

Approximating the Pathway Axis and the Persistence Diagrams of a Collection of Balls in 3-Space*

Eitan Yaffe Dan Halperin

School of Computer Science

Tel-Aviv University, Israel

{eitanyaf,danha}@post.tau.ac.il

Abstract

Given a collection \mathcal{B} of balls in three-dimensional space we wish to explore the cavities, voids and tunnels in the complement space of $\cup\mathcal{B}$. We introduce the *pathway axis* of \mathcal{B} as a useful subset of the medial axis of the complement of $\cup\mathcal{B}$, and prove that it satisfies several desirable geometric properties. We present an algorithm that constructs the *pathway graph* of $\cup\mathcal{B}$, a piecewise-linear approximation of the pathway axis. At the heart of our approach is an approximation scheme that constructs a collection \mathcal{K} of same-size balls that approximate \mathcal{B} such that the Hausdorff distance between $\cup\mathcal{B}$ and $\cup\mathcal{K}$ is bounded by a prescribed parameter. We prove a bound on the ratio between the number of balls in \mathcal{K} and the number of balls in \mathcal{B} . We employ this bound and the approximation scheme to show how to approximate the *persistence diagrams* of $\cup\mathcal{B}$, which can be used to extract major topological features such as the large voids and tunnels in the complement of $\cup\mathcal{B}$. We show that our approach is superior in terms of complexity to the standard point-sample approaches for the two problems that we address in this paper: approximating the pathway axis of \mathcal{B} , and approximating the persistence diagrams of $\cup\mathcal{B}$. In a companion paper we introduce MolAxis, a tool for the identification of channels in macromolecules, that demonstrates how the pathway graph and the persistence diagrams are used to identify plausible pathways in the complement of molecules.

*This work has been supported in part by the Israel Science Foundation (grant no. 236/06), by the German-Israeli Foundation (grant no. 969/07), by the IST Programme of the EU as Shared-cost RTD (FET Open) Project under Contract No IST-006413 (ACS - Algorithms for Complex Shapes), and by the Hermann Minkowski–Minerva Center for Geometry at Tel Aviv University.

A preliminary version of this paper appeared in Proceedings of the 24th ACM Symposium on Computational Geometry (SoCG), pp. 260–269, 2008.

1 Introduction

Let \mathcal{B} be a finite collection of balls in \mathbb{R}^3 , and let $\cup\mathcal{B}^c$ denote the complement of their union, namely $\cup\mathcal{B}^c = (\cup_{B \in \mathcal{B}} B)^c$. We assume, without loss of generality, that every ball in \mathcal{B} is not smaller than a unit ball. We wish to capture the shape or skeleton of $\cup\mathcal{B}^c$ by examining its medial axis. The *medial axis* of $\cup\mathcal{B}^c$ is the set of points in $\cup\mathcal{B}^c$ that have more than one closest ball in \mathcal{B} . The exact medial axis of the complement of the union of balls is a subset of the Voronoi diagram [5, 8] of the balls and can be computed in an exact manner as shown by Boissonnat and Delage [6]. We opt for a piecewise linear approximation approach since the implementation is simpler. The underlying structure we use is the Voronoi diagram of points in space for which a robust implementation is readily available [15]. Moreover, since we further manipulate the diagram in our motivating application, we anticipated that it would be easier for us to work with polygons (the facets of the approximate structure) than to work with the curved facets of the Voronoi diagram of the balls.

There are algorithms that approximate the medial axis of an object from a set of unorganized points sampled on the surface of the object [1, 17]. Oudot and Boissonnat [7] introduce an algorithm for computing the medial axis that has certified results for smooth shapes. However, the complement of the union of a collection of balls is not bounded by a smooth surface, making it difficult to directly apply the techniques (and hence have the topological guarantees) obtained in the papers cited above. Our approach is different from these aforementioned approaches since we sample a volume with balls instead of sampling a surface with points.

The medial axis of an object is not stable under Hausdorff distance perturbation of the object [3]. There are algorithms that focus on a *subset* of the medial axis, and obtain certified results. Giesen *et al.* [23] approximate a useful subset of the medial axis of a shape with smooth boundary that captures the topology of the shape. The λ -medial axis [10, 11], introduced by Chazal and Lieutier, is a subset of the medial axis that for some “regular” values of λ remains stable under Hausdorff distance perturbation. This leads to an algorithm [11] that constructs an approximation of the λ -medial axis of an object from a set of noisy unorganized points sampled on or close to the (not necessarily smooth) boundary surface of the object. We employ the definitions and algorithm described there, making use of specific properties of the complement of the union of balls.

A result by Lieutier [25] states that the medial axis of an object and the object itself have the same “shape” (they are *homotopy equivalent*), under certain mild conditions. Yet in order to capture the correct topology of an object from points sampled on its boundary surface the quality of the sample must depend on the *weak feature size* [10]. In our scenario the weak feature size of $\cup\mathcal{B}^c$ can be arbitrarily small, which would require a dense sample that will downgrade the performance of our approximation algorithm dramatically. Instead, we utilize the emerging concept of topological persistence in order to state topological properties of our approximation scheme. Edelsbrunner *et al.* [20] introduce the notion of topological persistence during a growth process of the union of balls. In that work an efficient algorithm is described that classifies topological changes during the growth process as topological features or topological noise depending on their lifetime during the process. The theoretical notion of topological persistence was extended

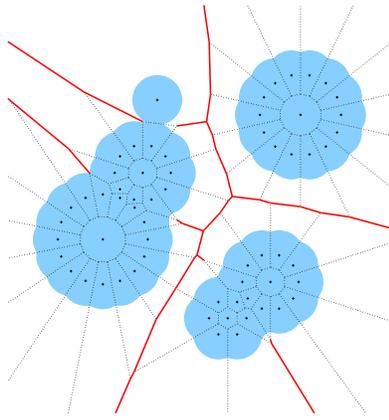


Figure 1: A pathway graph (bold lines) of a collection of discs. The pathway graph contains an approximation of the *core axis* of the discs, illustrated in Figure 2. The discarded portions of the Voronoi diagram are depicted using dotted lines.

independently by Carlsson *et al.* [9], by Chazal *et al.* [12], and by Cohen-Steiner *et al.* [14]. Dey *et al.* [16] harness topological persistence to characterize scalar functions defined on a topological space, and Chazal *et al.* [13] employ topological persistence to achieve an efficient algorithm for manifold reconstruction.

Cohen-Steiner *et al.* introduce in [14] the *persistence diagram* of a real-valued function f on a topological space. It encodes topological characteristics of f , giving a measure on the importance of topological features. More formally, given an integer k and a real-valued function f , the persistence diagram is a collection of points in the extended plane $\bar{\mathbb{R}}^2$ (defined in Section 2) that represent changes in the k -th singular homology group of $f^{-1}((-\infty, t])$, as we sweep $t \in \bar{\mathbb{R}}$ from $-\infty$ to ∞ . The main result in [14] asserts that the persistence diagram is stable under Hausdorff perturbation, making it suitable for approximation. We apply their general results to three specific persistence diagrams, by limiting ourselves to the distance function from the union of balls in 3-space, and to $k = 0, 1, 2$. We call these three diagrams the connected-components diagram (when $k = 0$), the tunnel diagram (when $k = 1$), and the pocket diagram (when $k = 2$) of the ball union. We collectively refer to these three diagrams as the *persistence diagrams* of the ball union.

In this paper we describe an approximation scheme in which we sample the volume of $\cup \mathcal{B}$ with unit balls, or alternatively, sample points that are distant at least 1 from the boundary surface of $\cup \mathcal{B}$. In Section 3 we define the *pathway axis*¹ of \mathcal{B} as a useful subset of the medial axis of $\cup \mathcal{B}^c$. We construct a collection \mathcal{K}_ε of unit balls such that the Hausdorff distance between $\cup \mathcal{B}$ and $\cup \mathcal{K}_\varepsilon$ is not larger than a prescribed ε . From the Voronoi diagram of the centers of \mathcal{K}_ε we construct the *pathway graph* of \mathcal{B} , a piecewise-linear approximation of the pathway axis. See Figure 1 for a two-dimensional illustration of a pathway graph of a collection of discs.

The presented algorithm is quite simple; most effort is invested in analyzing properties

¹For clarity of exposition and to avoid confusion with related terminology, we have revised some of the new terminology that we introduce relative to our previous publications. See Section 11 for a summary of the terms.

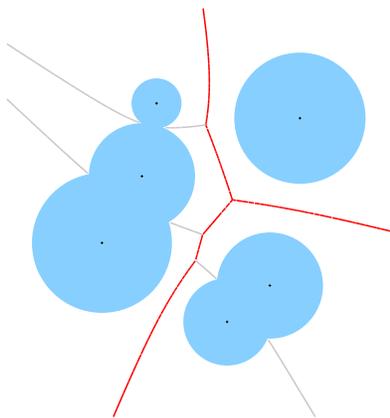


Figure 2: The core axis (bold curves) of the collection of discs that are illustrated in Figure 1. The discarded portions of the medial axis are depicted using light gray curves.

of the approximation scheme. In Section 4 we prove that the pathway graph approximates the pathway axis and converges to it as ε tends to 0. We prove that the pathway axis is contained in the medial axis of $\cup\mathcal{B}^c$ and contains the *core axis* of \mathcal{B} . The core axis of \mathcal{B} is the set of points in the medial axis of $\cup\mathcal{B}^c$ for which the collection of the closest balls in \mathcal{B} do not have a common intersection. It is the medial axis of $\cup\mathcal{B}^c$ without ‘dead-ends’ and therefore it is suitable for finding pathways that lie outside $\cup\mathcal{B}$. See Figure 2 for a two-dimensional illustration of the core axis of the discs depicted in Figure 1. Although the algorithm that we present pertains to balls in 3-space, we generalize some of the definitions and analysis of Section 4 where possible.

In Section 5 we derive a bound on the number of balls in \mathcal{K}_ε as a function of both the user-defined parameter ε , and the ratio between the radius of the largest and smallest balls in \mathcal{B} . In Section 6 we show how to approximate the persistence diagrams of $\cup\mathcal{B}$ using $O(|\mathcal{B}|/\varepsilon^{4/3})$ unit balls. In Section 7, we compare our approach with a point-sample approach, proving a multiplicative-factor gain of $\Omega(1/\varepsilon^{2/3})$ in the worst case in the number of sample entities over the point samples for the approximation of both the pathway axis and the persistence diagrams. In Section 8 we report on experimental results and show how the pocket diagram is used to identify the center of the largest chamber within a molecule, presented as a collection of balls. Section 10 describes implementation details and Section 11 is a glossary of the new terms.

In a companion paper [30] we present MolAxis, a new tool designed for the efficient identification of molecular channels. MolAxis was found to be a highly efficient and accurate tool in identifying transmembrane (TM) channels and pathways leading from buried cavities within enzymes to outside of their convex hull. More details on the results described in this paper can be found in the M.Sc. thesis [28]. The MolAxis web server [29] is available online at: <http://bioinfo3d.cs.tau.ac.il/MolAxis>.

2 Preliminaries

Our work builds on a large body of results concerning Voronoi diagrams and the medial axis. We borrow notation mainly from the work of Attali *et al.* [3] and the work of

Chazal and Lieutier [11]. For any set X we denote by \bar{X} , ∂X , X^c and $|X|$ the closure, the boundary, the complement and the cardinality of X , respectively. We denote by \mathbb{R} the union $\mathbb{R} \cup \{-\infty, +\infty\}$. $B(x, r)$ and $S(x, r)$ denote the closed ball and sphere of center x and radius r in \mathbb{R}^d , respectively. We denote the Euclidean distance between two points $x, y \in \mathbb{R}^d$ by $d(x, y)$. The distance between two subsets A, B of \mathbb{R}^d is defined to be $d(A, B) = \inf_{a \in A, b \in B} d(a, b)$. The *one-sided Hausdorff distance* between two compact subsets A and B of \mathbb{R}^d is:

$$d_H(A|B) = \sup_{x \in A} d(x, B) .$$

The (*symmetric*) *Hausdorff distance* between two compact subsets A and B of \mathbb{R}^d is the maximum of the two one-sided distances, namely $d_H(A, B) = \max(d_H(A|B), d_H(B|A))$. We say that A is a *Hausdorff approximation* of B with an *approximation resolution* of ε if the Hausdorff distance between A and B is not larger than ε . In such a case, we will say for short that A is an ε -approximation of B .

Let O be a bounded open subset of \mathbb{R}^d . For any point $x \in O$, we denote by $\Gamma_O(x)$ the set of closest points to x in the complement O^c , namely $\Gamma_O(x) = \{y \in O^c : d(x, y) = d(x, O^c)\}$. The *medial axis* $M[O]$ of O is the set of points in O that have at least two closest boundary points:

$$M[O] = \{x \in O : |\Gamma_O(x)| \geq 2\} .$$

λ -Medial Axis

We define $\mathcal{R}_O : O \rightarrow \mathbb{R}$ to be the distance function to the boundary, namely $\mathcal{R}_O(x) = d(x, O^c)$. \mathcal{R}_O is a strictly positive, real valued function. There always exists a unique closed ball with minimal radius enclosing $\Gamma_O(x)$ [11]. The real valued, positive function \mathcal{F}_O is defined as the radius of this smallest closed ball enclosing $\Gamma_O(x)$, or formally:

$$\mathcal{F}_O(x) = \inf\{r : \exists y \in \mathbb{R}^d, B(y, r) \supset \Gamma_O(x)\} .$$

We denote by $\Theta(x)$ the center of this smallest enclosing ball. Of course, when $x \notin M[O]$, we have $\Gamma_O(x) = \{\Theta(x)\}$ and $\mathcal{F}_O(x) = 0$. Given a real $\lambda \geq 0$ the λ -*medial axis* is defined to be:

$$M_\lambda[O] = \{x \in O : \mathcal{F}_O(x) \geq \lambda\} .$$

We say that λ is a *regular value* of O if the function that maps $\nu \in \mathbb{R}$ to $M_\nu[O]$ in \mathbb{R}^d is continuous under the Hausdorff metric at $\nu = \lambda$. Formally, λ is a regular value of O if for every $\delta > 0$ there exists $\psi > 0$ such that for any $\nu > 0$ that satisfies $|\lambda - \nu| < \psi$ it holds that $d_H(M_\lambda[O], M_\nu[O]) < \delta$. It is shown in [11] that if λ is a regular value of O , then the λ -medial axis transform is continuous at O for the Hausdorff distance:

Theorem 2.1 (Chazal and Lieutier [11]) *Let O be a bounded open subset of \mathbb{R}^d and λ be a regular value of O . For every $\delta > 0$, there exists $\mu > 0$ such that for every open subset \tilde{O} of \mathbb{R}^d ,*

$$d_H(O^c, \tilde{O}^c) \leq \mu \implies d_H(M_\lambda[O], M_\lambda[\tilde{O}]) \leq \delta .$$

The medial axis and λ -medial axis are formally defined on *bounded* open subsets of \mathbb{R}^d , yet the complement of a closed bounded set like $\cup\mathcal{B}$ is unbounded. To resolve this technicality we limit ourselves from this point onward to the domain of a large open ball Q that contains $\cup\mathcal{B}$. For example, when we refer to $\cup\mathcal{B}^c$ we actually refer to $Q \setminus \cup\mathcal{B}$.

λ -Voronoi Graph

Let E be a finite point set in \mathbb{R}^3 . We denote by $\mathcal{V}[E]$ the collection of *Voronoi cells* of E of dimension 0,1,2 or 3, and by $\mathcal{D}[E]$ its dual structure, the *Delaunay triangulation* of E which contains simplices of dimension 0,1,2 or 3. Note that $\mathcal{D}[E]$ is a *simplicial complex* (see, e.g., [5, 18]).

Let E be a finite point set and let $T \subset E$ be a collection of points such the simplicial complex σ_T (which is the convex hull of T) is in $\mathcal{D}[E]$. We say that σ_T is λ -*enclosed* if the points of T can be enclosed in a sphere of radius smaller than λ . The collection of λ -enclosed simplices is a simplicial complex, which is named the λ -*complex*. Note that a simplex σ_T is λ -enclosed if and only if the set of open balls centered at T with radius λ have a non-empty intersection. We call the dual of its complement, namely the collection of the dual Voronoi faces of simplices that are *not* in the λ -complex, the λ -*Voronoi graph* of E .

By definition, the λ -Voronoi graph of E and the λ -medial axis of E^c are the same. Therefore, we regard the λ -medial axis as an extension of the λ -Voronoi graph from the complement of finite point sets to general open bounded subsets. When dealing with finite point sets we will use the two terms interchangeably.

Approximating the λ -Medial Axis

We denote the boundary of O by Σ , namely $\Sigma = \partial O = \bar{O} \cap O^c$. The λ -medial axis of Σ^c is divided into an *inner* and *outer* λ -medial axis. The inner (resp. outer) λ -medial axis is contained in O (resp. O^c). A finite point set E is called a μ -*noisy sample* of Σ if the Hausdorff distance between Σ and E is less than μ . Chazal and Lieutier [11] provide an algorithm for approximating the λ -medial axis of Σ^c from the λ -Voronoi Graph of a μ -noisy point sample of Σ , which is based on Theorem 2.1.

The Persistence Diagram

We review here several definitions and a theorem which is needed in order to state our results. See [26] for an accessible introduction to Homology, and [14] for the definition of the persistence diagram and its properties.

Let f be a real function on a topological space X , and let k be a non-negative integer. The *persistence diagram* $D_k(f)$ of the pair (f, k) is a subset of $\bar{\mathbb{R}}^2$, which encodes the changes in the k -th singular homology group of $f^{-1}((-\infty, t])$, as we sweep $t \in \mathbb{R}$ from $-\infty$ to ∞ . In our scenario the topological space X is $\cup\mathcal{B}^c$ and f is the distance function from the boundary of $\cup\mathcal{B}$, denoted by $f_{\cup\mathcal{B}}$. We call $D_0(f_{\cup\mathcal{B}}), D_1(f_{\cup\mathcal{B}}), D_2(f_{\cup\mathcal{B}})$, the *connected-components diagram*, the *tunnel diagram* and the *pocket diagram* of $\cup\mathcal{B}$, respectively.

Let X, Y be two subsets of $\bar{\mathbb{R}}^2$. The *bottleneck distance* [14] between X and Y is

defined to be $d_\beta(X, Y) = \inf_\gamma \sup_x \|x - \gamma(x)\|_\infty$, where $x \in X$ range over all points and γ ranges over all bijections from X to Y . The main theorem proven by Cohen-Steiner *et al.* in [14] asserts that for any fixed k , small changes in f imply small changes in the persistence diagram of (f, k) under the bottleneck metric. We state here a weaker version of this theorem, which is sufficient for our needs.

Theorem 2.2 [14] *Let A, A' be two subsets of \mathbb{R}^3 such that $d_H(A, A') \leq \varepsilon$, let $f_A, f_{A'}$ denote the distance functions from A, A' , respectively. Given a fixed integer k , the persistence diagrams of $(f_A, k), (f_{A'}, k)$ satisfy*

$$d_\beta(D_k(f_A), D_k(f_{A'})) \leq \varepsilon.$$

3 Constructing the Pathway Graph

In this section we define the pathway graph and give an algorithm which constructs it. The algorithm is fairly simple and it proceeds in two steps. First, we construct a collection \mathcal{K}_ε of unit balls such that $\cup \mathcal{K}_\varepsilon$ constitutes an ε -approximation of $\cup \mathcal{B}$. In a second step we construct the Voronoi diagram of the centers of the balls in \mathcal{K}_ε and extract the pathway graph from it. We defer technical implementation details to Section 10. Properties of \mathcal{K}_ε , the pathway graph and pathway axis are presented and proved in the following sections.

3.1 Pathway Graph and Pathway Axis

Let X be a closed bounded subset of \mathbb{R}^3 , let $\varepsilon > 0$ be a real number, let \mathcal{K} be a collection of unit balls such that $\cup \mathcal{K}$ constitutes an ε -approximation of X , and let E denote the centers of the balls in \mathcal{K} .

Definition 3.1 (pathway graph) *We call the $(\lambda = 1)$ -Voronoi graph of E a pathway graph of X , with an approximation quality of ε .*

A pathway graph of X is a subset of the medial axis of $\cup \mathcal{K}^c$ and it contains only flat facets, i.e., patches of planes bounded by simple polygons. It is completely piecewise linear and easy to compute, sparing the need to construct facets that are bounded by arcs. We next define the pathway axis of \mathcal{B} . Recalling that $r \geq 1$, for each ball $B = B(c, r) \in \mathcal{B}$ we denote by B_H the concentric ball $B(c, r - 1)$, and let \mathcal{H} denote the collection of balls $\{B_H\}_{B \in \mathcal{B}}$.

Definition 3.2 (pathway axis of balls) *The pathway axis of \mathcal{B} is the $(\lambda = 1)$ -medial axis of $\cup \mathcal{H}^c$.*

The pathway graph converges to the pathway axis as ε tends to 0. We defer the convergence proof to Section 4, where we shall also prove several favorable properties of the pathway axis.

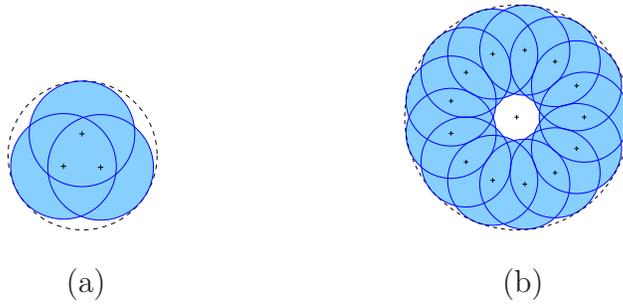


Figure 3: (a) An ε -flower (in the plane) without a void inside. (b) An ε -flower with a void inside. In both cases, the dashed line bounds the disc that is being approximated.

3.2 The ε -Flower

Let X be a closed bounded subset of \mathbb{R}^3 . We say that a finite point set E is a *point sample* of X if E is contained in ∂X . The set E is a ε -*point sample* of X if it is a point sample of X and $d_H(\partial X|E) \leq \varepsilon$. We extend the ε -sample concept from points to balls. We call a ball $B(x, r)$ a *sample ball* of X if it is contained in X and the distance of its center to the boundary of X is equal to its radius, namely $d(x, \partial X) = r$. A set K of balls is a *ball sample* of X if all balls in K are sample balls.

Definition 3.3 (ε -ball sample) *Given a set K of closed balls, a closed bounded subset X of \mathbb{R}^3 and a real parameter $\varepsilon > 0$ we say that K is an ε -ball sample of X if K is a ball sample of X and $d_H(\partial X|\bigcup K) \leq \varepsilon$.*

Note that if the balls in K have radius 0, then the definition of the ball sample coincides with the definition of the point set sample. Next we define the ε -flower, which is the primary primitive of our algorithm.

Definition 3.4 (ε -flower) *Let $B = B(x, r)$, $r \geq 1$ be a ball and $\varepsilon \geq 0$ be real a parameter. A set K of closed unit balls are called an ε -flower of B if they constitute an ε -ball sample of B .*

Note that the sample balls which constitute an ε -flower of a ball $B = B(x, r)$ are tangent to the surface of B from the inside and are centered on the boundary surface of $B(x, r - 1)$; see Figure 3 for examples of ε -flowers. Also note that since we wish to approximate all of $\cup \mathcal{B}$ and not only its boundary surface, we might need to use more than one ε -flower per each ball of \mathcal{B} (namely, several nested ε -flowers), as we shall see below.

3.3 The Algorithm

The input to the algorithm is a collection \mathcal{B} of three-dimensional balls such that every ball in \mathcal{B} is not smaller than a unit ball, and a real parameter $\varepsilon < 1/2$. The algorithm outputs a collection \mathcal{K}_ε of unit balls that constitute an ε -approximation of $\cup \mathcal{B}$, and a pathway graph P_ε of $\cup \mathcal{B}$, which is the ($\lambda = 1$)-medial axis of the balls in \mathcal{K}_ε .

Input:

A collection \mathcal{B} of balls each not smaller than a unit ball, and a real $\varepsilon < 1/2$.

Output:

\mathcal{K}_ε : A collection of unit balls such that $\cup \mathcal{K}_\varepsilon$ is an ε -approximation of $\cup \mathcal{B}$.

\mathcal{P}_ε : A pathway graph of $\cup \mathcal{B}$.

$E \leftarrow \emptyset, \mathcal{P}_\varepsilon \leftarrow \emptyset$

for all $B = B(x, r) \in \mathcal{B}$ **do**

$r' \leftarrow r$

while $r' > 0$ **do**

$E \leftarrow E \cup \text{FLOWER}(B(x, r'), \varepsilon)$

$r' \leftarrow r' - 1$

end while

end for

$\mathcal{K}_\varepsilon \leftarrow$ The collection of unit balls centered at the points of E

$\mathcal{D}[E] \leftarrow$ Delaunay triangulation of E

for all $\sigma \in \mathcal{D}[E]$ **do**

if (**not** $\text{IN_COMPLEX}(\sigma, 1)$) **then**

$\mathcal{P}_\varepsilon \leftarrow \mathcal{P}_\varepsilon \cup \text{DUAL}(\sigma)$

end if

end for

return $\mathcal{K}_\varepsilon, \mathcal{P}_\varepsilon$

Figure 4: Pseudocode of `PATHWAY_GRAPH`, the algorithm that constructs an ε -approximation and a pathway graph of $\cup \mathcal{B}$.

In the pseudocode of `PATHWAY_GRAPH` (see Figure 4) the procedure `DUAL(σ)` returns the dual Voronoi face of a simplex σ , `FLOWER(B, ε)` is a procedure that returns the centers of the unit balls in an ε -flower of a ball B , and `IN_COMPLEX(σ, x)` returns true if and only if σ is x -enclosed. In Section 10 we describe our implementation of these procedures. Note that the number of ε -flowers we construct per each $B = B(x, r) \in \mathcal{B}$ is $\lceil r \rceil$, yet in our motivating application r is usually small, as we discuss in the beginning of Section 8.

4 Geometric Properties of the Pathway Graph

Our goal in this section is to show that the pathway graph approximates the pathway axis. We achieve this by first proving a generalized result, showing that the λ -medial axis of the complement of the union of a collection of convex bodies in \mathbb{R}^d can be approximated using a sufficiently dense point sample (Theorem 4.4). We then employ the theorem to show that the pathway graph \mathcal{P}_ε , which algorithm `PATHWAY_GRAPH` constructs, approximates the pathway axis of \mathcal{B} (Corollary 4.5).

Let A be a collection of bounded, closed and convex bodies in \mathbb{R}^d , and let $\lambda \geq 0$ be a real constant. The Minkowski sum of a body $a \in A$ and the ball $B(0, \lambda)$, which we denote by $a^{+\lambda}$, is the collection of points which are distant not more than λ from a , or formally: $a^{+\lambda} = \{x \in \mathbb{R}^d : d(x, a) \leq \lambda\}$. The Minkowski sum of A with the ball $B(0, \lambda)$ is the collection $A^{+\lambda} = \{a^{+\lambda} : a \in A\}$. Recall that in Section 3 we defined the pathway axis of $\mathcal{B} = \mathcal{H}^{+1}$ to be the $(\lambda = 1)$ -medial axis of $\cup \mathcal{H}^c$. Here we generalize the definition from balls to convex bodies.

Definition 4.1 (pathway axis) *The pathway axis of $A^{+\lambda}$ is the λ -medial axis of $\cup A^c$.*

Note that for $\lambda = 1$ the pathway axis definition given in Section 3 coincides with this more general definition. In order to prove properties of the pathway axis we will define a more intuitive geometric entity called the *core axis*, and prove that the pathway axis contains it. Given a point $x \in \cup A^c$ and a body $a \in A$, we denote by $p_a(x)$ the single closest point of a to x , or formally, $p_a(x) = \operatorname{argmin}_{p \in a} \{d(p, x)\}$. Note that such a point exists since the bodies in A are convex. Recall that $\Gamma_{\cup A^c}(x)$ is the collection of points in $\cup A$ that are closest to x . We define $I_A(x)$ to be the set of bodies in A that are closest to x :

$$I_A(x) = \{a \in A : d(x, a) = d(x, \cup A)\} .$$

We define the *core axis* of A , denoted by $CA[A]$, to be the subset of $M[\cup A^c]$ for which the bodies of $I_A(\cdot)$ do not share a common point:

$$CA[A] = \{x \in M[\cup A^c] : \bigcap_{a \in I_A(x)} a = \emptyset\} .$$

We consider the core axis sufficient for finding pathways since it is a subset of the medial axis of $\cup A^c$ without ‘dead ends’. The following theorem states that the pathway axis of $A^{+\lambda}$ contains the core axis of $A^{+\lambda}$ and is contained in the medial axis of $(\cup A^{+\lambda})^c$. See Figure 5 for an illustration of the Theorem.

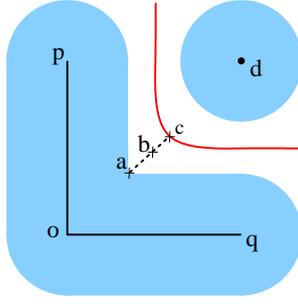


Figure 5: Illustration for Theorem 4.2. The points $o = (0, 0)$, $p = (0, 3)$, $q = (3, 0)$, and $d = (3, 3)$ are the vertices of a square. The collection of convex bodies A is composed of the segments $[o, p]$, $[o, q]$, and the point d . For $\lambda = 1$, A^λ is colored light gray, and the core axis of A is depicted using the bold curve which passes through c . The pathway axis of $A^{+\lambda}$ equals the union of the core axis and the segment $[b, c]$, and the medial axis of $(\cup A^{+\lambda})^c$ equals the union of the pathway axis and the segment $[a, b]$.

Theorem 4.2 *Let A be a collection of closed, bounded and convex bodies in \mathbb{R}^d and let $\lambda > 0$ be a real number. It holds that $CA[A^{+\lambda}] \subseteq M_\lambda[\cup A^c] \subseteq M[(\cup A^{+\lambda})^c]$.*

Proof: Let x be a point of $CA[A^{+\lambda}]$, and let Γ be the collection of closest points to x in $\cup A$. Assume that x is not in $M_\lambda[\cup A^c]$, i.e., Γ is contained in a ball $B = B(c, r)$ such that $r < \lambda$. Consider a body $a \in A$ such that $a^{+\lambda} \in I_{A^{+\lambda}}(x)$. It holds that $a \in I_A(x)$, because $d(x, a) = d(x, a^{+\lambda}) + \lambda$. Let us denote by p_a the single closest point to x in a : $p_a = \Gamma \cap a$. Note that p_a is a single point since a is convex. Now, from our assumption we know that $d(c, p_a) < \lambda$, or equivalently $c \in a^{+\lambda}$. This is true for every $a^{+\lambda}$ in $I_{A^{+\lambda}}(x)$; thus the intersection of all $a^{+\lambda}$ in $I_{A^{+\lambda}}(x)$ contains the point c and is not empty, or in other words x is not in $CA[A^{+\lambda}]$, in contradiction. It follows that x is in $M_\lambda[\cup A^c]$. Finally, $M_\lambda[\cup A^c]$ is contained in $M[(\cup A^{+\lambda})^c]$ since $\mathcal{R}_{\cup A^c}(x) \geq \mathcal{F}_{\cup A^c}(x) \geq \lambda$. \square

Corollary 4.3 *The pathway axis of \mathcal{B} contains the core axis of \mathcal{B} and is contained in the medial axis of $\cup \mathcal{B}^c$.*

The next theorem states that under specific conditions the λ -medial axis of $\cup A^c$ and the λ -Voronoi graph of a certain point sample E are close, See Figure 6 for an illustration.

Theorem 4.4 *Let A be a collection of closed bounded convex bodies in \mathbb{R}^d , let D denote the union $\cup A$, and let $\lambda \geq 0$ be a regular value of D^c . For every real number $\delta > 0$ there exists a real number $\mu > 0$ such that for every finite point set E in \mathbb{R}^d that satisfies:*

- (1) $E \subset D$,
- (2) $d_H(\partial D \mid E) < \mu$ and
- (3) $d_H(D \mid E) < \lambda$,

it holds that $d_H(M_\lambda[\cup A^c], M_\lambda[E^c]) < \delta$.

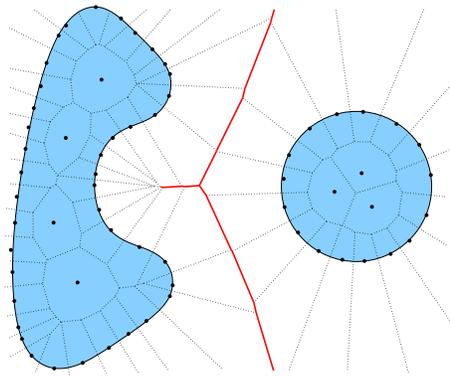


Figure 6: Illustration for Lemma 4.6. The closed and bounded set D is colored light gray. A finite point set E is depicted using small black dots (discs). The λ -Voronoi graph of E is depicted using a bold line, and the discarded portions of the Voronoi diagram of E is depicted using dotted lines. The λ -Voronoi graph of E serves as an approximation of the λ -medial axis of D^c , and its approximation quality relies only on the sample quality of the boundary surface of D .

We denote the pathway axis of \mathcal{B} by $PWA[\mathcal{B}]$. Before proving the theorem let us state a corollary which is an application of the theorem to our scenario. It states that the pathway graph that algorithm `PATHWAY_GRAPH` constructs approximates $PWA[\mathcal{B}]$.

Corollary 4.5 *Let \mathcal{B} be a finite collection of balls in general position that are each not smaller than a unit ball. For any $\delta > 0$ there exists an $\varepsilon > 0$ such that the pathway graph \mathcal{P}_ε , which algorithm `PATHWAY_GRAPH` constructs, satisfies $d_H(\mathcal{P}_\varepsilon, PWA[\mathcal{B}]) < \delta$.*

Proof: Let $\delta > 0$ be a real number. We observe that $\lambda = 1$ is a regular value of $\cup\mathcal{H}^c$ since the balls in \mathcal{B} are in general position, i.e., no two balls are tangent and no three balls meet in a single point. We choose the real μ whose existence is implied by Theorem 4.4. We set $\varepsilon = \frac{4}{5}\mu^2$ and let E be the centers of the unit balls \mathcal{K}_ε that are given as output by algorithm `PATHWAY_GRAPH`. Following Lemma 5.6, which we prove in Section 5, the point set E contains a μ -noisy sample of the boundary surface of $\cup\mathcal{H}$. Therefore E satisfies the three conditions of Theorem 4.4, and the corollary follows. \square

Getting back to the proof of Theorem 4.4, we wish to approximate the λ -medial axis of $\cup A^c$ using the λ -Voronoi graph of a point sample $E \subset \cup A$. Yet we require the points of E to constitute a *dense* sample of the boundary of $\cup A$ (Condition (2) of Theorem 4.4) and only a *constant* sample of the inside of $\cup A$ (Condition (3) of Theorem 4.4) since λ is constant. Therefore $\cup A$ and E are not as close under the Hausdorff distance as the conditions of Theorem 2.1 require. To resolve this problem we define an intermediate set \tilde{O} such that (1) \tilde{O}^c is close to $\cup A$, and (2) \tilde{O} has the same λ -medial axis as E^c . The lemma below defines \tilde{O} that satisfies the first condition and proves the latter condition. Given a closed bounded subset $D \subset \mathbb{R}^d$ and a finite point sample E , the lemma defines \tilde{O} and shows that if E satisfies a list of conditions then the λ -medial axis of \tilde{O} and the λ -Voronoi graph of E are equal. This lemma is a variant of Lemma 5.2 of Chazal and Lieutier [11]; see remark after the proof.

Lemma 4.6 *Let λ and μ be two real numbers such that $\lambda \geq \mu > 0$, let D be a closed bounded subset of \mathbb{R}^d , and let O denote the complement of D . Let E be a finite point set in \mathbb{R}^d that satisfies:*

- (1) $E \subset D$,
- (2) $d_H(\partial D \mid E) < \mu$ and
- (3) $d_H(D \mid E) < \lambda$.

Denoting by \tilde{O} the open set $O^{+\mu} \setminus E$, it holds that $M_\lambda[\tilde{O}] = M_\lambda[E^c]$.

Proof: First we prove that $M_\lambda[\tilde{O}] \subseteq M_\lambda[E^c]$. Let $x \in \tilde{O}$ be a point in $M_\lambda[\tilde{O}]$. Let $\Gamma = \Gamma_{\tilde{O}}(x)$ be the collection of (more than one) closest points to x in \tilde{O}^c . Γ is contained in $(O^{+\mu})^c \cup E$ since $\tilde{O}^c = (O^{+\mu} \setminus E)^c = (O^{+\mu} \cap E^c)^c = (O^{+\mu})^c \cup E$ by definition. If $\Gamma \subset E$ it holds that $x \in M_\lambda[E^c]$ and we are done with this direction. We therefore assume that there exists a point $z \in \Gamma$ such that $z \in (O^{+\mu})^c$. The point x is in O since if $x \in D$ we get $d_H(D \mid E) \geq d(x, E) = \mathcal{R}_{E^c}(x) \geq \mathcal{R}_{\tilde{O}^c}(x) \geq \mathcal{F}_{\tilde{O}}(x) \geq \lambda$ which contradicts the lemma's condition that $d_H(D \mid E) < \lambda$. Thus the segment $[x, z]$ intersects the boundary surface ∂D . Let y be a point in $[x, z] \cap \partial D$. It holds that $d(z, y) > \mu$ since $y \in \partial D$ and $z \notin O^{+\mu}$. Therefore the ball $B = B(y, \mu)$ is contained in the empty ball $B(x, \mathcal{R}_{\tilde{O}}(x))$. Thus $d(y, E) \geq \mu$ which contradicts the lemma's condition that $d_H(\partial D \mid E) < \mu$.

We prove that $M_\lambda[\tilde{O}] \supseteq M_\lambda[E^c]$. Let $x \in E^c$ be a point in $M_\lambda[E^c]$. Let $\Gamma = \Gamma_{\tilde{O}}(x)$ be the collection of (more than one) closest points to x in \tilde{O}^c . The point x is in \tilde{O} since if $x \in \tilde{O}^c = (O^{+\mu})^c \cup E \subseteq D$ we get $d_H(D \mid E) \geq d(x, E) = \mathcal{R}_{E^c}(x) \geq \mathcal{F}_{E^c}(x) \geq \lambda$ which contradicts the lemma's condition that $d_H(D \mid E) < \lambda$. If $\Gamma \subset E$ it holds that $x \in M_\lambda[\tilde{O}]$ and we are done with this direction. If on the other hand there exists a point $z \in \Gamma$ such that $z \in (O^{+\mu})^c$ we reach a contradiction in the same way as above. \square

Remark. The algorithm of Chazal *et al.* [11] does not distinguish between the inner or outer part of the λ -medial axis of ∂O . To distinguish the desired portion (inner or outer) an extra postprocessing stage is required. In contrast, the point sample that PATHWAY_GRAPH algorithm constructs contains a μ -sample of the boundary surface of $\cup \mathcal{H}$, and in addition contains surplus points that lie completely inside $\cup \mathcal{H}$ and 'cover' $\cup \mathcal{H}$, i.e., the Hausdorff distance between $\cup \mathcal{H}$ and E is bounded by $\lambda = 1$. This ensures us that the outer part of the λ -medial of $\cup \mathcal{H}^c$ (which is the part that lies within $\cup \mathcal{H}$) is an empty set, and therefore we do not need any postprocessing.

At last, we arrive at the proof of Theorem 4.4, which is now rather straightforward.

Proof (Theorem 4.4): Let A be a collection of closed bounded convex bodies in \mathbb{R}^d , let O denote $\cup A^c$, let $\lambda \geq 0$ be a regular value of O , and let $\delta > 0$ be a real number. There exists a real $\mu > 0$ such that $d_H(O^c, \tilde{O}^c) \leq \mu \implies d_H(M_\lambda[O], M_\lambda[\tilde{O}]) \leq \delta$ (according to Theorem 2.1). We choose a point sample $E \subset \cup A$ that satisfies the conditions of the theorem, and define \tilde{O} to be $O^{+\mu} \setminus E$. Since \tilde{O} satisfies $d_H(O^c, \tilde{O}^c) \leq \mu$ it holds that $d_H(M_\lambda[O], M_\lambda[\tilde{O}]) \leq \delta$. Therefore $d_H(M_\lambda[\cup A^c], M_\lambda[E^c]) \leq \delta$, since $\cup A^c = O$ (by definition) and $M_\lambda[\tilde{O}] = M_\lambda[E^c]$ (according to Lemma 4.6).

5 A Bound on the Number of Unit Balls

In this section we address the complexity aspect of our approximation scheme, focusing on the number of unit balls in \mathcal{K}_ε . We derive an upper bound on the ratio $|\mathcal{K}_\varepsilon|/|\mathcal{B}|$. In Section 7 we employ this bound to show the advantages of using ball samples over using point samples for approximating the medial axis of the complement of $\cup\mathcal{B}$ and approximating the persistence diagrams of $\cup\mathcal{B}$.

Based on the definitions of a ball sample and a point sample, presented in Section 3.2, we define the *approximation quality* of a sample. Let X be a closed bounded subset of \mathbb{R}^3 , let E be a finite point sample of X , and let K be a finite ball sample of X . We call the one-sided Hausdorff distance $d_H(\partial X|E)$ the *approximation quality* of E . In a similar fashion, we call the one-sided Hausdorff distance $d_H(\partial X|\cup K)$ the *approximation quality* of K . Given an integer $\chi > 0$, and denoting by μ the approximation quality of E , we say that E is χ -light if the number of sample points in any ball of radius μ is not greater than χ , namely $\forall x' \in \mathbb{R}^3, |B(x', \mu) \cap E| \leq \chi$.

The following theorem gives an upper bound on the number of unit balls needed to construct an ε -flower of a single ball $B \in \mathcal{B}$.

Theorem 5.1 *Let $B(x, r)$ be a ball with $r \geq 1$, let E be a finite point sample of $B(x, r-1)$. Let $K = K(E)$ be the collection of unit balls centered at E , which constitutes a ball sample of $B(x, r)$. Denoting by ε the approximation quality of K , if $\varepsilon < 1/2$ and E is χ -light then:*

$$|E| \leq \chi \frac{16r^2}{3\varepsilon^{4/3}}.$$

Next we state of corollary of the theorem, which give bounds on the ratio $|\mathcal{K}_\varepsilon|/|\mathcal{B}|$.

Corollary 5.2 *Let $\rho > 0$ and assume that for every ball $B(x, r) \in \mathcal{B}$ it holds that $r \leq \rho$. Assume that the centers of the of balls produced by the sampling procedure $FLOWER(B, \varepsilon)$ constitute a χ -light sampling for an integer χ . It holds that:*

$$|\mathcal{K}_\varepsilon|/|\mathcal{B}| \leq \chi \frac{16\lceil\rho\rceil^3}{3\varepsilon^{4/3}}.$$

If we regard ρ and χ as constants then $|\mathcal{K}_\varepsilon|/|\mathcal{B}| = O(1/\varepsilon^{4/3})$.

If we were interested only in an asymptotic bound, we could have required (without loss of generality) that all radii of balls in \mathcal{B} are at least 2 (while approximating with unit balls). In such a setting we get the following:

Proposition 5.3 *If the assumptions of Corollary 5.2 hold, and assuming that for every ball $B(x, r) \in \mathcal{B}$ it holds that $r \geq 2$, then $|\mathcal{K}_\varepsilon|/|\mathcal{B}| \in O(1/\varepsilon)$.*

Yet for practical reasons it is useful to allow the balls in \mathcal{B} to have a radius 1 or very close to 1, which results in much fewer sample unit balls. This requires a more refined analysis which results in the looser bound of Corollary 5.2.

In order to prove Theorem 5.1 we first prove several auxiliary claims. The following lemma gives upper and lower bounds on the number of points in a μ -sample which is χ -light.

Lemma 5.4 *Let E be a μ -sample of a ball $B = B(x, R)$. If E is χ -light and $\mu < \min(R, 1/2)$, then $16R^2/3\mu^2 \leq |E| \leq \chi(16R^2/3\mu^2)$.*

Proof: Let $S = S(x, R)$ be the boundary sphere of B . We denote by $\mathcal{A}(X)$ the surface area of a bounded surface X . Since E is χ -light it holds that:

$$\mathcal{A}(S) \leq \sum_{p \in E} \mathcal{A}(S \cap B(p, \mu)) \leq \chi \mathcal{A}(S).$$

It is not difficult to see that for any $p \in E$ it holds that $\frac{3}{4}\pi\mu^2 \leq \mathcal{A}(S \cap B(p, \mu)) \leq \pi\mu^2$. Therefore we get:

$$\frac{3}{4}\pi\mu^2|E| \leq \chi \mathcal{A}(S).$$

Since $\mathcal{A}(S) = 4\pi R^2$ we get:

$$\frac{16R^2}{3\mu^2} \leq |E| \leq \chi \frac{16R^2}{3\mu^2}.$$

□

In the next lemma we establish a relation between the approximation quality of a point sample E of $B(x, r - 1)$ and the approximation quality of the ball sample $K(E)$ of $B(x, r)$:

Lemma 5.5 *Let $B = B(x, r)$ be a ball with $r \geq 1$, and let E be a point sample of $B(x, r - 1)$. Let $\varepsilon = d_H(\partial B(x, r) | \cup K(E))$, and $\mu = d_H(\partial B(x, r - 1) | E)$. If $\varepsilon < \min(r - 1, 1/2)$ then:*

$$\mu^2 \geq 2\varepsilon \frac{r - 1}{r}.$$

Proof: Let $p \in \partial B$ and $B(c, 1) = k \in K(E)$ be a point and a ball such that $d(p, k) = \varepsilon$. We know such p and k exist since B and $\cup K$ are compact and ε is the approximation quality. Let $k = B(c, 1)$ be a ball in K that is ε distant from p . We denote by f the radial projection of p onto the ball $B(x, r - 1)$; see Figure 7 for an illustration. We note that $c \in E$ is the closest point of E to f , since k is the closest ball to p . Therefore $\mu \geq d(f, c)$ giving an implicit bound on μ .

We use the following notation. The intersection of the segment \overline{pc} with the boundary of k is marked with d , the closest point to c on the segment \overline{px} is marked with g , and the midpoint of \overline{cf} is marked with m . We denote the length of \overline{cf} by $\bar{\mu}$. Since $\mu \geq d(f, c) = \bar{\mu}$

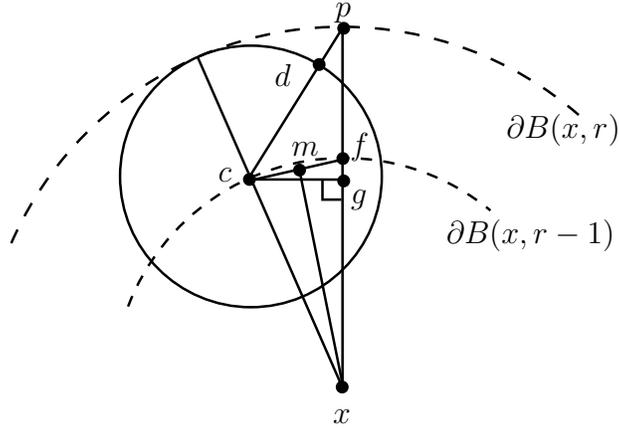


Figure 7: Illustration for the proof of Lemma 5.5 which depicts the geometric relation between μ and ε . The length of \overline{dp} is ε and the length of \overline{cf} bounds μ .

it suffice to prove the bound for $\bar{\mu}$. First, we note that $|\overline{cx}| = r - 1$, $|\overline{px}| = r$, $|\overline{pd}| = \varepsilon$. Note that Δxcf is an isosceles triangle, and therefore $\angle xmf = 90^\circ$. Since $\Delta fgc \sim \Delta fmx$ we deduce $|\overline{fg}|/|\overline{cf}| = |\overline{mf}|/|\overline{xf}|$ or in other words $|\overline{fg}| = \frac{(\bar{\mu})^2}{2(r-1)}$. Considering the triangle Δcgp and the equality $|\overline{pg}| = |\overline{fg}| + 1$ we can express ε as a function of $\bar{\mu}$:

$$\begin{aligned}
\varepsilon &= \sqrt{|\overline{pg}|^2 + |\overline{cg}|^2} - 1 \\
&= \sqrt{(|\overline{fg}| + 1)^2 + |\overline{cg}|^2} - 1 \\
&= \sqrt{|\overline{cf}|^2 + 2|\overline{fg}| + 1} - 1 \\
&= \sqrt{\bar{\mu}^2 + 1 + \frac{(\bar{\mu})^2}{r-1}} - 1.
\end{aligned}$$

We can now work towards expressing $\bar{\mu}$ as a function of ε :

$$\begin{aligned}
(\varepsilon + 1)^2 &= \bar{\mu}^2 + 1 + \frac{(\bar{\mu})^2}{r-1} \\
\varepsilon^2 + 2\varepsilon &= (\bar{\mu})^2 \frac{r}{r-1}.
\end{aligned}$$

Reorganizing the terms we obtain

$$\bar{\mu}^2 = (\varepsilon^2 + 2\varepsilon) \frac{r-1}{r} \geq 2\varepsilon \frac{r-1}{r}.$$

Recall that $\mu \geq d(f, c)$ or in other words $\mu \geq \bar{\mu}$. The bound asserted in the lemma follows. \square

We can finally prove Theorem 5.1, using the lemmas above.

Proof (Theorem 5.1): Let E be a sample as defined in the theorem and let μ be its approximation quality. We handle three cases, according to the approximation quality

ε of $K(E)$ and the radius r . First, the trivial case of a small ball. If $(r - 1) \leq \varepsilon$ then E obviously contains a constant number of points. In the two other cases we can use Lemma 5.5, which states that $\mu^2 \geq 2\varepsilon \frac{r-1}{r}$. We also use Lemma 5.4 (with $R \equiv r - 1$) which states that $|E| \leq \chi \frac{16(r-1)^2}{3\mu^2}$. In case $\varepsilon < (r - 1) \leq \varepsilon^{\frac{1}{3}}$ we get:

$$\mu^2 \geq 2\varepsilon \frac{r-1}{r} \geq 2\varepsilon - \frac{2\varepsilon}{\varepsilon+1} = \frac{2\varepsilon^2}{\varepsilon+1} \geq \varepsilon^2.$$

We use this bound in conjunction with Lemma 5.4 and get a bound on $|E|$:

$$|E| \leq \chi \frac{16(r-1)^2}{3\mu^2} \leq \chi \frac{16\varepsilon^{\frac{2}{3}}}{3\varepsilon^2} \leq \chi \frac{16}{3\varepsilon^{\frac{4}{3}}} \leq \chi \frac{16r^2}{3\varepsilon^{\frac{4}{3}}}.$$

In the remaining case, where $\varepsilon^{\frac{1}{3}} < (r - 1)$ we get:

$$\begin{aligned} \mu^2 &\geq 2\varepsilon \frac{r-1}{r} \geq 2\varepsilon - \frac{2\varepsilon}{r} \geq 2\varepsilon - \frac{2\varepsilon}{\varepsilon^{\frac{1}{3}}+1} = \varepsilon \left(2 - \frac{2}{\varepsilon^{\frac{1}{3}}+1}\right) \\ \mu^2 &\geq 2 \frac{\varepsilon^{\frac{4}{3}}}{\varepsilon^{\frac{1}{3}}+1} \geq \varepsilon^{\frac{4}{3}}. \end{aligned}$$

By using the above inequality together with the bound of Lemma 5.4 we get:

$$|E| \leq \chi \frac{16(r-1)^2}{3\mu^2} \leq \chi \frac{16r^2}{3\varepsilon^{\frac{4}{3}}}.$$

□

Upper Bound on Point Approximation Quality

Let B be a ball in \mathcal{B} , let E be a point sample of $B(x, r - 1)$, and let $K = K(E)$ be the collection of unit balls centered at E , which are a ball sample of $B(x, r)$. We Denote by ε the approximation quality of K , and by μ the approximation quality of E . To conclude this section we state a lemma that gives an upper bound on μ as a function of ε , which we have already used in Section 4.

Lemma 5.6 *Let $B = B(x, r)$ be a ball such that $r \geq 1$, and let E be a point sample of $B(x, r - 1)$. Let ε denote $d_H(\partial B(x, r) | \cup K(E))$, and μ denote $d_H(\partial B(x, r - 1) | E)$. If $\varepsilon < 1/2$ then:*

$$\mu^2 \leq \frac{5}{4}\varepsilon.$$

Proof: Let E be a point sample as above and let μ be its approximation quality. Let f be a point on $B(x, r - 1)$ where the distance from the nearest sample point is μ (see Figure 7). Let $c \in E$ be the closest point to f and p be the point on ∂B closest to f . Let

$k = B(c, 1)$ denote a ball in $K(E)$. The distance of p to $\cup K$ is achieved on k and is equal to $\bar{\varepsilon}$ for a real $\bar{\varepsilon} \leq \varepsilon$. Applying the same analysis as in Lemma 5.5, we get:

$$\mu^2 = (\bar{\varepsilon}^2 + 2\bar{\varepsilon})\frac{r-1}{r} \leq (\varepsilon^2 + 2\varepsilon)\frac{r-1}{r} \leq \frac{5}{4}\varepsilon.$$

□

6 Approximating the Persistence Diagrams

The collection \mathcal{K}_ε of unit balls can be used to construct an efficient approximation of the persistence diagrams of $\cup \mathcal{B}$. The set of unit balls \mathcal{K}_ε , which our algorithm constructs for \mathcal{B} , satisfies $d_H(\cup \mathcal{B}, \cup \mathcal{K}_\varepsilon) \leq \varepsilon$. According to Theorem 2.2, the three persistence diagrams of $\cup \mathcal{K}_\varepsilon$ approximate the three persistence diagrams of $\cup \mathcal{B}$. This implies that “major” topological features, i.e., large voids and tunnels are the same for both unions. The following theorem asserts the efficiency of our approximation, and its proof follows easily from Corollary 5.2.

Theorem 6.1 *Let \mathcal{B} be a finite collection of balls that are each not smaller than a unit ball, let k be an integer in $\{0, 1, 2\}$, and let $\varepsilon < 1/2$ be a real number. The collection \mathcal{K}_ε of unit balls, which algorithm `PATHWAY_GRAPH` constructs, satisfies $d_\beta(D_k(f_{\cup \mathcal{B}}), D_k(f_{\cup \mathcal{K}})) \leq \varepsilon$, and $|\mathcal{K}_\varepsilon| = O(|\mathcal{B}|/\varepsilon^{4/3})$.*

We approximate the pocket diagram of $\cup \mathcal{B}$ using the pocket diagram of $\cup \mathcal{K}_\varepsilon$, which we compute by applying the algorithm presented in [20]. We make use of the pocket diagram to identify the center of the largest chamber within a molecule; see Section 8.

7 Comparison with Point Sampling Techniques

In this section we compare our ball-sample approach with a standard point-sample approach (see, e.g., [7, 23]) for the two problems that we address in this paper: approximating the medial axis of the complement of $\cup \mathcal{B}$ and approximating the persistence diagrams of $\cup \mathcal{B}$. To solve both problems with point samples, the boundary surface of each of the balls in \mathcal{B} is sampled with a collection of points S_ε , such that S_ε contains an ε -sample of the boundary surface of $\cup \mathcal{B}$. As described in [3], the λ -Voronoi graph of S_ε for $\lambda > \varepsilon$ serves as an approximation of the medial axis, which tends to the medial axis as λ tends to 0. Given a fixed k and denoting by f_{S_ε} the distance function from S_ε , the persistence diagram of (f_{S_ε}, k) can serve as an approximation of the persistence diagram of $(f_{\cup \mathcal{B}}, k)$.

Let us focus on a single ball $B \in \mathcal{B}$. The advantage of our approach is the relatively small number of unit balls needed to construct an ε -flower of B , compared to the number of points needed to construct a point ε -sample of B . The number N_{ball} of unit balls needed by `PATHWAY_GRAPH` to constitute an ε -flower B is not more than $\chi(16r^2/3\varepsilon^{4/3})$, according to Theorem 5.1. On the other hand, the number N_{point} of points needed to constitute a point ε -sample of B is at least $16r^2/3\varepsilon^2$, as we show in Lemma 5.4. Therefore we get:

Proposition 7.1 *If we regard χ is a as constant then $N_{point}/N_{ball} = \Omega(1/\varepsilon^{2/3})$.*

8 Experimental Results

We now wish to present results that demonstrate the performance of the algorithm. (For ample biological results see the companion MolAxis paper [30].) All tests were carried out on a Pentium IV 3.0 GHz machine with 1GB of RAM running a LINUX native operating system. We report on a set of tests that were performed on a collection of 3251 balls, which model a P450 Enzyme isozyme (H atoms discarded). The ratio between the largest and smallest input balls is $\rho = 1.21$. (We note that a small value of ρ is the common case in proteins.) In Table 1 we report on a set of tests performed on these input balls with varying resolution. In Figure 8 we provide (a) a graph of the size of \mathcal{K}_ε as a function of ε , comparing the experimental results to the theoretical bound proved in Theorem 5.1 and (b) a graph of the number of Voronoi vertices in the pathway graph as a function of the number of balls in \mathcal{K}_ε . Note that as ε becomes smaller and the number of unit balls increases, the ratio between the number of Voronoi vertices and number of unit balls tends to one. We believe that the explanation of this phenomenon is that our sampled points behave like points sampled on a smooth surface [4] or points sampled on a polyhedral surface [2]. In these two cases the number of faces in the Voronoi diagrams is less than the worst case $\Theta(n^2)$ faces, and is $O(n \log n)$ and $O(n)$ respectively. In Table 2 we provide the runtime breakdown of three runs, for $\varepsilon = 0.4$, $\varepsilon = 0.1$, $\varepsilon = 0.01$.

Resolution	Max BPS	Unit Balls	Runtime (sec's)	File size (Mb)
0.4	1	3251	2.456	2.6
0.2	8	9585	5.120	4.5
0.1	12	19996	8.737	6.3
0.05	42	31287	12.229	7.9
0.02	162	114222	35.522	16.9
0.01	162	362262	101.830	37.7
0.005	642	439182	123.444	44.4

Table 1: Data for a set of runs with different resolutions on a single instance of the P450 enzyme. *Resolution* is the Hausdorff approximation quality ε . *Max BPS* is the maximal number of unit balls used to approximate a single ball. *Unit Balls* is the total number of unit balls used in the approximation (the number of balls in \mathcal{K}_ε). *Runtime* is the total runtime in seconds, including the construction of the pathway graph.

Phase	$\varepsilon = 0.4$	$\varepsilon = 0.1$	$\varepsilon = 0.01$
Construct ball sample	0.016	0.088	1.720
Compute triangulation	0.088	0.568	10.645
Compute alpha shape family	0.456	2.848	51.471

Table 2: Runtime breakdown on a single instance of the P450 enzyme (see the relevant rows in Table 1 for more details about these runs). The time is given in seconds. Note that the most time-consuming phase is the Alpha shape computation, and as ε decreases it becomes more dominant.

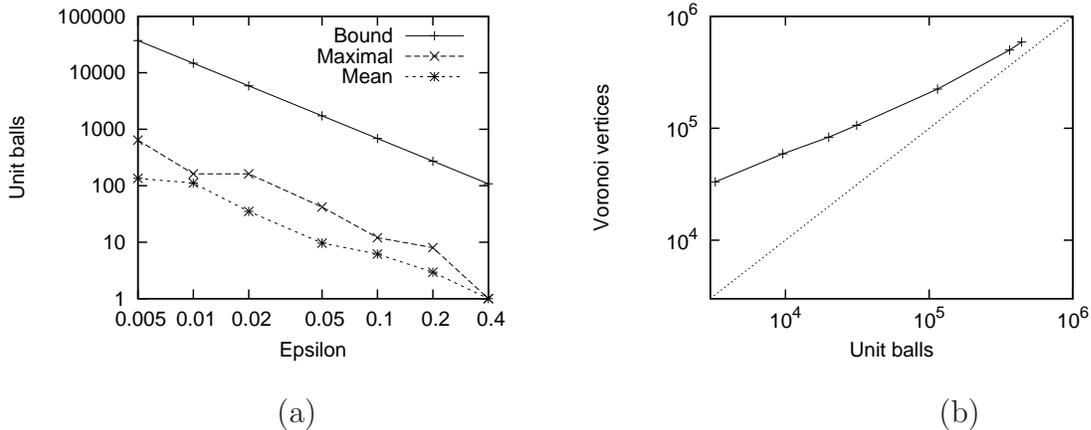


Figure 8: Graphs generated for the set of runs that are analyzed in Table 1. All scales are logarithmic. (a) Experimental and theoretical values of the ratio $|\mathcal{B}|/|\mathcal{K}_\varepsilon|$. *Maximal* is the maximal BPS, i.e., the maximal number of unit balls used to approximate a single ball. *Mean* is the mean BPS over all input balls. *Bound* is the theoretical upper bound as expressed in Theorem 5.1, with $\chi = 3$. Note that for $\varepsilon > 0.2$ we need exactly one unit ball to approximate each input ball. (b) The number of Voronoi vertices in the pathway graph as a function of the number of unit balls in \mathcal{K}_ε . We draw the function $y = x$ as a reference using a dotted line. Note that as the number of unit balls increases the ratio between the number of Voronoi vertices and the number of unit balls becomes closer to one.

Pocket Diagram

Recall that we call the persistence diagram of $(f_{\cup\mathcal{B}}, k)$ the *pocket diagram* of $\cup\mathcal{B}$. To give some intuition as to what the pocket diagram is imagine a continuous uniform growth process of the balls in \mathcal{B} . During the growth process the complement of the expanding $\cup\mathcal{B}$ is divided into connected components, with one unbounded component and zero or more *voids*. Voids are transient in the sense that they may appear and disappear during the growth process. Each void has a *birth time* and a *death time*. We call these temporal voids, which existed during the growth process, the *pockets* (or *chambers*) of \mathcal{B} , similar to the definition by Edelsbrunner *et al.* [19, 20], yet with a more natural geometric meaning.² The *lifetime* of a pocket is defined to be its death time minus its birth time. We extract the lifetime of pockets from the pocket diagram that we have computed. See Figure 9 for an example of the pocket diagram of a collection of balls that model a large molecule (the P450 Enzyme) that has one large inner chamber. An interesting direction for further investigation can be to compare the persistence homology under the Euclidean distance function which we compute, with the persistent homology under the power distance function which is presented in [19, 20].

Constructing the geometric location of persistent homology classes that represent wide tunnels from the tunnel diagram (when $k = 1$) is a non-trivial task. We believe this problem can be addressed using techniques as the one introduced by Freedman *et al.* (see, e.g., [22]).

²Edelsbrunner *et al.* [19] define pockets using the *power distance*, which means that in their growth process the balls are expanded in a non-uniform manner, i.e., larger balls grow faster.

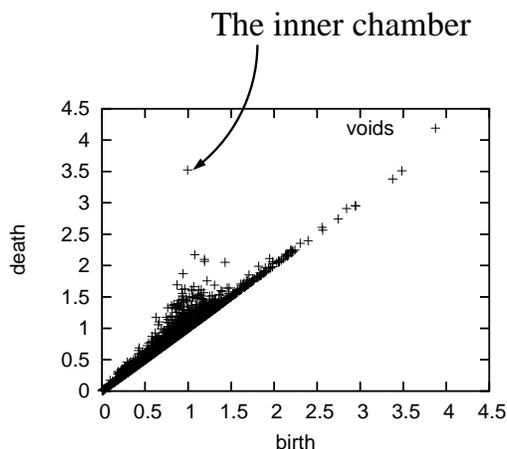


Figure 9: The pocket diagram of a collection of balls that model the P450 Enzyme with one large inner chamber. Notice that all pockets except the inner chamber are close to the $x = y$ diagonal and can be seen as topological noise.

9 MolAxis

MolAxis is a tool designed to assist the biologist or the biochemist in the identification of molecular channels. In a companion paper [30] we introduce MolAxis, and present analysis of channels in molecules achieved with MolAxis. A major advantage of our approach is that since the medial axis is composed of two-dimensional surface patches it reduces the dimension of the problem, i.e., it transforms a three-dimensional problem to a two-dimensional problem. This dimension reduction, combined with the novel sampling technique introduced in this paper, leads to a highly efficient algorithm. To the best of our knowledge it is the first attempt to approximate and analyze the medial axis of the complement of a molecule in order to construct channels.

Since the MolAxis web server [29] was put up on July 2008, it was used to analyze more than 400 distinct molecules. In Figure 10 we show the channels of the human CYP3A4, a P450 isozyme with 3251 atoms, as computed using MolAxis. The CYP3A4 protein has an active site that is located in an inner chamber, and is thought to have 8 channels that connect the active site to the outside of the protein. Being very fast, the tool has been successfully used to analyze channel dimensions and lining residues in hundreds of snapshots of a Molecular Dynamics simulation [24] of the human CYP3A4.

10 Implementation Details

In Section 10.1 we describe how we implemented the procedure $\text{FLOWER}(B, \varepsilon)$ which constructs an ε -flower for a single ball $B \in \mathcal{B}$. In Section 10.2 we explain how we construct the pathway graph \mathcal{P}_ε using ready-made tools. (See Section 3 for the definitions of these entities.)

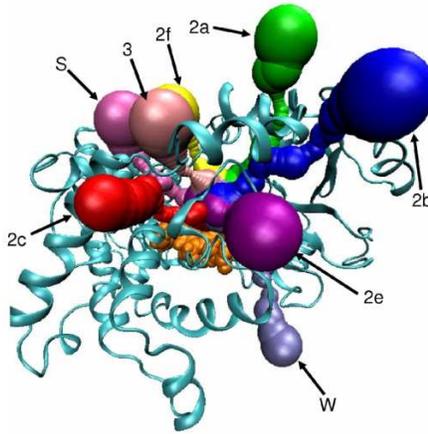


Figure 10: CYP3A4 channels as detected by MolAxis. CYP3A4 is represented by ribbons. The 8 channels are represented with different shades of gray and are tagged according to their biological name.

10.1 ε -Flower Construction

For each ball $B = B(c, r) \in \mathcal{B}$ we construct a set of unit balls K_B that are an ε -flower of B . Constructing an ε -flower with a minimal number of unit balls is an optimization problem closely related to the following problem: “how can n points be distributed on a unit sphere such that they maximize the minimum distance between any pair of points?” Such a configuration of points is called a *spherical code* and its construction has been intensively studied [27]. We employ two heuristic sampling techniques for producing an ε -flower.

The first heuristic is **Icosahedron refinement**. An *icosahedron* is the Platonic solid P_3 having 12 vertices, 30 edges, and 20 congruent equilateral triangular faces. We denote by $I_0(c, r)$ an icosahedron that has its vertices on the sphere $S = S(c, r-1)$. To get $I_1(c, r)$ we refine $I_0(c, r)$ as follows. We split each edge of $I_0(c, r)$ into two new edges, by adding a vertex in its midpoint. We split each triangle of $I_0(c, r)$ into four new smaller triangles, by connecting the new midpoint vertices. Finally we centrally project the new vertices onto the sphere S . In this way we get a *refined icosahedron* of degree one, denoted by $I_1(c, r)$, which comprises 42 vertices, 80 triangles and 120 edges. We iterate the refinement and define a sequence of refined icosahedra: $I_0(c, r), I_1(c, r), I_2(c, r)$, etcetera. For any natural number η we define $\varepsilon_{\text{ico}}(\eta)$ to be the one-sided Hausdorff distance between the boundary surface of B and the union of the collection of unit balls centered at the vertices of $I_\eta(c, r)$. We compute $\varepsilon_{\text{ico}}(\eta)$ by scanning all triangles of $I_\eta(c, r)$. Given the user-specified parameter $\varepsilon > 0$ we find the smallest η such that $\varepsilon_{\text{ico}}(\eta) \leq \varepsilon$, and return the vertices of $I_\eta(c, r)$.

The icosahedron refinement technique has a major drawback: The number of vertices in each icosahedron in the sequence jumps in large steps, i.e., 12, 42, ... Since the first icosahedron has 12 vertices, we have a ‘gap’ between 2 and 11 that we wish to bridge. Therefore we use a second heuristic, **Random points**, as described next.

We repeat the following procedure for each $i = 2, \dots, 11$. Using a typically large integer constant N_{rnd} , we generate $j = 1, \dots, N_{\text{rnd}}$ random point sets E_{ij} , each containing exactly

i points, such that the points of E_{ij} are located on the sphere $S(c, r - 1)$. Recall that for any finite point collection $E \subset \mathbb{R}^3$ we denote by $K(E)$ the collection of unit balls centered at E . We choose the set $E_{i\tilde{j}}$, $1 < \tilde{j} \leq N_{\text{rnd}}$, such that for any $1 < j \leq N_{\text{rnd}}$ it holds that $d_H(\partial B \cup K(E_{i\tilde{j}})) \leq d_H(\partial B \cup K(E_{ij}))$.

We denote $E_{i\tilde{j}}$ by E_i and denote the one-sided Hausdorff distance $d_H(\partial B \cup K(E_{i\tilde{j}}))$ by $\varepsilon_{\text{rnd}}(i)$. After completing the procedure for all $2 \leq i \leq 11$ we have 10 computed point sets $\{E_i\}_{i=2,\dots,11}$ with their respective one-sided Hausdorff distances $\{\varepsilon_{\text{rnd}}(i)\}_{i=2,\dots,11}$. Now, given the user-specified parameter ε , we choose the minimal $2 \leq i \leq 11$ such that $\varepsilon_{\text{rnd}}(i) \leq \varepsilon$. If i exists we use E_i as the centers of the ε -flower. If no i satisfies $\varepsilon_{\text{rnd}}(i) \leq \varepsilon$ we use icosahedron refinement.

10.2 Computing the Pathway Graph

In our implementation we define the pathway graph in a slightly different manner. Instead of defining it to be the $(\lambda = 1)$ -Voronoi graph of the point sample E we define it to be the $(\alpha = 1)$ -Voronoi graph of E (defined below), since the α -Voronoi graph can be easily computed using the 3D Alpha Shapes package [15] of the CGAL library.

α -Voronoi Graph

Let E be a finite point set in \mathbb{R}^3 and let σ be a Delaunay d -simplex of $\mathcal{D}[E]$, spanned by the $d + 1$ point set T . We say that a ball is *empty* in E if it does not contain any points of E in its interior. Let R_T denote the radius of the smallest empty ball that contains all points of T on its boundary surface. We say that the simplex σ is α -*exposed* if $\alpha > R_T$ [21]. The collection of α -exposed simplices is a simplicial complex, which is called the α -*complex* of E . We call the collection of the dual Voronoi faces of simplices that are *not* in the α -complex the α -*Voronoi graph* of E .

We define the $\text{IN_COMPLEX}(\sigma, x)$ predicate of the PATHWAY_GRAPH algorithm (Section 3) to be true if and only if σ is x -exposed. Below we show why this change in definition from 1-enclosed to 1-exposed makes little difference in practice.

λ -Voronoi Graph and α -Voronoi Graph

In contrast to the λ -Voronoi graph, the α -Voronoi graph is defined only for finite point sets and cannot be extended to general open subsets of \mathbb{R}^3 , such as the complement of $\cup \mathcal{B}$. For $\alpha = \lambda$, if a simplex $\sigma_T \in \mathcal{D}[E]$ is α -exposed it is necessarily λ -enclosed but not vice versa. This is because σ_T is λ -enclosed if the set of balls centered at T with radius λ have a non-empty intersection X_T , which is a weaker predicate than the α -exposed predicate which requires that the dual Voronoi face V_T intersects X_T as well ($(X_T \cap V_T) \neq \emptyset$). Thus the α -Voronoi graph of E contains the λ -Voronoi graph of E for $\alpha = \lambda$. The lemma below states this and more. We restate a lemma given in [3], using our terminology and notation:

Lemma 10.1 [3] *For any finite point set E , the $(\alpha = 1)$ -Voronoi graph of E contains the $(\lambda = 1)$ -Voronoi graph of E and the two are homotopy equivalent.*

Remark. Eventually, when we use the $(\alpha = 1)$ -Voronoi graph to find paths in the complement of molecules (see [28, 30] for details), we define a graph G (in the graph theoretic sense) embedded as a straight-edge graph in the $(\alpha = 1)$ -Voronoi graph. We then run an algorithm that looks for certain optimal paths in G . Having a superset of features in G (over what we would have extracted from the $(\lambda = 1)$ -Voronoi graph) does not hamper the quality of the found paths, and has negligible effect on the running time, which is very low to start with.

α -Voronoi Graph Computation

We compute the $(\alpha = 1)$ -Voronoi graph of E_ϵ using the 3D Alpha Shapes package [15] of the CGAL library. We use exact arithmetic to ensure that the algorithm is robust. The Alpha Shapes data structure allows to retrieve the α -complex for any α value. We set the pathway graph \mathcal{P}_ϵ to be the collection of dual Voronoi faces of the simplices that are *not* in the $(\alpha = 1)$ -complex of E_ϵ .

11 Glossary

We summarize the main terms used in this paper. As already mentioned, we have slightly revised the new terminology, relative to our previous publications [28, 31].³

- **Core Axis** (Section 4): The set of points in the medial axis of $\cup\mathcal{B}^c$ for which the collection of (more than one) closest balls in \mathcal{B} do not have a common intersection. It is a subset of the medial axis of $\cup\mathcal{B}^c$ without ‘dead-ends’, and it is denoted by $CA[\mathcal{B}]$. The core axis is suitable for finding pathways that lie outside $\cup\mathcal{B}$.
- **Pathway Axis** (Section 3): A useful subset of the medial axis of $\cup\mathcal{B}^c$, denoted by $PWA[\mathcal{B}]$. In Section 4 we prove that the pathway axis contains the core axis.
- **Pathway Graph** (Section 3): Piecewise linear approximation of the pathway axis, which is constructed by the `PATHWAY_GRAPH` algorithm.
- **ϵ -Flower** (Section 3): A primitive that approximates a ball by using unit balls, and is used to construct the pathway graph.

References

- [1] N. Amenta, S. Choi, and R. Kolluri. The power crust, unions of balls, and the medial axis transform. *Computational Geometry: Theory and Applications*, 19(2–3):127–153, 2001.

³The term ‘pathway diagram’ of previous work [28, 31] is replaced in this paper by the term ‘pathway graph’ to avoid confusion with the persistence diagram of Cohen-Steiner *et al.* [14]. The ‘pathway axis’ of previous work is called in this paper the ‘core axis’.

- [2] D. Attali and J.-D. Boissonnat. A linear bound on the complexity of the Delaunay triangulation of points on polyhedral surfaces. *Discrete & Computational Geometry*, 31(3):369–384, 2004.
- [3] D. Attali, J.-D. Boissonnat, and H. Edelsbrunner. Stability and computation of medial axes: A state of the art report. In B. H. T. Möller and B. Russell, editors, *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration*. Springer-Verlag, Mathematics and Visualization, 2007.
- [4] D. Attali, J.-D. Boissonnat, and A. Lieutier. Complexity of the Delaunay triangulation of points on surfaces: The smooth case. In *Proceedings of the Symposium on Computational Geometry*, pages 201–210, 2003.
- [5] F. Aurenhammer and R. Klein. Voronoi diagrams. In J.-R. Sack and J. Urrutia, editors, *Handbook of Computational Geometry*, pages 201–290. Elsevier Science Publishers B.V. North-Holland, Amsterdam, 2000.
- [6] J.-D. Boissonnat and C. Delage. Convex hull and Voronoi diagram of additively weighted points. In *Proceedings of the European Symposium on Algorithms*, pages 367–378, 2005.
- [7] J.-D. Boissonnat and S. Oudot. Provably good sampling and meshing of surfaces. *Graphical Models*, 67(5):405–451, 2005.
- [8] J.-D. Boissonnat and M. Yvinec. *Algorithmic Geometry*. Cambridge University Press, UK, 1998. Translated from the French version by H. Brönnimann.
- [9] G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. In *Proceedings of the Symposium on Computational Geometry*, pages 184–193, 2007.
- [10] F. Chazal and A. Lieutier. Stability and homotopy of a subset of the medial axis. In *Proceedings of the Symposium on Solid Modeling and Applications*, pages 243–248, 2004.
- [11] F. Chazal and A. Lieutier. The “Lambda-medial axis”. *Graphical Models*, 67(4):304–331, 2005.
- [12] F. Chazal and A. Lieutier. Weak feature size and persistent homology: Computing homology of solids in \mathbb{R}^n from noisy data samples. In *Proceedings of the Symposium on Computational Geometry*, pages 255–262, 2005.
- [13] F. Chazal and S. Oudot. Towards persistence-based reconstruction in Euclidean spaces. In *Proceedings of the Symposium on Computational Geometry*, pages 232–241, 2008.
- [14] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

- [15] T. K. F. Da and M. Yvinec. 3D Alpha Shapes. In CGAL Editorial Board, editor, *CGAL- 3.2 User and Reference Manual*. 2006. http://www.cgal.org/Manual/3.2/doc_html/cgal_manual/Alpha_shapes_3/Chapter_main.html.
- [16] T. K. Dey and R. Wenger. Stability of critical points with interval persistence. *Discrete & Computational Geometry*, 38(3):479–512, 2007.
- [17] T. K. Dey and W. Zhao. Approximate medial axis as a Voronoi subcomplex. *Computer-Aided Design*, 36(2):195–202, 2004.
- [18] H. Edelsbrunner. *Geometry and Topology for Mesh Generation*. Cambridge University Press, New York, NY, USA, 2001.
- [19] H. Edelsbrunner, M. A. Facello, and J. Liang. On the definition and the construction of pockets in macromolecules. *Discrete Applied Mathematics*, 88(1-3):83–102, 1998.
- [20] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4):511–533, 2002.
- [21] H. Edelsbrunner and E. P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, 1994.
- [22] D. Freedman and C. Chen. Measuring and localing homology classes. *The Computing Research Repository (CoRR)*, abs/0705.3061, 2007. <http://arxiv.org/abs/0705.3061>.
- [23] J. Giesen, E. A. Ramos, and B. Sadri. Medial axis approximation and unstable flow complex. In *Proceedings of the Symposium on Computational Geometry*, pages 327–336, 2006.
- [24] A. Leach. *Molecular Modelling: Principles and Applications*. Prentice Hall, 2001.
- [25] A. Lieutier. Any open bounded subset of \mathbb{R}^n has the same homotopy type as its medial axis. *Computer-Aided Design*, 36(11):1029–1046, 2004.
- [26] J. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, 1984.
- [27] E. W. Weisstein. Spherical code, from mathworld — a wolfram web resource, <http://mathworld.wolfram.com/sphericalcode.html>, 2000.
- [28] E. Yaffe. Efficient construction of pathways in the complement of the union of balls in \mathbb{R}^3 . *M.Sc. Tel-Aviv University*, September 2007. <http://www.cs.tau.ac.il/~eitanyaf/thesis.pdf>.
- [29] E. Yaffe, D. Fishelovitch, H. J. Wolfson, D. Halperin, and R. Nussinov. MolAxis: A server for identification of channels in macromolecule. *Nucleic Acids Research*, 36:W210–W215, 2008.
- [30] E. Yaffe, D. Fishelovitch, H. J. Wolfson, D. Halperin, and R. Nussinov. MolAxis: Efficient and accurate identification of channels in macromolecules. *Proteins: Structure, Function, and Bioinformatics*, 73.1:72–86, 2008.

- [31] E. Yaffe and D. Halperin. Approximating the pathway axis and the persistence diagram of a collection of balls in 3-space. In *Proceedings of the Symposium on Computational Geometry*, pages 260–269, 2008.