# Geodesic Delaunay Triangulation and Witness Complex in the Plane

Jie Gao* Leonidas J. Guibas† Steve Y. Oudot† Yue Wang*

**Abstract**

We introduce a novel feature size for bounded planar domains endowed with an intrinsic metric. Given a point $x$ in such a domain $X$, the *homotopy feature size* of $X$ at $x$, or $\mathrm{hfs}(x)$ for short, measures half the length of the shortest loop through $x$ that is not null-homotopic in $X$. The resort to an intrinsic metric makes $\mathrm{hfs}(x)$ rather insensitive to the local geometry of $X$, in contrast with its predecessors (local feature size, weak feature size, homology feature size). This leads to a reduced number of samples that still capture the topology of $X$. Under reasonable sampling conditions involving $\mathrm{hfs}$, we show that the geodesic Delaunay triangulation $\mathcal{D}_X(L)$ of a finite sampling $L$ of $X$ is homotopy equivalent to $X$. Moreover, $\mathcal{D}_X(L)$ is sandwiched between the geodesic witness complex $\mathcal{C}^W_X(L)$ and a relaxed version $\mathcal{C}^W_{X,\nu}(L)$, defined by a parameter $\nu$. Taking advantage of this fact, we prove that the homology of $\mathcal{D}_X(L)$ (and hence of $X$) can be retrieved by computing the persistent homology between $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$. We propose algorithms for estimating $\mathrm{hfs}$, selecting a landmark set of sufficient density, building its geodesic Delaunay triangulation, and computing the homology of $X$ using $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$. We also present some simulation results in the context of sensor networks that corroborate our theoretical statements.

## 1 Introduction

There are many situations where a topological domain or space $X$ is known to us only through a finite set of samples. Understanding the global topological and geometric properties of $X$ through its samples is important in a variety of applications, including surface parametrization in geometry processing, non-linear dimensionality reduction for manifold learning, routing and information discovery in sensor networks, etc. Recent advances in geometric data analysis and in sensor networks have made an extensive use of a *landmarking strategy*. Given a point cloud $W$ sampled from a hidden domain or space $X$, the idea is to select a subset $L \subset W$ of landmarks, on top of which some data structure is built to encode the geometry and topology of $X$ at a particular scale. Examples in data analysis include the topology estimation algorithm of [16] and the multi-scale reconstruction algorithm of [6, 27]. Both algorithms rely on the structural properties of the *witness complex*, a data structure specifically designed by de Silva [15] for use with the landmarking strategy. Examples in sensor networks include the GLIDER routing scheme and its variants [22, 21]. The idea underlying these techniques is that the use of sparse landmarks at different density levels enables us to reduce the size of the data structures, and to perform calculations on the input data set at different scales. Two questions arise naturally: (1) how many landmarks are necessary to capture the invariants of a given object $X$ at a given scale? (2) what data structures should be built on top of them?

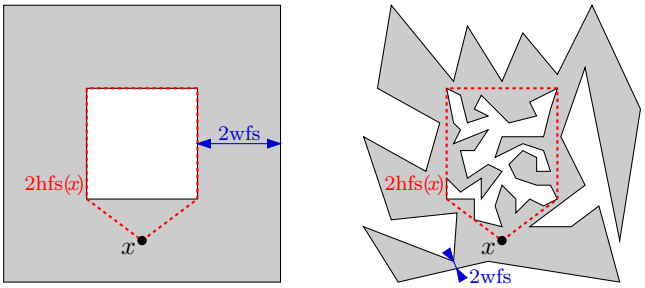



Figure 1: Two Lipschitz domains with very different weak feature sizes ($\mathrm{wfs}$), but similar homotopy feature sizes ($\mathrm{hfs}$).

Manifold sampling issues have been intensively studied in the past, independently of the context of landmarking. The first results in this vein were obtained by Amenta, Bern, and Eppstein, for the case where $X$ is a smoothly-embedded closed curve in the plane or surface in 3-space [1, 2]. Their bound on the landmark density depends on the local distance to the medial axis of $\mathbb{R}^2 \setminus X$ (the *local feature size*), and a data structure built on top of $L$ is the so-called *restricted Delaunay triangulation*. Several extensions of their result have been proposed, to deal with noisy data sets [17], sampled from closed manifolds of arbitrary dimensions [6, 13], smoothly or non-smoothly embedded in Euclidean spaces [7]. In parallel, others have focused on unions of congruent Euclidean balls and their topological invariants, which can be computed via the dual complex – known as the Čech complex. In a seminal paper [31], Niyogi *et al.* proved that, if $X$ is a smoothly-embedded closed manifold and $L$ a dense enough sampling of $X$, then, for a wide range of

*Department of Computer Science, Stony Brook University, Stony Brook, NY 11794. Email: {jgao,yuewang}@cs.sunysb.edu.

†Department of Computer Science, Stanford University, Stanford, CA 94305. Email: guibas@cs.stanford.edu, steve.oudot@stanford.edu

values of $r$, the union of the Euclidean open balls of radius $r$ about the points of $L$ deformation retracts onto $X$.

The above results hold only for closed manifolds. The presence of boundaries brings in some new issues and challenges. An interesting class of manifolds with boundaries is the one of bounded domains in $\mathbb{R}^n$, modeling the existence of natural obstacles to sampling certain areas. By studying the stability of distance functions to compact sets in $\mathbb{R}^n$, Chazal and others have extended the results of Niyogi *et al.* to a larger class of objects, including all bounded domains $X$ with piecewise-analytic boundaries [11]. Their bound on the landmark density depends on the so-called *weak feature size* of $X$, defined as the smallest positive critical value of the Euclidean distance to $\partial X$. This mild sampling condition makes the results of [11] valid in a very general setting. However, the weak feature size can be small compared to the size of the topological features of $X$, as illustrated in Figure 1 (right). As a result, many sample points are wasted satisfying the sampling condition of [11], when very few could suffice to capture the topology of $X$. In practice, this results in a considerable waste of memory and computation power.

The use of local/weak feature sizes in sampling a domain faces new challenges in certain classes of applications such as sensor networks for which the extrinsic metric is not readily available. In a sensor network, nodes may not know their locations, nor do they have any idea of the global picture such as whether there are holes in the network or the nodes on the hole boundaries. The only available metric is the wireless connectivity graph distances as measured by the hop count or shortest path distance metric. This strongly motivates us to use an intrinsic metric on the domain, instead of the extrinsic metric provided by the embedding for topology discovery and understanding. Intrinsic metrics have been studied in the context of Riemannian manifolds without boundary [30] and, from a more computational point of view, in the context of the so-called *intrinsic* Delaunay triangulations (iDT) of triangulated surfaces without boundary [5]. 2-D triangle meshes in 3-D that happen to coincide with the iDT of their vertices are known to have many attractive properties for PDE discretization [23], and generating such iDT meshes is a topic of considerable interest in geometry processing [18].

**Our contributions.** In the paper we focus on the special case of bounded domains in the plane — a setting which already raises numerous questions and finds important applications in sensor networks. We make the novel claim that resorting to an intrinsic metric instead of the Euclidean metric can result in very significant reductions in terms of the number of samples required to recover the homotopy type of a bounded domain. This is an especially appealing fact in the context of resource-constrained sensor nodes, as the number of samples directly translates to the storage requirement in the GLIDER routing scheme [22]. To this end, we introduce a new quantity, called the *homotopy feature size*, or $\mathrm{hfs}$ for short, which measures the size of the smallest topological feature (hole in this case) of the considered planar domain $X$. Specifically, given a point $x \in X$, $\mathrm{hfs}(x)$ is defined as half the length of the shortest loop through $x$ that is not null-homotopic in $X$ – see Figure 1 for an illustration. In particular, $\mathrm{hfs}(x)$ is infinite whenever $x$ lies in a simply connected component of $X$. In contrast with previous quantities, $\mathrm{hfs}$ depends essentially on the global topology of $X$, and it is only marginally influenced by the local geometry of the domain boundary. Under the assumption that $X$ has Lipschitz boundaries (the actual Lipschitz constant being unimportant in our context), we show that $\mathrm{hfs}$ is well-defined, positive, and 1-Lipschitz in the intrinsic metric. Moreover, if $L$ is a geodesic $\varepsilon\mathrm{hfs}$-sample of $X$, for some $\varepsilon < \frac{1}{3}$, then the cover of $X$ formed by the geodesic Voronoi cells of the points of $L$ satisfies the conditions of the Nerve theorem [8, 33], and therefore its dual Delaunay complex $\mathcal{D}_X(L)$ is homotopy equivalent to $X$. By geodesic $\varepsilon\mathrm{hfs}$-sample of $X$, we mean that every point $x \in X$ is at a finite geodesic distance to $L$, bounded from above by $\varepsilon\ \mathrm{hfs}(x)$. In the particular case when $X$ is simply connected, our sampling condition only requires that $L$ have at least one point on each connected component of $X$, regardless of the local geometry of $X$. In the general case, our sampling condition can be satisfied by placing a constant number of landmarks around each hole of $X$, and a number of landmarks in the remaining parts of $X$ that is logarithmic in the ratio of the geodesic diameter of $X$ to the geodesic perimeter of its holes. This is rather independent of the local geometry of the boundary $\partial X$ and can result in selecting far fewer landmarks than required by any of the earlier sampling conditions that guarantee topology capture.

The homotopy feature size is closely related to the concept of injectivity radius in Riemannian geometry. We stress this relationship in the paper, by showing that, for all point $x \in X$, $\mathrm{hfs}(x)$ is equal to the geodesic distance from $x$ to its cut-locus in $X$. This result also suggests a simple procedure for estimating $\mathrm{hfs}(x)$ at any point $x \in X$. Using this procedure, we devise a greedy algorithm for generating $\varepsilon\mathrm{hfs}$-samples of any given Lipschitz planar domain $X$, based on a packing strategy. The size of the output lies within a constant factor of the optimal, the constant depending on the doubling dimension of $X$. Our algorithm relies on two oracles whose actual implementations depend on the application considered. We provide some implementations in the context of sensor networks, based on pre-existing distributed schemes [22, 32].

Next, we focus on the structural properties of the so-called *geodesic witness complex*, an analog of the usual witness complex in the intrinsic metric. In many applications, computing $\mathcal{D}_X(L)$ can be hard, due to the difficulty of checking whether three or more geodesic Voronoi cells have a common intersection. This is especially true in sensor net-

works, where the intersections between the Voronoi cells of the landmarks can only be sought for among the set of nodes $W$, due to the lack of further information on the underlying domain $X$. Therefore, it is convenient to replace $\mathcal{D}_X(L)$ by the geodesic witness complex $\mathcal{C}_X^W(L)$, whose computation only requires us to perform geodesic distance comparisons, instead of locating points equidistant to multiple landmarks. Assuming that the geodesic distance can be computed exactly, we prove an analog of de Silva's theorem [15], which states that $\mathcal{C}_X^W(L)$ is included in $\mathcal{D}_X(L)$ under some mild sampling conditions. We also prove an analog of Lemma 3.1 of [27], which states that a relaxed version of $\mathcal{C}_X^W(L)$ (in which a simplex is $\nu$-witnessed by $w$ if its vertices belong to the $\nu+1$ nearest landmarks of $w$), denoted by $\mathcal{C}_{X,\nu}^W(L)$, contains $\mathcal{D}_X(L)$ under similar conditions. Unfortunately, as pointed out in [27], it is often the case that neither of them coincides with $\mathcal{D}_X(L)$. However, taking advantage of the fact that $\mathcal{D}_X(L)$ is sandwiched between $\mathcal{C}_X^W(L)$ and $\mathcal{C}_{X,\nu}^W(L)$, we show that computing the persistent homology between $\mathcal{C}_X^W(L)$ and $\mathcal{C}_{X,\nu}^W(L)$ gives the homology of $\mathcal{D}_X(L)$. This allows us to retrieve the homology of $X$ without computing $\mathcal{D}_X(L)$ in practice. Similar results have been proved for other types of filtrations [12, 14] and used in the context of sensor networks [26]. However, to the best of our knowledge, our result is the first one of this type for the witness complex filtration.

Finally, remark that when a bounded planar domain is given explicitly with its embedding, its topology is captured by its medial axis [4], and it can be computed efficiently by extracting homotopy bases [19, 20]. The work in this paper gives a way of extracting and learning the topology of the domain through its intrinsic geodesic metric, without the need for a geometric embedding.

## 2 The intrinsic metric

Let $I = [0,1]$. The ambient space is $\mathbb{R}^2$, endowed with the Euclidean metric, noted $\mathrm{d}_E$. Given a subset $X$ of $\mathbb{R}^2$, $\mathring{X}$, $\overline{X}$, and $\partial X$, stand respectively for the interior, closure, and boundary of $X$. Given $x \in \mathbb{R}^2$ and $r \in \mathbb{R}_+$, $B_E(x,r)$ denotes the Euclidean open ball of radius $r$ about $x$. Finally, $S^1$, $\mathbb{R}\times\{0\}$, and $\mathbb{R}^2_+$, denote respectively the unit circle, the abcissa line, and the closed upper half-plane in $\mathbb{R}^2$.

**Paths and loops.** Given a topological space $X$, a *path* in $X$ is a continuous map $I \to X$. For all $a, b \in I$ ($a \le b$), we call $\gamma_{|[a,b]}$ the path $s \mapsto \gamma(a + s(b-a))$, which can be viewed as the restriction of $\gamma$ to the segment $[a,b]$. In addition, $\bar{\gamma}$ denotes the inverse path $s \mapsto \gamma(1-s)$. Given another path $\gamma' : I \to X$ such that $\gamma'(0) = \gamma(1)$, we call $\gamma \cdot \gamma'$ their concatenation, defined by $\gamma\cdot\gamma'(s) = \gamma(2s)$ for $0 \le s \le \frac{1}{2}$ and $\gamma\cdot\gamma'(s) = \gamma'(2s-1)$ for $\frac{1}{2} \le s \le 1$. Given a point $x \in X$, a *loop* through $x$ in $X$ is a path $\gamma$ in $X$ that starts and ends at $x$, *i.e.* such that $\gamma(0) = \gamma(1) = x$. For simplicity, we write $\gamma : (I, \partial I) \to (X, x)$. Note that $\gamma$ can also be seen as a continuous map from the unit circle to $X$, and we write $\gamma : (S^1, 1) \to (X, x)$ to specify that $\gamma(1) = x$.

To any loop $\gamma : S^1 \to S^1$ corresponds a unique integer $\deg\gamma \in \mathbb{Z}$, called the *degree* of $\gamma$, such that $\deg\gamma = 0$ if $\gamma$ is a constant map $S^1 \to \{x\}$, and $\deg(\gamma\cdot\gamma') = \deg\gamma + \deg\gamma'$ for any loop $\gamma' : S^1 \to S^1$ satisfying $\gamma'(0) = \gamma(1)$. Moreover, it can be proved that $\deg\gamma = \deg\gamma'$ iff $\gamma$ and $\gamma'$ are homotopic in $S^1$ [28, Thm. 1.7], so that $\deg\gamma$ is a unique descriptor of the homotopy class of the loop $\gamma$. A similar concept exists for loops in the plane. Given $\gamma : S^1 \to \mathbb{R}^2$ and $x \in \mathbb{R}^2 \setminus \gamma(S^1)$, consider the map $\gamma_x = \pi_x \circ \gamma : S^1 \to S^1$, where $\pi_x : \mathbb{R}^2 \setminus \{x\} \to S^1$ is the radial projection: $\pi_x(y) = \frac{y-x}{\|y-x\|}$. Since $\pi_x$ is continuous over $\mathbb{R}^2 \setminus \{x\}$, $\gamma_x$ is a continuous loop in $S^1$. We then define the degree (aka winding number) of $\gamma$ with respect to $x$ as: $\deg_x\gamma = \deg\gamma_x$. If $\Gamma : S^1 \times I \to \mathbb{R}^2 \setminus \{x\}$ is a homotopy between two loops $\gamma, \gamma'$ in $\mathbb{R}^2\setminus\{x\}$, then $\pi_x \circ \Gamma$ is a homotopy between $\pi_x\circ\gamma$ and $\pi_x\circ\gamma'$ in $S^1$, and therefore we have: $\deg_x\gamma = \deg(\pi_x\circ\gamma) = \deg(\pi_x\circ\gamma') = \deg_x\gamma'$.

COROLLARY 2.1. *For any point $x \in \mathbb{R}^2$ and any loops $\gamma, \gamma' : S^1 \to \mathbb{R}^2 \setminus \{x\}$ that are homotopic in $\mathbb{R}^2 \setminus \{x\}$, we have $\deg_x\gamma = \deg_x\gamma'$. In particular, if $\gamma$ or $\gamma'$ is constant, then $\deg_x\gamma = \deg_x\gamma' = 0$.*

**Length structures and Lipschitz planar domains.** A good introduction to length spaces can be found in [9, Chap. 2]. Every subset $X$ of $\mathbb{R}^2$ inherits a length structure from $\mathbb{R}^2$, where admissible paths are all continuous paths $I \to X$, and where the length of a path $\gamma$ is defined by: $|\gamma| = \sup\{\sum_{i=0}^{n-1} \mathrm{d}_E(\gamma(t_i), \gamma(t_{i+1})),\ n \in \mathbb{N},\ 0 = t_0 \le t_1 \le \cdots \le t_n = 1\}$, the supremum being taken over all decompositions of $I$ into an arbitrary (finite) number of intervals. We clearly have $|\bar{\gamma}| = |\gamma|$. However, $|\gamma|$ is not always finite. Take for instance Koch's snowflake, a fractal curve defined as the limit of a sequence of polygonal curves in the plane. It can be easily shown that, at each iteration of the construction, the length of the curve is multiplied by $\frac{4}{3}$, so that the length of the limit curve is infinite. We say that $\gamma : I \to X$ is a *rectifiable* path if its length $|\gamma|$ is finite.

We make $X$ into a length space by defining an *intrinsic* (or *geodesic*) metric $\mathrm{d}_X$ as follows: $\forall x, y \in X$, $\mathrm{d}_X(x,y) = \inf\{|\gamma|,\ \gamma : I \to X,\ \gamma(0) = x,\ \gamma(1) = y\}$, the infimum being taken over all paths from $x$ to $y$ in $X$. Clearly, $\mathrm{d}_X(x,y) = +\infty$ whenever $x, y$ belong to different path-connected components of $X$. However, the converse is not always true. Take for instance a domain $X$ made of two disjoint disks connected by Koch's snowflake: if $x, y$ belong to different disks, then all curves connecting $x$ and $y$ go through Koch's snowflake and therefore have infinite length. As a consequence, the intrinsic topology induced by $\mathrm{d}_X$ on $X$ can be different from the Euclidean topology induced by $\mathrm{d}_E$. This is a critical issue because the geodesic Voronoi

diagram is bound to the intrinsic metric, whereas our goal is to retrieve the homotopy type of $X$ in the extrinsic metric. Another issue is that not all pairs of points $x, y \in X$ with $\mathrm{d}_X(x,y) < +\infty$ may have a shortest path in $X$, *i.e.* a path $\gamma : I \to X$ such that $\gamma(0) = x$, $\gamma(1) = y$, and $|\gamma| = \mathrm{d}_X(x,y)$. Take for instance two diametral points on the boundary of the unit closed disk, to which the closed disk of radius $\frac{1}{2}$ has been removed. These issues lead us to consider the special case of Lipschitz domains:

DEFINITION 2.1. *A* Lipschitz planar domain *is a compact embedded topological 2-submanifold of $\mathbb{R}^2$ with Lipschitz boundary. Formally, it is a compact subset $X$ of $\mathbb{R}^2$ such that, for all point $x \in \partial X$, there exists a neighborhood $V_x$ in $\mathbb{R}^2$ and a Lipschitz homeomorphism $\phi_x : \mathbb{R}^2 \to \mathbb{R}^2$, such that $\phi_x(0) = x$, $\phi_x(\mathbb{R} \times \{0\}) \cap V_x = \partial X \cap V_x$, and $\phi_x(\mathbb{R}^2_+) \cap V_x = X \cap V_x$.*

Observe that, for any neighborhood $V'_x \subseteq V_x$, we also have $\phi_x(0) = x$, $\phi_x(\mathbb{R} \times \{0\}) \cap V'_x = \partial X \cap V'_x$, and $\phi_x(\mathbb{R}^2_+) \cap V'_x = X \cap V'_x$. Therefore, $V_x$ can be assumed to be arbitrarily small. Moreover, since $\phi_x(0) = x$ and $\phi_x$ is continuous, $\phi_x^{-1}(V_x)$ is a neighborhood of the origin in $\mathbb{R}^2$, hence it contains an open Euclidean disk $B$ about the origin. By taking $\phi(B)$ as the new neighborhood $V_x$, we ensure that $\phi_x^{-1}(X \cap V_x)$ is the intersection of $\mathbb{R}^2_+$ with the open disk $B$, and therefore that it is convex.

Note that the actual Lipschitz constants of the charts $\phi_x$ in Definition 2.1 are unimportant: only the fact that the $\phi_x$ are Lipschitz counts. This makes the class of Lipschitz planar domains quite large: in particular, it contains all planar domains with piecewise-analytic boundaries. Moreover, the pathologies described above cannot occur on a Lipschitz domain, by the following theorem:

THEOREM 2.1. *For any Lipschitz planar domain $X$,*

- *(i) the intrinsic topology coincides with the Euclidean topology on $X$;*
- *(ii) every sequence of paths with uniformly bounded length contains a uniformly converging subsequence; therefore, all points $x, y \in X$ such that $\mathrm{d}_X(x,y) < +\infty$ have a shortest path in $X$;*
- *(iii) for any path $\gamma : I \to X$ and any real number $\varepsilon > 0$, there exists a rectifiable path $\gamma_\varepsilon : I \to X$, homotopic to $\gamma$ relative[1] to $\partial I$ in $X$, such that $\max_{s \in I} \min_{t \in I} \mathrm{d}_X(\gamma_\varepsilon(s), \gamma(t)) < \varepsilon$.*

The proof of the theorem is given in the full version of the paper [25]. It relies on the following facts: given a point $x \in \mathring{X}$, there is a small convex neighborhood $V_x \subseteq X$ inside which any given arc can be continuously deformed into a rectifiable arc. Now, given a point $x \in \partial X$, there is no such neighborhood as above. However, Definition 2.1 provides us with a neighborhood $V_x$ and a Lipschitz homeomorphism $\phi_x$ such that $\phi_x^{-1}(V_x \cap X)$ is convex. Then, inside $\phi_x^{-1}(V_x \cap X)$, we can deform any given arc into a rectifiable arc, whose image through $\phi_x$ is rectifiable and included in $X$.

[1] This means that the homotopy between $\gamma_\varepsilon$ and $\gamma$ is constant over $\partial I = \{0, 1\}$.

## 3 The homotopy feature size

DEFINITION 3.1. *Given a Lipschitz planar domain $X$ and a point $x \in X$, the* homotopy feature size *of $X$ at $x$ is the quantity:* $\mathrm{hfs}(x) = \frac{1}{2} \inf\{|\gamma|,\ \gamma : (S^1, 1) \to (X, x)$ *non null-homotopic in $X$*$\}$.

As illustrated in Figure 1, the resort to the intrinsic metric makes the homotopy feature size rather insensitive to the local geometry of the domain $X$.

LEMMA 3.1. *Let $X$ be a Lipschitz planar domain.*

- *(i) Given a point $x \in X$, if the path-connected component of $X$ that contains $x$ is simply connected, then $\mathrm{hfs}(x) = +\infty$. Else, $\mathrm{hfs}(x) < +\infty$, and there exists a non null-homotopic rectifiable loop $\gamma : (S^1, 1) \to (X, x)$ such that $\mathrm{hfs}(x) = \frac{1}{2}\,|\gamma| > 0$.*
- *(ii) The map $x \mapsto \mathrm{hfs}(x)$ is 1-Lipschitz in the intrinsic metric. As a consequence, it is continuous for the Euclidean topology, by Theorem 2.1, and $\mathrm{hfs}(X) = \inf\{\mathrm{hfs}(x),\ x \in X\}$ is positive.*

The proof of assertion (i) considers an arbitrary sequence of non null-homotopic rectifiable loops through $x$, whose lengths converge towards $2\ \mathrm{hfs}(x)$, and it applies Theorem 2.1 (ii) to this sequence. The proof of assertion (ii) takes two points $x, y$ lying in a same path-connected component of $X$ that is not simply connected, and it considers the non null-homotopic loop $\gamma_x$ through $x$ provided by assertion (i), as well as a shortest path $\gamma$ from $y$ to $x$. Then, $\gamma_y = \gamma \cdot \gamma_x \cdot \bar{\gamma}$ is a non null-homotopic rectifiable loop through $y$, of length $|\gamma_y| = 2\ \mathrm{hfs}(x) + 2\ \mathrm{d}_X(x,y)$. It follows that $\mathrm{hfs}(y) \leq \frac{1}{2}|\gamma_y| = \mathrm{hfs}(x) + \mathrm{d}_X(x,y)$.

LEMMA 3.2. *Let $X$ be a Lipschitz planar domain. For all point $x \in X$, every loop inside the geodesic open ball $B_X(x, \mathrm{hfs}(x))$ is null-homotopic in $X$.*

The proof is omitted in this abstract. Intuitively, a geodesic ball of center $x$ and radius less than $\mathrm{hfs}(x)$ cannot enclose any hole of $X$, therefore every loop inside such a ball must be null-homotopic in $X$. Note that Lemma 3.2 does not imply that $B_X(x, \mathrm{hfs}(x))$ itself is contractible. This fact is true nevertheless, but its proof requires some more work.

## 4 Structural results

Given a Lipschitz planar domain $X$, and a set of landmarks $L \subset X$ that is dense enough with respect to the homotopy feature size of $X$, we show in Section 4.1 that the geodesic Delaunay triangulation $\mathcal{D}_X(L)$ is homotopy equivalent to $X$

(Theorem 4.1), and in Section 4.2 that $\mathcal{D}_X(L)$ is sandwiched between the geodesic witness complex $\mathcal{C}_X^W(L)$ and its relaxed version $\mathcal{C}_{X,\nu}^W(L)$, for any set of witnesses $W \subseteq X$ that is dense enough compared to $L$ (Theorems 4.3 and 4.4).

**4.1 Geodesic Delaunay triangulation.** Consider a domain $X \subseteq \mathbb{R}^2$ and a finite set of sites $L \subset X$. The *geodesic Voronoi cell* of a site $p$ is the locus of the points $x \in X$ satisfying $\mathrm{d}_X(x,p) \leq \mathrm{d}_X(x,q)$ for all $q \in L$. The *geodesic Voronoi diagram* of $L$ in $X$, or $\mathcal{V}_X(L)$ for short, is the cellular decomposition of $X$ formed by the geodesic Voronoi cells of the sites. The nerve of $\mathcal{V}_X(L)$ is called the *geodesic Delaunay triangulation* of $L$ in $X$, noted $\mathcal{D}_X(L)$. The face of $\mathcal{V}_X(L)$ dual to a given simplex $\sigma \in \mathcal{D}_X(L)$ is noted $\mathrm{V}_X(\sigma)$.

Consider now a Lipschitz planar domain $X$, and a finite set of sites $L \subset X$ that is a *geodesic* $\varepsilon$hfs*-sample* of $X$, for some $\varepsilon < \frac{1}{3}$. This means that, for all point $x \in X$, the geodesic distance from $x$ to $L$ is finite and at most $\varepsilon\ \mathrm{hfs}(x)$. Note that $L$ has at least one point in every path-connected component of $X$, because geodesic distances to $L$ are required to be finite. We will see how to generate such point sets in Section 5.2.

THEOREM 4.1. *If $X$ is a Lipschitz planar domain, and if $L$ is a geodesic $\varepsilon$*hfs*-sample of $X$, for some $\varepsilon < \frac{1}{3}$, then $\mathcal{D}_X(L)$ and $X$ are homotopy equivalent.*

The rest of Section 4.1 is devoted to the proof of Theorem 4.1, which uses the Nerve theorem:

THEOREM 4.2. (FROM [8, 33]) *Let $\mathcal{U}$ be a finite closed cover of $X$, such that the intersection of any collection of elements of $\mathcal{U}$ is either empty or contractible. Then, the nerve of $\mathcal{U}$ is homotopy equivalent to $X$.*

In our case, we set $\mathcal{U}$ to be the collection of the geodesic Voronoi cells: $\mathcal{U} = \{\mathcal{V}_X(p),\ p \in L\}$. The nerve of this collection is precisely the geodesic Delaunay triangulation $\mathcal{D}_X(L)$. Thus, Theorem 4.2 reduces the proof of Theorem 4.1 to showing that the intersection of any arbitrary number of cells of $\mathcal{V}_X(L)$ is empty or contractible. We first show that the geodesic Voronoi cells are contractible:

LEMMA 4.1. *Under the hypotheses of Theorem 4.1, every cell of $\mathcal{V}_X(L)$ is contractible.*

*Proof.* Let $p \in L$. We first show that $\mathcal{V}_X(p)$ is path-connected. Let $x \in \mathcal{V}_X(p)$, and let $\gamma : I \to X$ be a shortest path from $p$ to $x$ in $X$. Such a path $\gamma$ exists by Theorem 2.1 (ii), since $x$ and $p$ lie in the same path-connected component of $X$, $\mathrm{d}_X(x,p)$ being finite due to the fact that $L$ is a geodesic $\varepsilon$hfs-sample of $X$. We will show that $\gamma(I) \subseteq \mathcal{V}_X(p)$. Assume for a contradiction that $\gamma(s) \notin \mathcal{V}_X(p)$ for some $s \in I$. This means that there exists a point $q \in L \setminus \{p\}$ such that $\mathrm{d}_X(\gamma(s),q) < \mathrm{d}_X(\gamma(s),p)$. By the triangle inequality, we have $\mathrm{d}_X(q,x) \leq \mathrm{d}_X(q,\gamma(s)) + \mathrm{d}_X(\gamma(s),x)$, where $\mathrm{d}_X(q,\gamma(s)) < \mathrm{d}_X(p,\gamma(s)) \leq |\gamma_{|[0,s]}|$ and $\mathrm{d}_X(\gamma(s),x) \leq |\gamma_{|[s,1]}|$. Hence, we have $\mathrm{d}_X(q,x) < |\gamma_{|[0,s]}| + |\gamma_{|[s,1]}| = |\gamma| = \mathrm{d}_X(p,x)$, which contradicts the assumption that $x \in \mathcal{V}_X(p)$. Therefore, $\gamma(I) \subseteq \mathcal{V}_X(p)$. This shows that $\mathcal{V}_X(p)$ is path-connected.

Assume now for a contradiction that $\mathcal{V}_X(p)$ is not simply connected. Then, since $\mathcal{V}_X(p) \subseteq X$ is a bounded subset of $\mathbb{R}^2$, its complement in $\mathbb{R}^2$ has at least two path-connected components, only one of which is unbounded, by the Alexander duality – see *e.g.* [28, Thm. 3.44]. Let $H$ be a bounded path-connected component of $\mathbb{R}^2 \setminus \mathcal{V}_X(p)$. $H$ can be viewed as a hole in $\mathcal{V}_X(p)$. We claim that $H$ is included in $X$. Indeed, consider a loop $\gamma : S^1 \to \mathcal{V}_X(p)$ that winds around $H$ – such a loop exists since $H$ is bounded by $\mathcal{V}_X(p)$. Take any point $x \in \mathcal{V}_X(p)$. For all $y \in \mathcal{V}_X(p)$, we have $\mathrm{d}_X(x,y) \leq \mathrm{d}_X(x,p) + \mathrm{d}_X(p,y) \leq \varepsilon\ \mathrm{hfs}(x) + \varepsilon\ \mathrm{hfs}(y)$, which is at most $\frac{2\varepsilon}{1-\varepsilon}\mathrm{hfs}(x)$ since hfs is 1-Lipschitz in the intrinsic metric (Lemma 3.1 (ii)). Thus, $\mathcal{V}_X(p)$ is included in the geodesic closed ball $B_X(x, \frac{2\varepsilon}{1-\varepsilon}\ \mathrm{hfs}(x))$, where $\frac{2\varepsilon}{1-\varepsilon} < 1$ since $\varepsilon < \frac{1}{3}$. Therefore, $\gamma : S^1 \to \mathcal{V}_X(p)$ is null-homotopic in $X$, by Lemma 3.2. Let $\Gamma : S^1 \times I \to X$ be a homotopy between $\gamma$ and a constant map in $X$. For any point $z \in H$, we have $\deg_z \gamma \neq 0$ since the loop $\gamma$ winds around $H$. If $z$ did not belong to $\Gamma(S^1 \times I)$, then $\Gamma$ would be a homotopy between $\gamma$ and a constant map in $\mathbb{R}^2 \setminus \{z\}$, thus by Corollary 2.1 we would have $\deg_z \gamma = 0$, thereby raising a contradiction. Hence, $\Gamma(S^1 \times I)$ contains all the points of hole $H$, which is therefore included in $X$.

It follows that the hole is caused by the presence of some sites of $L \setminus \{p\}$, whose geodesic Voronoi cells form $H$. Assume without loss of generality that there is only one such site $q$. We have $\mathcal{V}_X(q) = \overline{H}$, and $\partial H = \mathcal{V}_X(q) \cap \mathcal{V}_X(p)$. Consider the Euclidean ray $[p,q)$, and call $x$ its first point of intersection with $\partial H$ beyond $q$. Line segment $[q,x]$ is included in $\overline{H} \subseteq X$, therefore we have $\mathrm{d}_X(x,q) = \mathrm{d}_E(x,q)$, which yields: $\mathrm{d}_X(x,p) \geq \mathrm{d}_E(x,p) = \mathrm{d}_E(x,q) + \mathrm{d}_E(q,p) = \mathrm{d}_X(x,q) + \mathrm{d}_E(q,p) > \mathrm{d}_X(x,q)$. This contradicts the fact that $x$ belongs to $\partial H$ and hence to $\mathcal{V}_X(p)$. Thus, $\mathcal{V}_X(p)$ is simply connected. Since it is also path-connected, it is contractible. □

By very similar arguments, we can prove that the union of any two intersecting cells of $\mathcal{V}_X(L)$ is contractible. It follows then from Lemma 4.1 and from the following classical result of algebraic topology that their intersection is also contractible:

LEMMA 4.2.

*(i) The intersection of any $k$ simply connected subsets of $\mathbb{R}^2$ is either empty or simply connected.*

*(ii) If $X, Y$ are path-connected subsets of $\mathbb{R}^2$ such that $X \cup Y$ is simply connected, then $X \cap Y$ is either empty or path-connected.*

We will now extend the above results to the intersections of

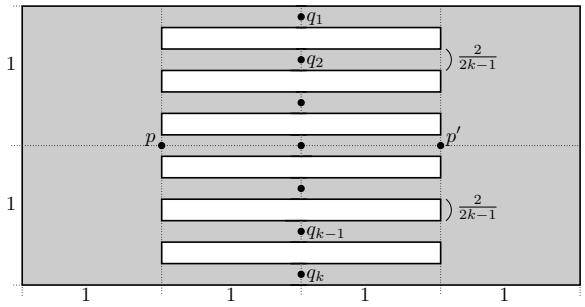



Figure 2: The size of $\{q_1, \cdots, q_k\}$ is $\frac{k}{2}$ times that of $\{p, p'\}$, although both point sets are sparse geodesic hfs-samples.

arbitrary numbers of cells of $\mathcal{V}_X(p)$, thereby concluding the proof of Theorem 4.1:

LEMMA 4.3. *Under the hypotheses of Theorem 4.1, for any $k$ sites $p_1, \cdots, p_k \in L$, the intersection $\mathcal{V}_X(p_1) \cap \cdots \cap \mathcal{V}_X(p_k)$ is either empty or contractible.*

*Proof.* The proof is by induction on $k$. Cases $k = 1$ and $k = 2$ have just been proved. Assume now that the result is true up to some $k \geq 2$, and consider $k+1$ sites $p_1, \cdots, p_{k+1} \in L$ such that $\mathcal{V}_X(p_1) \cap \cdots \cap \mathcal{V}_X(p_{k+1}) \neq \emptyset$. Notice first that $\mathcal{V}_X(p_1) \cap \cdots \cap \mathcal{V}_X(p_{k+1})$ is the intersection of $\bigcap_{i=1}^k \mathcal{V}_X(p_i)$ with $\mathcal{V}_X(p_{k+1})$, which by the induction hypothesis are simply connected. Hence, their intersection $\mathcal{V}_X(p_1) \cap \cdots \cap \mathcal{V}_X(p_{k+1})$ is also simply connected, by Lemma 4.2 (i). Observe now that $\left(\bigcap_{i=1}^k \mathcal{V}_X(p_i)\right) \cup \mathcal{V}_X(p_{k+1})$ can be rewritten as $\bigcap_{i=1}^k (\mathcal{V}_X(p_i) \cup \mathcal{V}_X(p_{k+1}))$. By the induction hypothesis (more precisely, according to the case $k = 2$), every $\mathcal{V}_X(p_i) \cup \mathcal{V}_X(p_{k+1})$ is simply connected, hence so is $\bigcap_{i=1}^k (\mathcal{V}_X(p_i) \cup \mathcal{V}_X(p_{k+1}))$, by Lemma 4.2 (i). It follows then from Lemma 4.2 (ii) that the intersection $\mathcal{V}_X(p_1) \cap \cdots \cap \mathcal{V}_X(p_{k+1})$ is path-connected, since by induction both $\bigcap_{i=1}^k \mathcal{V}_X(p_i)$ and $\mathcal{V}_X(p_{k+1})$ are, and since their union is simply connected. □

**4.2 Geodesic witness complex.** Consider a domain $X \subseteq \mathbb{R}^2$, as well as two finite subsets $L$ and $W$. Given a point $w \in W$ and a simplex $\sigma = [p_0, \cdots, p_l]$ with vertices in $L$, $w$ is a *witness* of $\sigma$ if for all $i = 0, \cdots, l$, $\mathrm{d}_X(w, p_i)$ is finite and bounded from above by $\mathrm{d}_X(w, q)$ for every $q \in L \backslash \{p_0, \cdots, p_l\}$. Observe that $w$ may only witness simplices whose vertices lie in the same path-connected component of $X$. The *geodesic witness complex* of $L$ relative to $W$, or $\mathcal{C}_X^W(L)$ for short, is the maximal abstract simplicial complex with vertices in $L$, whose faces are witnessed by points of $W$. The fact that $\mathcal{C}_X^W(L)$ is an abstract simplicial complex means that a simplex belongs to the complex only if all its faces do. In the sequel, $W$ will be referred to as the set of witnesses, and $L$ as the set of landmarks.

Our first result is an analog of de Silva's theorem [15] in the intrinsic metric. The proof uses the same machinery as in [3], and it relies on the intuitive fact that, when the set $L$ is a geodesic $\varepsilon$hfs-sample of $X$, the geodesic distances between a point $x \in X$ and its $k$ nearest landmarks in the intrinsic metric are at most $4^k \varepsilon \, \mathrm{hfs}(x)$, the exponent coming from the fact that hfs is 1-Lispchitz.

THEOREM 4.3. *Let $X$ be a Lipschitz planar domain, and $L$ a geodesic $\varepsilon$hfs-sample of $X$. If $\varepsilon \leq \frac{1}{4^{k+1}}$, for some integer $k \geq 0$, then the $k$-skeleton of $\mathcal{C}_X^W(L)$ is included in $\mathcal{D}_X(L)$ for all $W \subseteq X$.*

Our next result is an analog of Theorem 3.2 of [27], whose proof relies on a simple packing argument. It involves a relaxed version of the witness complex, defined as follows. Given an integer $\nu \geq 0$, a simplex $\sigma$ is $\nu$-witnessed by $w \in W$ if the vertices of $\sigma$ belong to the $\nu + 1$ landmarks closest to $w$ in the intrinsic metric. The geodesic $\nu$-witness complex of $L$ relative to $W$, or $\mathcal{C}_{X,\nu}^W(L)$ for short, is the maximum abstract simplicial complex made of $\nu$-witnessed simplices. Its dimension is at most $\nu$.

THEOREM 4.4. *Let $X$ be a Lipschitz planar domain, of doubling dimension $d$. Let $W$ be a geodesic $\delta$hfs-sample of $X$, and $L$ a geodesic $\varepsilon$hfs-sample of $X$ that is also $\frac{\varepsilon}{1+\varepsilon}$hfs-sparse. If $\varepsilon + 2\delta < 1$, then, for any integer $\nu \geq 2^{d+1} \, \frac{(1+\delta/\varepsilon)(1+\varepsilon)}{1-\varepsilon-2\delta}$, $\mathcal{D}_X(L)$ is included in $\mathcal{C}_{X,\nu}^W(L)$.*

The theorem assumes that $L$ is a $\frac{\varepsilon}{1+\varepsilon}$hfs-sparse geodesic $\varepsilon$hfs-sample of $X$, which means that every pair of landmarks $p, q$ satisfies: $\mathrm{d}_X(p, q) \geq \frac{\varepsilon}{1+\varepsilon} \min\{\mathrm{hfs}(p), \mathrm{hfs}(q)\}$. The lower bound on $\nu$ depends on the *doubling dimension* of $(X, \mathrm{d}_X)$, which is defined as the smallest integer $d$ such that every geodesic closed ball can be covered by a union of $2^d$ geodesic closed balls of half its radius. The doubling dimension measures the shape complexity of $X$, and it can be arbitrarily large. An example is given in Figure 2, where the $k$ geodesic balls $B_X(q_i, 1)$ are included in their respective branches, and therefore are disjoint. Moreover, they are packed inside the ball $B_X(p, 3)$, which therefore requires at least $k$ geodesic unit balls to be covered, by a result of [29]. It follows that the doubling dimension of $X$ is at least $\frac{1}{2} \log_2 k$, which can be made arbitrarily large.

It follows from Theorems 4.3 and 4.4 that, whenever $L$ and $W$ are dense enough, $\mathcal{D}_X(L)$ is sandwiched between $\mathcal{C}_X^W(L)$ and $\mathcal{C}_{X,\nu}^W(L)$, provided that $\nu$ is chosen sufficiently large. Our simulation results – see Section 6 – suggest that even small values of $\nu$ suffice in practice.

## 5 Algorithms

We will now describe high-level procedures for computing hfs, for generating geodesic $\varepsilon$hfs-samples, and for computing the homology of a Lipschitz planar domain. Our algorithms rely essentially on two oracles, whose implementations depend on the application considered.

**5.1 Computing the homotopy feature size.** As pointed out by Erickson and Whittlesey for Riemannian surfaces [20], the homotopy feature size is closely related to the concept of cut-locus. Given a path $\gamma : I \to X$, we call trajectory of $\gamma$ the set $\gamma(I)$. If $\gamma$ is a shortest path between $x = \gamma(0)$ and $y = \gamma(1)$, then $\gamma(I)$ is called a shortest trajectory between $x$ and $y$. Given a point $x \in X$, the cut-locus of $x$ in $X$, or $\mathrm{CL}_X(x)$ for short, is the locus of the points of $X$ having at least two distinct shortest trajectories to $x$ in $X$. The geodesic distance from $x$ to its cut-locus is denoted by $\mathrm{d}_X(x, \mathrm{CL}_X(x))$.

LEMMA 5.1. *If $X$ is a Lipschitz planar domain, then $\forall x \in X$, $\mathrm{hfs}(x) = \mathrm{d}_X(x, \mathrm{CL}_X(x))$.*

The proof is omitted in this abstract. Lemma 5.1 suggests a simple way of estimating hfs: given a point $x \in X$, grow a geodesic closed ball $B$ about $x$, starting with a radius of zero and ending when $B$ covers the path-connected component $X_x$ of $X$ containing $x$. Meanwhile, focus on the wavefront $\partial B$ as the radius of $B$ increases – this wavefront evolves as the iso-level sets of the map $y \mapsto \mathrm{d}_X(x, y)$. If at some stage the wavefront *self-intersects*, *i.e.* if there is a point $y \in \partial B$ with two or more distinct shortest trajectories to $x$, then interrupt the process and return the current value of the radius of $B$. Else, stop once $B$ covers $X_x$ and return $+\infty$.

By detecting the first self-intersection event in the growing process, the procedure finds a point of $\mathrm{CL}_X(x)$ closest to $x$ in the intrinsic metric, and therefore it returns $\mathrm{d}_X(x, \mathrm{CL}_X(x))$, which by Lemma 5.1 is equal to $\mathrm{hfs}(x)$. The procedure relies on two oracles: one that checks whether $B$ covers $X_x$ entirely, the other that checks whether the wavefront self-intersects at a given value of the radius of $B$, or rather between two given values of the radius of $B$.

**5.2 Generating geodesic $\varepsilon$hfs-samples.** Given a Lipschitz planar domain $X$ and a parameter $\varepsilon > 0$, we use a greedy packing strategy to generate geodesic $\varepsilon$hfs-samples of $X$. Initially, our algorithm selects an arbitrary point $p \in X$ and sets $L = \{p\}$. It also assigns to $p$ the geodesic open ball $B_p$ of center $p$ and radius $\frac{\varepsilon}{1+\varepsilon}\ \mathrm{hfs}(p)$, where $\mathrm{hfs}(p)$ is estimated using the procedure of Section 5.1. If $\mathrm{hfs}(p) = +\infty$, then $B_p$ coincides with the path-connected component of $X$ containing $p$. Then, at each iteration, the algorithm selects an arbitrary point $q \in X \setminus \bigcup_{p \in L} \overline{B}_p$, and it inserts this point in $L$. It also assigns a geodesic open ball $B_q$ to $q$, as detailed above for $p$. The process stops when $X \setminus \bigcup_{p \in L} \overline{B}_p = \emptyset$.

The algorithm uses a variant of an oracle of Section 5.1, which can tell whether a given union of geodesic balls covers $X$, and return a point outside the union in the negative. Upon termination, every point $x \in X$ lies in some closed ball $\overline{B}_p$, and we have $\mathrm{d}_X(x, L) \leq \mathrm{d}_X(x, p) \leq \frac{\varepsilon}{1+\varepsilon}\ \mathrm{hfs}(p)$, which is at most $\varepsilon\ \mathrm{hfs}(x)$ since hfs is 1-Lipschitz in the intrinsic metric. Moreover, $\mathrm{d}_X(x, p)$ is finite because $B_p$ is included in the path-connected component of $X$ containing $p$. Therefore, $L$ is a geodesic $\varepsilon$hfs-sample of $X$. Furthermore,

LEMMA 5.2. *For all $\varepsilon \in ]0, 1[$, the algorithm terminates, and the size its output is within $2^d\ \frac{3+3\varepsilon+2\varepsilon^2}{1-\varepsilon}$ times the size of any geodesic $\varepsilon$hfs-sample of $X$, where $d$ is the doubling dimension of $X$.*

The proof, omitted in this abstract, relies on a simple packing argument. The influence of the doubling dimension of $X$ is illustrated in Figure 2, where $P = \{p, p'\}$ and $Q = \{q_1, \cdots, q_k\}$ are geodesic hfs-samples of $X$, because $\mathrm{hfs}(x)$ is everywhere at least half the perimeter of a hole, namely $2 + \frac{2}{2k-1}$. Although $Q$ is hfs-sparse, its size is $|Q| = \frac{k}{2}\ |P|$, where $k$ is of the order of $2^d$, as emphasized in Section 4.2.

**5.3 Computing the homology of a Lipschitz domain.** Given a geodesic $\varepsilon$hfs-sample $L$ of a Lipschitz planar domain $X$, a variant of the procedure of Section 5.1 can be used to build $\mathcal{D}_X(L)$: grow geodesic balls around the points of $L$ at same speed, and report the intersections between the fronts. The homology of $\mathcal{D}_X(L)$ gives then the homology of $X$, by Theorem 4.1. But in many practical situations, $X$ is only known through a finite sampling $W$, which makes it hard to detect the intersections between more than two fronts. In this discrete setting, it is relevant to replace the construction of $\mathcal{D}_X(L)$ by the ones of $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$, which only require to compare geodesic distances at the points of $W$. The homology of $\mathcal{D}_X(L)$ can then be computed via the persistent homology between $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$.

More precisely, we use simplicial homology with coefficients in a field, which in practice will be $\mathbb{Z}/2$ – omitted in our notations. The inclusion map $i : \mathcal{C}^W_X(L) \hookrightarrow \mathcal{C}^W_{X,\nu}(L)$ induces a homomorphism $i* : H^\Delta_k(\mathcal{C}^W_X(L)) \to H^\Delta_k(\mathcal{C}^W_{X,\nu}(L))$. By applying the persistence algorithm [34] to the filtration $\mathcal{C}^W_X(L) \hookrightarrow \mathcal{C}^W_{X,\nu}(L)$, we can compute the rank of $i_*$. Thus, the goal is to relate the $\mathrm{rank}\ i_*$ to $\dim H^\Delta_k(\mathcal{D}_X(L))$, the $k$th Betti number of $\mathcal{D}_X(L)$. We know from Theorems 4.3 and 4.4 that $\mathcal{C}^W_X(L) \subseteq \mathcal{D}_X(L) \subseteq \mathcal{C}^W_{X,\nu}(L)$ under some sampling conditions, which we will assume from now on. The inclusion maps $j : \mathcal{C}^W_X(L) \hookrightarrow \mathcal{D}_X(L)$ and $j' : \mathcal{D}_X(L) \hookrightarrow \mathcal{C}^W_{X,\nu}(L)$ induce homomorphisms $j_*$, $j'_*$ on the homology groups, such that $i_* = (j' \circ j)_* = j'_* \circ j_*$. It follows that $\dim H^\Delta_k(\mathcal{D}_X(L)) \geq \mathrm{rank}\ j'_* \geq \mathrm{rank}\ i_*$, which means that every $k$-cycle that persists between $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$ is a non-trivial $k$-cycle of $\mathcal{D}_X(L)$. In fact, we even have:

THEOREM 5.1. *Assume that the hypotheses of Theorems 4.3 and 4.4 are satisfied, with $k = \nu$, $L \subseteq W$, and with* hfs *replaced by* $\min\{\mathrm{hfs}, \mathrm{d_M}\}$, *where* $\mathrm{d_M}$ *is the Euclidean distance to the medial axis* M *of* $\mathbb{R}^2 \setminus X$. *Then, the range space of $i_*$ is isomorphic to $H^\Delta_k(\mathcal{D}_X(L))$. In other words, we have:* $\mathrm{rank}\ i_* = \dim H^\Delta_k(\mathcal{D}_X(L))$.

This theorem guarantees that the persistent homology between $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$ gives the homology of $\mathcal{D}_X(L)$. The bounds on the densities of landmarks and witnesses depend on $\mathrm{d_M}$, which requires that $\mathrm{M} \cap X = \emptyset$. This is true if $X$ has smooth boundaries, but also if $\partial X$ only has convex corners (oriented outwards). The fact that $\mathrm{hfs}$ and $\mathrm{d_M}$ are both 1-Lipschitz in the intrinsic metric[2] implies that the densities deep inside the domain $X$ can be small, although they may have to be large near $\partial X$.

The proof of Theorem 5.1, omitted in this abstract, proceeds in two steps: first it shows that $j'_*$ is injective, then it shows that $j_*$ is surjective. It follows from the injectivity of $j'_*$ that $\dim H^{\Delta}_k(\mathcal{D}_X(L)) = \mathrm{rank}\ j'_*$, which is equal to $\mathrm{rank}\ i_*$ by the surjectivity of $j_*$.

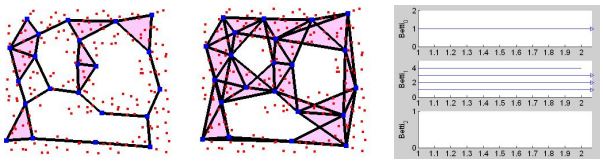
## 6 Application to sensor networks

We have implemented the algorithms of Section 5 in the context of sensor networks, where the nodes do not have geographic locations, and where the intrinsic metric is approximated by the hop-count distance in the connectivity graph (a unit disk graph in our case). The quality of the distance approximation can be guaranteed, provided that the node density is sufficiently high [25].

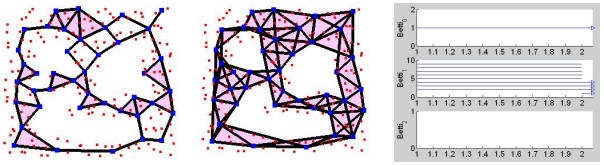
**Homotopy feature size computation.** Given a node $x$, we estimate the geodesic distance of $x$ to its cut-locus, which by Lemma 5.1 is equal to $\mathrm{hfs}(x)$. Wang *et al.* [32] proposed a distributed algorithm for detecting the cut-locus, which works as follows: the node $x$ sends a flood message with initial hop count 1; each node receiving the message forwards it after incrementing the hop count. Thus, every node learns its minimum hop count to the node $x$. Then, each pair of neighbors check whether their least common ancestor (LCA) is at hop-count distance at least $d$. If so, then they also check whether their two shortest paths to the LCA contain nodes at least $d$ away from each other (by looking at the $\frac{d}{2}$-ring neighborhoods of the nodes of the paths). Every pair satisfying these conditions is called a cut pair. As proved in [32], every hole of perimeter greater than $d$ yields a cut pair. Then, every cut node checks its neighbors, and if it has the minimum hop count, then it reports back to $x$ with the hop count value. Thus, $x$ gets a report from one node on each connected component of the cut-locus, and learns the homotopy feature size as the minimum hop value. For a weighted graph, the operation is similar.

**Landmark selection and witness complex computation.** The landmark selection implements the incremental algorithm of Section 5.2 in a distributed manner. A node has two states, *covered* and *uncovered*. A covered node lies inside the geodesic ball of some landmark. Initially, all the nodes are uncovered. They wait for different random periods of time, after which they promote themselves to the status of landmark. Each new landmark floods the network, computes its homotopy feature size, and informs all the nodes within its geodesic ball to be covered. Thus, every node eventually becomes covered or a landmark itself.

The witness complex is computed in a similar way as in [22]. The selected landmarks flood the network, and every node records its minimum hop counts to them. With this information, it determines which simplices it witnesses. A round of information aggregation collects all the simplices and constructs the witness complex. In a planar setting, where only the Betti numbers $\beta_0$ and $\beta_1$ are non-zero, we only need to build the 2-skeleton of the witness complex. Therefore, each node may store only its three nearest landmarks, and it may avoid forwarding messages from other landmarks, which reduces the message complexity.

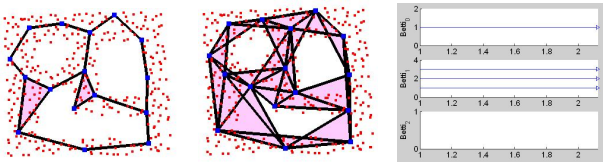


(a) $n = 217$, $d \approx 7.66$, $\varepsilon = 0.5$, $\nu = 2$, weighted distance.

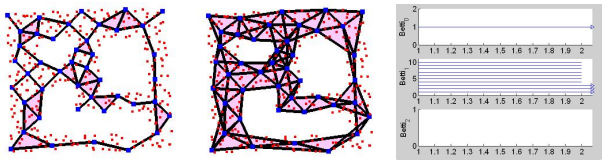


(b) $n = 217$, $d \approx 7.66$, $\varepsilon = 0.25$, $\nu = 2$, weighted distance.

Figure 3: From left to right: witness complex, relaxed witness complex, persistence barcode of the filtration [10].



(a) $n = 353$, $d \approx 7.66$, $\varepsilon = 0.5$, $\nu = 2$, weighted distance.

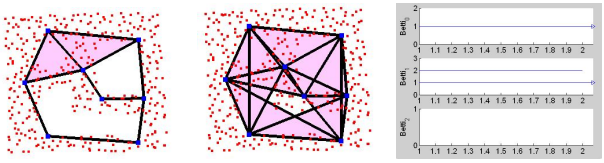


(b) $n = 353$, $d \approx 7.66$, $\varepsilon = 0.25$, $\nu = 2$, weighted distance.

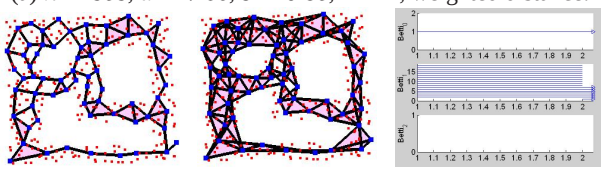
Figure 4: Same setting as above, with a higher node density.

**Simulation results and discussion.** Figures 3 to 7 present some simulation results showing the dependency of the landmark selection and homology computation on various parameters. We used $n$ sensor nodes randomly distributed in a Lipschitz planar domain. Two nodes within unit Euclidean distance of each other are connected. The resulting average

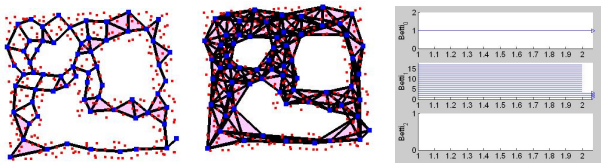
[2] Since $\mathrm{d_M}$ is 1-Lipschitz in the Euclidean metric, it is also 1-Lipschitz in the intrinsic metric, because $\mathrm{d}_E \leq \mathrm{d}_X$.



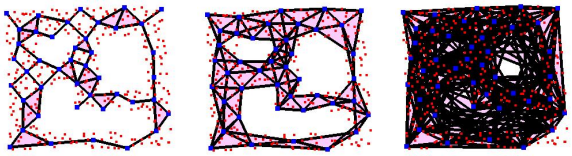
(a) $n = 353$, $d \approx 7.66$, $\varepsilon = 0.85$, $\nu = 2$, weighted distance.



(b) $n = 353$, $d \approx 7.66$, $\varepsilon = 0.15$, $\nu = 2$, weighted distance.

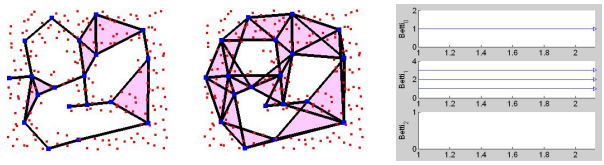


(c) $n = 353$, $d \approx 7.66$, $\varepsilon = 0.15$, $\nu = 4$, weighted distance.

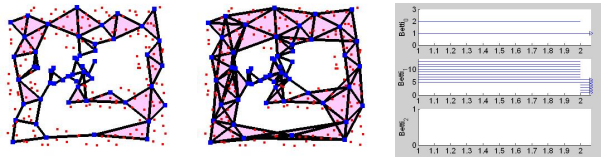


(d) $n = 353$, $d \approx 7.66$, $\varepsilon = 0.25$, weighted distance. Left: witness complex; Middle: $\nu = 2$; Right: $\nu = 11$.

Figure 5: Effect of varying $\nu$ versus the landmark density.



(a) $n = 217$, $d \approx 7.66$, $\varepsilon = 0.5$, $\nu = 2$, hop-count distance.



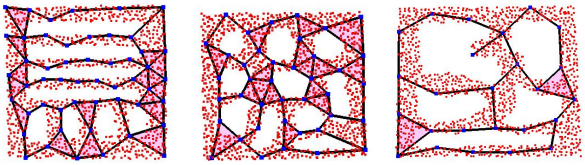
(b) $n = 217$, $d \approx 7.66$, $\varepsilon = 0.25$, $\nu = 2$, hop-count distance.

Figure 6: Same setting as above, with the weighted distance replaced by the hop-count distance.

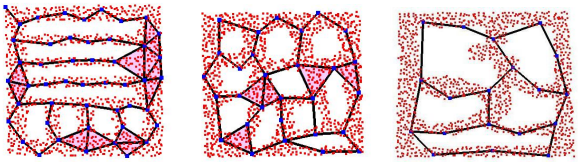
node degree is noted $d$. The intrinsic metric is approximated by the graph distance in the connectivity network, where each edge can be either unweighted (hop-count distance) or weighted by its Euclidean length (weighted distance).

Figure 3 shows a typical example, with (a) $\varepsilon = 0.5$ and (b) $\varepsilon = 0.25$. In both cases, only the genuine 3 holes persist and are therefore identified as non-trivial 1-cycles in the geodesic Delaunay triangulation.

- *Node density.* We vary the number of nodes from 217 to 355. The average degree remains the same. The result is shown in Figure 4. Again, the persistent homology between the witness complex and its relaxed version gives the homology of the domain. Thus, only the intrinsic geometry of the domain matters, not the scale of the network, as long as the latter remains sufficiently dense.
- *Landmark density.* Figure 5 shows our results on the same setup as above, with (a) $\varepsilon = 0.85$ and (b) $\varepsilon = 0.15$. In the first case, only two holes are captured, because of the low landmark density. In the second case, three non-genuine holes are not destroyed in the relaxed witness complex, because the value of the relaxation parameter $\nu$ is too small given the relatively low node density. Increasing $\nu$ from 2 to 4 produces the correct answer (c). But setting $\nu$ to too high a value ($\nu = 11$, $\varepsilon = 0.25$) destroys some of the genuine holes (d). Throughout our experiments, the algorithm produced correct results with small values of $\nu$ ($\nu \leq 4$), provided that the nodes and landmarks sets were reasonably dense. This demonstrates the practicality of our approach, despite the large theoretical bounds stated in Theorems 4.3, 4.4 and 5.1.
- *Weighted graph distance vs. hop-count distance.* Since the hop-count distance is a poorer approximation to the geodesic distance, the range of values of $\varepsilon$ that work with it is reduced. In Figure 6 for instance, the scheme works well with $\varepsilon = 0.5$, but not with $\varepsilon = 0.25$, in contrast with the results of Figure 3.
- *Packing strategy.* Figure 7 shows some of our sampling results. It appears that different packing strategies can produce samples of very different sizes, as predicted by Lemma 5.2. Maximizing the ratio $\frac{\mathrm{d}_X(q,L)}{\mathrm{hfs}(q)}$ at each iteration seems to be a very effective strategy in practice, but it is also time-consuming.



(a) $\varepsilon = \frac{1}{3}$, random landmark selection outside $\bigcup_{p \in L} \overline{B}_p$.



(b) $\varepsilon = \frac{1}{3}$, insertion of $\mathrm{argmax}\{\frac{\mathrm{d}_X(q,L)}{\mathrm{hfs}(q)},\ q \in W \setminus \bigcup_{p \in L} \overline{B}_p\}$.

Figure 7: Landmark sets obtained by two different packing strategies, and their geodesic witness complexes.

To conclude, let us emphasize that our approach turned out to be also quite robust under more realistic communication models, such as quasi-unit disk graph with link failures.

**Possible extensions.** This work focuses on the planar case, with applications in sensor networks. A natural extension would be to consider bounded domains in higher-

dimensional Euclidean spaces, with applications in robotics and geometric data analysis. Also, it would be relevant to generate homology bases whose elements isolate the various holes of $X$. There exists some work along this line, but for a slightly different context [24].

**Acknowledgements.** L. Guibas and S. Oudot wish to acknowledge the supports of DARPA grant HR0011-05-1-0007 and NSF grants FRG-0354543 and CNS-0626151. J. Gao and Y. Wang wish to acknowledge the support of NSF CAREER Award CNS-0643687.

## References

[1] N. Amenta and M. Bern. Surface reconstruction by Voronoi filtering. *Discrete Comput. Geom.*, 22(4):481–504, 1999.

[2] N. Amenta, M. Bern, and D. Eppstein. The crust and the $\beta$-skeleton: Combinatorial curve reconstruction. *Graphical Models and Image Processing*, 60:125–135, 1998.

[3] D. Attali, H. Edelsbrunner, and Y. Mileyko. Weak witnesses for Delaunay triangulations of submanifolds. In *Proc. ACM Sympos. on Solid and Physical Modeling*, pages 143–150, 2007.

[4] H. Blum. A transformation for extracting new descriptors of shape. In W. Wathen-Dunn, editor, *Models for the Perception of Speech and Visual Form*, pages 362–380. MIT Press, 1967.

[5] A. I. Bobenko and B. A. Springborn. A discrete laplace-beltrami operator for simplicial surfaces. Technical Report, 2005.

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

[7] J.-D. Boissonnat and S. Oudot. Provably good sampling and meshing of Lipschitz surfaces. In *Proc. 22nd Annu. Sympos. Comput. Geom.*, pages 337–346, 2006.

[8] K. Borsuk. On the imbeddings of systems of compacta in simplicial complexes. *Fund. Math.*, 35:217–234, 1948.

[9] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry*, volume 33 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2001.

[10] G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas. Persistence barcodes for shapes. In *Proc. Symp. Geom. Process.*, pages 127–138, 2004.

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

[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 *Proc. 21st Annual ACM Symposium on Computational Geometry*, pages 255–262, 2005.

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

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

[15] V. de Silva. A weak definition of Delaunay triangulation. Technical report, Stanford University, October 2003.

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

[17] T. K. Dey and S. Goswami. Provable surface reconstruction from noisy samples. *Computational Geometry: Theory and Applications*, 35(1-2):124–141, 2006.

[18] R. Dyer, H. Zhang, and T. Moeller. Delaunay mesh construction. In *Proc. Eurographics Sympos. Geometry Processing*, pages 273–282, 2007.

[19] J. Erickson and S. Har-Peled. Optimally cutting a surface into a disk. *Discrete and Computational Geometry*, 31(1):37–59, 2004.

[20] J. Erickson and K. Whittlesey. Greedy optimal homotopy and homology generators. In *Proc. 16th ACM-SIAM Sympos. on Discrete algorithms*, pages 1038–1046, 2005.

[21] Q. Fang, J. Gao, and L. J. Guibas. Landmark-based information storage and retrieval in sensor networks. In *Proc. IEEE INFOCOM'06*, pages 1–12, April 2006.

[22] Q. Fang, J. Gao, L. J. Guibas, V. de Silva, and L. Zhang. GLIDER: gradient landmark-based distributed routing for sensor networks. In *Proc. IEEE INFOCOM'05*, volume 1, pages 339–350, 2005.

[23] M. Fisher, B. Springborn, A. I. Bobenko, and P. Schröder. An algorithm for the construction of intrinsic delaunay triangulations with applications to digital geometry processing. In *SIGGRAPH Courses*, pages 69–74, 2006.

[24] D. Freedman and C. Chen. Measuring and localizing homology classes. Technical Report, Rensselaer Polytechnic Institute, May 2007.

[25] J. Gao, L. J. Guibas, S. Y. Oudot, and Y. Wang. Geodesic Delaunay triangulation and witness complex in the plane. Full version. `http://geometry.stanford.edu/papers/ggow-gdtwcp-08/ggow-gdtwcp-08-full.pdf`.

[26] R. Ghrist and A. Muhammad. Coverage and hole-detection in sensor networks via homology. In *Proc. the 4th International Symposium on Information Processing in Sensor Networks (IPSN'05)*, pages 254–260, 2005.

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

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

[29] A. Kolmogorov and V. Tikhomirov. $\epsilon$-entropy and $\epsilon$-capacity of sets of functions. *Translations of the AMS*, 17:277–364, 1961.

[30] G. Leibon and D. Letscher. Delaunay triangulations and Voronoi diagrams for Riemannian manifolds. In *Proc. 16th Annu. ACM Sympos. Comput. Geom.*, pages 341–349, 2000.

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

[32] Y. Wang, J. Gao, and J. S. B. Mitchell. Boundary recognition in sensor networks by topological methods. In *Proc. MobiCom*, pages 122–133, September 2006.

[33] W.-T. Wu. On a theorem of Leray. *Chinese Mathematics*, 2:398–410, 1962.

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