$-sample of X. Thus, all the vertices of σ are included in the Euclidean ball $B(v, 2\alpha + 6(\varepsilon + \delta))$. Moreover, for any vertex u of σ and any point $p \in \mathrm{V}_\omega(u) \cap X$, we have $\|p - u'\| \leq \varepsilon + \delta$, where u' is a landmark closest to p in the Euclidean metric. Combined with the fact that $\|p - u\|^2 - \omega(u)^2 \leq \|p - u'\|^2 - \omega(u')^2$, we get: $\|p - u\|^2 \leq \|p - u'\|^2 + \omega(u)^2 \leq 2(\varepsilon + \delta)^2$, since by Lemma 3.3 of [6] we have $\omega(u) \leq 2\,\bar{\omega}\left(\frac{\varepsilon}{\mathrm{rch}(X)}\right)(\varepsilon + \delta) \leq \varepsilon + \delta$. Hence, $\mathrm{V}_\omega(u) \cap X$ is included in $B(u, \sqrt{2}(\varepsilon + \delta)) \subset B(v, 2\alpha + (6 + \sqrt{2})(\varepsilon + \delta))$. Since this is true for every vertex u of σ, we get: $\tilde{h}_0(\sigma) \subseteq \mathrm{CH}\left(\bigcup_{u \text{ vertex of } \sigma} \mathrm{V}_\omega(u) \cap X\right) \subseteq B(v, 2\alpha + (6 + \sqrt{2})(\varepsilon + \delta))$. Now, v belongs to $L \subseteq W \subseteq X$, and by assumption we have $2\alpha + (6 + \sqrt{2})(\varepsilon + \delta) < \mathrm{rch}(X)$, therefore $\tilde{h}_0(\sigma)$ does not intersect the medial axis of X. □

Input: W, distances $\{l(w, w'),\ w, w' \in W\}$.
Init: Let $L := \emptyset$, $\varepsilon := +\infty$;
While $L \subsetneq W$ do
  Let $p := \mathrm{argmax}_{w \in W} \min_{v \in L} l(w, v)$;
  // p is chosen arbitrarily in W if $L = \emptyset$
  $L := L \cup \{p\}$;
  $\varepsilon := \max_{w \in W} \min_{v \in L} l(w, v)$;
  Update $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$;
  Compute persistence of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$;
End_while
Output: diagram showing the evolution of persistent Betti numbers versus ε.

Figure 2: Pseudo-code of the algorithm.

5. APPLICATION TO RECONSTRUCTION

Taking advantage of the structural results of Section 3, we devise a very simple yet provably-good algorithm for constructing nested pairs of complexes that capture the homology of a large class of compact subsets of $\mathbb{R}^d$. This algorithm is a variant of the greedy refinement technique of [26], which builds a set L of landmarks iteratively and in the meantime maintains a suitable data structure. In our case, the data structure is composed of a nested pair of simplicial complexes, which can be either $\mathcal{R}^\alpha(L) \hookrightarrow \mathcal{R}^{\alpha'}(L)$ or $\mathcal{C}^\alpha_W(L) \hookrightarrow \mathcal{C}^{\alpha'}_W(L)$, for specific values $\alpha < \alpha'$. Both variants of the algorithm enjoy similar theoretical guarantees, but the variant using witness complexes is likely to be more effective in practice. In the sequel we focus on the variant using Rips complexes because its analysis is somewhat simpler.

The algorithm.

The input is a finite point set W in an arbitrary metric space, together with the pairwise distances $l(w, w')$ between the points of W. Initially, we set $L = \emptyset$ and $\varepsilon = +\infty$.

At each iteration, the point of W lying furthest away[5] from L in the metric l is inserted in L, and ε is set to $\max_{w \in W} \min_{v \in L} l(w, v)$. Then, $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$ are updated, and the persistent homology of $\mathcal{R}^{4\varepsilon}(L) \hookrightarrow \mathcal{R}^{16\varepsilon}(L)$ is computed using the persistence algorithm [33].

The algorithm terminates when $L = W$. The output is the diagram showing the evolution of the persistent Betti numbers versus ε, which have been maintained throughout the process. As we will see below, with the help of this diagram the user can determine a relevant scale at which to process the data: it is then easy to generate the corresponding subset L of landmarks (the points of W have been sorted according to their order of insertion in L during the process), and to rebuild $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$. The pseudo-code of the algorithm is provided in Figure 2.

Guarantees on the output.

For any $i > 0$, let $L(i)$ and $\varepsilon(i)$ denote respectively L and ε at the end of the ith iteration of the main loop of

[5] At the first iteration, an arbitrary point of W is chosen, since L is empty.

the algorithm. Since $L(i)$ keeps growing with i, $\varepsilon(i)$ is a decreasing function of i. In addition, $L(i)$ is an $\varepsilon(i)$-sample of W, by definition of $\varepsilon(i)$. Hence, if W lies at Hausdorff distance δ of some compact set $X \subset \mathbb{R}^d$, then we have $\mathrm{d}_{\mathcal{H}}(L(i), X) \leq \delta + \varepsilon(i)$. Therefore, Theorem 3.6 provides us with the following theoretical guarantee:

THEOREM 5.1. *If the input W lies at Hausdorff distance δ of some compact set $X \subset \mathbb{R}^d$, with $\delta < \frac{1}{18}\mathrm{wfs}(X)$, then, at each iteration i such that $\delta < \varepsilon(i) < \frac{1}{18}\mathrm{wfs}(X)$, the persistent homology groups of $\mathcal{R}^{4\varepsilon(i)}(L(i)) \hookrightarrow \mathcal{R}^{16\varepsilon(i)}(L(i))$ are isomorphic to the homology groups of X^λ, $\forall \lambda \in (0, \mathrm{wfs}(X))$.*

Under mild conditions on the input, this theorem guarantees the existence of a plateau showing the homology of X^λ, of length at least $(\frac{1}{18}\mathrm{wfs}(X) - \delta)$, in the diagram of persistent Betti numbers. When δ is small enough compared to $\mathrm{wfs}(X)$, the plateau is large enough to be detected. In cases where W samples several compact sets with different weak feature sizes, the theorem ensures that several plateaus appear in the diagram, showing plausible reconstructions at various scales – see Figure 3. Once a relevant scale has been selected, the corresponding landmark set and nested complexes are easily rebuilt. Differently from the algorithm of [26], this outcome is not a single embedded simplicial complex but a nested pair of abstract complexes, whose images in $\mathbb{R}^d$ lie at Hausdorff distance[6] $O(\varepsilon)$ of X, and whose persistent homology gives the homology of X^λ.

Update of $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$.

We now describe how to maintain $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$. In fact, we settle for describing how to rebuild $\mathcal{R}^{16\varepsilon}(L)$ completely at each iteration, which is sufficient for achieving our complexity bounds, although it is clearly much preferable in practice to use more local rules to update the simplicial complexes. Consider the one-skeleton graph G of $\mathcal{R}^{16\varepsilon}(L)$. By definition, a simplex that is not a vertex belongs to $\mathcal{R}^{16\varepsilon}(L)$ if and only if all its edges are in G. Therefore, the simplices of $\mathcal{R}^{16\varepsilon}(L)$ are precisely the cliques of G. The simplicial complex can then be built as follows: (1.) build graph G, (2.) find all maximal cliques in G, and (3.) report the maximal cliques and all their subcliques. We perform Step 1. naively in $O(|L|^2)$ time, where $|L|$ denotes the size of L. For Step 2., we use the output-sensitive algorithm of [32], which finds all the maximal cliques of G in $O(k\ |L|^3)$ time, where k is the size of the answer. Finally, we report all the subcliques of the maximal cliques in a time that is linear in the total number of cliques, which is also the size of $\mathcal{R}^{16\varepsilon}(L)$. Therefore, at each iteration of the algorithm, $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$ are rebuilt within $O(|\mathcal{R}^{16\varepsilon}(L)|\ |L|^3)$ time, where $|\mathcal{R}^{16\varepsilon}(L)|$ denotes the size of $\mathcal{R}^{16\varepsilon}(L)$.

Running time of the algorithm.

Let $|W|$ denote the size of W. At each iteration of the algorithm, point p and parameter ε are computed naively by iterating over the points of W, and for each such point, by reviewing its distances to all the landmarks. This procedure takes $O(|W||L|)$ time. Once $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$ have been updated, the persistence algorithm runs in $O(|\mathcal{R}^{16\varepsilon}(L)|^3)$ time [21, 33]. Hence,

[6]Indeed, every simplex of $\mathcal{R}^{16\varepsilon}(L)$ has all its vertices in $X^{\varepsilon+\delta} \subseteq X^{2\varepsilon}$, and the lengths of its edges are at most 16ε.

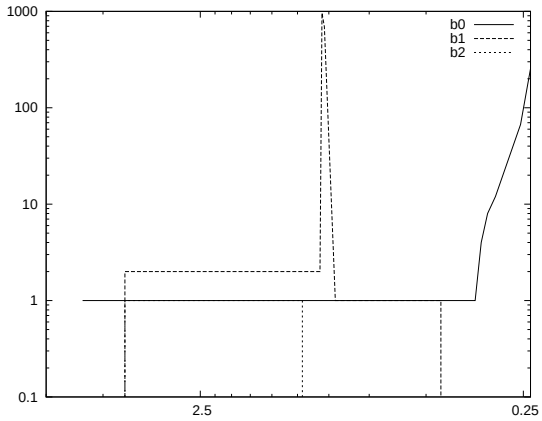

Figure 3: Output of our algorithm when applied blindly to the 1000-dimensional data set of Figure 1.

LEMMA 5.2. *The time complexity of every iteration of the algorithm is $O(|W||L| + |\mathcal{R}^{16\varepsilon}(L)||L|^3 + |\mathcal{R}^{16\varepsilon}(L)|^3)$.*

In addition, standard packing arguments (omitted in this extended abstract) provide the following tight upper bounds on the size of $\mathcal{R}^{16\varepsilon}(L)$ in Euclidean spaces:

LEMMA 5.3. *If L is a finite ε-sparse point set in $\mathbb{R}^d$, then $|\mathcal{R}^{16\varepsilon}(L)| \leq 2^{33^d}|L|$. If in addition L lies on a m-submanifold X of $\mathbb{R}^d$, with $\mathrm{rch}(X) > 256\varepsilon$, then $|\mathcal{R}^{16\varepsilon}(L)| \leq 2^{35^m}|L|$. These upper bounds are tight in order of magnitude.*

Whenever the input point cloud W lies on a smooth m-submanifold X of $\mathbb{R}^d$, the lemma suggests[7] that the algorithm goes through two consecutive phases. First, a transition phase where the landmark set L is too coarse for the dimensionality of X to have an influence on the shapes and sizes of the stars of the vertices of $\mathcal{R}^{16\varepsilon}(L)$. For instance, if X is an embedded curve that roughly fills in the unit ball in $\mathbb{R}^d$, then, for large values of ε, L is nothing but a sampling of the d-ball. Then comes a second, stable phase, where L is dense enough for the dimensionality of X to play a role. Denoting by i_0 the last iteration of the transition phase, we deduce from Lemmas 5.2 and 5.3 that the running time of the algorithm is $O(|W||L(i_0)|^2 + 8^{33^d}|L(i_0)|^5 + 8^{35^m}\ |W|^5)$. There remains to get rid of the term depending on d, which we do using a backtracking strategy. Specifically, we first run the algorithm without maintaining $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$, which simply sorts the points of W according to their order of insertion in L. Then, we run the algorithm backwards, starting with $L = L(|W|) = W$ and considering at each iteration j the landmark set $L(|W| - j)$. During this second phase, we do maintain $\mathcal{R}^{4\varepsilon}(L)$ and $\mathcal{R}^{16\varepsilon}(L)$ and compute their persistent Betti numbers. If W samples X densely enough, then Theorem 5.1 ensures that the relevant plateaus will be computed before the transition phase starts, and thus before the size of the data structure becomes independent of the dimension of X. It is then up to the user to stop the process when the space complexity becomes too large. This variant of the algorithm has the following complexity:

THEOREM 5.4. *If W is a point cloud in Euclidean space $\mathbb{R}^d$, then the running time of the algorithm is $O(8^{33^d}|W|^5)$,*

[7]Note that, at every iteration i of the process, $L(i)$ is an $\varepsilon(i)$-sparse point set, since the algorithm always inserts in L the point of W lying furthest away from L [26, Lemma 4.1].

where $|W|$ denotes the size of W. If in addition W is a δ-sample of some smooth m-submanifold of $\mathbb{R}^d$, with δ small enough, then the running time becomes $O(8^{35^m}\ |W|^5)$.

By Lemma 5.3, the bounds in Theorem 5.4 are tight in order of magnitude. Thus, the algorithm can have a doubly exponential complexity in m when the input point set is densely sampled from a m-dimensional smooth manifold. However, it can be shown that the m-skeleton of the Rips complex is in fact simply exponential in m. Hence, when a reasonable upper bound b on m is known, one can reduce the running time of the algorithm to $2^{O(m^2)}|W|^5$ by considering only the b-skeleton of the Rips complex. Similarly, the running time reduces to $2^{O(d^2)}|W|^5$ if only the d-skeleton is considered.

6. CONCLUSION

This paper makes effective the approach developped in [12, 15] by providing an efficient, provably-good and easy-to-implement algorithm for the topological and geometric analysis of point cloud data in arbitrary dimensions. Addressing a weaker version of the classical reconstruction problem, the algorithm ultimately outputs a nested pair of complexes at a user-defined scale, from which the homology of the underlying shape X can be inferred. When X is a smooth submanifold of $\mathbb{R}^d$, the complexity of the algorithm scales up with the intrinsic dimension of X and not with the ambient dimension d, assuming that the pairwise distances between the data points have been pre-computed. Thus, a new step is made towards reconstructing (low-dimensional) manifolds in high-dimensional spaces in reasonable time with guarantees. However, there still remains the challenging problem of constructing an embedded complex that is topologically equivalent and geometrically close to the sampled shape.

The theoretical framework developed in the paper can be used for the analysis of various persistence-based methods in Euclidean spaces. It can also virtually be applied in any metric space (provided that the result of [12, 15] on unions of balls can be extended), thanks to the genericity of Lemma 3.4 and of the arguments of Section 3.2.1. A class of spaces of particular interest to us is the class of compact Riemannian manifolds, possibly with boundaries, with applications in machine learning and sensor networks.

Acknowledgements.

The authors wish to thank Gunnar Carlsson, Vin de Silva, Leonidas Guibas, Tamal Dey, and the anonymous referees for their insightful comments and suggestions. Our experiments were carried out using Matlab and the Plex library. This work was partially supported by ANR grant GeoTopAl.

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] F. Aurenhammer. Voronoi diagrams: A survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23(3):345–405, September 1991.

[5] P. Bendich, D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov. Inferring local homology from sampled stratified spaces. In *Proc. 48th Annu. IEEE Sympos. Foundations of Computer Science*, pages 536–546, 2007.

[6] J.-D. Boissonnat, L. J. Guibas, and S. Y. Oudot. Manifold reconstruction in arbitrary dimensions using witness complexes. In *Proc. 23rd Sympos. on Comp. Geom.*, pages 194–203, 2007.

[7] G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian. On the local behavior of spaces of natural images. *International Journal of Computer Vision*, June 2007.

[8] F. Cazals and J. Giesen. Delaunay triangulation based surface reconstruction. In J.D. Boissonnat and M. Teillaud, editors, *Effective Computational Geometry for Curves and Surfaces*, pages 231–273. Springer, 2006.

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

[10] F. Chazal and A. Lieutier. 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] F. Chazal and A. Lieutier. Stability and computation of topological invariants of solids in $\mathbb{R}^n$. *Discrete Comput. Geom.*, 37(4):601–617, 2007.

[13] S.-W. Cheng, T. K. Dey, H. Edelsbrunner, M. A. Facello, and S.-H. Teng. Sliver exudation. *Journal of the ACM*, 47(5):883–904, 2000.

[14] S.-W. Cheng, T. K. Dey, and E. A. Ramos. Manifold reconstruction from point samples. In *Proc. 16th Sympos. Discrete Algorithms*, pages 1018–1027, 2005.

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

[16] V. de Silva. A weak definition of Delaunay triangulation. Technical report, Stanford University, October 2003. To appear in *Geometriae Dedicata*.

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

[18] V. de Silva and R. Ghrist. Coverage in sensor networks via persistent homology. *Algebraic & Geometric Topology*, 7:339–358, 2007.

[19] J. Dugundji. An extension of Tietze's theorem. *Pacific J. Math.*, 1:353–367, 1951.

[20] H. Edelsbrunner. The union of balls and its dual shape. *Discrete Comput. Geom.*, 13:415–440, 1995.

[21] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28:511–533, 2002.

[22] H. Edelsbrunner and N. R. Shah. Triangulating topological spaces. *Int. J. on Comp. Geom.*, 7:365–378, 1997.

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

[24] R Ghrist. Barcodes: The persistent topology of data. *Bull. Amer. Math. Soc.*, October 2007.

[25] K. Grove. Critical point theory for distance functions. In *Proc. of Symposia in Pure Mathematics*, volume 54, 1993.

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

[27] A. Hatcher. *Algebraic Topology*. Cambridge Univ. Press, 2001.

[28] T. Kaczynski, K. Mischaikow, and M. Mrozek. *Computational Homology*. Number 157 in Applied Mathematical Sciences. Springer-Verlag, 2004.

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

[30] S. Y. Oudot. On the topology of the restricted Delaunay triangulation and witness complex in higher dimensions. Technical report, Stanford University, November 2006. LANL arXiv:0803.1296v1 [cs.CG], `http://arxiv.org/abs/0803.1296`.

[31] V. Robins. Towards computing homology from approximations. *Topology*, 24:503–532, 1999.

[32] S. Tsukiyama, M. Ide, H. Ariyoshi, and I. Shirakawa. A new algorithm for generating all the maximal independent sets. *SIAM J. on Computing*, 6:505–517, 1977.

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