Mesh Enhanced Persistent Homology*

Benoît Hudson
bhudson@tti-c.edu

Gary L. Miller
glmiller@cs.cmu.edu

Steve Y. Oudot
steve.oudot@inria.fr

Donald R. Sheehy
dsheehy@cs.cmu.edu

November 6, 2009

Abstract

We apply ideas from mesh generation to improve the time and space complexity of computing the persistent homology of a point set in $\mathbb{R}^d$. The traditional approach to persistence starts with the α-complex of the point set and thus incurs the $O(n^{\lfloor d/2 \rfloor})$ size of the Delaunay triangulation. The common alternative is to use a Rips complex and then to truncate the filtration when the size of the complex becomes prohibitive, possibly before discovering relevant topological features. Given a point set P of n points in $\mathbb{R}^d$, we construct a new filtration, the α-mesh, of size $O(n)$ in time $O(n^2)$ with persistent homology approximately the same as that of the α-shape filtration. This makes it possible to compute the complete persistence barcode in $O(n^3)$ time, where n is the number of points. Previously, this bound was only achievable (with exponentially worse constants) for computing partial barcodes from uniform samples from manifolds. The constants in this paper are all singly exponential in d, making them suitable for medium dimensions.

1 Introduction

Persistent homology is a powerful tool for inferring the abstract shape underlying a set of geometric points [17, 9]. Unfortunately, the persistence algorithm does not scale well, even into medium (say 4-10) dimensions, because the size can blow up to $n^{O(d)}$. In this paper, we show how Delaunay refinement meshing technology can dramatically improve the time and space complexity of computing the persistent homology of a point sample in $\mathbb{R}^d$.

The first step in computing persistent homology is to construct a nested sequence of simplicial complexes called a filtration. Traditionally, the α-shape filtration is used, a sequence of subcomplexes of the Delaunay triangulation of the input points. Then, the persistent homology is computed in time $O(m^3)$ where m is the overall size of the complex. The complexity of the Delaunay triangulation can be as bad as $O(n^{\lfloor d/2 \rfloor})$ in d dimensions.

To get around this problem, other complexes have been proposed, including the Rips complex or the witness complex. Unfortunately, these complexes do not actually avoid the complexity blowup of the Delaunay triangulation. In fact, the complexity of these alternative complexes is greater than that of the Delaunay complex. They are computationally feasible only because they can be constructed incrementally in the order of the filtration. Thus, it is possible to truncate the filtration when some size limit has been reached, possibly before the relevant topological structure has been discovered. This can happen even for quite simple examples. Consider the simple example for which persistent homology is expected to provide interesting information shown in Figure 1. Features at dramatically different scales lead to an explosion in the size of the complex.

*This work was partially supported by the National Science Foundation under grant number CCF-0635257.

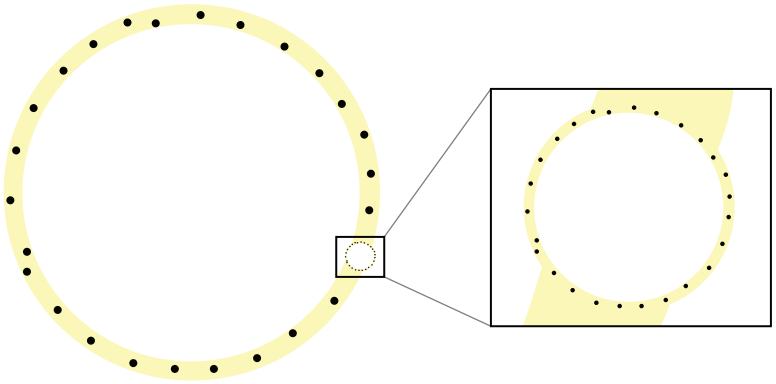

Figure 1: When two persistent cycles appear at dramatically different scales, the Rips and Cěch complex filtrations will reach $O(n^d)$ complexity before detecting the larger one.

Delaunay refinement meshes have many advantages over these complexes. In particular, the number of simplices around any given vertex is bounded by a constant that is $2^{O(d)}$, which though large, is a significant improvement over $n^{O(d)}$. To achieve this, new points called Steiner points are added to the input. The number of points added is $2^{O(d)}n$. We provide the relevant meshing background in Section 2.4.

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 algorithm[14]. Second, we partition the input into *well-paced* sets which guarantees that the complexity of the filtration stays linear and the runtime stays quadratic as the dimension grows (see Section 4).

Our main contribution is a new filtration, the α-mesh filtration, of size $2^{O(d)}n$ that can be computed in $2^{O(d)}n^2$ time. We first present a simplified version of the algorithm that runs in $2^{O(d)}n^2 \log(\Delta)$ time, where Δ is the spread of the input (Section 3). We then show how to eliminate the dependence on the spread by a recursive decomposition in Section 4.

The α-mesh filtration allows us to compute the full persistence diagram in $2^{O(d)}n^3$ time.

2 Preliminaries

2.1 Simplicial Complexes and Nerves

Given an arbitrary set S, an *abstract simplicial complex* C on S is a hereditary family of subsets $\sigma \subseteq S$. That is, if $\sigma_1 \subset \sigma_2 \in C$, then $\sigma_1 \in C$ as well. The subsets σ are called *simplices* and the value $|\sigma| - 1$ is called the *dimension* of σ. The simplicial complex is a link from combinatorics to topology.

An *embedded simplicial complex* is an abstract simplicial complex C on $S \subset \mathbb{R}^d$ such that each simplex σ is identified with it's convex hull, $\operatorname{conv}\sigma$. Moreover, for every pair of simplices, we have $\operatorname{conv}\sigma_1 \cap \operatorname{conv}\sigma_2 = \operatorname{conv}(\sigma_1 \cap \sigma_2)$. This requirement can be understood to mean that the simplices of an embedded simplicial complex glue together at faces and do not intersect otherwise. The embedded simplicial complex is a link from topology to affine geometry.

When we have an embedded simplicial complex, we sometimes care only about the space that it covers. We denote the covered space as $|C| = \bigcup_{\sigma \in C} \operatorname{conv}\sigma \subset \mathbb{R}^d$.

A special class of simplicial complex arises from covering spaces with sets. Let S be a topological subspace of $\mathbb{R}^d$ such that $S = \bigcup_{U \subset \mathcal{U}} U$, where $\mathcal{U}$ is a collection of closed sets. We call $\mathcal{U}$ a *closed cover* of S. The nerve of $\mathcal{U}$ is the abstract simplicial complex $\mathcal{NU}$ on $\mathcal{U}$ such that $\sigma \subset \mathcal{U} \in \mathcal{NU}$ if and only if $\bigcap_{U \in \sigma} U$ is nonempty. A mapping $\mathcal{U} \to \mathbb{R}^d$ induces a mapping from $\mathcal{NU} \to \mathbb{R}^d$. If such a mapping yields $\mathcal{NU}$ as an embedded simplicial complex,

then we have a geometric representation of $\mathcal{NU}$. We will not distinguish between $\mathcal{NU}$ and its geometric representation.

The Nerve Theorem is a classical result in combinatorial topology [13, Cor. 4G.3]. It states that given a space S and a corresponding cover $\mathcal{U}$, S is homotopy equivalent to $\mathcal{NU}$ if every intersection of sets $U, U' \in \mathcal{U}$ is empty or contractible. If $\mathcal{U}$ satisfies the conditions of the Nerve Theorem, we call it a *good closed cover*.

2.2 Persistent Homology

Given a simplicial complex, it is possible to compute the homology groups of the underlying space. The homology groups over a field are vector spaces, so their computation reduces to a sequence of simple operations in linear algebra [11]. This is one of the few topological invariants on a space that can be computed. If the input is a point sample from a topological space rather than a simplicial complex, the problem is less straightforward. Persistent homology is an approach to this problem that has been the focus of much research in recent years. The intuitive idea is to compute a growing sequence of simplicial complexes on the input known as a filtration and observe which topological features "persist" during the course of the filtration. Persistence has been used for reconstruction [1, 5, 12], analysis of scalar fields [3], and nonlinear dimensionality reduction [8].

In building a filtration, the geometric properties of the point set are used to infer the structure of the complexes. Several different complexes have been used for persistence filtrations in the literature. The original work on persistence used the α-complex, which is the nerve of $\mathrm{Vor}(P)$ restricted to the α-balls centered at points of P [10, 18]. The α-complex is a subset of the Delaunay triangulation, $\mathrm{Del}(P)$. Later work used other filtrations in an attempt to avoid the $O(n^{\lfloor d/2 \rfloor})$ worst-case behavior of $\mathrm{Del}(P)$. These included the Cĕch complex, the Rips complex and the witness complex [5], all of which connect nearby points together into simplices. These complexes can also exhibit $n^{O(d)}$ blowups in size, but can be computed in practice if the filtration is truncated. For the witness complex, a subsample of the input is used, making it possible to bound the size of the complex by $2^{O(d^2)}n$ (see [5]). This method is suitable for inputs sampled from a manifold and achieves the best known running time of $2^{O(d^2)}n^4$. In this paper, we will show that this bound can be improved to $2^{O(d)}n^3$, even for non-manifold inputs.

Given a filtration on a simplicial complex with m simplices, the running time of the persistence algorithm is $O(m^3)$[18]. For this reason, bounding the size of the complex represents a large win for computing persistent homology in medium dimensions.

Persistence Diagrams and Stability.

Def. persistence diagram

The widespread use of persistent homology is often justified by the so-called "stability of persistence diagrams". This means that similar inputs give rise to similar persistence diagrams.

Much work has been done to bound the difference between two persistence diagrams in terms of some distance measure on the inputs [6, 4, 2, 7]. The setting for such work must necessarily give meaning to the relationship between the two distance measures. In the current context, we want small distances between persistence diagrams to correspond to a similar evolution of topological features in the filtration.

In particular, we want such a distance measure to be invariant under scaling of the inputs. This is important if we are to study multi-scale phenomena. Thus, it makes sense to consider persistence diagrams on a log-scale. There are two natural interpretations to this. On the one hand, we can simply scale the x and y axes of the persistence diagram. On the other hand, we can reparameterize the filtration used to compute the persistence on a log scale. The only difference is in how we interpret the filtration parameter α and how we interpret the axes of the persistence diagram.

The offset filtration is the sequence of sublevel sets of the distance function. Thus, the filtration parameter α corresponds to a geometric distance in the original space. Recall that both axes of the persistence diagram are based on α. We will now redraw the persistence diagrams so that the axes correspond to $\log \alpha$.

The original work on stability considered the context of sublevel sets of well-behaved functions[6]. They proved that for two functions f and g, if $||f - g||_\infty \leq \varepsilon$ then $||Df - Dg||_B \leq \varepsilon$. That is, if the two functions differed by at most ε at any point in space, then there is a matching between persistent homological features such that representative points on the diagram do not differ by more that ε.

In [2], Chazal et al give a general theory of proximity of persistence diagrams. They first define what it means for two filtrations to be *strongly ε-interleaved*. This definition is more general than we need. It will suffice for this work to observe that two filtrations $\mathcal{F}_R$ and $\mathcal{G}_R$ such that $F_{\alpha-\varepsilon} \subset G_\alpha \subset F_{\alpha+\varepsilon}$ for all α are strongly ε-interleaved. We reproduce the main theorem of [2] below.

Theorem 1. *Let $\mathcal{F}_R$ and $\mathcal{G}_R$ be two tame persistence modules. If $\mathcal{F}_R$ and $\mathcal{G}_R$ are strongly ε-interleaved, then $d_B^\infty(D\mathcal{F}_R, D\mathcal{G}_R) \leq \varepsilon$.*

For the filtrations we will see in this paper, the above theorem implies that strongly interleaved filtrations have similar persistence diagrams.

The *log diagram distance* between two filtrations $\{A^\alpha\}$ and $\{B^\alpha\}$ is defined as follows.

$$d_D^{\log}(A^\alpha, B^\alpha) = d_B^\infty(D\{A^{\log \alpha}\}_{\alpha \in \mathbb{R}_{\geq 0}}, D\{B^{\log \alpha}\}_{\alpha \in \mathbb{R}_{\geq 0}}) \tag{1}$$

Using this definition, we get the following corollary to Theorem 1.

Corollary 2. *If two filtrations $\{A^\alpha\}, \{B^\alpha\}$ have the property that $A^{\alpha/\gamma} \subseteq B^\alpha \subseteq A^{\alpha\gamma}$ then $d_D^{\log}(\{A^\alpha\}, \{B^\alpha\}) \leq log\gamma$.*

Persistent Nerves. In the sequel we will make extensive use of the following variant of the Nerve Theorem, introduced in [5]:

Lemma 3 (Persistent Nerve)**.** *Let $X \subset X'$ 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 the same finite parameter set A, such that $U_\alpha \subseteq U'_\alpha$ for all $\alpha \in A$. Then, there exist homotopy equivalences $NU \to X$ 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.*

2.3 Distance Functions

Let $d_P(x)$ denote the distance from the point $x \in \mathbb{R}^d$ to the nearest vertex in the input set P. That is, $d_P(x) = \min_{p \in P} |x - p|$. The function d_P is a distance function on the space, mapping $\mathbb{R}^d \to R$ with well defined sublevel sets

$$P^\alpha = d_P^{-1}[0, \alpha].$$

Observe that the set P^α is simply the union of balls of radius α centered at vertices in P. The family of such sublevel sets parameterized by α is the *offset filtration*. This filtration has many nice properties and is used often in persistent homology.

We will also be considering an alternative distance function that is common to meshing, the Ruppert local feature size. We define this function as $f_P(x) = \min_{p_1, p_2 \in P} \max\{|x - p_1|, |x - p_2|\}$. That is, $f_P(x)$ is the distance from x to the second nearest vertex in P. We define the *Ruppert filtration* using the sublevel sets of f_P. That is,

$$R^\alpha = f_P^{-1}[0, \alpha].$$

The Ruppert local feature size has the advantage that it doesn't go to 0 for finite point sets, thus bounding from below the minimum scale. Also, it is slightly more robust to outliers,

since it requires two points rather than one to drive the value down. For our purposes, it is helpful to look at the Ruppert filtration because of its connection to standard meshing algorithms.

2.4 Sparse Voronoi Refinement

Let P be a finite point sets in Euclidean space $\mathbb{R}^d$. We denote by $\mathrm{Vor}(P)$ the *Voronoi diagram* of P, defined as a collection of closed cells $\{\mathrm{Vor}(p) : p \in P\}$, where each cell $\mathrm{Vor}(p)$ 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. Since each cell is convex, the collection $\mathrm{Vor}(P)$ forms a *good closed cover* of its underlying space $|\mathrm{Vor}(P)| = \bigcup_{p\in P} \mathrm{Vor}(p) = \mathbb{R}^d$. Its nerve, the *Delaunay triangulation* $\mathrm{Del}(P)$, is therefore homotopy equivalent to $\mathbb{R}^d$, by the Nerve Theorem. In fact, the canonical inclusion $P \hookrightarrow \mathbb{R}^d$ yields $\mathrm{Del}(P)$ as an embedded simplicial complex. The *underlying space of* $\mathrm{Del}(P)$, $|\mathrm{Del}(P)|$, coincides with the convex hull of the point cloud P and is thus homeomorphic to $\mathbb{R}^d$.

We assume that the input points are in general position so that at most $d+1$ Voronoi cells intersect at a point. We need this assumption only to guarantee that the dual Delaunay complex is indeed a simplicial complex. This is equivalent to assuming that no $d+2$ points lie on a common sphere.

For a Voronoi cell, $\mathrm{Vor}(v)$, let R_v denote the radius of the smallest ball centered at v that contains all of $\mathrm{Vor}(v)$. Let r_v denote the largest ball centered at v that is entirely contained in $\mathrm{Vor}(v)$. We define the *aspect ratio* of the Voronoi cell to be R_v/r_v. We assume that the input points lie in a reasonably sized bounding box of side length $O(\mathrm{diam}(P))$ and clip Voronoi cells to lie in this box. Thus, we may assume that all (clipped) Voronoi cells have finite aspect ratio.

For any finite set $S \subset \mathbb{R}^d$ let $NN_S(x)$ denote the distance from the point $x \in \mathbb{R}^d$ to its nearest neighbor in $S \setminus \{x\}$. The aspect ratio of a Voronoi cell $\mathrm{Vor}(v)$ in a set P can also be written as $2R_v/NN_P(v)$. The quantity $R_v/NN_P(v)$ is known as the *radius-edge* ratio of the Delaunay simplex dual to the furthest vertex of $\mathrm{Vor}(v)$ to v.

A common variant of the meshing problem takes the set P as input and returns a superset M with the property that all of the Voronoi cells have aspect ratio bounded by some constant ρ. The mesh vertices $S = M \setminus P$ are known as *Steiner points*. In a slight overload of notation, we refer to the point set M as the *mesh*, assuming the Delaunay topology. The Sparse Voronoi Refinement algorithm (Svr) can produce M in near-optimal $O(|P| \log(\Delta(P) + |M|))$ time, where $\Delta(P)$ is the ratio of the largest to smallest interpoint distances among the points of P, also known as "the spread of P" [14]. The algorithm we present in this paper uses Svr as a black box but only on subsets of the input so as to avoid the dependence on the spread. Also, it reduces the output sensitive term, $|M|$ to $O(n)$ using a technique similar to the one in [16].

3 The α-mesh filtration

Give an overview here: the rationale is that we add Steiner points outside the sampled object in order to reduce the complexity of the Delaunay; the consequence is that the alpha-complex filtration no longer has the nice properties it enjoys on sampled shapes. We therefore have to change the way the Delaunay triangulation is filtered, and in fact our filter is very simple and naturally related to the offsets of the input point cloud.

3.1 Construction

Let $\mathrm{ball}(p, \alpha)$ denote the closed ball of radius α centered at p. Let P^α be the α-offset of the input points P. Formally,

$$P^\alpha = \bigcup_{p\in P} \mathrm{ball}(p, \alpha).$$

The nested family of sets, P^α, parameterized by α is known as the *offset filtration on* P.

Given a mesh M, we want to construct a filtration on the mesh that approximates the offset filtration. The filtration is defined by assigning a time $t(\sigma)$ to each simplex $\sigma \in \mathrm{Del}(M)$. The *$\alpha$-mesh filtration*, D_M^α, is the set of all simplices $\sigma \in \mathrm{Del}(M)$ such that $t(\sigma) \leq \alpha$. To define t we also consider the time $s(v)$ when a vertex v may appear in larger simplices in the filtration. This differs from previous methods in which all input points are assumed to enter the filtration at time 0.

The Svr algorithm adds a bounding box around the input points. If $\mathrm{diam}(P)$ is the diameter of the input set P, then the bounding box used by Svr has side length $O(\mathrm{diam}(P))$. As shown in [15], the extra work needed to fill out the rest of the bounding box is negligible.

$$s(v) = \begin{cases} \frac{1}{2}NN_P(v) & \text{if } v \in P \\ NN_P(v) & \text{if } v \in S \end{cases} \tag{2}$$

$$t(\{v_1, \ldots, v_k\}) = \begin{cases} 0 & \text{if } k = 1 \text{ and } v_1 \in P \\ \max\limits_i \{s(v_i)\} & \text{otherwise} \end{cases} \tag{3}$$

The α-mesh filtration is defined as the closed sublevel sets of t^{-1}. That is,

$$D_M^\alpha = t^{-1}[0, \alpha].$$

If σ' is a face of a simplex σ then $t(\sigma') \leq t(\sigma)$, so the complexes in the filtration are all proper simplicial complexes.

Persistent homology of the complement. Describe the *reverse filter* here: instead of filtering $\mathrm{Del}(M)$ by the (closed) sublevel-sets of t, we filter the complex by the (open) superlevel-sets of t.

3.2 Theoretical Guarantees

Our goal in this section is to show that the α-mesh filtration, D_M^α, has the similar persistent homology to the offset filtration, P^α. We do the analysis in terms of a dual filtration, V_M^α based on the clipped Voronoi diagram. The set $\mathrm{Vor}_M(v)$ will always refer to the Voronoi cell of v clipped to the bounding box.

$$\mathrm{Vor}_M(v) = \{x \in BB \mid |x - v| \leq d_P(x)\}.$$

Associate a closed convex set $U_\alpha(v)$ to each vertex v in the mesh M as follows.

$$U_\alpha(v) = \begin{cases} \emptyset & \text{if } t(\{v\}) > \alpha \\ \mathrm{ball}(v, \alpha) & \text{if } v \in P \text{ and } s(v) > \alpha \\ \mathrm{Vor}_M(v) & \text{otherwise} \end{cases} \tag{4}$$

The Voronoi filtration is defined as

$$V_M^\alpha = \bigcup_{v \in M} U_\alpha(v) \tag{5}$$

The collection of sets $\mathcal{U}_\alpha = \{U_\alpha(v) \mid v \in M, t(\{v\}) \leq \alpha\}$ forms a closed cover of V_M^α. Let $\mathcal{NU}_\alpha$ denote the nerve of this cover. The following Lemma describes the relationship between $\mathcal{NU}_\alpha$ and the α-mesh, D_M^α.

Lemma 4. *For all $\alpha \geq 0$, the complex D_M^α is isomorphic to $\mathcal{NU}_\alpha$.*

Proof. The desired isomorphism is $\phi : D_M^\alpha \to \mathcal{NU}_\alpha$ defined by $\phi(\{v_1, \ldots, v_k\}) = \{U_\alpha(v_i)\}_{i=1\ldots k}$. To prove that this map is indeed an isomorphism, it suffices to show that a simplex $\sigma \subset M$ is in D_M^α if and only if $\cap_{v\in\sigma} U_v \neq \emptyset$.

First, consider the case where $\sigma = \{v\}$ is a singleton. If $v \in P$, then $U_\alpha(v)$ is in the $\mathcal{U}_\alpha$ and $\{v\}$ is in D_M^α for all $\alpha \leq \mathrm{diam}(P)$. For $v \in S$, $\sigma = \{v\} \in D_M^\alpha$ if and only if $NN_P(v) \leq \alpha$. This is also the criteria for which $U_\alpha \in \mathcal{U}_\alpha$.

For larger simplices σ, the sets $U_\alpha(v)$ corresponding to the vertices $v \in \sigma$ are all Voronoi cells of $\mathrm{Vor}(M)$. Thus, $\sigma \in D_M^\alpha$ if and only if $\phi(\sigma) = \cap_{v\in\sigma} U_v \neq \emptyset$.

□

Lemma 5. *For all $\beta \geq \alpha \geq 0$, the k-dimensional persistence diagram of the filtrations $\{V_M^\alpha\}_{\alpha\geq 0}$ and $\{D_M^\alpha\}_{\alpha\geq 0}$ are the same for all $k = 0 \ldots d$. Equivalently, $d_D^{\log}(D_M^\alpha, V_M^\alpha) = 0$.*

Proof. The topological spaces defined by V_M^α and D_M^α are homotopy equivalent. This follows directly from Lemma 4, the Nerve Theorem, and the observation that the cells in the cover of V_M^α are all convex.

In addition, the Persistent Nerve Lemma (Lemma 3) implies that the following diagram induced at the kth homology level by canonical inclusions $V_M^\alpha \hookrightarrow V_M^\beta$ and $D_M^\alpha \hookrightarrow D_M^\beta$ and by homotopy equivalences commutes for all $\beta \geq \alpha \geq 0$ and all $k \in \mathbb{N}$:

$$\begin{array}{ccc} H_k(V_M^\alpha) & \to & H_k(V_M^\beta) \\ \cong\downarrow & & \downarrow\cong \\ H_k(D_M^\alpha) & \to & H_k(D_M^\beta) \end{array}$$

It follows that the ranks of the homomorphisms $H_k(V_M^\alpha) \to H_k(V_M^\beta)$ and $H_k(D_M^\alpha) \to H_k(D_M^\beta)$ are the same. Since this is true for all $\beta \geq \alpha \geq 0$, the k-dimensional persistence diagrams of the filtrations $\{V_M^\alpha\}_{\alpha\geq 0}$ and $\{D_M^\alpha\}_{\alpha\geq 0}$ are the same as desired.

□

Let the clipped offsets be defined in analogy with the clipped Voronoi cells as follows.

$$P_\square^\alpha = \{x \in BB \mid d_P(x) \leq \alpha\}.$$

Lemma 6. *For all $\alpha \geq 0$,*

$$V_M^{\alpha/\rho} \subseteq P_\square^\alpha \subseteq V_M^{2\alpha}.$$

Proof. Let x be a point of $V_M^{\alpha/\rho}$. Let v be the nearest mesh vertex to x. If v is an input point with $NN_M(v) > 2\alpha/\rho$, then $x \in \mathrm{ball}(v, \alpha/\rho)$ and thus $x \in P_\square^{\alpha/\rho} \subseteq P_\square^\alpha$. If v is an input point with $NN_M(p) \leq 2\alpha/\rho$, then the quality of $\mathrm{Vor}(v)$ guarantees that $|x-v| \leq \alpha$ and thus $x \in P_\square^\alpha$. If v is a Steiner point then there exists $p \in P$ such that $\alpha/\rho \geq |v-p| \geq NN_M(v)$. Since the Voronoi aspect ratio of v, $\rho(v)$, is bounded by ρ, it is true that for all $y \in \mathrm{Vor}(v)$, $|y-v| \leq \alpha/2$. In particular, this holds for x and thus $|x-p| \leq |x-v|+|v-p| \leq (1+\rho/2)\alpha/\rho$. Since $\rho > 2$, $|x-p| < \alpha$, and thus $x \in P_\square^\alpha$. So, we have $V_M^{\alpha/\rho} \subset P_\square^\alpha$.

For the second inclusion, let x be a point of $P_\square^\alpha$. We wish to show that $x \in V_M^{2\alpha}$. Let $v \in M$ be the nearest mesh vertex to x. The three cases to consider are when (1) $v \in P$, (2) $v \in S \cap P_\square^\alpha$, and (3) $v \in S \setminus P_\square^\alpha$, where S denotes the Steiner points as before.

If $v \in P$ then $x \in \mathrm{ball}(v, \alpha) \subseteq V_M^\alpha$ and thus $x \in V_M^{2\alpha}$. If $v \in S \cap P_\square^\alpha$ then $\mathrm{Vor}(v) \subset V_M^\alpha$ and thus $x \in\subseteq V_M^{2\alpha}$.

The interesting case is when $v \in S \setminus P_\square^\alpha$. Let $p \in P$ be such that $x \in \mathrm{ball}(p, \alpha)$. Observe that $|v-p| \leq |v-x| + |x-p|$, $|x-v| \leq |x-p|$, and $|x-p| \leq \alpha$. Thus $|v-p| \leq 2\alpha$. It follows that $\mathrm{Vor}(v) \subset V_M^{2\alpha}$ and consequently, $x \in V_M^{2\alpha}$.

□

Theorem 7. $d_D^{\log}(D_M^\alpha, P^\alpha) \leq \log \rho$.

Proof. P^α is a union of balls, all of which are centered inside BB. There is a natural deformation retraction defined by the fibers of $f : P^\alpha \to P^\alpha_\square$ mapping points in P^α to the nearest point in $P^\alpha_\square$. Thus, P^α and $P^\alpha_\square$ have the same persistence diagrams and $d_D^{\log}(P^\alpha, P^\alpha_\square) = 0$. This fact, combined with Lemma 6 and Corollary 2, implies $d_D^{\log}(V_M^\alpha, P^\alpha_\square) \leq \log \rho$. Since $d_D^{\log}(D_M^\alpha, V_M^\alpha) = 0$ (Lemma 5), we conclude $d_D^{\log}(D_M^\alpha, P^\alpha) \leq \log \rho$ as desired. □

4 Recursively Well-Paced Subsets

4.1 Well-Paced Points

Let B be the vertices of a bounding box around a set P. Recall that $NN_{P'}(p)$ is the distance to the nearest neighbor of p in the set $P' \subseteq P$. Similarly, let $SN_{P'}$ be the distance to the second nearest neighbor of p in P'. Given an ordering $(p_1, \ldots, p_n)$ of P, let $P_i = \{p_1, \ldots, p_i\}$ and define $P_0 = \emptyset$. We say that a set of points P is *θ-well-paced* with respect to B if there is an ordering P such that

$$NN_{P_{i-1} \cup B}(p_i) \geq \theta SN_{P_{i-1} \cup B}(p_i),$$

for all $i = 1 \ldots n$. Note that $0 < \theta < 1$.

The well-paced criteria is a loose generalization of many sampling conditions on the spacing of an input set used in the literature, and may be viewed as an unstructured analogue of an unbalanced quadtree (see [15] for other examples and applications). When the bounding box is clear, we simply say P is *θ-well-paced*. When the particular value of θ is understood or unimportant, we just call P *well-paced*.

The output of a good aspect ratio meshing algorithm such as Svr has linear size when the input is a well-paced set. This result, first proven in [16], is a generalization of the linear cost of balancing a quadtree to the case of Delaunay refinement meshes, and captures the usefulness of well-pacing. Below, we paraphrase Theorem 2 of [16] in the terminology of this paper.

Theorem 8. *If P is θ-well-paced for some constant θ and m is the size of the mesh generated by the* Svr *algorithm, then $m = O(n)$.*

Observe that if P is well-paced then the minimum interpoint distance goes down at most by a factor of $(1 + \frac{1}{\theta})$ between P_i and P_{i+1}. Consequently, the spread, $\Delta(P)$, is upper bounded by $(1 + \frac{1}{\theta})^n$ and therefore $\log(\Delta(P)) = O(n)$. This fact combined with Theorem 8 imply that the $O(n \log(\Delta(P)) + m)$ running time of Svr is $O(n^2)$ on well-paced inputs.

4.2 Recursive Construction

Many inputs will not be well-paced. In such cases it suffices to construct a tree of well-paced subsets. Suppose P is our non-well-paced input that contains its own bounding box. A naïve greedy algorithm constructs a maximal θ-well-paced subset $Q \subseteq P$ of a point set in $O(n^2)$ time. The subset Q has the property that for all $p \in P \setminus Q$, $NN_Q(p) < \theta SN_Q(p)$ for otherwise Q would not be maximal. In other words, for every point p not selected by the algorithm there is a point q in Q that is much closer to p than all of the other points in Q. In fact, we can pick θ so that the points in $P \setminus Q$ are not even well paced with respect to the vertices of a quality mesh on Q.

Let q be the nearest point in Q to some non-well-paced point in $P \setminus Q$. Let R be the set of all $p \in P$ whose nearest neighbor in Q is q. Note that R includes the point q. We can add an appropriately size bounding box around R and again find a maximal well-paced subset. This recursive procedure yields a family $P_1, \ldots, P_k \subseteq P$ of well-paced subsets (each with respect to its own bounding box). The recursion tree has the property that a set P_i shares exactly one point with each of its children.

For each set P_i, let p_i denote the point inherited from its parent in the recursive construction. For the root set P_1, let p_1 be the first point added by the greedy algorithm. Let r_i be the maximum distance of a point in P_i to p_i. We call the point p_i the center and r_i the radius of P_i.

We construct a series of meshes $M_1, \ldots, M_k$ on the sets P_i augmented with bounding boxes. We modify the algorithm to include Steiner points s as long as $|s - p_i| \leq \frac{1}{\theta} r_i$. To compute the persistent homology of P, we will work on each mesh independently and combine the answers. Algorithmically, this is straightforward. In the rest of this section, we show how the union of the independent meshes can be modeled by a single filtration that can be intertwined with P^α.

4.3 Topological Consistency

Let $\{P_i\}$ be the tree of well-paced sets and let $\{M_1, ldots, M_k\}$ be the corresponding family of quality meshes.

$$D^\alpha_{M_*} = \bigcup_{i=1}^{k} D^\alpha_{M_i}$$

$$V^\alpha_{M_*} = \bigcup_{i=1}^{k} V^\alpha_{M_i}$$

Unlike D^α_M, the complex $D^\alpha_{M_i}$ is not an embedded simplicial complex. We rectify this situation with the following lemma.

Lemma 9. *For all $\alpha \geq 0$, there exists an embedded subcomplex $E^\alpha_{M_*} \subseteq D^\alpha_{M_*}$ that is a deformation retraction of $D^\alpha_{M_*}$.*

Proof. For a mesh M_i, let $r(M_i)$ be the smallest radius such that $\text{ball}(p_i, r(M_i))$ contains the entire bounding box of M_i. Define $E^\alpha_{M_*}$ as follows.

$$E^\alpha_{M_*} = \bigcup_{i | r(M_i) > \alpha} D^\alpha_{M_i}.$$

Clearly, $E^\alpha_{M_*}$ is a subset of $D^\alpha_{M_*}$.

The deformation retraction is defined by collapsing any mesh M_i in $D^\alpha_{M_*}$ that is not in $E^\alpha_{M_*}$ to a single point. All omitted meshes have $r(M_i) \leq \alpha$ and thus $|D^\alpha_{M_i}|$ is just the convex closure of the bounding box. Such a mesh is simply connected and we can therefore retract it to $p_i = D^\alpha_{M_i} \cap D^\alpha_{M_j}$.

□

Lemma 10. $d_D^{\log}(E^\alpha_{M_*}, V^\alpha_{M_*}) = 0$.

Proof. The meshes omitted from $E^\alpha_{M_*}$ are exactly those that are covered by a single ball of radius alpha of one of their points. Thus, if we consider the cover $\mathcal{U} = \{U^{(i)}_\alpha \mid D^\alpha_{M_i} \subset E^\alpha_{M_*}\}$, we find that $E^\alpha_{M_*}$ is exactly the nerve of this cover. Moreover, $\bigcup_{U \in \mathcal{U}} U = V^\alpha_{M_*}$, so the Persistent Nerve Lemma implies the Lemma. □

We can define the clipped offsets for the recursive meshes in the obvious way.

$$P^\alpha_{*\square} = \bigcup_{i=1}^{k} P^\alpha_{i\square},$$

where $P^\alpha_{i\square}$ is the clipped offset of mesh M_i as before.

Lemma 11. $d_D^{\log}(V^\alpha_{M_*}, P^\alpha_{*\square}) \leq \log \rho$.

Proof. This follows from Corollary 2 and Lemma 6 applied to the individual meshes M_i. □

Lemma 12. $d_D^{\log}(P^\alpha, P^\alpha_{*\square}) \leq \log(1+\theta)$.

Proof. It suffices to prove that $d_D^{\log}(\mathcal{C}^\alpha, \mathcal{C}^\alpha_\square) \leq \log(1+\theta)$, where $\mathcal{C}^\alpha_\square$ is the nerve of the cover of $P^\alpha_{*\square}$ induced by the clipped α-balls. Choose a basis B for $H_k(\mathcal{C}^\alpha)$. For any $b \in B$, there exists a cycle $z(b) \in \mathcal{C}^{\alpha(1+\theta)}$ such that all vertices in b are contained in a single mesh M_i. Moreover $z(b)$ is homology equivalent to b in $\mathcal{C}^{\alpha(1+\theta)}$. This follows from the fact that if any simplex σ in b uses vertices from more than one mesh, then $\alpha > \frac{1}{\theta} r(M_j)$ for all but one of the meshes, call it M_i. Since all of the vertices of $z(b)$ are contained in M_i, $z(b) \in \mathcal{C}^{\alpha(1+\theta)}_\square$. Thus, z and the canonical inclusions induce a family of homomorphisms $\{\phi_\alpha : H_k(\mathcal{C}^\alpha) \to H_k(\mathcal{C}^{\alpha(1+\theta)}_\square)\}$. There is trivial family of homomorphisms $\{\psi_\alpha : H_k(\mathcal{C}^\alpha_\square) \to H_k(\mathcal{C}^{\alpha(1+\theta)})\}$ induced by the canonical inclusions $\mathcal{C}^\alpha_\square \hookrightarrow \mathcal{C}^\alpha$ and $\mathcal{C}^\alpha \hookrightarrow \mathcal{C}^{\alpha(1+\theta)}$. It is straightforward to show that ϕ_α and ψ_α commute as required for strong interleaving, and thus $d_D^{\log}(\mathcal{C}^\alpha, \mathcal{C}^\alpha_\square) \leq \log(1+\theta)$ as desired.

□

Theorem 13. $d_D^{\log}(D^\alpha_{M_*}, P^\alpha) \leq \log((1+\theta)\rho)$.

Proof. The Theorem follows directly from the preceding Lemmas and the triangle inequality. □

5 Tighter Intertwining via Overmeshing

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a sizing function. As long as $f < f_P$, SVR can return a mesh such that the radii of all Voronoi cells $\mathrm{Vor}(v)$ are bounded to be within a constant factor of $f(v)$. We can define $f(x) = \frac{f_P(x)}{k}$ for any constant k. The standard mesh size analysis implies that the output size m can be bounded as follows.

$$m = O\left(\int_B \frac{1}{f(z)^d} dz\right) \tag{6}$$

For our particular choice of f, we have that

$$m = O\left(k^d \int_B \frac{1}{f_P(z)^d} dz\right). \tag{7}$$

Thus, we can scale down the feature size everywhere by a factor of k at a cost of k^d in our overall mesh size.

Using this method of overmeshing, it is possible to prove a tighter intertwining of the α-mesh filtration with the Ruppert filtration. We no longer leave the input points as a special case. Instead, we simply compute f at each vertex in the mesh. A Voronoi cell $\mathrm{Vor}(v)$ enters the filtration V'^α_M at time $\alpha = f(v)$. A simplex σ enters the filtration D'^α_M at time $\alpha = \max_{v \in \sigma} f(v)$.

Lemma 14. *For all* $\alpha \in [0, diam(P)]$.

$$V'^{k\alpha/\rho}_M \subset R^\alpha \subset V'^{2\alpha/k}_M.$$

Proof. We proceed in a manner similar to before. □

The same idea can be applied to get a tighter intertwining with the offset filtration. In such an approach, the birth times of Voronoi cells significantly nearer to one vertex than any other, must be inserted only when that near vertex is within $\alpha/2$ of its nearest neighbor. We omit the exact construction and the proof of its intertwining.

6 Conclusion

In this work we bring together mesh generation and persistent homology. The result is a new method for computing persistent homology. The points are preprocessed into a linear size mesh in worst-case quadratic time. We have shown that the persistent homology of a filtration on this mesh matches that of the offset filtration. Thus, we can run the traditional persistence algorithm on this complex and achieve a dramatic improvement in the asymptotic runtime.

References

[1] J.-D. Boissonat, L. J. Guibas, and S. Y. Oudot. Manifold reconstruction in arbitrary dimensions using witness complexes. In *Proceedings of the 23rd ACM Symposium on Computational Geometry*, 2007.

[2] 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.

[3] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba. Analysis of scalar fields over point cloud data. In *SODA: ACM-SIAM Symposium on Discrete Algorithms*, 2009.

[4] F. Chazal and A. Lieutier. Stability and computation of topological invariants of solids in R^n. *GEOMETRY: Discrete & Computational Geometry*, 37(4):601–617, 2007.

[5] F. Chazal and S. Y. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. Comput. Geom.*, pages 232–241, 2008.

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

[7] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and D. Morozov. Persistent homology for kernels, images, and cokernels. In *SODA: ACM-SIAM Symposium on Discrete Algorithms*, 2009.

[8] V. de Silva and M. Vejdemo-Johansson. Persistent cohomology and circular coordinates. In *Proceedings of the 25th ACM Symposium on Computational Geometry*, 2009.

[9] H. Edelsbrunner and J. Harer. Persistent homology – a survey. In J. E. Goodman, J. Pach, and R. Pollack, editors, *Twenty Years After*. AMS Press, 2007.

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

[11] J. Friedman. Computing Betti numbers via combinatorial laplacians. In *STOC: ACM Symposium on Theory of Computing*, pages 386–391, 1996.

[12] L. J. Guibas and S. Y. Oudot. Reconstruction using witness complexes. In *Proceedings 18th ACM-SIAM Symposium: Discrete Algorithms*, pages 1076–1085, 2007.

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

[14] 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.

[15] 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.

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

[17] A. Zomorodian. *Topology for Computing*. Cambridge Univ. Press, 2005.

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

A Tighter Constants for Mesh Size Analysis

In [16], it is proven that mesh refinement algorithms such as SVR will produce meshes of size $O(n)$ for well-paced inputs. In the context of that result, the dimension was taken to be a constant. The $O(n)$ reported hides constants that are $d^{O(d)}$. In this section, we prove that the constants are only $2^{O(d)}$ by a more careful analysis. The proof will be almost identical to that given in [16], with the exception that all constants in the proof will be independent of d.

We start with the following basic fact about the output size, m, for optimal meshing algorithms.

$$m \leq 2^{c_1 d} \int_{x \in \Omega} \frac{1}{\mathrm{lfs}^{(n)}(x)^d} d\Omega \tag{8}$$

Let P be a set of well-paced points with respect to a bounding box B. The proof will be by induction on $n = |P|$. Let lfs_i be the local feature size function induced by $B \cup \{p_1, \ldots, p_i\}$. Let $\Psi_i = 2^{c_1 d} \int_{x \in \Omega} \frac{1}{\mathrm{lfs}_i(x)^d} d\Omega$ where c_1 is the constant from the upper bound in Equation 8. In general, c_1 will depend on the particular meshing algorithm used.

We want to prove that $\Psi_n \leq 2^{c_2 d} n$ for some constant c_2 and $n > 0$.

The base of the induction is $\Psi_0 = 2^{c_1(d+1)}$ can be computed explicitly from the observation that $\mathrm{lfs}^{(0)}(x) \geq \frac{s}{2}$ for any point x in a bounding box with side length s.

By induction, we assume $\Psi_{n-1} \leq 2^{c_2}(n-1) + \Psi_0$ for some constant c_2. It will suffice to show that $\Psi_n - \Psi_{n-1} < c_2$. We can split the Ruppert sizing integral as follows.

$$\Psi_n = 2^{c_1 d} \int_{x \in \Omega} \frac{1}{\mathrm{lfs}_n(x)^d} d\Omega \tag{9}$$

$$\leq \Psi_{n-1} + 2^{c_1 d} \int_{x \in U} \frac{1}{\mathrm{lfs}_n(x)^d} - \frac{1}{\mathrm{lfs}_{n-1}(x)^d} d\Omega \tag{10}$$

where $U \subseteq \Omega$ is the set of all points for which the local feature size was changed by the insertion of p_n. Let $R = r_{p_n}$. The following two inequalities hold for all $x \in U$, the first is trivial and the second follows from the definition of well-paced points.

$$\mathrm{lfs}_n(x) \geq |p_n - x|, \text{ and} \tag{11}$$

$$\mathrm{lfs}_{n-1}(x) \leq |p_n - x| + \frac{R}{\theta}. \tag{12}$$

We use these inequalities to compute the integral above using spherical coordinates assuming the new point p_n is at the origin. Since the integrand is positive everywhere, we can upper

bound the integral by integrating over all of $\mathbb{R}^d$ instead of just U:

$$\Psi_n - \Psi_{n-1} \leq 2^{c_1 d} \int_{x \in U} \frac{1}{|x|^d} - \frac{1}{(|x| + \frac{R}{\theta})^d} dV, \tag{13}$$

$$\leq 2^{c_1 d} \int_0^\infty \int_{S_r} \left(\frac{1}{r^d} - \frac{1}{(r + \frac{R}{\theta})^d} \right) dA dr, \tag{14}$$

$$\leq 2^{c_1 d} s_d \int_0^\infty \left(\frac{1}{r^d} - \frac{1}{(r + \frac{R}{\theta})^d} \right) r^{d-1} dr, \tag{15}$$

where S_r is the sphere of radius r and s_d is the surface area of the unit d-sphere. Note the rough bound, $s_d < \pi^{d/2} < 2^d$. In the ball of radius $\frac{R}{2}$ around p_n the lfs is at least $\frac{R}{2}$, so the contribution of this region to Ψ_n is less than some constant c_3.

$$\Psi_n - \Psi_{n-1} \leq 2^{c_1 d + 1} \left(c_3 + \int_{\frac{R}{2}}^\infty \left(\frac{1}{r^d} - \frac{1}{(r + \frac{R}{\theta})^d} \right) r^{d-1} dr \right) \tag{16}$$

By the change variable $yR/\theta = r$ and simplifying we get:

$$\Psi_n - \Psi_{n-1} \leq 2^{c_1 d + 1} \left(c_3 + \int_{\frac{\theta}{2}}^\infty \left(\frac{(y+1)^d - y^d}{y(y+1)^d} \right) dy \right) \tag{17}$$

$$\leq 2^{c_1 d + 1} \left(c_3 + \sum_{i=0}^{d-1} \binom{d}{i} \int_{\frac{\theta}{2}}^\infty \frac{y^i}{y^{d+1}} dy \right) \tag{18}$$

$$= 2^{c_1 d + 1} \left(c_3 + \sum_{i=0}^{d-1} \binom{d}{d-i} \frac{1}{d-i} \left(\frac{2}{\theta} \right)^{d-i} \right) \tag{19}$$

$$\leq 2^{c_1 d + 1} \left(c_3 + (2/\theta + 1)^d \right) \tag{20}$$

Observing that the constant on the last inequality is $2^{O(d)}$ completes the proof.