Topological Inference via Meshing [*] [†]

Benoît Hudson
Autodesk, Inc.

Gary L. Miller
Computer Science Dept.
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15217
glmiller@cs.cmu.edu

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

Donald R. Sheehy
Computer Science Dept.
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15217
dsheehy@cs.cmu.edu

ABSTRACT

We apply ideas from mesh generation to improve the time and space complexities of computing the full persistent homological information associated with a point cloud P in Euclidean space $\mathbb{R}^d$. Classical approaches rely on the Čech, Rips, α-complex, or witness complex filtrations of P, whose complexities scale up very badly with d. For instance, the α-complex filtration incurs the $n^{\Omega(d)}$ size of the Delaunay triangulation, where n is the size of P. The common alternative is to truncate the filtrations when the sizes of the complexes become prohibitive, possibly before discovering the most relevant topological features. In this paper we propose a new collection of filtrations, based on the Delaunay triangulation of a carefully-chosen superset of P, whose sizes are reduced to $2^{O(d^2)}n$. Our filtrations interleave multiplicatively with the family of offsets of P, so that the persistence diagram of P can be approximated in $2^{O(d^2)}n^3$ time in theory, with a near-linear observed running time in practice. Thus, our approach remains tractable in medium dimensions, say 4 to 10.

Categories and Subject Descriptors

F.2.2 [**Analysis of Algorithms and Problem Complexity**]: Nonnumerical Algorithms and Problems—*Computations on discrete structures, Geometrical problems and computations*

General Terms

Algorithms, Theory

Keywords

topological inference, persistent homology, mesh generation, sparse Voronoi refinement

1. INTRODUCTION

The goal of topological inference is to infer topological invariants of a space given only a sample of the space. In the setting of geometric points, it is natural to consider the point set at different scales by blowing the points up to balls. Persistent homology is a powerful tool for understanding the topological structure of a point cloud across these different scales. Given a point cloud P in Euclidean space $\mathbb{R}^d$, one can build a simplicial complex over P, then associate a time of appearance $t(\sigma)$ to each simplex σ (full-dimensional or lower-dimensional) in the complex. This defines a nested family of simplicial complexes indexed by time, known as a *filtration.* As simplices are added to the complex, the set of topological features (connected components, holes, tunnels, voids, etc.) changes: new ones are created, and old ones destroyed. The persistence algorithm [13, 26] takes the filtration and produces a *persistence diagram* that encodes the lifespans of these topological features. Given an appropriate choice of filtration, one can prove that short-lived features correspond to sampling noise, while long-lived features correspond to relevant topological structures underlying the input set.

Several filtrations have been used with success in the past, including the α-complex [12] and witness complex [10, 11] filtrations, which are based on the Delaunay triangulation of P or an approximation of it, and the Čech [14] and Vietoris-Rips [25] filtrations, which are derived from the nerves of collections of congruent balls centered at the data points. The ability of these filtrations to capture the homological information associated with a point cloud is certified by a well-founded theory [3], which largely explains their success from a theoretical point of view. In practice however, the cost to build them makes their use prohibitive, even in medium dimensions, say 4 to 10. When α becomes large, the size of

the α-complex approaches that of the Delaunay triangulation: $n^{\Omega(d)}$ for n data points in d dimensions, even for some relatively "nice" inputs [15, 16]. The sizes of the Čech, Rips and (relaxed) witness complexes grow even more quickly, as $2^{\Omega(n)}$.

To avoid this issue, researchers usually resort to truncating the filtrations at a prescribed size limit. This truncation is equivalent to looking at the data at small scales only, and can make the algorithm miss relevant structures at larger scales. A typical scenario, inspired from [17], is described in Figure 1 (left): it consists of a point cloud sampled evenly from a helicoidal curve drawn on the Clifford torus in $\mathbb{R}^4$. In this case, the point cloud admits at least three candidate underlying spaces: at a small scale, the curve; at a larger scale, the torus; and at an even larger scale, the 3-sphere of radius $\sqrt{2}$ on which the Clifford torus is naturally embedded. One might also add the point cloud itself and $\mathbb{R}^4$ at either ends of the spectrum.

In order to analyze such data sets at different scales using only truncated filtrations, Chazal and Oudot [6] proposed a *landmarking* strategy in the spirit of [17], which maintains a growing subset of the data points, on which the simplicial complexes are built. However, their approach produces a weaker form of data representation than persistence diagrams, which does not explicitly correlate the features visible at different scales. As a result, they can get false positives when retrieving the set of persistent topological features. See e.g. [17, Fig. 7] for an example.

A simple algorithm. In this paper, we use techniques from finite element mesh generation to build a complex on the input set with some extra vertices that is guaranteed to have small size. We then order the simplices in a natural and easy to compute way in order to generate a filtration. This complex can then be input directly to standard algorithms to discover the persistent topological features. In its simplest form our algorithm can be implemented using both the mesh generator and the persistence computation as black boxes. Thus, the only new code to implement is the filtering of the mesh.

Consequently, the work in this paper is primarily in the analysis of the theoretical guarantees. We show how varying the mesh generation techniques, one can get different guarantees with respect to the size of the complex and the tightness of the approximation.

Enter Sparse Voronoi Refinement. We preprocess the point cloud P using techniques inspired by Delaunay refinement, iteratively inserting new points of $\mathbb{R}^d$ called *Steiner points* into P until some quality criterion is met. Here, quality will be measured by the aspect ratios of the Voronoi cells, so as to guarantee that the size of the Delaunay triangulation of the augmented set $P \cup S$ drops down to $2^{O(d^2)}(n + |S|)$ when the criterion is met. Furthermore, the number of Steiner points needed to meet the criterion is $2^{O(d \log d)}n$, which makes the size of the final triangulation only $2^{O(d^2)}n$. In order to realize the benefits of the refined mesh, we compute it without first constructing the Delaunay triangulation, as is possible using the Sparse Voronoi Refinement (SVR) algorithm [19]. In addition, we partition the input into *well-paced* sets which guarantees that the size of the filtration

stays linear in n, modulo a constant factor that depends exponentially on d [21].

Once the augmented point cloud $P \cup S$ has been computed, we order the simplices of its Delaunay triangulation according to a filter $t : \text{Del}(P \cup S) \to \mathbb{R}$. Several different filters are analyzed in the paper, yielding filtrations with different properties: some are easier to build, others come with better approximation guarantees. The choice of a particular filter depends on the application considered and is therefore left to the user. Note that all our filters are based on distances to the input point cloud P, as illustrated in Figure 2. This enables us to show that the corresponding filtrations are interleaved on a logarithmic scale with the filtration of the offsets of P, in the sense of [4], from which we can deduce that they produce accurate (approximate) persistence diagrams. Computing the persistence diagram takes time cubic in the number of simplices and thus dominates our worst-case $2^{O(d^2)}n^3$ overall runtime. This bound, though large, is still a significant improvement over $n^{\Omega(d)}$. Moreover, in practice, the persistence diagram computation takes near-linear time (on an input with $2^{O(d^2)}n$ simplices), which makes our approach tractable in small to medium dimensions (4-10) for moderate input sizes (thousands to tens of thousands of points). A preliminary implementation bears out these predictions (see Section 6).

Layout of the paper. In Section 2 we recall the necessary background on Sparse Voronoi Refinement and on persistent homology. The rest of the paper is devoted to the description of our approach. We first present a simplified version in Section 3 that produces a filtration that is $\log(\rho)$-interleaved with the offsets filtration of P, for some constant $\rho \geq 2$. The size of this filtration is $2^{O(d^2)}n \log(\Delta(P))$, where $\Delta(P)$ denotes the *spread* of P. We then show in Section 4 how the interleaving between our filtration and the offsets filtration can be tightened, so that we can produce persistence diagrams that are accurate within any arbitrarily small error. Finally, in Section 5 we concentrate on the size of the filtration and show how to eliminate its dependence on the spread by a recursive decomposition of the input.

2. PRELIMINARIES

Throughout the paper, the ambient space is $\mathbb{R}^d$, endowed with the standard Euclidean norm, noted $|\cdot|$. We use singular homology with coefficients in a field, omitted in our notations for simplicity. We refer the reader to [18] for a thorough introduction to homology theory.

2.1 Clipped Voronoi Diagrams and Sparse Voronoi Refinement

Let P be a finite set of distinct points lying in general position in $\mathbb{R}^d$. We denote by $\text{Vor}(P)$ the *Voronoi diagram* of P, defined as a collection of closed cells $\{\text{Vor}(p) : p \in P\}$, where each cell $\text{Vor}(p)$ is the locus of the points of $\mathbb{R}^d$ that are at least as close to p as to any other point of P. Its dual complex is known as the *Delaunay triangulation* $\text{Del}(P)$. Since P lies in general position, $\text{Del}(P)$ is an embedded simplicial complex in $\mathbb{R}^d$, whose underlying space coincides with the convex hull of P.

Given an axis-aligned box BB containing P, we consider the restrictions of the Voronoi diagram and Delaunay triangulation to BB. Specifically, given a point $p \in P$, we
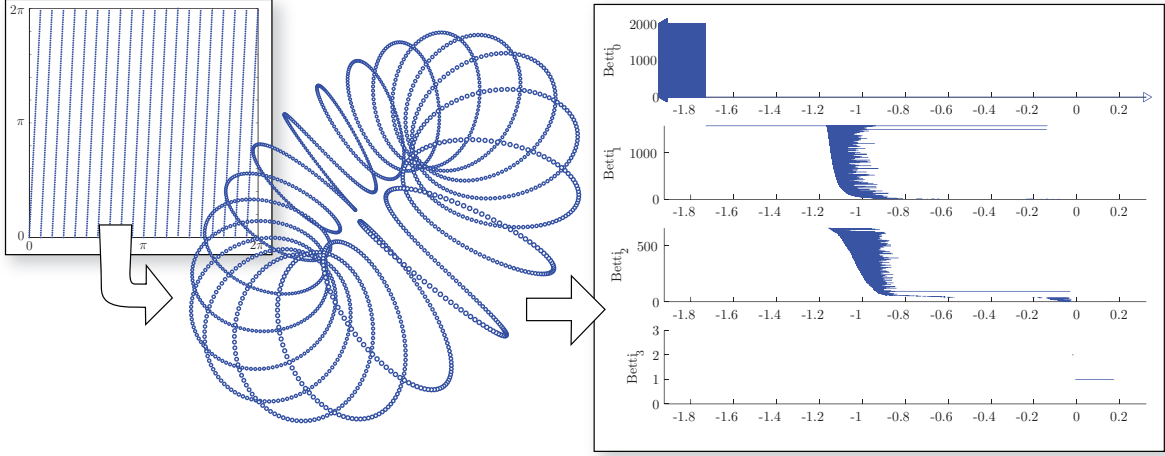
Figure 1: The Clifford data set. Left: point cloud sampled uniformly along a periodic curve in $[0, 2\pi]^2$, then mapped onto a helicoidal curve drawn on the Clifford torus in $\mathbb{R}^4$ via the canonical embedding $(u, v) \mapsto (\cos u, \sin u, \cos v, \sin v)$. Right: log-scale barcode obtained on this data set using the filtration of Section 3.2.

call $\mathrm{Vor}_\square(p)$ its Voronoi cell clipped to BB: $\mathrm{Vor}_\square(p) = \mathrm{Vor}(p) \cap BB$. We call $\mathrm{Vor}_\square(P)$ the Voronoi diagram clipped to BB, and $\mathrm{Del}_\square(P)$ its dual complex, which is a subcomplex of $\mathrm{Del}(P)$. For a clipped Voronoi cell $\mathrm{Vor}_\square(p)$, we let R_p be the radius of the smallest Euclidean ball centered at p that contains all of $\mathrm{Vor}_\square(p)$, and we let r_p be the radius of the largest Euclidean ball centered at p that is entirely contained in $\mathrm{Vor}_\square(v)$. We define the *aspect ratio* of the clipped Voronoi cell to be R_p/r_p.

*Sparse Voronoi Refinement (*SVR*).* The SVR algorithm takes a finite point cloud P as input and returns a finite superset M of P that satisfies the following properties:

(i) M is a point sampling of some axis-aligned bounding box BB of side length $O(\mathrm{diam}(P))$ around P,

(ii) $\mathrm{Del}_\square(M) = \mathrm{Del}(M)$, where $\mathrm{Del}_\square(M)$ denotes the Delaunay triangulation clipped to BB,

(iii) the clipped Voronoi cells have aspect ratios bounded from above by an absolute constant $\rho \geq 2$,

(iv) The size of $\mathrm{Del}(M)$ is $2^{O(d^2)}|M|$,

(v) The size of M is $2^{O(d)}n\log(\Delta(P))$, where $\Delta(P)$ denotes the *spread* of P, i.e. the ratio of the largest to smallest interpoint distances among the points of P.

As shown in [20], the extra work needed to fill in the entire bounding box BB with point samples is negligible. The points of $S = M \setminus P$ added by the algorithm are called *Steiner points*. We will refer to the point set M along with its Delaunay triangulation as the *mesh*. The SVR algorithm can produce M in near-optimal $2^{O(d)}n\log(\Delta(P)) + O(|M|)$ time [19]. As shown in [21], it is possible to reduce the output-sensitive term $|M|$ to $2^{O(d\log d)}n$ by applying SVR to well-chosen subsets of the input called *well-paced sets*. This technique will be used in Section 5 to improve the complexity of our method, which uses SVR as a black box.

2.2 Filtrations, persistence diagrams and stability

A *filtration* is a one-parameter family $\mathcal{F} = \{F_\alpha\}_{\alpha \geq 0}$ of topological spaces that is nested with respect to inclusion,

that is: $F_\alpha \subseteq F_\beta$ for all $\beta \geq \alpha \geq 0$. Persistence theory describes the evolution of the homology of the sets F_α as α ranges from 0 to $+\infty$. This is done through a special type of planar representation called a *persistence diagram*. When the parameter of $\mathcal{F}$ is clear and varies over all of $[0, +\infty]$, we write simply $\{F_\alpha\}$ rather than $\{F_\alpha\}_{\alpha \geq 0}$. Given a discrete subset $A = \{\cdots, \alpha_i, \alpha_j, \alpha_k, \cdots\}$ of $[0, +\infty)$ that has no accumulation point, the canonical inclusions $\cdots \hookrightarrow F_{\alpha_i} \hookrightarrow F_{\alpha_j} \hookrightarrow F_{\alpha_k} \hookrightarrow \cdots$ induce a directed system of vector spaces involving the r-dimensional homology groups:

$$\cdots \longrightarrow H_r(F_{\alpha_i}) \longrightarrow H_r(F_{\alpha_j}) \longrightarrow H_r(F_{\alpha_k}) \longrightarrow \cdots$$

If all the vector spaces are finite-dimensional, then the filtration, $\mathcal{F}$, is said to be *tame*, and the algebraic structure of this system can be encoded as a multi-set of points in $[0, +\infty]^2$. The *r-th persistence diagram* of $\mathcal{F}$ is then obtained by considering a growing family $\{A_i\}_{i \in \mathbb{N}}$ of discrete sets, whose union is dense in $[0, +\infty)$, and by taking the well-defined limit of their corresponding multi-sets, which does not depend on the choice of the family $\{A_i\}_{i \in \mathbb{N}}$. The union of all such diagrams for r ranging over $\mathbb{N}$ is called the *persistence diagram* of $\mathcal{F}$, noted $\mathrm{D}\mathcal{F}$. Intuitively, each point $p \in \mathrm{D}f$ encodes the lifespan of some homological feature (connected component, hole, void, etc.) appearing at time p_x and dying at time p_y in the filtration. For formal developments on this topic, please refer to [4], whose framework is adopted here.

An important property of persistence diagrams is to be stable under small perturbations of the filtrations. Proximity between filtrations is defined in terms of mutual nesting: specifically, two tame filtrations $\mathcal{F}, \mathcal{G}$ are said to be ε-*interleaved* if we have $F_\alpha \subseteq G_{\alpha+\varepsilon}$ and $G_\alpha \subseteq F_{\alpha+\varepsilon}$ for all $\alpha \geq 0$. Under this condition, it is known that the persistence diagrams $\mathrm{D}\mathcal{F}$ and $\mathrm{D}\mathcal{G}$ are ε-close in the bottleneck distance [4, 8]. Recall that the bottleneck distance $\mathrm{d}_B^\infty(A, B)$ between two multi-sets $A, B \subset [0, +\infty]^2$ is defined as the quantity $\min_\gamma \max_{p \in A} \|p - \gamma(p)\|_\infty$, where $\|\cdot\|_\infty$ denotes the l_∞-norm and γ ranges over all bijections from A to B. To make sure that such bijections always exist, all persistence diagrams

are enriched with the diagonal $\{(x,x) : x \in [0, +\infty]\}$, whose multiplicity is set to infinity. The formal statement goes as follows:

THEOREM 2.1. (STABILITY [4, 8]). *If two tame filtrations $\mathcal{F}, \mathcal{G}$ are ε-interleaved, then $d_B^\infty(D\mathcal{F}, D\mathcal{G}) \leq \varepsilon$.*

Multiplicative interleaving. In this paper we consider filtrations $\mathcal{F}, \mathcal{G}$ that are ε-interleaved *multiplicatively*, that is: $F_\alpha \subseteq G_{\alpha\varepsilon}$ and $G_\alpha \subseteq F_{\alpha\varepsilon}$ for all $\alpha \geq 0$. Consider $\log \mathcal{F}$ and $\log \mathcal{G}$, the reparametrizations of the filtrations $\mathcal{F}$ and $\mathcal{G}$ on the natural logarithmic scale[1]: $\forall \alpha \in \mathbb{R},\ \log F_\alpha = F_{\exp(\alpha)}$ and $\log G_\alpha = G_{\exp(\alpha)}$. Multiplicative ε-interleaving of $\mathcal{F}$ and $\mathcal{G}$ implies additive $\log(\varepsilon)$-interleaving of their reparametrizations, $\log \mathcal{F}$ and $\log \mathcal{G}$, that is: $\forall \alpha \in \mathbb{R},\ \log F_\alpha \subseteq \log G_{\alpha + \log \varepsilon}$ and $\log G_\alpha \subseteq \log F_{\alpha + \log \varepsilon}$.

As a result, multiplicative interleaving of filtrations implies the following weaker form of proximity between their persistence diagrams, where the notation $d_D^{\log}(\mathcal{F}, \mathcal{G})$ stands for the quantity $d_B^\infty(D \log \mathcal{F}, D \log \mathcal{G})$:

COROLLARY 2.2. *If two filtrations $\mathcal{F}, \mathcal{G}$ are ε-interleaved multiplicatively, then $d_D^{\log}(\mathcal{F}, \mathcal{G}) \leq \log \varepsilon$.*

The persistence diagram of a simplicial filtration (i.e. a finite family of nested finite abstract simplicial complexes) can be computed using the persistence algorithm [13, 26]. Its runtime is determined by the total number of simplices m in the filtration; in the worst case, it can take $O(m^3)$ time, although in practice near-linear times are common. Constructing the filtration itself will take us near-linear time as well: reducing the size of the filtration represents a large win for computing persistent homology.

2.3 Filters and projections

In addition to filtering simplicial complexes, a filtration can also be defined by considering a topological space X and a non-negative function $t : X \to [0, +\infty)$, called a *filter*, which encodes the time at which each point of X appears in the filtration. The spaces forming the filtration are the sublevel-sets of t, i.e. the sets of the form $F_\alpha = t^{-1}([0, \alpha])$. In this setting, a filtered simplicial complex is a piecewise constant function that maps the interior of a simplex σ to $t(\sigma)$.

Distance functions are an important class of filters considered in the paper. Given a compact set P in Euclidean space $\mathbb{R}^d$, let $d_P(x)$ denote the distance from $x \in \mathbb{R}^d$ to the nearest point in P:

$$d_P(x) = \min_{p \in P} |x - p|.$$

For any $\alpha \geq 0$, the sublevel-set $P^\alpha = d_P^{-1}([0, \alpha])$ is the union of the closed Euclidean balls of same radius α about the points of P, called the α-*offset* of P. The family of sublevel-sets of d_P is thus known as the *offsets filtration* of P in the literature. This filtration has played an important role in topological inference from point cloud data, where it has been used as a central theoretical tool for proving the correctness of existing methods [5, 6, 8, 22].

We will also be using a related distance function induced by $P \subset \mathbb{R}^d$. The Ruppert local feature size, $f_P(x)$, is the distance from x to its second nearest neighbor in P [23]:

$$f_P(x) = \min_{p_1 \neq p_2 \in P} \max\{|x - p_1|, |x - p_2|\}.$$

[1]Throughout the paper, log denotes the natural logarithm.

The triangle inequality implies that f_P is 1-Lipschitz. The function f_P is a standard tool in mesh generation as it bounds the size of Voronoi cells locally in an optimal, bounded aspect ratio mesh.

To conclude this preliminary section, let us mention another important function derived from a compact set P: the metric projection $\pi_P : \mathbb{R}^d \to \mathbb{R}^d$, defined by $\pi_P(x) = \operatorname{argmin}_{y \in P} |x - y|$ for all $x \in \mathbb{R}^d$. This function will be instrumental in showing topological equivalence between various families of spaces. In particular, we will rely on the following classical result of convex geometry, whose proof is recalled for completeness in the full version of the paper:

LEMMA 2.3. *If $P \subset \mathbb{R}^d$ is compact and convex, then the projection π_P is well-defined and 1-Lipschitz over $\mathbb{R}^d$.*

3. THE α-MESH FILTRATION

Our strategy is to build a superset M of the input point set P, and then to filter the Delaunay triangulation of M in order to obtain a filtration that can be related to the offsets filtration of P. In Section 3.1 we present a simplified version of our filter that yields a partial approximation (Theorem 3.5). This version may not always be useful in practice, however it is more intuitive and will enable us to emphasize the key ingredients of our approach. In Section 3.2 we will reuse these ingredients and expand upon them to improve our filter and obtain a full approximation guarantee (Theorem 3.7).

3.1 Basic filter

Our input is a finite set P of points in general position in $\mathbb{R}^d$. We first apply the SVR algorithm to construct a superset $M \supseteq P$ that satisfies conditions (i) through (v) of Section 2.1. We then define the filter $t : \text{Del}(M) \to \mathbb{R}$ as follows, identifying each k-simplex $\sigma \in \text{Del}(M)$ with its vertex set $\{v_0, \ldots, v_k\} \subseteq M$:

$$t(\sigma) = \max_{v \in \sigma} d_P(v).$$

We define the α-*mesh filtration* $\{D_M^\alpha\}$ as the sublevel-sets filtration of t: for every $\alpha \geq 0$, let D_M^α be the subcomplex of $\text{Del}(M)$ made of all simplices σ with $t(\sigma) \leq \alpha$. Note that if τ is a face of σ then $t(\tau) \leq t(\sigma)$, so the spaces forming the filtration are proper simplicial complexes. Furthermore, $D_M^\alpha \subseteq D_M^\beta$ for all $\beta \geq \alpha \geq 0$.
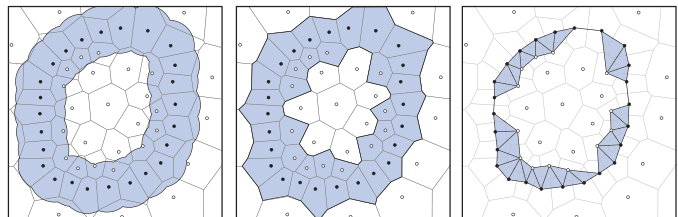
Figure 2: From left to right: the offset P^α, the α-Voronoi V_M^α and its dual α-mesh D_M^α.

Intuitively, our basic filter sorts the simplices of $\text{Del}(M)$ according to their distances to P, in order to simulate within $\text{Del}(M)$ the growth of the offsets of P — see Figure 2 for an illustration. Our analysis shows that this simulation process works thanks to the fact that Voronoi cells have bounded aspect ratios.

Theoretical analysis. Our goal is to relate $\{D_M^\alpha\}$ to the offsets filtration $\{P^\alpha\}$. We do the analysis in terms of a dual filtration, $\{V_M^\alpha\}$, based on the clipped Voronoi diagram $\mathrm{Vor}_\square(M)$ — see Figure 2 (center) for an example. To each point $v \in M$ we assign a closed convex set $U_\alpha(v)$ as follows:

$$U_\alpha(v) = \left\{ \begin{array}{ll} \emptyset & \text{if } \alpha < t(v), \\ \mathrm{Vor}_\square(v) & \text{otherwise.} \end{array} \right.$$

The filtration $\{V_M^\alpha\}$ is defined by $V_M^\alpha = \bigcup_{v \in M} U_\alpha(v)$ for all $\alpha \geq 0$, and the collection $\mathcal{U}_\alpha = \{U_\alpha(v)\}_{v \in M}$ forms a closed cover of V_M^α. Let $\mathcal{N}\mathcal{U}_\alpha$ denote its nerve, i.e. the simplicial complex containing one k-simplex for each non-empty intersection of $k + 1$ elements in $\mathcal{U}_\alpha$. Both D_M^α and $\mathcal{N}\mathcal{U}_\alpha$ are embedded as subcomplexes of the full simplex 2^M over the vertex set M, and the following lemma stresses their relationship:

LEMMA 3.1. *For all $\alpha \geq 0$, the subcomplexes D_M^α and $\mathcal{N}\mathcal{U}_\alpha$ of the full simplex 2^M are equal.*

PROOF. Consider first the case of a zero-dimensional simplex $\sigma = \{v\}$. The definition of D_M^α states that $\{v\} \in D_M^\alpha$ if and only if $\mathrm{d}_P(v) \leq \alpha$, which is also the criterion for which $U_\alpha(v)$ is not empty and hence belongs to the collection $\mathcal{U}_\alpha$. Thus, $\{v\} \in D_M^\alpha \Leftrightarrow \{v\} \in \mathcal{N}\mathcal{U}_\alpha$.

Consider now the case of a higher-dimensional simplex $\sigma = \{v_0, \cdots, v_k\}$. The definition of D_M^α states that $\sigma \in D_M^\alpha$ if and only if $\sigma \in \mathrm{Del}(M)$ and $\max_i t(v_i) \leq \alpha$, which is equivalent to $\bigcap_{i=0}^k \mathrm{Vor}(v_i) \neq \emptyset$ and $\max_i t(v_i) \leq \alpha$. By assertion (ii) of Section 2.1, we have $\bigcap_{i=0}^k \mathrm{Vor}(v_i) \neq \emptyset \Leftrightarrow \bigcap_{i=0}^k \mathrm{Vor}(v_i) \cap BB \neq \emptyset \Leftrightarrow \bigcap_{i=0}^k \mathrm{Vor}_\square(v_i) \neq \emptyset$. Hence, $\sigma \in D_M^\alpha$ if and only if $\sigma \in \mathcal{N}\mathcal{U}_\alpha$. $\square$

Since the sets $U_\alpha(v)$ in the cover of V_M^α are convex, the Nerve Theorem [18, §4G] implies that V_M^α and its nerve $\mathcal{N}\mathcal{U}_\alpha$ are homotopy equivalent. Furthermore, since the sets $U_\alpha(v)$ are monotonically increasing with α, it follows from the Persistent Nerve Theorem [6] that the filtration $\{V_M^\alpha\}$ and the family of nerves $\{\mathcal{N}\mathcal{U}_\alpha\}$ have identical persistence diagrams (see the full version of the paper for the details of the argument). Combined with Lemma 3.1, this fact draws a connection between the persistence diagrams of $\{V_M^\alpha\}$ and $\{D_M^\alpha\}$:

LEMMA 3.2. *The persistence diagrams of the filtrations $\{V_M^\alpha\}$ and $\{D_M^\alpha\}$ are identical.*

Let the clipped offsets be defined as follows, in analogy with the clipped Voronoi cells: $P_\square^\alpha = P^\alpha \cap BB = \{x \in BB \mid \mathrm{d}_P(x) \leq \alpha\}$. Let also $r_P = \frac{1}{2} \max_{p \in P} \mathrm{d}_{M\setminus\{p\}}(p)$. Using the fact that the clipped Voronoi cells have bounded aspect ratios, we can show that the clipped offsets filtration is multiplicatively interleaved with $\{V_M^\alpha\}$, starting at time $\alpha = r_P$:

LEMMA 3.3. *For all $\alpha \geq r_P$, we have $V_M^{\alpha/\rho} \subseteq P_\square^\alpha \subseteq V_M^{\rho\alpha}$.*

PROOF. Let $x \in V_M^{\alpha/\rho} \subseteq BB$, and let $v \in M$ be such that $x \in U_{\alpha/\rho}(v)$. Let also $p \in P$ be closest to v. If $v \in P$, then assertion (iii) of Section 2.1 implies that $\mathrm{d}_P(x) \leq |x - v| \leq \frac{\rho}{2}\mathrm{d}_{M\setminus\{v\}}(v) \leq r_P \leq \alpha$. If $v \in M \setminus P$, then the fact that $U_{\alpha/\rho}(v) \neq \emptyset$ implies that $|v - p| \leq \alpha/\rho$ and $U_{\alpha/\rho}(v) = \mathrm{Vor}_\square(v)$. This implies that $|x - p| \leq |x - v| + |v - p| \leq |x - v| + \alpha/\rho$. Now, assertion (iii) of Section 2.1 guarantees that the aspect ratio of $\mathrm{Vor}_\square(v)$ is at most ρ, implying that $|x - v| \leq \frac{\rho}{2}\mathrm{d}_{M\setminus\{v\}}(v) \leq \frac{\rho}{2}|v - p| \leq \frac{\alpha}{2}$. Thus, $\mathrm{d}_P(x) \leq |x - p| \leq \alpha(\frac{1}{2} + \frac{1}{\rho})$. Since $\rho \geq 2$, we conclude that

$\mathrm{d}_P(x) \leq \alpha$. Hence, in all cases we have $\mathrm{d}_P(x) \leq \alpha$, which means that $x \in P_\square^\alpha$.

Let now $x \in P_\square^\alpha$, and let $v \in M$ and $p \in P$ be closest to x. Then, x belongs both to $\mathrm{Vor}_\square(v)$ and to the Euclidean ball of center p and radius α. It follows that $\mathrm{d}_P(v) \leq |v - p| \leq |v - x| + |x - p| \leq 2|x - p| \leq 2\alpha \leq \rho\alpha$. This means that $U_{\rho\alpha}(v) = \mathrm{Vor}_\square(v)$, which contains x. As a consequence, we have $x \in V_M^{\rho\alpha}$. $\square$

Finally, we relate the clipped offsets filtration to the real offsets filtration:

LEMMA 3.4. *For all $\alpha \geq 0$, the canonical inclusion $P_\square^\alpha \hookrightarrow P^\alpha$ is a homotopy equivalence.*

PROOF. Let π_{BB} denote the metric projection onto BB, that is: $\pi_{BB}(x) = \mathrm{argmin}_{y \in BB}|x - y|$. Since BB is compact and convex, Lemma 2.3 ensures that π_{BB} is well-defined and 1-Lipschitz over the entire space $\mathbb{R}^d$. Let x be a point of P^α, and let $x' = \pi_{BB}(x)$. We will show that the line segment $[x, x']$ is included in P^α. Let $p \in P$ be such that $|x - p| \leq \alpha$. Since BB contains P, we have $\pi_{BB}(p) = p$, and therefore $|p - x'| \leq |p - x|$ since π_{BB} is 1-Lipschitz. It follows that both x and x' belong to the ball of center p and radius α. Since this ball is convex, it contains in fact the whole line segment $[x, x']$.

Let now $F : [0, 1] \times P^\alpha \to \mathbb{R}^d$ be defined by $F(t, x) = (1 - t)x + t\pi_{BB}(x)$. Since π_{BB} is 1-Lipschitz, F is continuous. In addition, the above discussion shows that $F(t, P^\alpha) \subseteq P^\alpha$ for all $t \in [0, 1]$. Also, since $P_\square^\alpha \subseteq BB$, the restriction of π_{BB} to $P_\square^\alpha$ is the identity, therefore so is the restriction of F. Finally, for all $x \in P^\alpha$ we have $F(1, x) = \pi_{BB}(x) \in P^\alpha \cap BB = P_\square^\alpha$. Hence, F is a deformation retraction of P^α onto $P_\square^\alpha$, which implies that the canonical inclusion $P_\square^\alpha \hookrightarrow P^\alpha$ is a homotopy equivalence. $\square$

With the above results we can compare the diagrams of the *truncated*[2] filtrations $\{P^\alpha\}_{\alpha \geq r_P}$ and $\{D_M^\alpha\}_{\alpha \geq r_P}$:

THEOREM 3.5. *On the natural logarithmic scale, the persistence diagrams of the truncated filtrations $\{P^\alpha\}_{\alpha \geq r_P}$ and $\{D_M^\alpha\}_{\alpha \geq r_P}$ are $\log \rho$-close in the bottleneck-distance, i.e.*
$$\mathrm{d}_\mathrm{D}^{\log}(\{P^\alpha\}_{\alpha \geq r_P}, \ \{D_M^\alpha\}_{\alpha \geq r_P}) \leq \log \rho.$$

PROOF. By Lemma 3.4, the canonical inclusions $P_\square^\alpha \hookrightarrow P^\alpha$ and $P_\square^\beta \hookrightarrow P^\beta$ are homotopy equivalences that commute with the inclusions $P_\square^\alpha \hookrightarrow P_\square^\beta$ and $P^\alpha \hookrightarrow P^\beta$ for all $\beta \geq \alpha \geq 0$, so the filtrations $\{P^\alpha\}$ and $\{P_\square^\alpha\}$ have identical persistence diagrams. In addition, Lemma 3.2 implies that $\{D_M^\alpha\}_{\alpha \geq 0}$ and $\{V_M^\alpha\}$ have identical diagrams. The result follows then from the interleaving of the truncated filtrations $\{P_\square^\alpha\}_{\alpha \geq r_P}$ and $\{V_M^\alpha\}_{\alpha \geq r_P}$ (Lemma 3.3) and its consequences on the proximity of their diagrams (Corollary 2.2). $\square$

Intuitively, Theorem 3.5 means that homological features appearing in the offsets filtration after time $\alpha = r_P$ are captured by the α-mesh filtration, with approximately the same birth and death times on the natural logarithmic scale; features appearing before r_P and dying after r_P are also captured, but starting at times potentially as late as r_P,

---

[2]Although these filtrations are only indexed over a subinterval of $[0, +\infty)$, their persistence diagrams can be defined using the same process as in Section 2.2, and the proofs of Theorem 2.1 and Corollary 2.2 carry over.

the death times remaining approximately the same; finally, features appearing and dying before r_P may not be captured at all by the α-mesh filtration.

3.2 Full filter

The basic filter only enables us to approximate the persistence diagram of the offsets filtration after a certain time $\alpha = r_P$ (Theorem 3.5). The reason for this appears clearly in the proof of Lemma 3.3: even though we have $P_\square^\alpha \subseteq V_M^{\rho\alpha}$ for all $\alpha \geq 0$, the symmetric inclusion $V_M^\alpha \subseteq P_\square^{\rho\alpha}$ only holds when $\alpha \geq r_P$, since the clipped Voronoi cells of the input points appear in V_M^α at time $\alpha = 0$ and are not covered by P^α before $\alpha = \rho r_P$. In the dual α-mesh, this phenomenon translates into the appearance of edges between the points of P at time $\alpha = 0$, whereas such edges should normally appear when α-balls around these points touch one another.

In this section we propose a solution to this issue, which consists in modifying the filter of $\mathrm{Del}(M)$ so as to somewhat delay the appearances of the simplices incident to the points of P in the α-mesh filtration. The rest of the approach remains unchanged: we apply SVR on the input point cloud P, to get a vertex set $M \supseteq P$, then we define a modified filter $\tilde{t} : \mathrm{Del}(M) \to \mathbb{R}$ and build its sublevel-sets filtration $\{\tilde{D}_M^\alpha\}$.

Filter modification. We follow the above recommendations and modify our filter to be a little more careful in the region close to the input points. Let $s(v) = \frac{1}{2} f_M(v)$ if $v \in P$ and $s(v) = \mathrm{d}_P(v)$ otherwise, where f_M denotes the Ruppert local feature size of M. Then, we modify our filter as follows:

$$\tilde{t}(\sigma) = \begin{cases} d_P(v) & \text{if } \sigma \text{ is a vertex } v \in M, \\ \max_{v \in \sigma} s(v) & \text{otherwise.} \end{cases}$$

In other words, we delay the time at which Delaunay simplices incident to the points of P appear, although the points themselves still appear as vertices at time $\alpha = 0$. The amount of the delay is based on the Ruppert local feature size of the augmented point set M.

Theoretical analysis. We resort again to a dual filtration $\{\tilde{V}_M^\alpha\}$ with $\tilde{V}_M^\alpha = \bigcup_{v \in M} \tilde{U}_\alpha(v)$, where the sets $\tilde{U}_\alpha(v)$ are slightly modified versions of the sets $U_\alpha(v)$ from Section 3.1:

$$\tilde{U}_\alpha(v) = \begin{cases} \emptyset & \text{if } \alpha < \tilde{t}(v), \\ \mathrm{ball}(v, \alpha) & \text{if } v \in P \text{ and } \tilde{t}(v) \leq \alpha < s(v), \\ \mathrm{Vor}_\square(v) & \text{otherwise.} \end{cases}$$

Making $\tilde{U}_\alpha(v)$ equal to $\mathrm{ball}(v, \alpha)$ instead of the empty set when $v \in P$ and $\alpha < s(v)$ enables us to obtain an interleaving between the entire filtrations $\{\tilde{V}_M^\alpha\}$ and $\{P_\square^\alpha\}$:

LEMMA 3.6 (ANALOG OF LEMMA 3.3). *For all $\alpha \geq 0$, we have $\tilde{V}_M^{\alpha/\rho} \subseteq P_\square^\alpha \subseteq \tilde{V}_M^{\rho\alpha}$.*

PROOF. Let x be a point of $\tilde{V}_M^{\alpha/\rho} \subseteq BB$, and let $v \in M$ be such that $x \in \tilde{U}_{\alpha/\rho}(v)$. If $v \in P$ with $\mathrm{d}_{M \setminus \{v\}}(v) > 2\alpha/\rho$, then $x \in \mathrm{ball}(v, \alpha/\rho)$ and thus $x \in P_\square^{\alpha/\rho} \subseteq P_\square^\alpha$. If $v \in P$ with $\mathrm{d}_{M \setminus \{v\}}(v) \leq 2\alpha/\rho$, then condition (iii) of Section 2.1 guarantees that $|x - v| \leq \alpha$ and thus $x \in P_\square^\alpha$. If $v \in S$, then the analysis is exactly the same as in the proof of Lemma 3.3. So, we have $\tilde{V}_M^{\alpha/\rho} \subseteq P_\square^\alpha$.

For the second inclusion, let x be a point of $P_\square^\alpha$. We want to show that $x \in \tilde{V}_M^{\rho\alpha}$. Let $v \in M$ be closest to x. If $v \in P$, then $x \in \mathrm{ball}(v, \alpha) \cap \mathrm{Vor}_\square(v)$, which is included in $\tilde{U}_\alpha(v)$

by definition (recall that we have $\tilde{t}(v) = 0$). As a result, $x \in \tilde{V}_M^{\rho\alpha}$. If $v \in S$, then the analysis is the same as in the proof of Lemma 3.3. So, we have $P_\square^\alpha \subseteq \tilde{V}_M^{\rho\alpha}$. $\square$

The rest of the analysis is roughly the same as in Section 3.1, and we refer the reader to the full version of the paper for the details:

THEOREM 3.7 (ANALOGUE OF THEOREM 3.5). *On the natural logarithmic scale, the persistence diagrams of $\{P^\alpha\}$ and $\{\tilde{D}_M^\alpha\}$ are $\log \rho$-close in the bottleneck-distance, that is,*
$$\mathrm{d}_D^{\log}\left(\{P^\alpha\}, \{\tilde{D}_M^\alpha\}\right) \leq \log \rho.$$

4. TIGHTER INTERLEAVING VIA OVERMESHING

It follows from Theorem 3.7 that the approximation error induced by our approach diminishes as parameter ρ shrinks. However, recall from assertion (iii) of Section 2.1 that the mesher requires $\rho \geq 2$ in order to terminate. In this section we show how we can reduce the approximation error to any arbitrarily small $\varepsilon > 0$ by inserting a controlled number of extra Steiner points and by using a more elaborate filter, which rids our analysis of its dependence on parameter ρ. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a sizing function. As long as f is bounded from above by the Ruppert local feature size f_P, SVR can return a mesh M such that the radius R_v of every Voronoi cell $\mathrm{Vor}(v)$ is at most $f(v)$. Given a parameter $\varepsilon > 0$, let $f(x) = \frac{1}{3(1+1/\varepsilon)} f_P(x)$. This means that for any $v \in M$ and any $x \in \mathrm{Vor}_\square(v)$, $f_P(v) \geq 3(1 + \frac{1}{\varepsilon})|x - v|$. The standard mesh size analysis implies that the size m of our output mesh M will be bounded as follows:

$$m = O\left(\int_{BB} \frac{1}{f(z)^d} dz\right) = O\left(\left(3(1 + \frac{1}{\varepsilon})\right)^d \int_{BB} \frac{1}{f_P(z)^d} dz\right).$$

In other words, our new sizing function f will only increase the mesh size by a factor of $\left(3(1 + \frac{1}{\varepsilon})\right)^d$.

Modified α-mesh filtration. As before, we run SVR on the input point set P, but this time using the sizing function f described above. Letting M denote the output superset of P, we modify the filter on $\mathrm{Del}(M)$ in such a way that the Voronoi cells of mesh vertices that are significantly closer to a given point $p \in P$ than to the others appear only once p lies within $\alpha/2$ of its nearest neighbor in $P \setminus \{p\}$.

More precisely, for every point $x \in \mathbb{R}^d$ let n_x denote the point of P closest to x — if there are two or more such points, then choose either of them as n_x. We define the following function on the mesh vertices:

$$\forall v \in M, \ s'(v) = \max\left\{\mathrm{d}_P(v), \frac{1}{2} f_P(n_v)\right\}. \quad (1)$$

Note that when v belongs to P, we have $n_v = v$ and $s'(v) = \frac{1}{2} f_P(v)$. Also, if v is equidistant to two vertices $p, q \in P$, then $\mathrm{d}_P(v) \geq \frac{1}{2} f_P(p)$ and $\mathrm{d}_P(v) \geq \frac{1}{2} f_P(q)$, so the choice of which serves as n_v is irrelevant. Our new filter $t' : \mathrm{Del}(M) \to \mathbb{R}$ is defined as follows:

- for every vertex $v \in M$, let $t'(v) = 0$ if $v \in P$ and $t'(v) = s'(v)$ if $v \in M \setminus P$,
- for every higher-dimensional simplex $\sigma = \{v_0, \cdots, v_k\} \in \mathrm{Del}(M)$, let $t'(\sigma) = \max_{i \in \{0, \cdots, k\}} s'(v_i)$.

The modified α-mesh filtration $\{D_M'^\alpha\}_{\alpha \geq}$ is defined as the sublevel-sets filtration of t', so once again each space $D_M'^\alpha$ is a subcomplex of Del(M).

Approximation guarantee. Here again the analysis is done in terms of a dual filtration $\{V_M'^\alpha\}$, defined by $V_M'^\alpha = \bigcup_{v \in M} U_\alpha'(v)$, where

$$U_\alpha'(v) = \begin{cases} \emptyset & \text{if } v \in M \setminus P \text{ and } \alpha < s'(v), \\ \text{ball}(v, \frac{\alpha}{1+\varepsilon}) & \text{if } v \in P \text{ and } \alpha < s'(v), \\ \text{Vor}_\square(v) & \text{otherwise.} \end{cases}$$

Let $\mathcal{U}_\alpha'$ denote the collection of sets $\{U_\alpha'(v)\}_{v \in M}$. In contrast with Section 3, the sets $U_\alpha'(v) \in \mathcal{U}_\alpha'$ are not monotonically increasing with α, the problem being that $U_\alpha'(v) \not\subseteq U_\beta'(v)$ when $v \in P$ and $\alpha < s'(v) \leq \beta$. Nevertheless, for our choice of sizing field f the family $\{V_M'^\alpha\}$ is still a valid filtration and $D_M'^\alpha$ is the nerve of its cover. These two facts are proved respectively in Lemmas A.3 and A.4 of Appendix A.

As in the previous section, the filtration $\{V_M'^\alpha\}$ is interleaved multiplicatively with $\{P_\square^\alpha\}$, however this time with an arbitrarily small interleaving factor:

LEMMA 4.1. *Given $\varepsilon \leq \frac{1}{2}$, for all $\alpha \geq 0$, $V_M'^{\alpha/(1+\varepsilon)} \subseteq P_\square^\alpha \subseteq V_M'^{\alpha(1+\varepsilon)}$.*

PROOF. First we prove $V_M'^{\alpha/(1+\varepsilon)} \subseteq P_\square^\alpha$. Let x be a point in $V_M'^{\alpha/(1+\varepsilon)}$, and let $v \in M$ be such that $x \in U_{\alpha/(1+\varepsilon)}'(v)$. There are two cases to consider, depending on the value of α and on the location of v. In each case, the goal is to show that $d_P(x) \leq \alpha$.
Case $\alpha/(1+\varepsilon) < s'(v)$:
In this case we have $v \in P$ and $U_{\alpha/(1+\varepsilon)}'(v) = \text{ball}(v, \alpha/(1+\varepsilon)^2)$, which gives $d_P(x) \leq |x - v| \leq \frac{\alpha}{(1+\varepsilon)^2} \leq \alpha$.
Case $\alpha/(1+\varepsilon) \geq s'(v)$:
Since f_P is 1-Lipschitz, we have $f_P(v) \leq f_P(n_v) + d_P(v)$, which by (1) is at most $3s'(v)$. Hence,

$$d_P(x) \leq d_P(v) + |x - v| \leq s'(v) + f(v)$$

$$\leq s'(v) + \frac{\varepsilon}{3(1+\varepsilon)} f_P(v) \leq s'(v)(1+\varepsilon) \leq \alpha.$$

Now we prove the other inclusion, namely $P_\square^\alpha \subseteq V_M'^{\alpha(1+\varepsilon)}$. Let x be a point in $P_\square^\alpha$, and let $v \in M$ be closest to x. Then, $x \in \text{Vor}_\square(v)$, and we will show that either $x \in U_{\alpha(1+\varepsilon)}'(v)$ or $x \in U_{\alpha(1+\varepsilon)}'(n_v)$. If $\alpha(1+\varepsilon) \geq s'(v)$ then $U_{\alpha(1+\varepsilon)}'(v) = \text{Vor}_\square(v)$, which contains x, so we may assume $\alpha(1+\varepsilon) < s'(v)$. Once again we distinguish between several cases:
Case $v \in P$:
In this case we have $U_{\alpha(1+\varepsilon)}'(v) = \text{ball}(v, \alpha)$, which contains x by hypothesis.
Case $v \in M \setminus P$ and $s'(v) = \frac{1}{2} f_P(n_v)$:
Our assumption that $\alpha < \frac{s'(v)}{1+\varepsilon}$ implies $|x - n_x| \leq \alpha < \frac{f_P(n_v)}{2(1+\varepsilon)}$. It turns out that points x and v have a common nearest neighbor in P. Indeed, otherwise we can derive the following contradiction:

$$f_P(n_v) \leq |n_v - n_x| \leq |n_v - v| + |v - x| + |x - n_x|$$

$$\leq d_P(v) + \frac{\varepsilon}{3(1+\varepsilon)} f_P(v) + \alpha$$

$$< \left(\frac{1}{2} + \frac{\varepsilon}{2(1+\varepsilon)} + \frac{1}{2(1+\varepsilon)}\right) f_P(n_v) = f_P(n_v).$$

So, $n_x = n_v$. Since in addition $\alpha(1+\varepsilon) < s'(v) = \frac{1}{2} f_P(n_v) \leq s'(n_v)$, we deduce that $U_{\alpha(1+\varepsilon)}'(n_v) = \text{ball}(n_v, \alpha) = \text{ball}(n_x, \alpha)$, which contains x.
Case $v \in M \setminus P$ and $s'(v) = d_P(v)$:
In fact, this case cannot occur. Indeed, our assumption that $\alpha < \frac{s'(v)}{1+\varepsilon}$ implies $|x - n_x| < \frac{s'(v)}{1+\varepsilon}$, which enables us to derive the following contradiction:

$$s'(v) = d_P(v) \leq |v - x| + |x - n_x|$$

$$< \frac{\varepsilon}{3(1+\varepsilon)} f_P(v) + \frac{1}{1+\varepsilon} s'(v)$$

$$\leq \frac{\varepsilon}{3(1+\varepsilon)} f_P(n_v) + \frac{\varepsilon}{3(1+\varepsilon)} d_P(v) + \frac{1}{1+\varepsilon} s'(v)$$

$$< \frac{2\varepsilon}{3(1+\varepsilon)} s'(v) + \frac{\varepsilon}{3(1+\varepsilon)} s'(v) + \frac{1}{1+\varepsilon} s'(v) = s'(v).$$

$\square$

The usual approach to relating the persistence diagrams of $\{D_M'^\alpha\}$ and $\{V_M'^\alpha\}$ cannot be used directly here because, although $D_M'^\alpha$ is the nerve of a good closed cover of $V_M'^\alpha$, the sets $U_\alpha'(v)$ in that cover do not grow monotonically with α.

Consequently, we use an intermediate filtration, $\{\mathcal{N}\mathcal{U}_\alpha''\}$, where each space $\mathcal{N}\mathcal{U}_\alpha''$ is defined as the nerve of the collection of sets $\mathcal{U}_\alpha'' = \{U_\alpha''(v)\}_{v \in M}$ where $U_\alpha''(v) = U_\alpha'(v) \cap \text{Vor}_\square(v)$. It turns out that $\mathcal{N}\mathcal{U}_\alpha''$ and $D_M'^\alpha$ are isomorphic cell complexes. The proof of this fact is a simple exercise, recalled in Lemma A.5 of Appendix A for completeness.

Let now $V_M''^\alpha = \bigcup_{v \in M} U_\alpha''(v)$. Obviously, we have $V_M''^\alpha \subseteq V_M'^\alpha$ since $U_\alpha''(v) \subseteq U_\alpha'(v)$ for all $v \in M$. It turns out that the canonical inclusion $V_M''^\alpha \hookrightarrow V_M'^\alpha$ is a homotopy equivalence, as proved in Lemma A.6 of Appendix A.

Equipped with these definitions and properties, we can prove the main result of the section:

THEOREM 4.2. *Given a user-defined parameter $\varepsilon \in (0, \frac{1}{2}]$ controlling the sizing function for M, the persistence diagrams of $\{P^\alpha\}$ and $\{D_M'^\alpha\}$ on the natural logarithmic scale are ε-close in the bottleneck-distance, i.e. $d_D^{\log}(\{P^\alpha\}, \{D_M'^\alpha\}) \leq \varepsilon$.*

PROOF. The persistence diagrams of $\{D_M'^\alpha\}$ and $\{\mathcal{N}\mathcal{U}_\alpha''\}$ are identical because $\mathcal{N}\mathcal{U}_\alpha''$ and $D_M'^\alpha$ are isomorphic cell complexes. In addition, since the sets $U_\alpha''(v)$ are convex and monotonically increasing with α, the same standard arguments of algebraic topology as the ones used in Section 3.1 show that the persistence diagrams of $\{\mathcal{N}\mathcal{U}_\alpha''\}$ and $\{V_M''^\alpha\}$ are identical. Finally, the persistence diagrams of $\{V_M''^\alpha\}$ and $\{V_M'^\alpha\}$ are the same because the canonical inclusion $V_M''^\alpha \hookrightarrow V_M'^\alpha$ is a homotopy equivalence for all $\alpha \geq 0$. Combining these equivalences, we obtain that $\{D_M'^\alpha\}$ and $\{V_M'^\alpha\}$ have identical persistence diagrams. Now, it follows from Lemma 4.1 and Corollary 2.2 that $d_D^{\log}(\{V_M'^\alpha\}, \{P_\square^\alpha\}_{\alpha \geq 0}) \leq \log(1 + \varepsilon) \leq \varepsilon$. And to complete the proof, recall from Lemma 3.4 that $\{P_\square^\alpha\}$ and $\{P^\alpha\}$ have identical persistence diagrams. $\square$

5. RECURSIVELY WELL-PACED SUBSETS

Under some restrictions on P, we can guarantee that $|M| = 2^{O(d)}|P|$ [20]; however, in general, the mesh can have size up to $2^{O(d)}|P| \log(\Delta(P))$, and the bound is tight. We now define a filtration over the point cloud that recursively partitions the points so that each mesh has linear size, and

the same interleavings as before still hold. Let BB be a bounding box around the input point set P. Given an ordering $(p_1, \ldots, p_n)$ of P, let $P^{(0)} = P \cap BB$ and $P^{(i)} = \{p_1, \ldots, p_i\} \cap BB$ for all $i = 1, \cdots, n$. For any $\theta \in (0, 1)$, we say that P is θ-*well-paced* with respect to BB if there is an ordering of P such that $d_{P^{(i)}}(p_{i+1}) \geq \theta f_{P^{(i)}}(p_{i+1})$ for all $i = 0, \cdots, n-1$, where $f_{P^{(i)}}$ is the Ruppert feature size with respect to $P^{(i)}$ as defined in Section 2.3. The output of SVR has linear size when the input is a well-paced set [21].

A simple greedy algorithm can be used to find a maximal θ-well-paced subset P_1 of the input P. The remaining points of $P \setminus P_1$ are assigned to clusters around the points in P_1 based on their nearest neighbor in P_1. The greedy algorithm is repeated recursively on the clusters, each given a reasonably-sized bounding box, producing a collection of θ-well-paced point sets $P_1, \ldots, P_k$. As in Section 4, we construct a mesh M_i for each P_i as well as the corresponding filtration $D_{M_i}'^\alpha$. We can then define a new (abstract) filtration $D_{M_*}^\alpha = \bigcup_{i=1}^k D_{M_i}'^\alpha$, and by applying the proof methods of Section 4 to the individual meshes M_i, we obtain the following theorem, the detailed proof of which can be found in the full version of the paper:

THEOREM 5.1. *The persistence diagrams of $\{D_{M_*}^\alpha\}$ and $\{P^\alpha\}$ on the natural logarithmic scale are $(3\theta + \varepsilon)$-close in the bottleneck distance, i.e. $d_D^{\log}(\{D_{M_*}^\alpha\}, \{P^\alpha\}) \leq 3\theta + \varepsilon$.*

The advantage of this slightly more complicated construction is that it only runs SVR on well-paced points and thus, the size of the filtration does not depend on the spread of the input. There is a tradeoff however, since the size can increase by a factor of $1/\theta^{O(d)}$.

6. EXPERIMENTS

As a proof of concept, we experimented on 2,000 points chosen on the 4-dimensional Clifford torus (see Figure 1). We modified a pre-existing SVR implementation [1] to run in 4D and compute the filtration of Section 3.2. We used the Plex library [27] to compute the persistence diagram. To the 2,000 input points, SVR added approximately 71,000 Steiner points to achieve an aspect ratio $\rho = 3.08$ (a value chosen for technical reasons). The mesh contained 2 million pentahedra, or 12 million simplices. It took approximately 1 hour to compute the mesh and filtration, and another 7 hours to compute the persistence diagram.

Figure 1 (right) displays the persistence diagram thus obtained as a *persistence barcode* [2] on the natural logarithmic scale, where homological features are sorted first by their dimension, then by their birth time, and drawn as intervals. Intervals with arrow heads extend to infinity. The qualitative interpretation of the barcode is straightforward: scanning through the scales from smallest to largest, we see the disconnected point cloud, the helicoidal curve, the Clifford torus, the 3-sphere of radius $\sqrt{2}$, and finally the ambient space $\mathbb{R}^4$, represented simply as a space with trivial reduced homology groups. Note that the topological noise appearing in the 2-dimensional barcode between -0.2 and 0 is made of many short intervals of length less than 0.05. The 3-sphere structure is of particular interest because it had never been observed before, being too far from the beginning of the filtration for Rips or Čech filtrations to capture it.

Quantitatively, the curve appears at time $\log \alpha = -1.73$, which corresponds roughly to half the distance between consecutive points along the curve. The second 1-cycle of the

torus appears around $\log \alpha = -1.2$, which is only slightly sooner than the time ($\log \alpha = -1.16$) at which consecutive periods of the curve start being connected in the offsets filtration. The 2-cycle of the torus appears soon afterwards, since the square $[0, 2\pi]^2$ gets filled in rapidly once consecutive periods of the curve start to connect. The isolines $u = C^t$ and $v = C^t$ are mapped to unit circles in $\mathbb{R}^4$, so both 1-cycles as well as the 2-cycle should disappear at $\log \alpha = \log 1 = 0$ in the barcode, which is close to being the case. Now, the Euclidean distance to the Clifford torus over the 3-sphere reaches its maximum at point $(\sqrt{2}, 0, 0, 0)$, where its value is $\sqrt{4 - 2\sqrt{2}} \approx 1.08$, so the 3-sphere should appear around $\log \alpha = \log 1.08 \approx 0.08$ in the barcode, which is indeed the case in our result. Finally, at the end of the barcode the approximation quality worsens slightly: since the 3-sphere has radius $\sqrt{2}$, the 3-cycle should disappear at $\log \sqrt{2} \approx 0.35$, but in reality it does so sooner, around $\log \alpha = 0.18$. Nevertheless, the absolute error is still within $\log 1.18$, meaning that our result is as good as if a multiplicative 1.18-interleaving had been obtained, whereas the aspect ratio bound ρ used by the SVR algorithm was 3.08. These observations suggest that our approach can perform better in practice than expected from the theory.

Comparison. The 4-skeleton of the Rips filtration of P reaches an equivalent size (2 million pentahedra) as early as $\log \alpha = -0.75$, which makes it difficult with this budget to detect the torus, and impossible to detect the 3-sphere. Increasing the limit to a mere $\log \alpha = -0.5$ already raises the size of the 4-skeleton to more than 10 million pentahedra. The Clifford torus is not a worst case for the α-complex filtration. However, as mentioned previously, the α-complex is susceptible to pathological behaviour on some other very reasonable inputs.

Engineering issues. Our implementation is very preliminary and would benefit from substantial engineering. In particular, the SVR implementation we adapted creates a bounding box to avoid dealing with boundary issues. The box includes a number of vertices that grows proportionally to $1/(\rho - 1)^{d-1}$; with $\rho = 3.08$ in 4D, this is 2,800 points. No research is needed to solve this issue. In addition, since we did not have access to efficient staged predicates and constructions in 4D, we used exact rational arithmetic, which in 3D slows SVR down by a factor of worse than 20. Despite this, meshing was not the bottleneck.

7. DISCUSSION

Steiner point choice. Since all our filtrations are derived from the mesh Del(M), their sizes (and therefore the complexity of the whole approach) depend heavily on the size of M. Some work has been done in two and three dimensions to optimize point placement (e.g. [24]), reducing the mesh size for any requested quality. Furthermore, there is a huge industry in mesh smoothing, which in practice improves the quality of a mesh as a post-processing step. Reductions in the number of Steiner points are particularly important as the dimension increases; meanwhile, improving the mesh quality reduces the approximation error.

Higher dimensions. A major limitation of our approach lies in the fact that it is tied to the ambient space $\mathbb{R}^d$, which is fine only in small to moderate dimensions. One possibility for improvement would be to refine the approach and its analysis, so as to make its complexity depend on the dimensionality of the topological features the user is interested in. For instance, in scenarios where the data are high-dimensional but the user is only interested in low-dimensional structures, it would be interesting to devise a mechanism that captures low-dimensional topological features at all scales, at a cost that does not depend exponentially on the ambient dimension. Some work has been done in this direction [6], but many problems remain open. It would be interesting to see if meshing techniques could help in this context.

Approximating other filtrations. Our family of filtrations enables us to approximate the persistence diagrams of other filtrations besides the offsets filtration. An interesting example is the *Ruppert filtration*, i.e. the sublevel-sets filtration of the Ruppert local feature size f_P. Using the same machinery as in Section 3, we can devise a filtration of the mesh $\mathrm{Del}(M)$ that is interleaved with the Ruppert filtration of P, thus making it possible to approximate the persistence diagram of f_P through meshing. The details are provided in the full version of the paper. We believe the Ruppert local feature size has a role to play in topological inference, due to its close connections to distance functions (f_P is the distance to the second nearest neighbor in P) and to meshing algorithms.

Superlevel-sets filtrations. Given a point cloud $P \subset \mathbb{R}^d$, our approach can also yield filtrations that are interleaved with the family of superlevel-sets of d_P. However, in $\mathbb{R}^d$ the persistence diagram of the superlevel-sets filtration of d_P provides in fact the same information as the diagram of the sublevel-sets filtration [9][3]. It would be interesting to see if our approach can be extended to other ambient spaces where approximating the superlevel-sets of distance functions would make sense.

8. REFERENCES

[1] U. A. Acar, B. Hudson, G. L. Miller, and T. Phillips. SVR: Practical engineering of a fast 3D meshing algorithm. In *Proc. 16th International Meshing Roundtable*, pages 45–62, 2007.

[2] G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas. Persistence barcodes for shapes. *Internat. Journal of Shape Modeling*, 11:149–187, 2005.

[3] F. Chazal and D. Cohen-Steiner. *Geometric Inference.* to appear as a book chapter, Springer, 2007.

[4] F. Chazal, D. Cohen-Steiner, M. Glisse, L. J. Guibas, and S. Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the 25th ACM Symposium on Computational Geometry*, 2009.

[5] F. Chazal and A. Lieutier. Stability and computation of topological invariants of solids in R^n. *GEOMETRY:*

[6] F. Chazal and S. Y. Oudot. Towards persistence-based reconstruction in euclidean spaces. In *Proceedings of the 24th ACM Symposium on Computational Geometry*, 2008.

[7] D. Cohen-Steiner. Private communication.

[8] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. In *Proceedings of the 21st ACM Symposium on Computational Geometry*, 2005.

[9] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Extending persistence using Poincaré and Lefschetz duality. *Found. Comput. Math.*, 2008. To appear.

[10] V. de Silva. A weak characterisation of the Delaunay triangulation. *Geometriae Dedicata*, 2008. to appear.

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

[12] H. Edelsbrunner. The union of balls and its dual shape. In *SCG '93: Proceedings of the ninth annual symposium on Computational geometry*, pages 218–231, New York, NY, USA, 1993. ACM.

[13] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Compututational Geometry*, 28:511–533, 2002.

[14] S. Eilenberg and N. Steenrod. *Foundations of Algebraic Topology.* Princeton University Press, 1952.

[15] J. Erickson. Nice point sets can have nasty delaunay triangulations. In *Symposium on Computational Geometry*, pages 96–105, 2001.

[16] J. Erickson and S. Kim. Arbitrarily large neighborly families of congruent symmetric convex 3-polytopes. *Discrete Geometry* (A. Bezdek, editor), pages 267–278, 2003.

[17] L. J. Guibas and S. Y. Oudot. Reconstruction using witness complexes. *Discrete and Computational Geometry*, 40:325–356, 2008.

[18] A. Hatcher. *Algebraic Topology.* Cambridge University Press, 2001.

[19] B. Hudson, G. Miller, and T. Phillips. Sparse Voronoi Refinement. In *Proceedings of the 15th International Meshing Roundtable*, pages 339–356, Birmingham, Alabama, 2006. Long version available as Carnegie Mellon University Technical Report CMU-CS-06-132.

[20] B. Hudson, G. L. Miller, T. Phillips, and D. R. Sheehy. Size complexity of volume meshes vs. surface meshes. In *SODA: ACM-SIAM Symposium on Discrete Algorithms*, 2009.

[21] G. L. Miller, T. Phillips, and D. R. Sheehy. Linear-size meshes. In *CCCG: Canadian Conference in Computational Geometry*, 2008.

[22] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *Discrete Comput. Geom.*, 39(1-3):419–441, March 2008.

[23] J. Ruppert. A Delaunay refinement algorithm for quality 2-dimensional mesh generation. *J. Algorithms*, 18(3):548–585, 1995. Fourth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) (Austin, TX, 1993).

Discrete & Computational Geometry, 37(4):601–617, 2007.

---

[3]The result of [9] is stated for continuous real-valued functions over compact manifolds without boundary. In our setting, even though $\mathbb{R}^d$ is not compact, the result of [9] carries over by an easy reduction [7].

[24] A. Üngör. Off-centers: A new type of steiner points for computing size-optimal quality-guaranteed delaunay triangulations. *Comput. Geom.*, 42(2):109–118, 2009.

[25] L. Vietoris. Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen. *Mathematische Annalen*, 97(1):454–472, 1927.

[26] A. Zomorodian and G. Carlsson. Computing persistent homology. *GEOMETRY: Discrete & Computational Geometry*, 33(2):249–274, 2005.

[27] PLEX 2.5. See http://comptop.stanford.edu/programs/plex.html.

A. APPENDIX: MISSING PROOFS FROM SECTION 4

The proofs of Lemmas A.3 through A.6 rely on the following technical results:

LEMMA A.1. *For all $v \in P$,*

$$\text{ball}\left(v, \frac{s'(v)}{1+\varepsilon}\right) \subseteq \bigcup_{\substack{u \in M: \\ |u-v| \leq s'(v)}} \text{Vor}_\square(u),$$

where $\varepsilon \leq \frac{1}{2}$ is a user defined parameter that controls the sizing function for M.

PROOF. Assume for a contradiction that $\text{ball}(v, \frac{s'(v)}{1+\varepsilon})$ intersects $\text{Vor}_\square(u)$ for some $u \in M$ such that $|u-v| > s'(v) = \frac{1}{2}f_P(v)$, and let x be a point in the intersection. Using the triangle inequality, the sizing function $f = \frac{\varepsilon}{3(1+\varepsilon)}f_P$, and the 1-Lipschitz property of f_P, we obtain:

$$|u-v| \leq |v-x| + |x-u|$$
$$\leq \frac{s'(v)}{1+\varepsilon} + \frac{\varepsilon}{3(1+\varepsilon)}f_P(u)$$
$$\leq \frac{1}{2(1+\varepsilon)}f_P(v) + \frac{\varepsilon}{3(1+\varepsilon)}f_P(u)$$
$$\leq f_P(v)\left(\frac{1}{2(1+\varepsilon)} + \frac{\varepsilon}{3(1+\varepsilon)}\right) + \frac{\varepsilon}{3(1+\varepsilon)}|u-v|,$$

which implies that $|u-v| \leq f_P(v)\left(\frac{\frac{1}{2(1+\varepsilon)} + \frac{\varepsilon}{3(1+\varepsilon)}}{1 - \frac{\varepsilon}{3(1+\varepsilon)}}\right) = \frac{1}{2}f_P(v)$, which contradicts our hypothesis. □

LEMMA A.2. *For all $\alpha \geq 0$ and all $v \in P$, if $s'(v) > \alpha$ then $U'_\alpha(v) \cap U'_\alpha(u) = \emptyset$ for all other $u \in M$.*

PROOF. Suppose for a contradiction that there exists some $u \in M$ such that $U'_\alpha(v) \cap U'_\alpha(u) \neq \emptyset$. If $U'_\alpha(u) = \text{ball}(u, \frac{\alpha}{1+\varepsilon})$ then $u \in P$ and we get the following contradiction:

$$f_P(v) \leq |u-v| \leq \frac{2\alpha}{1+\varepsilon} < 2\alpha < 2s'(v) = f_P(v).$$

If $U'_\alpha(u) = \text{Vor}_\square(u)$ then $s'(u) \leq \alpha < s'(v)$. By Lemma A.1, if the Voronoi cell $\text{Vor}_\square(u)$ intersects $U'_\alpha(v) = \text{ball}(v, \frac{\alpha}{1+\varepsilon})$ then $|u-v| \leq s'(v)$. In this case, we get the following contradiction:

$$s'(v) = \frac{1}{2}f_P(v) \leq \frac{1}{2}|v - n_u| \leq \frac{1}{2}(|v-u| + |u-n_u|)$$
$$\leq \frac{1}{2}(s'(v) + d_P(u)) \leq \frac{1}{2}(s'(v) + s'(u)) < s'(v).$$

□

With these technical results at hand, we can prove Lemmas A.3 through A.6:

LEMMA A.3. *Given $\varepsilon \leq \frac{1}{2}$, the family $\{V'^\alpha_M\}_{\alpha \geq 0}$ is a valid filtration.*

PROOF. Let $v \in M$ and $\beta \geq \alpha \geq 0$. By definition, we have $U'_\alpha(v) \subseteq U'_\beta(v)$ unless $v \in P$ and $\alpha < s'(v) \leq \beta$, which is the case we will now address. In this case, we have $U'_\alpha(v) = \text{ball}(v, \frac{\alpha}{1+\varepsilon})$ and $U'_\beta(v) = \text{Vor}_\square(v)$. Let S denote the set $M \cap \text{ball}(v, s'(v))$. For every $u \in S$ we must have $v = n_u$, for otherwise the triangle inequality would imply that $f_P(v) \leq |v - n_u| < 2s'(v) = f_P(v)$, a contradiction. As a result, $s'(u) = s'(v) \leq \beta$, and thus $U'_\beta(u) = \text{Vor}_\square(u)$. Then, Lemma A.1 implies that $U'_\alpha(v) \subseteq \bigcup_{u \in S} \text{Vor}_\square(u) = \bigcup_{u \in S} U'_\beta(u) \subseteq V'^\beta_M$. □

LEMMA A.4. *Given $\varepsilon < \frac{3}{2}$, the complex D'^α_M coincides with the nerve of the cover $\mathcal{U}'_\alpha$ of V'^α_M.*

PROOF. For an input vertex $p \in P$, we have $\{p\} \in D'^\alpha_M$ and $U'_\alpha(p) \neq \emptyset$ for any $\alpha \geq 0$. For a Steiner vertex $v \in M \setminus P$, we have $\{v\} \in D'^\alpha_M \Leftrightarrow s'(v) \leq \alpha \Leftrightarrow U'_\alpha(v) \neq \emptyset$. So, for all $\alpha \geq 0$ the vertices of D'^α_M and of the nerve of $\mathcal{U}'_\alpha$ coincide.

Before moving on to higher-dimensional simplices, let us make the following claim. Let $v \in P$ be such that $s'(v) > \alpha$, and thus $U'_\alpha(v) = \text{ball}(v, \frac{\alpha}{1+\varepsilon})$. By Lemma A.2, $\{v\}$ is the only simplex of the nerve of $\mathcal{U}'_\alpha$ that contains v. We claim that $\{v\}$ is also the only simplex of D'^α_M that contains v. Indeed, let $u \in M$ be any neighbor of v in $\text{Del}(M)$. First, we have $u \in M \setminus P$. Indeed, $|v-u| \leq f(v) + f(u) \leq \frac{2\varepsilon}{3(1+\varepsilon)}(f_P(v) + f_P(u))$, which is at most $\frac{2\varepsilon}{3(1+\varepsilon)}f_P(v) + \frac{\varepsilon}{3(1+\varepsilon)}|u-v|$ since f_P is 1-Lipschitz. It follows that $|u-v| \leq \frac{2\varepsilon}{3+2\varepsilon}f_P(v) < f_P(v)$, which means that $u \notin P$. Second, we have $n_u = v$. Indeed, $|v-n_u| \leq |v-u| + |u-n_u| \leq 2|v-u| \leq \frac{4\varepsilon}{3+2\varepsilon}f_P(v)$, which is less than $f_P(v)$ since $\varepsilon < \frac{3}{2}$. This means that $n_u = v$. It follows that $s'(u) \geq \frac{1}{2}f_P(n_u) = \frac{1}{2}f_P(v) = s'(v) > \alpha$, which proves our claim.

Our claim implies that any simplex $\sigma = \{v_0, \ldots, v_k\}$ with $k \geq 1$ that appears in D'^α_M or in the nerve of $\mathcal{U}'_\alpha$ satisfies $U'_\alpha(v_i) = \text{Vor}_\square(v_i)$ for all $i = 0 \ldots k$. Hence, $\sigma \in D'^\alpha_M$ if and only if σ belongs to the nerve of $\mathcal{U}'_\alpha$. □

LEMMA A.5. *For all $\alpha \geq 0$, $\mathcal{N}\mathcal{U}''_\alpha = D'^\alpha_M$.*

PROOF. Observe that $U''_\alpha(v) \subseteq U'_\alpha(v)$ for all points $v \in M$, so $\mathcal{N}\mathcal{U}''_\alpha$ is naturally included in the nerve of $\mathcal{U}'_\alpha$, which by Lemma A.4 coincides with D'^α_M. For the other inclusion, we observe that $U'_\alpha(v) \subseteq U''_\alpha(v)$ unless $v \in P$ and $\alpha < s'(v)$. However, by Lemma A.2, such vertices v only appear in 0-simplices of D'^α_M. These 0-simplices also appear in $\mathcal{N}\mathcal{U}''_\alpha$ since $\text{Vor}_\square(v) \cap \text{ball}(v, \frac{\alpha}{1+\varepsilon}) \neq \emptyset$, so indeed, $D'^\alpha_M \subseteq \mathcal{N}\mathcal{U}''_\alpha$. □

LEMMA A.6. *For all $\alpha \geq 0$, the canonical inclusion $V''^\alpha_M \hookrightarrow V'^\alpha_M$ is a homotopy equivalence.*

PROOF. We will exhibit a deformation retraction of V'^α_M onto V''^α_M on each connected component of V'^α_M separately. By Lemma A.2, every vertex $v \in P$ with $s'(v) > \alpha$ has the property that $U'_\alpha(v)$ is disjoint from all other sets $U'_\alpha(u)$ and thus forms a separate connected component. On this component the deformation retraction is easily defined using the metric projection onto the convex set $U''_\alpha(v)$, as in Lemma 3.4. All other connected components of V'^α_M can be expressed as unions of $U'_\alpha(u)$'s, each of which is equal to $\text{Vor}_\square(u)$. For these components the identity map is a trivial deformation retraction. □