Alpha-Beta Witness Complexes ⋆

Dominique Attali[1], Herbert Edelsbrunner[2], John Harer[3], and Yuriy Mileyko[4]

[1] LIS-CNRS, Domaine Universitaire, BP 46, 38402 Saint Martin d'Hères, France
[2] Departments of Computer Science and Mathematics, Duke University, Durham, and Geomagic, Research Triangle Park, North Carolina
[3] Department of Mathematics and Center for Computational Science, Engineering, and Medicine, Duke University, Durham, North Carolina
[4] Department of Computer Science, Duke University, Durham, North Carolina

Abstract. Building on the work of Martinetz, Schulten and de Silva, Carlsson, we introduce a 2-parameter family of witness complexes and algorithms for constructing them. This family can be used to determine the gross topology of point cloud data in $\mathbb{R}^d$ or other metric spaces. The 2-parameter family is sensitive to differences in sampling density and thus amenable to detecting patterns within the data set. It also lends itself to theoretical analysis. For example, we can prove that in the limit, when the witnesses cover the entire domain, witness complexes in the family that share the first, scale parameter have the same homotopy type.

1 Introduction

The analysis of large data sets is a paradigm of growing importance in the sciences. Broad advances in technology are leading to ever larger data sets capturing information in unprecedented detail. Examples are micro-arrays that probe gene activity for entire genomes and sensor networks that challenge our ability to integrate time-series of distributed measurements. After distilling such data and giving it a geometric interpretation as a point cloud in possibly high-dimensional ambient space, we are faced with the problem of extracting properties of that cloud, such as its gross topology, various patterns within it, or its geometric shape. We see the study of these point clouds as an extension of the reconstruction of surfaces from point clouds in $\mathbb{R}^3$; see the recent monograph on this topic [1].

In this paper we adopt the point of view that the goal is not the reconstruction of a unique shape but rather a hierarchy that captures the data at different scale levels. In this we are inspired by the work on alpha shapes where scale is captured by the radius of the spherical neighborhoods defined around the data points [2]. Our point of departure is in the method of reconstruction. Instead of appealing to the metric of the ambient space we use the data itself to drive the formation of the family of complexes. Specifically, we distinguish data points by the way we use them: the *landmarks* form the vertices of the complexes we build and the *witnesses* provide support for simplices we add to connect the vertices. This idea can be traced back to the topology adapting networks

⋆ Research by the authors is partially supported by DARPA under grant HR0011-05-1-0007 and by CNRS under grant PICS-3416.

of Martinetz and Schulten [3], who draw an edge between two landmarks if there is a witness for which they are the two nearest. We may interpret the witness as a proof for the edge to belong to the Delaunay triangulation of the landmark points. Unfortunately, a witness is not proof for its three nearest landmarks forming a triangle in the Delaunay triangulation. The resulting impasse was overcome for ordinary Delaunay triangulations by de Silva [4]. He proved that if for every subset of $p+1$ landmarks there is a witness for which the points in the subset are at least as close as any other landmarks, then this is a proof for the $p+1$ landmarks to form a p-simplex in the Delaunay triangulation. This insight motivated de Silva and Carlsson to introduce a generalization of the Martinetz-Schulten networks to two- and higher-dimensional complexes [5]. They used their new tool to study the picture collection of van Hateren and van der Schaaf [6], also considered by Lee, Pedersen and Mumford [7]. The main insight from their work is that a majority of small pixel subarrays can be parametrized on a (two-dimensional) Klein bottle in 7-dimensional ambient space [8].

If the witness complex is patterned after the Delaunay triangulation, why do we not just construct the latter? There is a variety of reasons, including

- the size of the complex can be controlled by choosing the landmarks while not ignoring the information provided by the possibly many more sample points;
- distances are easier to compute than the primitives required to construct Delaunay triangulations;
- extending the definition of witness complexes to metric spaces different from Euclidean spaces is comparatively straightforward;

all already mentioned in [5]. There are also significant drawbacks, such as the locally imperfect reconstruction caused by the finiteness of the witness set. The main purpose of this paper is to present methods that cope with the mentioned drawback of witness complexes. Our main contributions are theoretical, in understanding the family of witness complexes and its algorithms. Specifically,

(i) we introduce a 2-parameter family that contains prior witness complexes as subfamilies;
(ii) we generalize de Silva's result for Delaunay triangulations to witness complexes in the limit;
(iii) we analyze the structure of the family of witness complexes by subdividing its parameter plane;
(iv) we give algorithms to construct this subdivision, compute homology within it, and visualize the result.

Methods that extract the multi-scale topological information along with computational experiments will be presented elsewhere.

Outline. Section 2 presents the complexes after which we model our witness complexes. Section 3 introduces the 2-parameter family of witness complexes. Section 4 studies the family through subdivisions of the parameter plane. Section 5 describes algorithms constructing alpha-beta witness complexes. Section 6 concludes the paper.

2 Complexes

In this section, we introduce the family of complexes that provide the intuition for our witness complexes. The family contains the 1-parameter families of Čech and alpha complexes and uses a second parameter to interpolate between them. We begin with definitions from algebraic topology.

Simplicial Complexes. The geometric notion of a *simplex*, σ, is the convex hull of a collection of affinely independent points in $\mathbb{R}^d$. We say the points *span* the simplex. If there are $p+1$ points in the collection, we call σ a *p-simplex* and $p = \dim\sigma$ its *dimension*. Any subset of the $p+1$ points defines another simplex, $\tau \leq \sigma$, and we call τ a *face* of σ and σ a *coface* of τ. A *simplicial complex* is a finite collection of simplices, K, that is closed under the face relation and satisfies the extra condition that any two of its simplices are either disjoint or their intersection is a face of both. A *subcomplex* is a simplicial complex $K' \subseteq K$. It is *full* if it contains all simplices in K exclusively spanned by vertices in K'. We often favor the abstract view in which a p-simplex is just a collection of $p+1$ points, a face is simply a subset, and a simplicial complex is a finite system of such collections closed under the subset relation. For every finite abstract simplicial complex, there is a large enough finite dimension, d, such that the complex can be realized as a simplicial complex in $\mathbb{R}^d$. For example, d equal to one plus twice the largest dimension of any simplex is always sufficient. The primary use of a simplicial complex is to construct or represent a topological space. Its *underlying space* is the subset of $\mathbb{R}^d$ covered by the simplices, together with the topology inherited from $\mathbb{R}^d$. Finally, K *triangulates* a topological space if its underlying space is homeomorphic to that topological space.

A computationally efficient approach to classifying topological spaces is based on *homology groups* [9]. For a given space, there is one group for each dimension p capturing, in some sense, the holes with p-dimensional boundaries. We use modulo-2 arithmetic and thus get homology groups isomorphic to $\mathbb{Z}/2\mathbb{Z}$ to some non-negative integer power. That power is the *rank* of the group and the *p-th Betti number* of the topological space. The classification of spaces by homology groups is strictly coarser than by homotopy type. It follows that two spaces with the same homotopy type have isomorphic homology groups, of all dimensions. Building a simplicial complex incrementally and writing down the result at every stage, we get a nested sequence of complexes,

$$\emptyset = K_0 \subset K_1 \subset \ldots \subset K_n = K,$$

which we refer to as a *filtration* of K. The inclusion $K_i \subset K_j$ induces a homomorphism from the p-th homology group of K_i to the p-th homology group of K_j, for every $p \geq 0$. We refer to the image of the homomorphism as a *persistent homology group* and to its rank as a *persistent Betti number*. For more information on these groups we refer to [10, 11].

Čech and Alpha Complexes. There are but a few complexes that have been used to turn a finite set of points into a multi-scale representation of the space from which

the points are sampled. Perhaps the oldest construction is the nerve of a collection of spherical neighborhoods, one about each data point. To formalize this idea, let $L \subseteq \mathbb{R}^d$ be a finite set of points.

Definition. For any real number $\alpha \geq 0$, the *Čech complex* of L, $\check{\mathrm{C}}\mathrm{ech}(\alpha)$, consists of all simplices $\sigma \subseteq L$ for which there exists a point $x \in \mathbb{R}^d$ such that $\|x - k\| \leq \alpha$, for all vertices $k \in \sigma$.

The Nerve Lemma implies that $\check{\mathrm{C}}\mathrm{ech}(\alpha)$ has the same homotopy type as the union of the balls with radius α and centered at points in L [12]. A similar construction requires, in addition, that x be closest to and equally far from the relevant data points [2].

Definition. For any real number $\alpha \geq 0$, the *alpha complex* of L, $\mathrm{Alpha}(\alpha)$, consists of all simplices $\sigma \subseteq L$ for which there exists a point $x \in \mathbb{R}^d$ such that $\|x - k\| \leq \alpha$ and $\|x - k\| \leq \|x - \ell\|$, for all $k \in \sigma$ and all $\ell \in L$.

Equivalently, $\mathrm{Alpha}(\alpha)$ is the nerve of the collection of balls of radius α, each clipped to within the Voronoi cell of its center. The Nerve Lemma implies that $\mathrm{Alpha}(\alpha)$ also has the homotopy type of the union of balls. In summary, $\mathrm{Alpha}(\alpha)$ is a subcomplex of $\check{\mathrm{C}}\mathrm{ech}(\alpha)$ and the two have the same homotopy type, for every $\alpha \geq 0$. Alpha complexes are more efficient than Čech complexes but require the evaluation of a more complicated geometric primitive. For $\alpha = \infty$, we have the nerve of the collection of Voronoi cells, also known as the *Delaunay complex* of L, $\mathrm{Delaunay} = \mathrm{Alpha}(\infty)$ [13].

Almost Alpha Complexes. We interpolate between Čech and alpha complexes using a second parameter, β.

Definition. For any real numbers $\alpha, \beta \geq 0$, the *almost alpha complex*, $\mathrm{AA}(\alpha, \beta)$, consists of the simplices $\sigma \subseteq L$ for which there exists a point $x \in \mathbb{R}^d$ such that $\|x - k\| \leq \alpha$ and $\|x - k\|^2 \leq \|x - \ell\|^2 + \beta^2$, for all $k \in \sigma$ and all $\ell \in L$.

As suggested by the name, these complexes are similar to but different from the almost Delaunay complexes introduced in [14]. For $\beta \geq \alpha$, the second constraint is redundant, and for $\beta = 0$, it requires that x be equidistant from all $k \in \sigma$. In other words, $\mathrm{AA}(\alpha, \alpha) = \check{\mathrm{C}}\mathrm{ech}(\alpha)$ and $\mathrm{AA}(\alpha, 0) = \mathrm{Alpha}(\alpha)$. Letting r be the maximum distance between x and a vertex $k \in \sigma$, the conditions in the above definition can be rewritten as $r^2 - \beta^2 \leq \|x - k\|^2 \leq \alpha^2$, for all $k \in \sigma$, and $r^2 - \beta^2 \leq \|x - \ell\|^2$, for all $\ell \in L - \sigma$. As illustrated in Fig.1, this is equivalent to all vertices of σ belonging to an annulus centered at x and all points of L lying on or outside the inner sphere of this annulus.

There is a dual geometric interpretation that is also useful. Let $a_k(\alpha)$ be the closed ball with center k and radius α, and write $a_\sigma(\alpha)$ for the common intersection of the balls $a_k(\alpha)$, for $k \in \sigma$. Similarly, let $b_{k,\ell}(\beta)$ be the closed half-space of points whose square distance to k exceeds that to ℓ by at most β^2, and write $b_{\sigma,\upsilon}(\beta)$ for the common intersection of the half-spaces $b_{k,\ell}(\beta)$, for $k \in \sigma$ and $\ell \in \upsilon$. Then σ belongs to $\mathrm{AA}(\alpha, \beta)$ iff $\mathrm{region}_\sigma(\alpha, \beta) = a_\sigma(\alpha) \cap b_{\sigma,L}(\beta)$ is non-empty. But this region is the intersection of the regions of its vertices, $\mathrm{region}_\sigma(\alpha, \beta) = \bigcap_{k \in \sigma} \mathrm{region}_k(\alpha, \beta)$, as illustrated in Fig.1. Hence, $\mathrm{AA}(\alpha, \beta)$ is the nerve of the regions of the vertices. Independent of β, the union of these regions is the union of balls of radius α, same as for the Čech and the

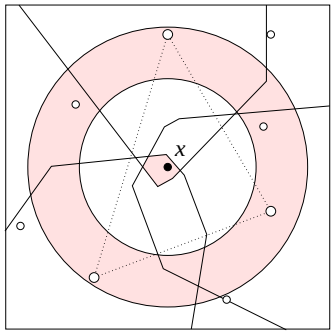

Fig. 1: The outer circle of the annulus centered at x has radius α and the inner circle has square radius $r^2 - \beta^2 \leq \alpha^2 - \beta^2$. The region of points x that witness the membership of the highlighted triangle in $\mathrm{AA}(\alpha, \beta)$ is the common intersection of the regions of its three vertices, each the intersection of the outer disk with a convex polygon.

alpha complexes. Indeed, β only controls the amount of overlap between the regions, which increases with increasing β. Since the regions are convex, the Nerve Lemma implies that the homotopy type of $\mathrm{AA}(\alpha, \beta)$ is the same as that of the union of balls. We summarize,

$$\mathrm{Alpha}(\alpha) \subseteq \mathrm{AA}(\alpha, \beta) \subseteq \check{\mathrm{C}}\mathrm{ech}(\alpha), \tag{1}$$
$$\mathrm{Alpha}(\alpha) \simeq \mathrm{AA}(\alpha, \beta) \simeq \check{\mathrm{C}}\mathrm{ech}(\alpha), \tag{2}$$

for all $\alpha, \beta \geq 0$.

3 Alpha-Beta Witness Complexes

The almost alpha complexes have witness versions obtained by collecting all simplices whose regions contain at least one of a finite set of sampled points. This construction is problematic for small values of β, for which the regions of the vertices have only small overlap. Following de Silva [4], we introduce the concept of a weak witness and show that the resulting witness complexes are better approximations of the complexes than the mentioned witness versions.

Weak and Strong Witnesses. The general set-up consists of a finite set $X \subseteq \mathbb{R}^d$ of *witnesses* and another, usually smaller finite set $L \subseteq \mathbb{R}^d$ of *landmarks*. We consider complexes over L consisting of simplices that have the backing of witnesses in X. Specifically, we call $x \in X$ a *weak* (α, β)*-witness* of $\sigma \subseteq L$ if

[I] $\|x - k\| \leq \alpha$, for all $k \in \sigma$, and
[II] $\|x - k\|^2 \leq \|x - \ell\|^2 + \beta^2$, for all $k \in \sigma$ and all $\ell \in L - \sigma$.

Equivalently, x belongs to $a_\sigma(\alpha) \cap b_{\sigma, L-\sigma}(\beta)$. We call a weak (α, β)-witness a *strong* (α, β)*-witness* if the inequality in Condition [II] holds for all $k \in \sigma$ and all $\ell \in L$ or,

equivalently, if $x \in a_\sigma(\alpha) \cap b_{\sigma,L}(\beta)$. The difference is in the set of landmarks that compete with the vertices of σ. For a weak witness this set excludes the vertices of σ which therefore do not compete with each other. This subtle difference has important consequences.

Definition. For any real numbers $\alpha, \beta \geq 0$, the *alpha-beta witness complex*, $\mathrm{Witness}(\alpha, \beta)$, consists of the simplices $\sigma \subseteq L$ such that every face $\tau \leq \sigma$ has a weak (α, β)-witness in X.

Condition [II] is redundant unless α exceeds β so we restrict the 2-parameter family to $0 \leq \beta \leq \alpha \leq \infty$. With increasing value of α and, independently, of β, the requirements for being a weak witness get more tolerant, which implies $\mathrm{Witness}(\alpha, \beta) \subseteq \mathrm{Witness}(\alpha', \beta')$ whenever $\alpha \leq \alpha'$ and $\beta \leq \beta'$.

Witness Complexes in the Limit. Similar to almost alpha complexes, the alpha-beta witness complexes have a nice geometric interpretation. We describe it in the full version of the paper, where we also show how to extend de Silva's result on Delaunay triangulations to almost alpha complexes. In particular, we prove that the existence of a weak (α, β)-witness for each face implies the existence of a strong (α, β)-witness for the simplex. In other words, if $X = \mathbb{R}^d$ then the alpha-beta witness complex is the same as the almost alpha complex.

Weak Almost Alpha Theorem. If $X = \mathbb{R}^d$ then $\mathrm{Witness}(\alpha, \beta) = \mathrm{AA}(\alpha, \beta)$.

For finite sets X, the alpha-beta witness complex can only be smaller than for $X = \mathbb{R}^d$, which implies $\mathrm{Witness}(\alpha, \beta) \subseteq \mathrm{AA}(\alpha, \beta)$. This should be contrasted with the fact that a strong witness for a simplex is a weak witness for all faces of the simplex. Hence, the witness version of the almost alpha complex, which collects all simplices with strong (α, β)-witnesses in X, is a subcomplex of $\mathrm{Witness}(\alpha, \beta)$. By (2), the homotopy type of the almost alpha complex does not depend on β. Any variation in the homotopy type of the alpha-beta witness complex for fixed value of α must therefore be attributed to insufficient sampling.

4 2-parameter Family

In this section, we focus on the family of witness complexes, describing properties in terms of subdivisions of the parameter plane. In this plane of points (α, β) the balls grow from left to right and the Voronoi cells grow from bottom to top. Potentially interesting sub-families arise as horizontal and vertical lines but also as 45-degree lines along which the balls and cells grow at the same rate.

Comparison with Prior Notions. Several versions of witness complexes have been defined in [5]. We compare them with the 2-parameter family, limiting ourselves to Čech-like constructions. We start with the first version introduced by de Silva and Carlsson.

Definition. The *strict witness complex*, W_∞, consists of the simplices $\sigma \subseteq L$ whose faces belong to W_∞ and for which there exists a witness $x \in X$ such that

[S] $\|x-k\| \leq \|x-\ell\|$, for all $k \in \sigma$ and all $\ell \in L-\sigma$.

Condition [S] is the same as Condition [II] for $\beta = 0$. There is no counterpart to [I] but we can make this condition redundant by setting $\alpha = \infty$. In other words, $W_\infty = \mathrm{Witness}(\infty, 0)$ in our family, as indicated in Fig.2. To introduce the other three constructions in [5], let p be the dimension of σ and $\mathrm{dist}_j(x)$ the distance of $x \in X$ from its j-nearest landmark point. Using a non-negative real parameter R, we get three 1-parameter families of witness complexes, each obtained by substituting one of

[0] $\|x-k\| \leq R$, for all vertices $k \in \sigma$;
[1] $\|x-k\| \leq R + \mathrm{dist}_1(x)$, for all vertices $k \in \sigma$;
[Δ] $\|x-k\| \leq R + \mathrm{dist}_{p+1}(x)$, for all vertices $k \in \sigma$;

for Condition [S] in the definition of W_∞. Following [5], we denote the members of the three families as $W(R,0)$, $W(R,1)$, and $W(R,\Delta)$. The members of the first family are the witness versions of the Čech complex, $W(R,0) = \mathrm{Witness}(R,R)$. For $R=0$ in the second family, we get a p-simplex σ iff there is a witness in the intersection of the $p+1$ Voronoi cells of its vertices, which happens with probability 0 unless $p=0$. As R increases, we get more tolerant about the precise location of the witness. Equivalently, we can think of growing the Voronoi cells and adding a simplex whenever we find a witness in the common intersection of the enlarged cells. The effect of increasing R is therefore similar to that of increasing β in Condition [II], although the enlarged cells have different shape. Condition [Δ] is less restrictive than Condition [**1**] so we have $W(R,1) \subseteq W(R,\Delta)$. We can interpret [$\Delta$] in terms of growing order-$(p+1)$ Voronoi cells. This makes the complexes in the third family rather similar to alpha-beta witness complexes for $\alpha = \infty$, although the geometric details are again different. The growth prescribed by Condition [II] is milder and more controlled than that prescribed by Condition [Δ]. Indeed, we have $\mathrm{Witness}(\infty, R) \subseteq W(R,\Delta)$, for all $R \geq 0$. To see this, consider Conditions [II] and [Δ] for a witness x and a p-simplex σ. If the $p+1$ vertices of σ are the $p+1$ closest landmarks then x and σ satisfy both conditions for all values of β and R. Otherwise, the smallest distance from x to a landmark ℓ not in σ is at most $\mathrm{dist}_{p+1}(x)$. For $\beta = R$, Condition [II] is equivalent to $\|x-k\|^2 \leq R^2 + \|x-\ell\|^2$ for all $\ell \in L-\sigma$. It follows that $\|x-k\|^2 \leq R^2 + \mathrm{dist}^2_{p+1}(x)$ which implies Condition [Δ]. The containment relation cannot be reversed, meaning there is no positive constant c such that $W(R,\Delta)$ is necessarily a subcomplex of $\mathrm{Witness}(\infty, cR)$.

Birthline Subdivision. We decompose the parameter plane into maximal regions within which the alpha-beta witness complexes are the same. For this purpose, we introduce two collections of functions, $A_\sigma, B_{\sigma,\upsilon} : \mathbb{R}^d \to \mathbb{R}$, defined by

$$A_\sigma(x) = \max_{k \in \sigma} \|x-k\|^2;$$
$$B_{\sigma,\upsilon}(x) = \max_{k \in \sigma} \|x-k\|^2 - \min_{\ell \in \upsilon} \|x-\ell\|^2.$$

Both are convex. It follows that their sublevel sets are convex regions, namely the intersections of balls and half-spaces used earlier, $A_\sigma^{-1}(-\infty, \alpha^2] = a_\sigma(\alpha)$ and $B_{\sigma,\upsilon}^{-1}(-\infty, \beta^2] = b_{\sigma,\upsilon}(\beta)$. Hence, a point $x \in X$ is a weak (α,β)-witness for σ iff

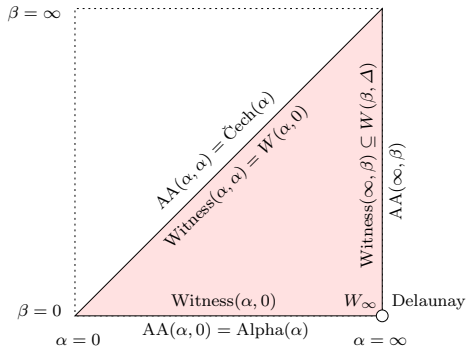

Fig. 2: The parameter plane of alpha-beta witness complexes. We find the Čech and alpha complexes and the witness complexes of de Silva and Carlsson along the edges of the triangle.

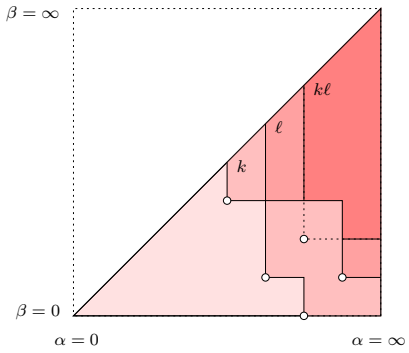

Fig. 3: Since vertices have no proper faces, $Q(k, X)$ and $Q(\ell, X)$ are unions of quadrants. For the edge, $Q(k\ell, X)$ is the portion of its union of quadrants inside $Q(k, X)$ and $Q(\ell, X)$.

$A_\sigma(x) \leq \alpha^2$ and $B_{\sigma,L-\sigma}(x) \leq \beta^2$. The two conditions are independent implying the set of points (α^2, β^2) whose coordinates satisfy them form an upper right quadrant which we denote $Q(\sigma, x)$. Since σ can have more than one weak witness, we consider the union of quadrants they define, and since we require all faces of σ to have weak witnesses, we take the intersection of these unions,

$$Q(\sigma, X) = \bigcap_{\tau \leq \sigma} \left(\bigcup_{x \in X} Q(\tau, x) \right),$$

calling its boundary the *birthline* of σ. It decomposes the parameter plane into two regions such that σ belongs to $\mathrm{Witness}(\alpha, \beta)$ iff the point (α^2, β^2) lies on or to the upper right of the birthline; see Fig.3.

The birthlines decompose the parameter plane into the *birthline subdivision* consisting of maximal regions within which the alpha-beta witness complexes are the same. Neighboring regions are separated by curves, each belonging to one or more birthlines. Curves meet at common endpoints where birthlines merge or cross. Curves that belong to two or more birthlines are common, even in the generic case. In a typical example, the witness complexes in two neighboring regions differ by a collapse, which consists of all faces of a simplex that are cofaces of a proper face of that simplex. A collapse does not affect the homotopy type of the complex, implying that we get isomorphic homology groups in the two regions, for all dimensions.

5 Algorithms

We focus on algorithms that construct the family rather than individual alpha-beta witness complexes. We begin by constructing the birthline subdivision of the parameter plane, which we use as a representation of the family. We then discuss an algorithm for

computing the homology of the complexes in the family. To extract patterns we consider classes that persist while we vary the two parameters.

Constructing Birthlines. Recall that a p-simplex σ and a witness x define a quadrant above and to the right of its corner point. The first coordinate of the corner is $A_\sigma(x) = \max_{k\in\sigma} \|x-k\|^2$. To get the second coordinate, we find the set of $p+1$ landmarks closest to x and distinguish between two cases. If this set is σ then x is a weak witness of σ for all values of β so the second coordinate of the corner is zero. Else this set contains a closest landmark ℓ not in σ and we get the second coordinate as $B_{\sigma,L-\sigma}(x) = A_\sigma(x) - \|x-\ell\|^2$. Clearly these computations benefit from a data structure that provides fast access to the landmarks near a query point. There are many data structures available for this task and we refer to Indyk [15] for a recent survey of this literature. The union of the quadrants $Q(\sigma,x)$, over all witnesses x, is the lower staircase of their corner points. Constructing this staircase is another classic problem in computational geometry [16]. There are many fast methods including a plane-sweep algorithm that constructs the staircase from left to right. This algorithm is convenient for our purposes since it can be reused to compute the birthline of σ as the upper envelope of the staircases of all faces of σ. Finally, we use the plane-sweep algorithm a third time to convert the collection of birthlines into the birthline subdivision. Alternatively, we can do all three plane-sweeps in one, constructing the birthline subdivision directly from the corner points of the quadrants.

What we described is hardly the most efficient method to construct the birthline subdivision. In particular, we expect that most of the quadrants are redundant. It would be interesting to prove bounds on the output size, the number of edges in the birthline subdivision, and to find an algorithm that avoids looking at redundant quadrants and achieves a running time sensitive to the output size.

Computing Homology. We now describe an algorithm that computes the p-th Betti number for each region in the subdivision. It does this for all values of p. The main idea is to explore the parameter plane in a topological sweep that advances a directed path connecting the start-point, $(0,0)$, with the end-point, (∞,∞), while remaining monotonically non-decreasing in both parameters at all times. Initially, the path follows the lower edge of the parameter plane, from $(0,0)$ to $(\infty,0)$, and then the right edge, from $(\infty,0)$ to (∞,∞). We represent this combinatorially by the sequence of simplices labeling the birthlines the path crosses. If m denotes the number of landmarks, we go from the empty complex at $(0,0)$ to the m-simplex at (∞,∞), which implies that the sequence contains all $M = 2^m$ simplices spanned by the landmark points. An elementary move pushes the path locally across a vertex of the subdivision. This corresponds to locally reordering the simplices, which we do one transposition at a time. After processing all transpositions, we arrive at the final path, which follows the diagonal from $(0,0)$ to (∞,∞). The purpose of the sweep is to compute the Betti numbers of the regions, which we do using the algorithm in [10] for the initial sequence and the algorithm in [17] to update the information for each transposition. In the worst case, the initialization takes time cubic in M and each transposition takes time linear in M.

The algorithm's biggest impediment is the large size of the complex at (∞, ∞). To make it feasible for landmark sets that are not very small, we choose an upper bound b for β. Shrinking the parameter domain this way seems appropriate since α and β play fundamentally different roles. The first parameter, α, controls the resolution of the reconstruction, allowing small features to form for small α and letting gross features take over for large α. The second parameter, β, controls how tolerantly we interpret witnesses. The strict interpretation at $\beta = 0$ combined with occasional gaps in the distribution of witnesses leads to holes caused by sporadically missing simplices. The findings in [5] suggest that small non-zero values of β suffice to repair these holes. Although our mathematical formulation of tolerance is different from that paper, we expect the same holds for alpha-beta witness complexes.

Persistence. We now address the question of how to read the Betti numbers of the family represented by the birthline subdivision. We are not after finding the "best" complex since we cannot expect that a single complex would contain all interesting patterns in the data. Since these patterns are expressed at different scale levels a simultaneous representation may indeed be impossible. Instead, we are looking for homology classes that persist while α and β vary. Ideally, we would like to define a notion of two-parameter persistence but there are algebraic difficulties [18]. We therefore fall back on the one-parameter notion introduced in [10] which measures the length of the interval in a path along which a homology class persists. Since the scale level is controlled solely by α it makes sense to draw the path horizontally in the parameter plane so that persistence captures scale. In other words, the directed path used in the computation of homology sweeps the parameter plane from bottom to top. More precisely, we gradually increase β from 0 to b and restrict the path to two turns, one at (β^2, β^2) and the other at (∞, β^2), with a horizontal line in between. To simulate monotonicity, which is necessary to reduce the sweep to transpositions, we advance the horizontal line by processing the simultaneous elementary moves from right to left. For each value of β we can visualize the persistence information in a two-dimensional diagram as defined in [19]. Each homology class is represented by a point whose first coordinate marks its birth and whose second coordinate marks its death. Since birth occurs before death this point lies above the diagonal and its vertical distance from the diagonal is its *persistence*.

As proved by Cohen-Steiner et al. [19], small changes in the function cause only small changes in the diagram. In the case at hand, the function is the value of α at which a simplex is added to the witness complex. As β increases the value of α at which the simplex enters stays the same or decreases. The changes correspond to the steps in the birthlines and are therefore not continuous. Most of the time the steps are small but not always. In particular the first step at which a simplex is introduced can be large. Nevertheless it is useful to stack up the persistence diagrams and to describe the evolution of a homology class as a possibly discontinuous curve in three-dimensional space. In a context in which these curves are continuous they have been referred to as *vines* forming a collection called a *vineyard* [17]. The vineyard of the family of alpha-beta complexes enhances the visualization of persistent homology classes by showing how the persistence changes with varying β, the amount of tolerance with which we recognize a witness of a simplex.

6 Questions and Extensions

We conclude this paper with a list of open questions and suggestions for further research motivated by our desire to improve the algorithms.

Can we take advantage of the hole repairing quality of β without paying the high price of exploding numbers of simplices? Evidence in support of this possibility is that an overwhelming majority of changes caused by increasing β are collapses, which preserve the homotopy type. This is consistent with our observation that in the limit, for $X = \mathbb{R}^d$, the homotopy type of $\mathrm{Witness}(\alpha, \beta)$ is independent of β.

Under reasonable assumptions on the distribution of witnesses and landmarks, what is the expected size of the alpha-beta witness complex as a function of α and β? Similarly, what is the expected number of corners per birthline and what is the expected size of the birthline subdivision?

There are strong parallels between work on witness complexes and on surface and shape reconstruction. Are there versions of witness complexes analogous to the Wrap complex [20], which may be viewed as following Forman's theory of discrete Morse functions [21]? Similarly, are there relaxations of the alpha-beta witness complexes akin to the independent complexes studied in [22]?

Data sets are often contained in subspace of Euclidean space. Recent work in this direction proves that every smoothly embedded compact manifold of dimension 1 or 2 in $\mathbb{R}^d$ has sufficiently fine samplings of landmarks and witnesses such that $\mathrm{Witness}(\infty, 0)$ is homeomorphic to the manifold [23]. A counterexample to extending this result to manifolds of dimension 3 or higher is described in [24]. The counterexample is based on slivers, very flat tetrahedra in the Delaunay triangulation, suggesting the use of sliver exudation methods to remedy the situation [25]. It would be interesting to extend these results to samplings of submanifolds in which density variations encode important information about the data.

Acknowledgments

The authors thank David Cohen-Steiner and Dmitriy Morozov for helpful technical discussions.

References

[1] Dey, T.K.: *Curve and Surface Reconstruction.* Cambridge Univ. Press, England, 2007.

[2] Edelsbrunner, H. and Mücke, E.P.: Three-dimensional alpha shapes. *ACM Trans. Comput. Graphics* **13** (1994), 43–72.

[3] Martinetz, T. and Schulten, K.: Topology representing networks. *Neural Networks* **7** (1994), 507–522.

[4] de Silva, V.: A weak definition of Delaunay triangulation. Manuscript, Dept. Mathematics, Pomona College, Claremont, California, 2003.

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

[6] van Hateren, J.H. and van der Schaaf, A.: Independent component filters of natural images compared with simple cells in primary visual cortex. *Proc. Royal Soc. London B* **265** (1998), 359–366.

[7] Lee, A.B., Pedersen, K.S. and Mumford, D.: The nonlinear statistics of high-contrast patches in natural images. *Intl. J. Comput. Vision* **54** (2003), 83–103.
[8] Carlsson, G., Ishkhanov, T., de Silva, V. and Zomorodian, A.: On the local behavior of spaces of natural images. Manuscript, Dept. Mathematics, Stanford Univ., Stanford, California, 2006.
[9] Munkres, J.R.: *Elements of Algebraic Topology.* Addison-Wesley, Redwood City, California, 1984.
[10] Edelsbrunner, H., Letscher, D. and Zomorodian, A.: Topological persistence and simplification. *Discrete Comput. Geom.* **28** (2002), 511–533.
[11] Zomorodian, A. and Carlsson, G.: Computing persistent homology. *Discrete Comput. Geom.* **33** (2005), 249–274.
[12] Leray, J.: Sur la forme des espaces topologiques et sur les point fixes des représentations. *J. Math. Pures Appl.* **24** (1945), 95–167.
[13] Delaunay, B.: Sur la sphère vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskikh i Estestvennykh Nauk* **7** (1934), 793–800.
[14] Bandyopadhyay, D. and Snoeyink, J.: Almost-Delaunay simplices: nearest neighbor relations for imprecise points. In: Proc. 15th Ann. ACM-SIAM Sympos. Discrete Alg., 2004, 403–412.
[15] Indyk, P.: Nearest neighbors in high-dimensional space. In Goodman, J.E., O'Rourke, J., eds.: *CRC Handbook of Discrete and Computational Geometry*, second edition, Chapman & Hall/CRC, Baton Rouge, Louisiana (2004) 877–892.
[16] H. T.Ǩung, F.L. and Preparata, F.P.: On finding the maxima of a set of vectors. *J. Assoc. Comput. Mach.* **22** (1975), 469–476.
[17] D. Cohen-Steiner and H.E., Morozov, D.: Vines and vineyards by updating persistence in linear time. In: Proc. 22nd Ann. Sympos. Comput. Geom., 2006, 119–126.
[18] Carlsson, G. and Zomorodian, A.: The theory of multidimensional persistence. In: Proc. 23rd Ann. Sympos. Comput. Geom., 2007, to appear.
[19] Cohen-Steiner, D., Edelsbrunner, H. and Harer, J.: Stability of persistence diagrams. *Discrete Comput. Geom.* **37** (2007), 103–120.
[20] Edelsbrunner, H.: Surface reconstruction by wrapping finite sets in space. In Aronov, B., Basu, S., Pach, J., Sharir, M., eds.: *Discrete and Computational Geometry — The Goodman-Pollack Festschrift*. Springer-Verlag, Berlin (2003), 379–404.
[21] Forman, R.: Combinatorial differential topology and geometry. In Billera, L.J., Björner, A., Greene, C., Simion, R., Stanley, R.P., eds.: *New Perspective in Geometric Combinatorics*, volume 38, Cambridge Univ. Press (1999), 177–206.
[22] Attali, D. and Edelsbrunner, H.: Inclusion-exclusion formulas from independent complexes. *Discrete Comput. Geom.* **37** (2007), 59–77.
[23] Attali, D., Edelsbrunner, H. and Mileyko, Y.: Weak witnesses for Delaunay triangulations of submanifolds. In: ACM Sympos. Solid Phys. Modeling, 2007, to appear.
[24] Oudot, S.Y.: On the topology of the restricted Delaunay triangulation and witness complexes in higher dimensions. Manuscript, Dept. Comput. Sci., Stanford Univ., Stanford, California, 2007.
[25] Boissonnat, J.D., Guibas, L.J. and Oudot, S.Y.: Manifold reconstruction in arbitrary dimensions using witness complexes. In: Proc. 23rd Ann. Sympos. Comput. Geom., 2007, to appear.