Geodesic Delaunay Triangulations in Bounded Planar Domains

Steve Y. Oudot and Leonidas J. Guibas
Stanford University
and
Jie Gao and Yue Wang
Stony Brook University

We introduce a new feature size for bounded domains in the plane endowed with an intrinsic metric. Given a point x in a domain X, the *systolic feature size* of X at x measures half the length of the shortest loop through x that is not null-homotopic in X. The resort to an intrinsic metric makes the systolic feature size rather insensitive to the local geometry of the domain, in contrast with its predecessors (local feature size, weak feature size, homology feature size). This reduces the number of samples required to capture the topology of X, provided that a reliable approximation to the intrinsic metric of X is available. Under sufficient sampling conditions involving the systolic feature size, we show that the geodesic Delaunay triangulation $\mathcal{D}_X(L)$ of a finite sampling L is homotopy equivalent to X. Under similar conditions, $\mathcal{D}_X(L)$ is sandwiched between the geodesic witness complex $\mathcal{C}^W_X(L)$ and a relaxed version $\mathcal{C}^W_{X,\nu}(L)$. In the conference version of the paper, we took advantage of this fact and proved that the homology of $\mathcal{D}_X(L)$ (and hence the one of X) can be retrieved by computing the persistent homology between $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$. Here, we investigate further and show that the homology of X can also be recovered from the persistent homology associated with inclusions of type $\mathcal{C}^W_{X,\nu}(L) \hookrightarrow \mathcal{C}^W_{X,\nu'}(L)$, under some conditions on the parameters $\nu \leq \nu'$. Similar results are obtained for Vietoris-Rips complexes in the intrinsic metric. The proofs draw some connections with recent advances on the front of homology inference from point cloud data, but also with several well-known concepts of Riemannian (and even metric) geometry. On the algorithmic front, we propose algorithms for estimating the systolic feature size of a bounded planar domain X, selecting a landmark set of sufficient density, and computing the homology of X using geodesic witness complexes or Rips complexes.

Categories and Subject Descriptors: I.3.5 [**Computer Graphics**]: Computational Geometry and Object Modeling—*Curve, surface, solid, and object representation*; C.2.1 [**Computer Communication Networks**]: Network Architecture and Design—*Network topology*

General Terms: Algorithms, Theory

This work was done while Steve Oudot was a post-doctoral scholar in the Computer Science Department at Stanford University, under the supervision of Prof. Leonidas Guibas. A preliminary version of the paper was presented at the 19th ACM-SIAM Symposium on Discrete Algorithms (SODA 2008). The authors acknowledge the supports of DARPA grant HR0011-05-1-0007, NSF grants FRG-0354543 and CNS-0626151, and NSF CAREER Award CNS-0643687.
Authors' addresses: Steve Y. Oudot, INRIA Saclay – Île-de-France, 4 rue Jacques Monod, 91893 Saclay cedex, France, email: steve.oudot@inria.fr ; Leonidas J. Guibas, Computer Science Departement, Gates Building, Stanford University, Stanford, CA 94305, email: guibas@cs.stanford.edu ; Jie Gao, Yue Wang, 1415 Computer Science Building, Stony Brook University, Stony Brook, NY 11794, emails: {jgao, yuewang}@cs.sunysb.edu.

Additional Key Words and Phrases: Alexandrov space, Delaunay triangulation, feature size, intrinsic metric, systole, witness complex

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 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 [de Silva and Carlsson 2004] and the multi-scale reconstruction algorithm of [Boissonnat et al. 2007; Guibas and Oudot 2007]. Both algorithms rely on the structural properties of the *witness complex*, a data structure specifically designed by de Silva [de Silva 2008] for use with the landmarking strategy. Examples in sensor networks include the GLIDER routing scheme and its variants [Fang et al. 2005; Fang et al. 2006]. 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?

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 [Amenta and Bern 1999; Amenta et al. 1998]. Their bound on the landmarks density depends on the local distance to the medial axis of $\mathbb{R}^2 \setminus X$ (the *local feature size*), and the 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 [Dey and Goswami 2006], sampled from closed manifolds of arbitrary dimensions [Boissonnat et al. 2007; Cheng et al. 2005], smoothly or non-smoothly embedded in Euclidean spaces [Boissonnat and Oudot 2006]. In parallel, others have focused on unions of congruent Euclidean balls and their topological invariants. In a seminal paper [Niyogi et al. 2008], 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 values of r, the union of the open Euclidean balls of radius r about the points of L deformation retracts onto X.

The above results only hold for manifolds without boundary. 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 $\mathbb{R}^n$. These naturally arise in the configuration spaces of motion planning problems in robotics, in monitoring complex domains with sensor networks, and in many other contexts where natural

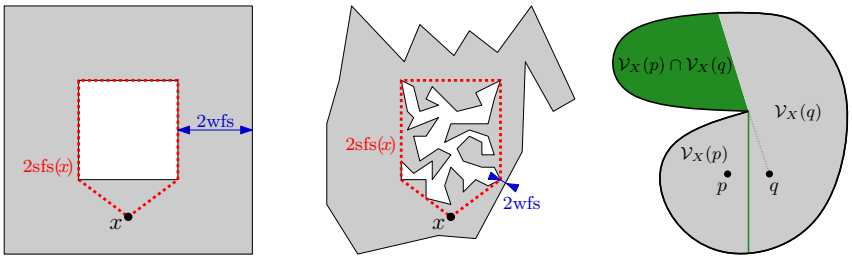


Fig. 1. Left and center: two Lipschitz domains with very different weak feature sizes (wfs), but similar systolic feature sizes. Right: a geodesic Voronoi edge with non-zero Lebesgue measure.

obstacles to sampling certain areas exist. By studying the stability of distance functions to compact sets in R^n, Chazal and Lieutier [Chazal and Lieutier 2007] have extended the sampling theory to a much larger class of objects, including some non-smooth non-manifold compact sets. Their bound on the landmarks density depends on the so-called *weak feature size* of X, defined as the smallest positive critical value of the Euclidean distance to ∂X. This mild sampling condition is shown to be sufficient for the recovery of the homology and homotopy groups of X. Although the results of [Chazal and Lieutier 2007] are valid in a very general setting, in many cases the weak feature size is small compared to the size of the topological features of X, because it is bound to extrinsic quantities — see Figure 1 (center). As a result, many landmarks are wasted satisfying the sampling condition of [Chazal and Lieutier 2007], whereas very few would suffice[1] to capture the topology of X. In practice, this results in a considerable waste of memory and computation power.

The case of bounded domains suggests the use of an intrinsic metric on the domain, instead of the extrinsic metric provided by the embedding. This is essential for certain classes of applications, such as sensor networks, where node location information may not be available and only the geodesic distance can be approximated via wireless connectivity graph distances. Intrinsic metrics have been studied in the context of Riemannian manifolds without boundary [Leibon and Letscher 2000] and, from a more computational point of view, in the context of the so-called *intrinsic* Delaunay triangulations (iDT) of triangulated surfaces without boundary [Bobenko and Springborn 2007]. 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 [Fisher et al. 2006], and generating such iDT meshes is a topic of considerable interest in geometry processing [Dyer et al. 2007].

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

[1]Here we are only discussing the number of landmarks, and not the number of sample points. Indeed, for our approach to work in practice, an accurate approximation to the geodesic distance in X must be provided, which may be given for free in some situations (*e.g.* in robotics), but which may as well require many sample points in other cases (*e.g.* in sensor networks, see Section 7). In all situations, the main advantage of our approach is to build data structures on top of a very small set of landmarks.

applications in sensor networks. We make the novel claim that resorting to an intrinsic metric instead of the Euclidean metric can result in significant reductions in terms of the number of landmarks required to recover the homotopy type of a bounded domain – an appealing fact in the context of resource-constrained nodes used in sensor networks. To this end, we introduce a new quantity, called the *systolic feature size*, or sfs 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 $x \in X$, $\mathrm{sfs}(x)$ is defined as half the length of the shortest loop through x that is not null-homotopic in X – see Figure 1 (left and center) for an illustration. In particular, $\mathrm{sfs}(x)$ is infinite when x lies in a simply connected component of X. The term *systolic feature size* is coined after the concept of *systole*, first introduced by Loewner around 1949 and later developed by Berger, Gromov and others [Gromov 1996]. The systole at x is the length of the shortest non-contractible loop in X that passes through x, therefore it is precisely equal to $2\mathrm{sfs}(x)$.

In contrast with previous quantities, sfs 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 sfs is well-defined, positive, and 1-Lipschitz in the intrinsic metric. Moreover, if L is a geodesic εsfs-sample of X, for some $\varepsilon \leq \frac{1}{3}$, then the cover of X formed by the geodesic Voronoi cells of the points of L satisfies the conditions of the Nerve theorem [Borsuk 1948; Wu 1962], and therefore its dual Delaunay complex $\mathcal{D}_X(L)$ is homotopy equivalent to X. By geodesic εsfs-sample of X, we mean that every point $x \in X$ is at a finite geodesic distance less than $\varepsilon \cdot \mathrm{sfs}(x)$ to L. In the particular case when X is simply connected, our sampling condition only requires that L has 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 ∂X and can result in selecting far fewer landmarks than required by any of the earlier sampling conditions that guarantee topology recovery.

The systolic 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 $x \in X$, $\mathrm{sfs}(x)$ is equal to the geodesic distance from x to its cut-locus in X. This result also suggests a simple procedure for estimating $\mathrm{sfs}(x)$ at any point $x \in X$. Using this procedure, we devise a greedy algorithm for generating εsfs-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 [Fang et al. 2005; Wang et al. 2006].

We also 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 networks, 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 [de Silva 2008], 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 [Guibas and Oudot 2007], which states that a relaxed version of $\mathcal{C}_X^W(L)$ contains $\mathcal{D}_X(L)$ under similar conditions. The relaxation consists in allowing a simplex to be ν-witnessed by w if its vertices belong to the $\nu+1$ nearest landmarks of w, and the relaxed complex is denoted by $\mathcal{C}_{X,\nu}^W(L)$. Unfortunately, as pointed out in [Guibas and Oudot 2007], it is often the case that neither $\mathcal{C}_X^W(L)$ nor $\mathcal{C}_{X,\nu}^W(L)$ coincides with $\mathcal{D}_X(L)$. In the conference version of this paper [Gao et al. 2008], we took advantage of the fact that $\mathcal{D}_X(L)$ is sandwiched between $\mathcal{C}_X^W(L)$ and $\mathcal{C}_{X,\nu}^W(L)$, and we proved that the homology of $\mathcal{D}_X(L)$ (and hence the one of X) can be retrieved by computing the persistent homology between $\mathcal{C}_X^W(L)$ and $\mathcal{C}_{X,\nu}^W(L)$. Thus, the homology of X can be recovered without the need for constructing $\mathcal{D}_X(L)$ in practice. The drawback of the approach is that the proof of correctness requires the sampling density to be driven by the distance to the medial axis of $\mathbb{R}^2 \setminus X$, which can be arbitrarily small compared to the systolic feature size and requires some more stringent conditions on the regularity of the domain boundary [Gao et al. 2008].

In the present paper we consider a different approach, based on recent advances on the front of homology inference from point cloud data [Chazal and Oudot 2008]. Focusing on the one-parameter family of relaxed geodesic witness complexes $\mathcal{C}_{X,\nu}^W(L)$, where parameter ν ranges over $\mathbb{N}$, we show that this family is interleaved with the one-parameter family of *Čech complexes* $\mathcal{C}_\alpha(L)$, where parameter α ranges over $\mathbb{R}_+$. The interleaving of the two families of spaces implies that the persistent homological information they carry is similar [Chazal et al. 2009]. Now, $\mathcal{C}_\alpha(L)$ is the nerve of the union of the open geodesic balls of same radius α about the points of L, and that its homology is related to the one of its dual union of balls via the Nerve theorem. This union of geodesic balls covers the whole domain X and therefore shares the same topological invariants as long as α is large enough. Thus, via unions of open geodesic balls and their dual Čech complexes, a connection is drawn between the homology of X and the persistent homology of the one-parameter family of relaxed witness complexes. The weak point of this connection resides in the application of the Nerve theorem, which requires the geodesic balls to satisfy certain local conditions detailed in Definition 4.4 below. These conditions are automatically satisfied by small enough geodesic balls on Riemannian manifolds. Nevertheless, Lipschitz planar domains are not Riemannian manifolds, and the main point of our analysis is to show that geodesic balls of radii at most a fraction of the systolic feature size do satisfy the conditions of the Nerve theorem (Lemma 5.5). Our proof draws connections between the systolic feature size and the distance to the cut locus on the one hand (Lemma 5.6), as well as between Lipschitz planar domains and a class

of length spaces called *Alexandrov spaces* on the other hand (Theorem 5.10).

The paper is organized as follows: after recalling the necessary background in Section 2, we introduce the systolic feature size and give some of its basic properties in Section 3. Then, in Section 4, we study the topological structure of the geodesic Delaunay triangulation. We also relate the geodesic Delaunay triangulation to the geodesic witness complex. In Section 5 we turn the focus to the study of small geodesic balls in Lipschitz planar domains, from which theoretical guarantees on the homological structure of geodesic witness complexes are derived. In Section 6, we detail our algorithms for sampling Lipschitz domains in the plane, estimating their systolic feature size, and computing their homology. These algorithms are adapted to the sensor networks setting in Section 7.

2. BACKGROUND AND DEFINITIONS

The ambient space is $\mathbb{R}^2$, endowed with the Euclidean metric, noted d_E. Given a subset X of $\mathbb{R}^2$, $\mathring{X}$, $\overline{X}$, and ∂X, stand respectively for the interior, the closure, and the boundary of X. For all $x \in \mathbb{R}^2$ and all $r \in \mathbb{R}_+$, $B_E(x,r)$ denotes the open Euclidean ball of center x and of radius r. We also set $I = [0,1]$. 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.

2.1 Algebraic tools

Paths and loops. Given a subset X of $\mathbb{R}^2$, a *path* in X is a continuous map $I \to X$. For all $a, b \in I$ $(a \leq b)$, $\gamma_{|[a,b]}$ denotes the path $s \mapsto \gamma(a + s(b-a))$, which can be seen as the restriction of γ to the segment $[a,b]$. In addition, $\bar{\gamma}$ denotes the path $s \mapsto \gamma(1-s)$, which can be seen as the inverse of γ. Given two paths $\gamma, \gamma' : I \to X$ such that $\gamma(1) = \gamma'(0)$, $\gamma \cdot \gamma'$ denotes their concatenation, defined by $\gamma \cdot \gamma'(s) = \gamma(2s)$ for $0 \leq s \leq \frac{1}{2}$ and $\gamma \cdot \gamma'(s) = \gamma'(2s-1)$ for $\frac{1}{2} \leq s \leq 1$. A space X where all pairs of points are connected by at least one path is said to be *path-connected*.

Given a point $x \in X$, a *loop* through x in X is a path γ in X that starts and ends at x, *i.e.* such that $\gamma(0) = \gamma(1) = x$. For simplicity, we write $\gamma : (I, \partial I) \to (X, x)$. An equivalent representation[2] for γ is as a continuous map from the unit circle to X, and in this case we write $\gamma : (S^1, 1) \to (X, x)$ to specify that $\gamma(1) = x$. The concatenation operation gives a monoid structure to the set of loops through a same basepoint $x \in X$, the identity element being the constant loop $I \to \{x\}$ (or, equivalently, $S^1 \to \{x\}$).

Homotopy of maps and spaces. Given two topological spaces X and Y, two continuous maps $f, g : X \to Y$ are said to be *homotopic* if there exists a continuous map $F : X \times I \to Y$ such that, for all $x \in X$, we have $F(x,0) = f(x)$ and $F(x,1) = g(x)$. The map F is called a *homotopy* between f and g. It can be viewed as a path between f and g in the space of continuous maps from X to Y. Two spaces X and Y are said to be *homotopy equivalent* if there exist two maps $f : X \to Y$ and $g : Y \to X$, such that $g \circ f$ is homotopic to the identity in X and $f \circ g$ is homotopic to the identity in Y. Homotopy equivalent spaces have similar topological

[2]The choice of a particular representation for loops depends on the context, and it is always made explicit in the sequel.

invariants, such as Betti numbers, homology groups, or homotopy groups.

Suppose that a homotopy $F : X \times I \to Y$ between two maps $f, g : X \to Y$ keeps a certain subspace $X' \subseteq X$ fixed, that is: $\forall x' \in X'$, $\forall t \in I$, $F(x', t) = f(x') = g(x')$. Then, F is called a homotopy between f and g relative to X', and f, g are said to be homotopic relative to X'. A special case of interest is when $X = S^1$ and $X' = \{1\}$. Then, the maps f and g are two loops through a same basepoint $y \in Y$ that remains fixed throughout the homotopy F. If g is the constant loop $S^1 \to \{y\}$, then f is said to be *null-homotopic* in Y. The relation of homotopy relative to ∂I between loops through a same basepoint $y \in Y$ is an equivalence relation. The quotient monoid, endowed with the binary operation induced by concatenation, has in fact a group structure, and it is called the *fundamental group* of Y at basepoint y. If Y is path-connected, then its fundamental group is independent (up to isomorphism) of the chosen basepoint. And if moreover the fundamental group is trivial (*i.e.* all loops through any fixed basepoint are homotopic to the constant loop), then Y is said to be *simply connected*. We refer the reader to Chapter 1 of [Hatcher 2001] for further reading on homotopy theory with fixed basepoint.

Degrees of loops. To any loop $\gamma : S^1 \to S^1$ in the unit circle corresponds a unique integer $\deg \gamma \in \mathbb{Z}$, called the *degree* of γ, such that $\deg(\gamma \cdot \gamma') = \deg \gamma + \deg \gamma'$ for all loops $\gamma, \gamma' : S^1 \to S^1$, and that $\deg \gamma = 0$ for any constant map $\gamma : S^1 \to \{x\}$. It is easily seen that $\deg \bar{\gamma} = -\deg \gamma$. Moreover, it can be proved that the degree is invariant over each homotopy class of loops in S^1, so that $\deg \gamma$ encodes the homotopy class of the loop γ – see *e.g.* [Hatcher 2001, Theorem 1.7]. We can define a similar concept for loops in the plane. Given a loop $\gamma : S^1 \to \mathbb{R}^2$ and a point $x \in \mathbb{R}^2 \setminus \gamma(S^1)$, 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 onto the unit circle centered at x, define by $\pi_x(y) = \frac{y-x}{\mathrm{d}_E(y,x)}$. Since π_x is continuous over $\mathbb{R}^2 \setminus \{x\}$, the map γ_x is a continuous loop in S^1. We then define the degree of γ with respect to x as: $\deg_x \gamma = \deg \gamma_x$. It is also known as the winding number of γ about x. Given a point $x \in \mathbb{R}^2$, if Γ is a homotopy between two loops γ, γ' 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, hence 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 γ or γ' is constant, then $\deg_x \gamma = \deg_x \gamma' = 0$.*

Other useful results. We now recall two standard results of algebraic topology that relate the unions and intersections of planar sets that are *absolute neighborhood retracts* (ANR). A subset X of a topological space Y is a neighborhood retract if there exist an open set $X \subseteq \Omega \subseteq Y$ and a retraction $\Omega \to X$, *i.e.* a continuous map $\Omega \to X$ whose restriction to X is the identity. A topological space X is an ANR if every embedding of X as a closed subset of a normal space is a neighborhood retract [Borsuk 1967]. The proofs of the two results are given in Appendix A for completeness.

Proposition 2.2.

(i) Let $X_1, \cdots, X_k$ be compact planar sets such that the intersection of any arbitrary collection of the X_i's is a non-empty ANR. If $X_1, \cdots, X_k$ are simply connected, then so are the path-connected components of $X_1 \cap \cdots \cap X_k$.

(ii) Let X, Y be compact planar sets such that X, Y and $X \cap Y$ are non-empty ANR's. If X, Y are path-connected and $X \cup Y$ is simply connected, then $X \cap Y$ is path-connected.

2.2 Length structures

Most of the material of this section comes from Chapter 2 of [Burago et al. 2001]. The Euclidean space $\mathbb{R}^2$ is naturally endowed with a *length structure*, where admissible paths are all continuous paths $I \to \mathbb{R}^2$, and where the length of a path γ is defined by:

$$|\gamma| = \sup \left\{ \sum_{i=0}^{n-1} \mathrm{d}_E(\gamma(t_i), \gamma(t_{i+1})),\ n \in \mathbb{N},\ 0 = t_0 \leq t_1 \leq \cdots \leq t_n = 1 \right\}, \qquad (1)$$

where the supremum is 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. Therefore, we have $|\cdot| : C^0(I, \mathbb{R}^2) \to \mathbb{R}_+ \cup \{+\infty\}$. When the length of γ is finite, we say that γ is a *rectifiable* path. Note also that $|\cdot|$ may not be continuous with respect to the uniform topology over $C^0(I, \mathbb{R}^2)$. Take for instance the sequence of piecewise-linear curves $\gamma_i : I \to \mathbb{R}^2$ defined by $\gamma_i(t) = \big(t,\ t \mod \frac{1}{i}\big)$ if $\lfloor \frac{t}{i} \rfloor$ is even, and $\gamma_i(t) = \big(t,\ \frac{1}{i} - (t \mod \frac{1}{i})\big)$ if $\lfloor \frac{t}{i} \rfloor$ is odd. This sequence converges uniformly to the unit segment $t \mapsto (t, 0)$, yet every γ_i has length $\sqrt{2}$ therefore the limit length is $\sqrt{2}$. Nevertheless, $|\cdot|$ is lower semi-continuous [Burago et al. 2001, Proposition 2.3.4], which means that the limit length (here, $\sqrt{2}$), if it exists, must be at least the length of the limit path (here, 1).

Any subset X of $\mathbb{R}^2$ inherits a length structure from $\mathbb{R}^2$, where the class of admissible paths is $C^0(I, X)$, and where the length function is the same as above. We define an *intrinsic* (or *geodesic*) metric d_X over X as follows:

$$\forall x, y \in X,\ \mathrm{d}_X(x, y) = \inf \{|\gamma|,\ \gamma : I \to X,\ \gamma(0) = x,\ \gamma(1) = y\}, \qquad (2)$$

where the infimum is taken over all paths in X connecting x and y. It is clear that we have $\mathrm{d}_X(x, y) = +\infty$ whenever x, y belong to different path-connected components of X. However, the converse is not always true. Take for instance a set 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. This raises a critical issue, which is that the topology induced by d_X on X – also called intrinsic topology – may not always coincide[3] with the topology induced by d_E – also called Euclidean topology. This is a problem since the geodesic Voronoi diagram is closely related to the intrinsic metric d_X, whereas the goal is to capture the topology of X for the extrinsic metric

[3]In particular, a map $\gamma : I \to X$ that is continuous for the Euclidean topology may not always be continuous for the intrinsic topology. For instance, for any point $x \in \gamma(I)$ that lies on Koch's snowflake, the geodesic distance between x and any other point of X is infinite, which implies that, for any $r > 0$, the open geodesic ball $B_X(x, r)$ is reduced to $\{x\}$, and hence its pre-image through γ is a closed subset of I, and not an open subset of I.

d_E. In order to bridge the gap between the two topologies, we will make further assumptions on the subspace X in the next section.

Another issue is that some pairs of points $x, y \in X$ may not have a shortest path connecting them, *i.e.* a path $\gamma : I \to X$ such that $\gamma(0) = x$, $\gamma(1) = y$, and $|\gamma| = \mathrm{d}_X(x, y)$. This means that the infimum in Eq. (2) is not always a minimum. As an example, take for X the closed unit disk $\overline{B_E}(0, 1)$, and remove the closed disk $\overline{B_E}(0, \frac{1}{2})$ from it: points $(-1, 0)$ and $(1, 0)$ have no shortest path connecting them in X. Nevertheless, when X is compact, the following variant of the Arzela-Ascoli theorem applies:

THEOREM 2.3 THM. 2.5.14 AND PROP. 2.5.19 OF [BURAGO ET AL. 2001]. *If X is compact, then every sequence of paths with uniformly bounded length contains a uniformly converging subsequence. As a consequence, every pair of points connected by a rectifiable path in X has a shortest path in X.*

2.3 Lipschitz domains in the plane

To deal with the issues of the previous section, we make further assumptions on our domain X.

DEFINITION 2.4. *A* Lipschitz domain *in the plane 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 points $x \in \partial X$, 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 of x included in 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 ϕ_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 around x, we ensure that $\phi_x^{-1}(X \cap V_x)$ is the intersection of $\mathbb{R}^2_+$ with the open disk B. This makes the pre-image of $X \cap V_x$ through ϕ_x convex.

The concept of Lipschitz domain is related to the classical notion of smooth submanifold with boundary – see *e.g.* Chapter 8 of [Lee 2002], the only difference being that the local charts ϕ are only required to be Lipschitz, and not C^1-continuous. As a result, the boundary of X may not be smooth. This makes the class of Lipschitz domains quite large: in particular, it contains all smooth or polygonal domains.

Since a Lipschitz domain X is a compact subset of $\mathbb{R}^2$, Theorem 2.3 applies, and therefore any pair of points of X connected by a rectifiable path in X has a shortest path in X. Moreover, according to Rademacher's theorem [Federer 1996, §3.1.6], the boundary ∂X is differentiable almost everywhere. But the property of Lipschitz domains that is most interesting to us is that their boundaries are rectifiable, since they are locally images of Lipschitz maps [Federer 1996, §2.10.11]. This enables to show that the pathological cases mentioned in Section 2.2 cannot occur with a Lipschitz domain[4], as stated in Theorem 2.5 below.

[4]In particular, the boundary of a Lipschitz domain cannot coincide locally with a fractal curve such as Koch's snowflake, whose length is infinite.

Bibliographical note. Lipschitz domains are sometimes called *weakly Lipschitz manifolds* [Axelsson and M$^{\text{c}}$Intosh 2004] in the literature, as opposed to *strongly Lipschitz manifolds* [Boissonnat and Oudot 2006], for which it is further assumed that the boundary of the domain coincides locally with the graph of some univariate Lipschitz function. Notice also that, in contrast with [Boissonnat and Oudot 2006], we do not make any assumption on the Lipschitz constants of the local charts. All we need to know is that the latter are Lipschitz, so that their images are rectifiable [Federer 1996, §2.10.11].

THEOREM 2.5. *If X is a Lipschitz domain in the plane, then the intrinsic topology coincides with the Euclidean topology on X.*

PROOF. First, Eq. (2) implies that $\mathrm{d}_E(x,y) \leq \mathrm{d}_X(x,y)$ for all $x, y \in X$. It follows that every open Euclidean ball centered in X contains the open geodesic ball of same center and same radius. As a consequence, every open set in (X, d_E) is also open in (X, d_X). This means that the intrinsic topology is finer than the Euclidean topology. To show that, conversely, the Euclidean topology is also finer than the intrinsic topology, we will use the following technical result:

CLAIM 2.5.1. *If X is a Lipschitz domain in the plane, then, for all point $x \in X$, the map $y \mapsto \mathrm{d}_X(x,y)$ is continuous for the Euclidean topology on X.*

PROOF. Let $x, y \in X$. We will prove that, for all $\varepsilon > 0$, there exists a $\delta > 0$ such that $\forall y' \in B_E(y,\delta) \cap X$, $|\mathrm{d}_X(x,y') - \mathrm{d}_X(x,y)| < \varepsilon$.

- Assume first that $y \in \mathring{X}$. Then there exists $\varepsilon' > 0$ such that $B_E(y,\varepsilon') \subseteq \mathring{X}$. Let $\delta = \min\{\varepsilon, \varepsilon'\}$. For all $y' \in B_E(y,\delta)$, the line segment $[y,y']$ lies in $\mathring{X}$, hence $\mathrm{d}_X(y,y') = \mathrm{d}_E(y,y') < \varepsilon$. It follows then from the triangle inequality that $|\mathrm{d}_X(x,y') - \mathrm{d}_X(x,y)| \leq \mathrm{d}_X(y,y') < \varepsilon$.

- Assume now that $y \in \partial X$. There exists a neighborhood V_y of y in $\mathbb{R}^2$ such that $X \cap V_y = \phi_y(\mathbb{R}^2_+) \cap V_y$, for some Lipschitz homeomorphism ϕ_y. Let c_y be the Lipschitz constant of ϕ_y. As mentioned after Definition 2.4, we can assume without loss of generality that $\phi_y^{-1}(X \cap V_y)$ is the intersection of $\mathbb{R}^2_+$ with an open disk centered at the origin of radius at most $\frac{\varepsilon}{c_y}$. Then, for all point $y' \in X \cap V_y$, consider the path $\gamma : s \mapsto \phi_y\left(s\, \phi_y^{-1}(y')\right)$. Since $\phi_y^{-1}(X \cap V_y)$ is convex, $\gamma(I)$ is included in $X \cap V_y$, and hence in X. Moreover, the length of the line segment $[0, \phi_y^{-1}(y')]$ is less than $\frac{\varepsilon}{c_y}$, hence the length of γ is less than ε, since ϕ_y is c_y-Lipschitz [Federer 1996, §2.10.11]. It follows that $\mathrm{d}_X(y,y') < \varepsilon$, which implies that $|\mathrm{d}_X(x,y') - \mathrm{d}_X(x,y)| \leq \mathrm{d}_X(y,y') < \varepsilon$, by the triangle inequality. This concludes the proof of the claim. □

We can now show that the Euclidean topology is finer than the intrinsic topology on X, which will end the proof of Theorem 2.5. Consider any open geodesic ball $B_X(x,\varepsilon)$, where $x \in X$ and $\varepsilon > 0$. Observe that $B_X(x,\varepsilon) = \mathrm{d}_X(x,.)^{-1}([0,\varepsilon[)$, where $\mathrm{d}_X(x,.)$ denotes the map $y \mapsto \mathrm{d}_X(x,y)$. Since $[0,\varepsilon[$ is open in $\mathbb{R}_+$ and $\mathrm{d}_X(x,.)$ is continuous for the Euclidean topology, $B_X(x,\varepsilon)$ is open in (X, d_E). And since the open geodesic balls form a basis for the intrinsic topology, every open set in (X, d_X) is also open in (X, d_E). This means that the Euclidean topology is finer than the geodesic topology on X.

From now on, X will be endowed with the Euclidean topology by default. Thanks to Theorem 2.5, this topology will coincide with the intrinsic topology whenever X is a Lipschitz domain.

The next result states that every path in X can be approximated within any accuracy by a homotopic rectifiable path. This implies that the homotopy classes of paths in X coincide with the homotopy classes of rectifiable paths. In particular, every pair of points lying in the same path-connected component of X is connected by a rectifiable path, and hence it has a shortest path in X, by Theorem 2.3.

LEMMA 2.6. *For any continuous path $\gamma : I \to X$ and any real number $\varepsilon > 0$, there exists a rectifiable path $\gamma_\varepsilon : I \to X$, homotopic to γ relative*[5] *to ∂I in X, such that $\max_{s\in I} \min_{t\in I} \mathrm{d}_X(\gamma_\varepsilon(s), \gamma(t)) < \varepsilon$.*

The quantity $\max_{s\in I} \min_{t\in I} \mathrm{d}_X(\gamma_\varepsilon(s), \gamma(t))$ is nothing but the semi-Hausdorff distance from $\gamma_\varepsilon(I)$ to $\gamma(I)$ in the intrinsic metric. The basic idea of the proof is to define γ_ε as a piecewise-linear curve whose vertices lie on $\gamma(I)$. This is possible far away from the boundary of X, but not in its vicinity, where the shape of ∂X might prevent $\gamma_\varepsilon(I)$ from being included in X. However, in the vicinity of ∂X, we can map $\gamma(I)$ to parameter space through one of the local charts ϕ introduced in Definition 2.4. Since the pre-image of X is convex, we can define a piecewise-linear curve approximating $\phi^{-1}(\gamma(I))$ in parameter space, which we then map back to a rectifiable curve in X through ϕ. The rest of the section is devoted to the details of the proof and can therefore be skipped in a first reading.

PROOF. Let η be an arbitrary positive real number. According to Definition 2.4, for all $x \in \partial X$, there exists some neighborhood $V_x \subseteq \mathbb{R}^2$ such that, inside V_x, X coincides with the image of $\mathbb{R}^2_+$ through some Lipschitz homeomorphism ϕ_x. As mentioned after Definition 2.4, we can assume without loss of generality that V_x is included in $B_E(x, \frac{\eta}{2})$, and that the pre-image of $X \cap V_x$ through ϕ_x is convex. Consider the collection of open sets $\{V_x\}_{x\in\partial X}$. This is an open cover of ∂X, which is compact, hence there exist $x_1, \cdots, x_k$ such that $V_{x_1} \cup \cdots \cup V_{x_k}$ covers ∂X. For simplicity of notations, for all $i = 1, \cdots, k$ we rename V_{x_i} as V_i and ϕ_{x_i} as ϕ_i. The open sets V_i will be used to *shield* the boundary ∂X.

For all $s \in I$, we consider an open Euclidean disk B_s about $\gamma(s)$, of radius r_s defined as follows:

- if $B_E(\gamma(s), \frac{\eta}{2}) \cap \partial X = \emptyset$, then $r_s = \frac{\eta}{2}$;
- else, if $\gamma(s) \notin \partial X$, then $r_s = \mathrm{d}_E(\gamma(s), \partial X)$, where $\mathrm{d}_E(\gamma(s), \partial X) > 0$ denotes the Euclidean distance of $\gamma(s)$ to the closed set ∂X;
- else, we have $\gamma(s) \in \partial X$, therefore $\gamma(s)$ belongs to some neighborhood V_i, and we choose $r_s > 0$ such that $B_s \subseteq V_i$.

By construction, we have $B_s \subseteq \mathring{X}$ if $\gamma(s) \notin \partial X$, and $B_s \subseteq V_i$ for some i otherwise. Since γ is continuous, the pre-image of $\gamma(I) \cap B_s$ through γ is an open subset of I. Therefore, it is a disjoint union of open intervals in I. Consider the collection of all these open intervals, for s spanning I. This collection of intervals forms an open cover of I, which is compact, hence there are l intervals in the collection, $I_1, \cdots, I_l$,

[5] As mentioned in Section 2.1, this means that the homotopy between γ_ε and γ is constant over $\partial I = \{0, 1\}$.

such that $I = I_1 \cup \cdots \cup I_l$. Observe that, by construction, for all $i = 1, \cdots, l$ we have that $\gamma(I_i)$ is included in B_{s_i} for some $s_i \in I$.

We can assume without loss of generality that the family $\{I_i\}_{1 \leq i \leq l}$ is minimal, in the sense that the removal of any element would destroy the cover: $\forall i = 1, \cdots, l$, $\bigcup_{j \neq i} I_j \not\supseteq I$. If it is not so, then we can always remove elements from the family until the property is satisfied. Let us now re-order the elements of the family such that the left endpoint of I_i is smaller than the left endpoint of I_{i+1}, for all i. Since the family is minimal, the ordering on the left endpoints of the I_i is the same as the ordering on their right endpoints. As a consequence, each I_i intersects only I_{i-1} and I_{i+1}. Let $t_1 = 0$, $t_{l+1} = 1$, and $t_i \in I_{i-1} \cap I_i \ \forall i = 2, \cdots, l$. We will approximate γ by a piecewise Lipschitz curve connecting the $\gamma(t_i)$. For simplicity, we rename $\gamma_{|[t_i, t_{i+1}]}$ as γ_i.

By construction, for all $i = 1, \cdots, l$ we have $[t_i, t_{i+1}] \subseteq I_i$, hence $\gamma_i(I) = \gamma([t_i, t_{i+1}]) \subseteq \gamma(I_i)$, which is included in B_{s_i}.

- Assume first that $s_i \notin \partial X$, which implies that $B_{s_i} \subseteq \mathring{X}$ and $r_{s_i} \leq \frac{\eta}{2}$. Define γ_η^i as the linear interpolation between $\gamma(t_i)$ and $\gamma(t_{i+1})$, namely: $\gamma_\eta^i : s \mapsto (1 - s)\gamma(t_i) + s\gamma(t_{i+1})$. Since B_{s_i} is convex, $\gamma_\eta^i(I)$ is included in B_{s_i} and hence in X. Moreover, we have $\gamma_\eta^i(0) = \gamma(t_i) = \gamma_i(0)$, $\gamma_\eta^i(1) = \gamma(t_{i+1}) = \gamma_i(1)$, and the Hausdorff distance $\mathrm{d}_{\mathcal{H}}(\gamma_\eta^i(I), \gamma([t_i, t_{i+1}]))$ in the Euclidean metric is less than the diameter of B_{s_i}, which is bounded by η. Furthermore, the map $\Gamma : I \times I \to \mathbb{R}^2$ defined by $\Gamma(s, t) = (1 - t)\gamma_\eta^i(s) + t\gamma_i(s)$ is a homotopy relative to ∂I between γ_η^i and γ_i in $\mathbb{R}^2$. Since it is a linear interpolation between two maps whose images lie in B_{s_i}, which is convex, the image of Γ is also included in B_{s_i}, and hence in $\mathring{X}$. It follows that Γ is a homotopy relative to ∂I between γ_η^i and γ_i in X.

- Assume now that $s_i \in \partial X$, which implies that B_{s_i} is included in some V_j. Because of the presence of ∂X in the vicinity of $\gamma([t_i, t_{i+1}])$, we can no longer guarantee that the linear interpolation between $\gamma(t_i)$ and $\gamma(t_{i+1})$ remains in X. This is why we use the chart ϕ_j to map the arc $\gamma([t_i, t_{i+1}])$ to parameter space $\phi_j^{-1}(X \cap V_j)$, which is convex. Specifically, we define γ_η^i as the image through ϕ_j of the linear interpolation between the pre-images of $\gamma(t_i)$ and $\gamma(t_{i+1})$ in $\phi_j^{-1}(X \cap V_j)$, namely: $\gamma_\eta^i : s \mapsto \phi_j\left((1 - s)(\phi_j^{-1} \circ \gamma)(t_i) + s(\phi_j^{-1} \circ \gamma)(t_{i+1})\right)$. As in the previous case, we have $\gamma_\eta^i(0) = \gamma_i(0)$ and $\gamma_\eta^i(1) = \gamma_i(1)$. Moreover, since $\phi_j^{-1}(X \cap V_j)$ is convex, we have $(1 - s)(\phi_j^{-1} \circ \gamma)(t_i) + s(\phi_j^{-1} \circ \gamma)(t_{i+1}) \in \phi_j^{-1}(X \cap V_j)$ for all $s \in I$, hence $\gamma_\eta^i(I)$ is included in $X \cap V_j$. It follows that $\gamma_\eta^i : I \to X$, and that the Hausdorff distance $\mathrm{d}_{\mathcal{H}}(\gamma_\eta^i(I), \gamma_i(I))$ in the Euclidean metric is less than the diameter of V_j, which is bounded by η since $V_j \subseteq B_E(x_j, \frac{\eta}{2})$. Notice also that γ_η^i is a Lipschitz map, hence it is rectifiable, by [Federer 1996, §2.10.11]. Finally, the map $\Gamma : I \times I \to \mathbb{R}^2$ defined by $\Gamma(s, t) = \phi_j\left((1 - t)(\phi_j^{-1} \circ \gamma_\eta^i)(s) + t(\phi_j^{-1} \circ \gamma_i)(s)\right)$ is a homotopy relative to ∂I between γ_η^i and γ_i in $\mathbb{R}^2$. Since $\phi_j^{-1} \circ \Gamma$ is a linear interpolation between maps $\phi_j^{-1} \circ \gamma_\eta^i$ and $\phi_j^{-1} \circ \gamma_i$ in $\phi_j^{-1}(X \cap V_j)$, which is convex, the image of Γ is included in $X \cap V_j$. It follows that Γ is a homotopy relative to ∂I between γ_η^i and γ_i in X.

We now define γ_η as the concatenation of the γ_η^i, namely: $\gamma_\eta = \gamma_\eta^1 \cdot \gamma_\eta^2 \cdots \gamma_\eta^l$. By concatenating the homotopies relative to ∂I between the γ_η^i and the γ_i, we obtain a homotopy relative to ∂I between γ_η and γ in X. Moreover, since the

γ_η^i are rectifiable, so is γ_η. We also have $\gamma_\eta(0) = \gamma_\eta^1(0) = \gamma_1(0) = \gamma(0)$, and $\gamma_\eta(1) = \gamma_\eta^l(1) = \gamma_l(1) = \gamma(1)$. Finally, the Hausdorff distance $\mathrm{d}_{\mathcal{H}}(\gamma_\eta(I), \gamma(I))$ in the Euclidean metric is bounded by the maximum of the $\mathrm{d}_{\mathcal{H}}(\gamma_\eta^i(I), \gamma_i(I))$, which is less than η.

To conclude the proof of the lemma, we need to show that bounding the Hausdorff distance between γ and its approximation in the Euclidean metric is sufficient for bounding the semi-Hausdorff distance from the approximation to γ in the intrinsic metric. Let ε be an arbitrary positive real number. Since by Theorem 2.5 the Euclidean and geodesic topologies are equal on X, for all $s \in I$ there exists an $\eta_s > 0$ such that $B_E(\gamma(s), \eta_s)$ is included in $B_X(\gamma(s), \varepsilon)$. The balls $B_E(\gamma(s), \eta_s)$ form an open cover of $\gamma(I)$. Hence, for all $s \in I$, the Euclidean distance from $\gamma(s)$ to the complement of the cover in $\mathbb{R}^2$ is positive. Since γ and the distance to the complement are continuous, while I is compact, the infimum η of the distances of the $\gamma(s)$ to the complement is in fact a minimum, and therefore it is positive. Now, according to the previous paragraphs, there exists a curve $\gamma_\eta : I \to X$, homotopic to γ relative to ∂I in X, such that $\mathrm{d}_{\mathcal{H}}(\gamma(I), \gamma_\eta(I)) < \eta$. It follows that $\gamma_\eta(I) \subset \bigcup_{s\in I} B_E(\gamma(s), \eta)$, which is included in $\bigcup_{s\in I} B_E(\gamma(s), \eta_s) \subseteq \bigcup_{s\in I} B_X(\gamma(s), \varepsilon)$. This concludes the proof of Lemma 2.6, with $\gamma_\varepsilon = \gamma_\eta$. □

Observe that, in the proof of Lemma 2.6, the family of balls $\{B_s\}_{s\in I}$ forms an open cover of $\gamma(I)$. Letting ζ be the quantity $\inf\left\{\mathrm{d}_E(x, \gamma(I)) \mid x \in X \setminus \bigcup_{s\in I} B_s\right\} > 0$, the second part of the proof shows in fact that every path $\gamma' : I \to X$ such that $\gamma'(0) = \gamma(0)$, $\gamma'(1) = \gamma(1)$, and $\mathrm{d}_E(\gamma'(s), \gamma(s)) < \zeta$ for all $s \in (0, 1)$, is homotopic to γ relative to ∂I. Thus, we obtain the following guarantee:

LEMMA 2.7. *For any path $\gamma : I \to X$, there exists a quantity $\zeta > 0$ such that every path $\gamma' : I \to X$ with same endpoints as γ that satisfies $\mathrm{d}_E(\gamma'(s), \gamma(s)) < \zeta$ for all $s \in (0, 1)$ is homotopic to γ relative to ∂I.*

3. THE SYSTOLIC FEATURE SIZE

DEFINITION 3.1. *Let X be a Lipschitz domain in the plane. The* systolic feature size *of X at a given point $x \in X$ is the quantity:* $\mathrm{sfs}(x) = \frac{1}{2} \inf\{|\gamma|,\ \gamma : (S^1, 1) \to (X, x)$ *non null-homotopic in X}.

As illustrated in Figure 1 (left and center), the resort to an intrinsic metric makes the systolic feature size rather insensitive to the local geometry of the domain X. Indeed, sfs depends on the geodesic perimeters of the holes of X, which depend on the geometry of X at a more global scale.

The rest of this section is devoted to the proof of some useful basic properties of the systolic feature size.

LEMMA 3.2. *Let X be a Lipschitz domain in the plane, and let x be a point in X. If the path-connected component of X that contains x is simply connected, then* $\mathrm{sfs}(x) = +\infty$. *Else,* $\mathrm{sfs}(x) < +\infty$, *and there exists a non null-homotopic rectifiable loop $\gamma : (S^1, 1) \to (X, x)$ such that* $\mathrm{sfs}(x) = \frac{1}{2} |\gamma| > 0$.

PROOF. Let $x \in X$. Call X_x the path-connected component of X that contains x. Every loop through x in X is a loop in X_x. If X_x is simply connected, then the set $\{\gamma : (S^1, 1) \to (X_x, x)$ non null-homotopic in $X_x\}$ is empty, and therefore its

lower bound $\mathrm{sfs}(x)$ is infinite. Assume now that X_x is not simply connected. Then, there exists at least one non null-homotopic loop $\gamma_0 : (S^1, 1) \to (X_x, x)$. By Lemma 2.6, we can assume without loss of generality that γ_0 is rectifiable. We then have $\mathrm{sfs}(x) \leq \frac{1}{2} |\gamma_0| < +\infty$.

Consider now a sequence $(\gamma_i)_i$ of non null-homotopic loops through x in X_x, such that $(|\gamma_i|)_i$ converges to $2\mathrm{sfs}(x)$. Such a sequence exists, since $2\mathrm{sfs}(x) < +\infty$ is the infimum of the set of lengths of non null-homotopic loops through x. By convergence, we know that there exists a rank n such that, for all $i \geq n$, γ_i is a rectifiable curve of length $|\gamma_i| \leq 2\mathrm{sfs}(x)+1$. Thus, the sequence $(\gamma_{n+i})_i$ is uniformly bounded by $2\mathrm{sfs}(x)+1$, which implies by Theorem 2.3 that it contains a subsequence converging uniformly to some loop $\gamma : (I, \partial I) \to (X_x, x)$. It follows from Lemma 2.7 that, after a certain rank, every element in the subsequence is homotopic to γ relative to ∂I. As a consequence, γ is not null-homotopic in X, and therefore $|\gamma|$ is positive and at least $2\mathrm{sfs}(x)$. In addition, since $(|\gamma_i|)_i$ converges to $2\mathrm{sfs}(x)$, the lower semi-continuity of $|\cdot|$ implies that $|\gamma| \leq 2\mathrm{sfs}(x)$. As a conclusion, we have $|\gamma| = 2\mathrm{sfs}(x) > 0$. □

LEMMA 3.3. *Let X be a Lipschitz domain in the plane. The map $x \mapsto \mathrm{sfs}(x)$ is 1-Lipschitz in the intrinsic metric. Hence, it is continuous for the Euclidean topology, and $\mathrm{sfs}(X) = \inf\{\mathrm{sfs}(x),\ x \in X\}$ is positive.*

PROOF. Let $x, y \in X$. If x, y belong to different path-connected components of X, then we have $\mathrm{d}_X(x, y) = +\infty$. It follows that $|\mathrm{sfs}(x) - \mathrm{sfs}(y)| \leq \mathrm{d}_X(x, y)$. Assume now that x, y belong to the same path-connected component X_i of X. Let γ be a shortest path between x and y in X. We are guaranteed by Theorem 2.3 and Lemma 2.6 that such a path exists. If X_i is simply connected, then sfs is constant and equal to $+\infty$ over X_i. Else, consider a loop $\gamma_x : (S^1, 1) \to (X, x)$ such that $|\gamma_x| = 2\mathrm{sfs}(x) < +\infty$. Such a loop exists, by Lemma 3.2. Then, the path $\gamma_y = \bar{\gamma} \cdot \gamma_x \cdot \gamma$ is a loop through y in X. Its length is $|\gamma_x| + 2|\gamma| = 2\mathrm{sfs}(x) + 2\mathrm{d}_X(x, y)$. Moreover, the map $\gamma_x \mapsto \bar{\gamma} \cdot \gamma_x \cdot \gamma$ is known to induce an isomorphism between the fundamental groups of X_i at basepoints x and y — see *e.g.* [Hatcher 2001, Proposition 1.5]. Therefore, the loop γ_y is not null-homotopic in X, which implies that $\mathrm{sfs}(y) \leq \frac{1}{2} |\gamma_y| = \mathrm{sfs}(x) + \mathrm{d}_X(x, y)$. This proves that the map $x \mapsto \mathrm{sfs}(x)$ is 1-Lipschitz in the intrinsic metric, and hence continuous for the intrinsic topology, but also for the Euclidean topology, by Theorem 2.5. Since X is compact, there exists some point $x \in X$ such that $\mathrm{sfs}(X) = \mathrm{sfs}(x)$, which is positive, by Lemma 3.2. □

LEMMA 3.4. *Let X be a Lipschitz domain in the plane. For all point $x \in X$, every loop inside the open geodesic ball $B_X(x, \mathrm{sfs}(x))$ is null-homotopic in X.*

PROOF. Assume for a contradiction that there exists some point $x \in X$ and some loop $\gamma_x : S^1 \to B_X(x, \mathrm{sfs}(x))$ that is not null-homotopic in X. Since we have $\max_{s \in I} \mathrm{d}_X(x, \gamma_x(s)) < \mathrm{sfs}(x)$, Lemma 2.6 ensures that there exists a rectifiable loop $S^1 \to X$ that is homotopic to γ_x in X, and that is still included in $B_X(x, \mathrm{sfs}(x))$. Hence, we can assume without loss of generality that γ_x is rectifiable. Let ζ be a shortest path between x and $y = \gamma_x(0)$. The path $\gamma = \zeta \cdot \gamma_x \cdot \bar{\zeta}$ is a loop through x, included in $B_X(x, \mathrm{sfs}(x))$, of length $|\gamma| \leq |\gamma_x| + 2\mathrm{d}_X(x, y) < +\infty$. Moreover, γ is non null-homotopic in X, since it is homotopic to γ_x. It follows that $|\gamma| \geq 2\mathrm{sfs}(x)$.

For all $s \in I$, we define γ_s and ζ_s to be respectively the path $\gamma_{|[0,s]}$ and a shortest path between x and $\gamma(s)$. Let $s_0 = \inf\{s \mid \gamma_s \cdot \bar{\zeta}_s \text{ non null-homotopic in } X\}$. This means that, for all $s < s_0$, $\gamma_s \cdot \bar{\zeta}_s$ is null-homotopic in X, whereas for all $\eta > 0$ there exists some $s \in [s_0, s_0 + \eta[$ such that $\gamma_s \cdot \bar{\zeta}_s$ is not null-homotopic in X.

- If $s_0 = 0$, then there are arbitrarily short non null-homotopic loops through x in X, which contradicts the fact that $\mathrm{sfs}(x) > 0$ (Lemma 3.2).
- If $s_0 = 1$, then for s arbitrarily close to 1, $\gamma_{|[s,1]} \cdot \zeta_s$ is non null-homotopic in X, and of length arbitrarily close to $|\zeta_s| < \mathrm{sfs}(x)$, which contradicts the definition of $\mathrm{sfs}(x)$ (Definition 3.1).

It follows that $s_0 \in]0, 1[$. For all $\eta > 0$, there exist $s_{-\eta}, s_{+\eta} \in I$ such that $s_0 - \eta < s_{-\eta} < s_0 < s_{+\eta} < s_0 + \eta$, and that $\gamma_{s_{-\eta}} \cdot \bar{\zeta}_{s_{-\eta}}$ is null-homotopic in X whereas $\gamma_{s_{+\eta}} \cdot \bar{\zeta}_{s_{+\eta}}$ is not. Then, $\zeta_{s_{-\eta}}$ is homotopic to $\gamma_{s_{-\eta}}$ relative[6] to ∂I, which implies that $\zeta_{s_{-\eta}} \cdot \gamma_{|[s_{-\eta}, s_{+\eta}]}$ is homotopic to $\gamma_{s_{+\eta}}$ relative to ∂I. As a result, the loop $\gamma' = \left(\zeta_{s_{-\eta}} \cdot \gamma_{|[s_{-\eta}, s_{+\eta}]}\right) \cdot \bar{\zeta}_{s_{+\eta}}$ is homotopic to $\gamma_{s_{+\eta}} \cdot \bar{\zeta}_{s_{+\eta}}$, which is not null-homotopic in X. Hence, we have $|\gamma'| \geq 2\mathrm{sfs}(x)$, by definition of $\mathrm{sfs}(x)$.

Now, the length of γ' is $|\zeta_{s_{-\eta}}| + |\gamma_{|[s_{-\eta}, s_{+\eta}]}| + |\bar{\zeta}_{s_{+\eta}}| \leq 2 \max_{s \in I} \mathrm{d}_X(x, \gamma(s)) + |\gamma_{|[s_{-\eta}, s_{+\eta}]}|$. Since η is arbitrarily small, so is $|\gamma_{|[s_{-\eta}, s_{+\eta}]}|$, therefore $|\gamma'|$ is arbitrarily close to $2 \max_{s \in I} \mathrm{d}_X(x, \gamma(s))$, which is less than $2\mathrm{sfs}(x)$. This contradicts the fact that $|\gamma'| \geq 2\mathrm{sfs}(x)$, as proved in the previous paragraph. □

Note that Lemma 3.4 does not imply that the ball $B_X(x, \mathrm{sfs}(x))$ itself is contractible. It turns out that open geodesic balls of radius at most a fraction of the systolic feature size are contractible. The proof of this fact requires some more work though — see Section 5.

4. GEODESIC DELAUNAY TRIANGULATION AND WITNESS COMPLEX

Given a Lipschitz domain X in the plane, and a set of landmarks $L \subset X$ that is dense enough with respect to the systolic feature size of X, we show in Section 4.1 that the geodesic Delaunay triangulation $\mathcal{D}_X(L)$ has the same homotopy type as X (Theorem 4.3). Furthermore, for any set of witnesses $W \subseteq X$ that is dense enough compared to L, we prove 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 $\mathcal{C}_{X,\nu}^W(L)$ (Theorems 4.14 and 4.17). Densities of point clouds are measured according to the following definition, where the scalar field h will be chosen to be either a constant function or a fraction of the systolic feature size:

Definition 4.1. *Given a Lipschitz planar domain X and a function $h : X \to \mathbb{R}_+ \cup \{+\infty\}$, a set $L \subseteq X$ is a* geodesic h-sample *of X if we have $\mathrm{d}_X(x, L) < h(x)$ for all points $x \in X$. In addition, L is h-sparse if we have $\mathrm{d}_X(p, q) \geq \min\{h(p), h(q)\}$ for all points $p \neq q \in L$.*

It follows from the definition that any geodesic h-sample L of X must have points in every path-connected component of X, because geodesic distances to L are required to be finite ($\mathrm{d}_X(x, L) < h(x)$). We will see in Section 6.2 how to generate geodesic εsfs-samples of Lipschitz planar domains.

[6] As mentioned in Section 2.1, this means that the homotopy between $\zeta_{s_{-\eta}}$ and $\gamma_{s_{-\eta}}$ is constant over $\partial I = \{0, 1\}$.

4.1 Geodesic Delaunay triangulations

Geodesic Voronoi diagrams are nothing but Voronoi diagrams in the intrinsic metric:

DEFINITION 4.2. *Given a subset X of $\mathbb{R}^2$, and a finite subset L of X, the* geodesic Voronoi diagram *of L in X, or $\mathcal{V}_X(L)$ for short, is a cellular decomposition of X, where the cell of a point $p \in L$ is defined as the locus of all the points $x \in X$ such that $\mathrm{d}_X(x,p) \leq \mathrm{d}_X(x,q)\ \forall q \in L$. The nerve of $\mathcal{V}_X(L)$ is called the* geodesic Delaunay triangulation *of L in X, noted $\mathcal{D}_X(L)$.*

Given a simplex $\sigma \in \mathcal{D}_X(L)$, we call $\mathrm{V}_X(\sigma)$ its dual Voronoi face. Note that, in contrast with the Euclidean case, $\mathrm{V}_X(\sigma)$ does not always have Lebesgue measure zero when the dimension of σ is non-zero, as illustrated in Figure 1 (right).

THEOREM 4.3. *If X is a Lipschitz domain in the plane, and L a geodesic ε*sfs*-sample of X, for some $\varepsilon \leq \frac{1}{3}$, then $\mathcal{D}_X(L)$ and X are homotopy equivalent.*

The rest of Section 4.1 is devoted to the proof of Theorem 4.3. The proof relies on the so-called Nerve theorem, stated as Theorem 4.5 below, which relies on the concept of *good cover*:

DEFINITION 4.4. *Let $\mathcal{U}$ be a finite collection of closed (resp. open) subsets of X whose union covers X. Then, $\mathcal{U}$ is a* good closed (resp. open) cover *of X if for any non-empty subset $\mathcal{V} \subseteq \mathcal{U}$ the common intersection between the elements of $\mathcal{V}$ is either empty or contractible.*

THEOREM 4.5 FROM [BORSUK 1948; WU 1962], SEE ALSO [HATCHER 2001]. *The nerve of a good closed (resp. open) cover of X is homotopy equivalent to X.*

Here, we take $\mathcal{U}$ 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, proving Theorem 4.3 comes down to showing that any collection of cells of $\mathcal{V}_X(L)$ has an empty or contractible intersection, and then invoking Theorem 4.5. Our proof proceeds in three steps: first, we show that every single Voronoi cell is contractible (Section 4.1.1); then, we show that any pair of Voronoi cells has an empty or contractible intersection (Section 4.1.2); finally, we show inductively that any arbitrary collection of Voronoi cells has an empty or contractible common intersection (Section 4.1.3).

Along the way, our proof uses several results of algebraic topology (including the ones of Proposition 2.2) that require non-empty intersections of geodesic Voronoi cells to be ANR's. This fact turns out to be true in any Lipschitz planar domain, and it can be shown using the local continuity of the geodesic flow, proved[7] in Sections 5.1 and 5.2, as well as some nesting properties of neighborhood retracts, stated in Theorem III.3 of [Daverman 2007]. This minor and rather technical aspect of our proof does not bring any particular insights into the problem. Therefore, it is omitted for the convenience of exposition, and in the sequel non-empty intersections of geodesic Voronoi cells are admitted to be ANR's.

[7]The statements and proofs from Sections 5.1 and 5.2 do not rely on the results of this section, therefore they can be invoked here.

4.1.1 *Voronoi cells*

LEMMA 4.6. *Under the hypotheses of Theorem 4.3, every cell of $\mathcal{V}_X(L)$ is path-connected.*

PROOF. Let $p \in L$, and let $x \in \mathcal{V}_X(p)$. Let $\gamma : I \to X$ be a shortest path from p to x in X. Such a path γ exists 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 εsfs-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, $\gamma(I) \subseteq \mathcal{V}_X(p)$, and x is path-connected to p in $\mathcal{V}_X(p)$. □

LEMMA 4.7. *Under the hypotheses of Theorem 4.3, every cell of $\mathcal{V}_X(L)$ is simply connected.*

PROOF. Let $p \in L$. By Lemma 4.6, $\mathcal{V}_X(p)$ is path-connected. Assume 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.* [Hatcher 2001, Theorem 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 $\mathrm{d}_X(x,y) \leq \mathrm{d}_X(x,p) + \mathrm{d}_X(p,y) \leq \varepsilon\ \mathrm{sfs}(x) + \varepsilon\ \mathrm{sfs}(y)$, which is at most $\frac{2\varepsilon}{1-\varepsilon}\mathrm{sfs}(x)$ since sfs is 1-Lipschitz in the intrinsic metric. Thus, $\mathcal{V}_X(p)$ is included in the open geodesic ball $B_X(x, \frac{2\varepsilon}{1-\varepsilon}\ \mathrm{sfs}(x))$, where $\frac{2\varepsilon}{1-\varepsilon} \leq 1$ since $\varepsilon \leq \frac{1}{3}$. Therefore, $\gamma : S^1 \to \mathcal{V}_X(p)$ is null-homotopic in X, by Lemma 3.4. Let $\Gamma : S^1 \times I \to X$ be a homotopy between γ and a constant map in X. For any point $x \in H$, we have $\deg_x \gamma \neq 0$ since the loop γ winds around H. If x did not belong to $\Gamma(S^1 \times I)$, then Γ would be a homotopy between γ and a constant map in $\mathbb{R}^2 \setminus \{x\}$, thus by Corollary 2.1 we would have $\deg_x \gamma = 0$, thereby raising a contradiction. It follows that $\Gamma(S^1 \times I)$ contains all the points of H, which is therefore included in X.

As a consequence, the hole is caused by the presence of some sites of $L \setminus \{p\}$, whose geodesic Voronoi cells form H. Assume for simplicity that there is only one such site q, the case of several sites being similar. We then 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 ∂H beyond q. The 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 ∂H and hence to $\mathcal{V}_X(p)$. □

Since planar sets are aspherical [Cannon et al. 2002], their homotopy groups of dimension 2 or more are trivial. As a consequence, geodesic Voronoi cells have

the same homotopy groups as a point, up to isomorphism. Since in addition they are ANR's, they are homotopy equivalent to CW-complexes [James 1995, Chapter 26,§2]. Therefore, by Whitehead's theorem, they are homotopy equivalent to a point. Hence,

PROPOSITION 4.8. *Under the hypotheses of Theorem 4.3, every cell of $\mathcal{V}_X(L)$ is contractible.*

4.1.2 *Intersection of pairs of Voronoi cells.* We will now prove that the geodesic Voronoi cells have pairwise empty or contractible intersections. Given two sites $p, q \in L$ whose cells intersect, we first study the topological type of their union $\mathcal{V}_X(p) \cup \mathcal{V}_X(q)$, from which we can deduce the topological type of their intersection $\mathcal{V}_X(p) \cap \mathcal{V}_X(q)$.

LEMMA 4.9. *Under the hypotheses of Theorem 4.3, the union of any pair of intersecting cells of $\mathcal{V}_X(L)$ is simply connected.*

PROOF. Let $p, q \in L$ be such that $\mathcal{V}_X(p) \cap \mathcal{V}_X(q) \neq \emptyset$. The outline of the proof is the same as for Lemma 4.7. First, since by Lemma 4.6 $\mathcal{V}_X(p)$ and $\mathcal{V}_X(q)$ are path-connected, so is their union. Assume now for a contradiction that $\mathcal{V}_X(p) \cup \mathcal{V}_X(q)$ is not simply connected, and consider a hole H in $\mathcal{V}_X(p) \cup \mathcal{V}_X(q)$. Let $x \in \mathcal{V}_X(p) \cap \mathcal{V}_X(q)$. For any point $y \in \mathcal{V}_X(p)$, we have $\mathrm{d}_X(x,y) \leq \mathrm{d}_X(x,p) + \mathrm{d}_X(p,y) < \varepsilon\,\mathrm{sfs}(x) + \varepsilon\,\mathrm{sfs}(y)$, which is at most $\frac{2\varepsilon}{1-\varepsilon}\mathrm{sfs}(x)$ since sfs is 1-Lipschitz in the intrinsic metric. Idem for the points of $\mathcal{V}_X(q)$. As a consequence, $\mathcal{V}_X(p) \cup \mathcal{V}_X(q)$ is included in the open geodesic ball $B_X(x, \frac{2\varepsilon}{1-\varepsilon}\,\mathrm{sfs}(p))$, where $\frac{2\varepsilon}{1-\varepsilon} \leq 1$ since $\varepsilon \leq \frac{1}{3}$. Therefore, by the same argument as in the proof of Lemma 4.7, H is included in X.

It follows that the hole is caused by the presence of some sites of $L \setminus \{p, q\}$, whose geodesic Voronoi cells form H. Assume for simplicity that there is only one such site u, the case of several sites being similar. We then have $\mathcal{V}_X(u) = \overline{H}$, and $\partial H = \mathcal{V}_X(u) \cap (\mathcal{V}_X(p) \cup \mathcal{V}_X(q))$. Consider the Euclidean line l passing through u and perpendicular to (p, q). Let x, y be the first points of intersection of l with ∂H in each direction, starting from u. Since angles $\widehat{xup}$ and $\widehat{puy}$ sum up to $\pm\pi$, one of them (say $\widehat{xup}$) is obtuse. This implies that $\widehat{xuq}$ is also obtuse. Assume without loss of generality that $\mathrm{d}_X(x,p) \leq \mathrm{d}_X(x,q)$. Since the line segment $[u, x]$ is included in $\overline{H} \subseteq X$, we have $\mathrm{d}_X(x,u) = \mathrm{d}_E(x,u)$. Hence, using Pythagoras' theorem together with the fact that $\widehat{xup}$ is obtuse, we get:

$$\mathrm{d}_X(x,p)^2 \geq \mathrm{d}_E(x,p)^2 \geq \mathrm{d}_E(x,u)^2 + \mathrm{d}_E(u,p)^2 = \mathrm{d}_X(x,u)^2 + \mathrm{d}_E(u,p)^2 > \mathrm{d}_X(x,u)^2.$$

Now, x belongs to ∂H and hence to $\mathcal{V}_X(p) \cup \mathcal{V}_X(q)$. Moreover, we assumed without loss of generality that $\mathrm{d}_X(x,p) \leq \mathrm{d}_X(x,q)$, therefore x belongs to $\mathcal{V}_X(p)$, which contradicts the above equation. It follows that $\mathcal{V}_X(p) \cup \mathcal{V}_X(q)$ is simply connected, which concludes the proof of the lemma. □

Using the above result, we can now show that $\mathcal{V}_X(p) \cap \mathcal{V}_X(q)$ is contractible:

PROPOSITION 4.10. *Under the hypotheses of Theorem 4.3, the intersection of any pair of cells of $\mathcal{V}_X(L)$ is either empty or contractible.*

PROOF. Let $p, q \in L$ be such that $\mathcal{V}_X(p) \cap \mathcal{V}_X(q) \neq \emptyset$. Proposition 2.2 (i) tells us that every path-connected component of $\mathcal{V}_X(p) \cap \mathcal{V}_X(q)$ is simply connected,

since by Lemma 4.7 $\mathcal{V}_X(p)$ and $\mathcal{V}_X(q)$ are. Moreover, Proposition 2.2 (ii) tells us that $\mathcal{V}_X(p) \cap \mathcal{V}_X(q)$ is path-connected, since by Lemma 4.6 $\mathcal{V}_X(p)$ and $\mathcal{V}_X(q)$ are, and since by Lemma 4.9 their union is simply connected. It follows then from the asphericity of planar sets and from Whitehead's theorem that $\mathcal{V}_X(p) \cap \mathcal{V}_X(q)$ is contractible. □

4.1.3 *Intersection of arbitrary numbers of Voronoi cells.* The following result, combined with Theorem 4.5, concludes the proof of Theorem 4.3:

PROPOSITION 4.11. *Under the hypotheses of Theorem 4.3, 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 $k = 2$ were proved in Sections 4.1.1 and 4.1.2 respectively. 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 $\mathcal{V}_X(p_1) \cap \cdots \cap \mathcal{V}_X(p_{k+1}) \neq \emptyset$.

Observe 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 both simply connected. Hence, each path-connected component of their intersection $\mathcal{V}_X(p_1) \cap \cdots \cap \mathcal{V}_X(p_{k+1})$ is also simply connected, by Proposition 2.2 (i).

Consider now the union $\left(\bigcap_{i=1}^{k} \mathcal{V}_X(p_i)\right) \cup \mathcal{V}_X(p_{k+1})$, which is path-connected since both $\bigcap_{i=1}^{k} \mathcal{V}_X(p_i)$ and $\mathcal{V}_X(p_{k+1})$ are. Observe that the union can be rewritten as follows:

$$\left(\bigcap_{i=1}^{k} \mathcal{V}_X(p_i)\right) \cup \mathcal{V}_X(p_{k+1}) = \bigcap_{i=1}^{k} \left(\mathcal{V}_X(p_i) \cup \mathcal{V}_X(p_{k+1})\right).$$

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 $\bigcap_{i=1}^{k} \left(\mathcal{V}_X(p_i) \cup \mathcal{V}_X(p_{k+1})\right)$, by Proposition 2.2 (i). It follows then from Proposition 2.2 (ii) that the intersection $\mathcal{V}_X(p_1) \cap \cdots \cap \mathcal{V}_X(p_{k+1})$ is path-connected, since 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.

Thus, $\mathcal{V}_X(p_1) \cap \cdots \mathcal{V}_X(p_{k+1})$ is simply connected, and it follows from the asphericity of planar sets and from Whitehead's theorem that $\mathcal{V}_X(p_1) \cap \cdots \mathcal{V}_X(p_{k+1})$ is contractible. □

4.2 Geodesic witness complexes

Witness complexes in the intrinsic metric are defined in the same way as in the Euclidean metric:

DEFINITION 4.12. *Given a subset X of $\mathbb{R}^2$, and two subsets W, L of X such that L is finite,*

- *given a point $w \in W$ and a simplex $\sigma = [p_0, \cdots, p_l]$ with vertices in L, w is a* witness *of σ 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 all $q \in L \setminus \{p_0, \cdots, p_l\}$;*
- *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.*

Observe that a point $w \in W$ may only witness simplices whose vertices lie in the same path-connected component of X as 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 is called the set of witnesses, while L is referred to as the set of landmarks.

As in the Euclidean case, there exists a stronger notion of witness complex, where each witness is required to be equidistant to the vertices of the simplex σ. In this case, σ is a Delaunay simplex, and therefore the strong witness complex is included in the Delaunay triangulation. In his seminal work [de Silva 2008], de Silva shows that the weak witness complex is also included in the Delaunay triangulation, in the Euclidean metric. Below we give an equivalent of this result in the intrinsic metric – see Theorem 4.14. The proof uses the same kind of machinery as in [Attali et al. 2007], and it relies on the following fact:

LEMMA 4.13. *Let X be a Lipschitz domain in the plane, and L a geodesic εsfs-sample of X, for some $\varepsilon \leq 1$. Let x be a point of X, and p its $(k+1)$th nearest point of L in the intrinsic metric. If x and p lie in the same path-connected component of X, then $\mathrm{d}_X(x,p) < \left(\frac{3+\varepsilon}{1-\varepsilon}\right)^k \varepsilon\ \mathrm{sfs}(x)$. Else, $\mathrm{d}_X(x,p) = +\infty$.*

PROOF. The proof is by induction on k. We call X_x the path-connected component of X that contains x.

- Case $k = 0$: by definition, p is a nearest neighbor of x in L for the geodesic distance. Since L is a geodesic εsfs-sample of X, we have $\mathrm{d}_X(x,p) < \varepsilon\ \mathrm{sfs}(x) = \left(\frac{3+\varepsilon}{1-\varepsilon}\right)^k \varepsilon\ \mathrm{sfs}(x)$.

- General case: assume that the result holds up to some $k \geq 0$. Let $p_0, \cdots, p_{k+1}$ denote the $k+2$ points of L closest to x in the intrinsic metric, ordered according to their geodesic distances to x. If $p_{k+1} \notin X_x$, then we have $\mathrm{d}_X(x,p_{k+1}) = +\infty$, which proves the result for $k+1$. Assume now that $p_{k+1} \in X_x$. This implies that all the p_i also belong to X_x, since their geodesic distances to x are bounded by $\mathrm{d}_X(x,p_{k+1}) < +\infty$. By the induction hypothesis, we have $\mathrm{d}_X(x,p_0) \leq \cdots \leq \mathrm{d}_X(x,p_k) < \left(\frac{3+\varepsilon}{1-\varepsilon}\right)^k \varepsilon\ \mathrm{sfs}(x)$. Since p_{k+1} lies in X_x, the latter is not covered by $\mathcal{V}_X(p_0) \cup \cdots \cup \mathcal{V}_X(p_k)$. Therefore, there is a point $p \in L \setminus \{p_0, \cdots, p_k\}$ such that $\mathcal{V}_X(p)$ intersects the geodesic Voronoi cell of p_i, for some $i \in \{0, \cdots, k\}$. Note that p may or may not be p_{k+1} itself. Let $y \in \mathcal{V}_X(p_i) \cap \mathcal{V}_X(p)$. Since L is a geodesic εsfs-sample of X, we have $\mathrm{d}_X(y,p_i) = \mathrm{d}_X(y,p) < \varepsilon\ \mathrm{sfs}(y)$. Thus, by the triangle inequality and the induction hypothesis, we get:

$$\mathrm{d}_X(x,y) \leq \mathrm{d}_X(x,p_i) + \mathrm{d}_X(p_i,y) < \left(\frac{3+\varepsilon}{1-\varepsilon}\right)^k \varepsilon\ \mathrm{sfs}(x) + \varepsilon\ \mathrm{sfs}(y).$$

Since sfs is 1-Lipschitz in the intrinsic metric, we have $\mathrm{sfs}(y) \leq \mathrm{sfs}(x) + \mathrm{d}_X(x,y)$, which, by the above equation, is at most $\left(1 + \varepsilon\left(\frac{3+\varepsilon}{1-\varepsilon}\right)^k\right)\mathrm{sfs}(x) + \varepsilon\ \mathrm{sfs}(y)$. It follows that $\mathrm{sfs}(y) \leq \frac{(1-\varepsilon)^k + \varepsilon(3+\varepsilon)^k}{(1-\varepsilon)^{k+1}}\ \mathrm{sfs}(x)$. Now, since $p \notin \{p_0, \cdots, p_k\}$, we have $\mathrm{d}_X(x,p_{k+1}) \leq \mathrm{d}_X(x,p)$, which by the triangle inequality is at most $\mathrm{d}_X(x,p_i) + 2\mathrm{d}_X(y,p_i)$. By the induction hypothesis, $\mathrm{d}_X(x,p_i)$ is bounded by $\left(\frac{3+\varepsilon}{1-\varepsilon}\right)^k \varepsilon\ \mathrm{sfs}(x)$,

while our previous computations show that $2\mathrm{d}_X(y,p_i) < 2\varepsilon\frac{(1-\varepsilon)^k+\varepsilon(3-\varepsilon)^k}{(1-\varepsilon)^{k+1}}\ \mathrm{sfs}(x)$. In the end, we obtain:

$$\mathrm{d}_X(x,p_{k+1}) < \varepsilon\left(\left(\frac{3+\varepsilon}{1-\varepsilon}\right)^k + 2\frac{(1-\varepsilon)^k+\varepsilon(3+\varepsilon)^k}{(1-\varepsilon)^{k+1}}\right)\mathrm{sfs}(x) \leq \varepsilon\left(\frac{3+\varepsilon}{1-\varepsilon}\right)^{k+1}\mathrm{sfs}(x),$$

thus proving the result for $k+1$. □

In the special case where the point cloud L is a geodesic ε-sample of X, with a uniform bound ε on its density, the upper bound on the geodesic distance between x and its kth nearest point of L drops down to $(1+2k)\varepsilon$, by the same proof. It is worth pointing out the influence of the sampling regularity on the upper bound, which becomes exponential in k when the sampling is non-uniform, whereas it remains linear in k when the sampling is uniform. While it is clear that the linear bound in the uniform sampling case is tight, it is still unknown at this time whether the exponential bound in the non-uniform sampling case is tight or not.

THEOREM 4.14. *Let X be a Lipschitz domain in the plane, and L a geodesic ε*sfs*-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 $W \subseteq X$.*

PROOF. The proof is by induction on k. There will be in fact two inductions, therefore we call this one Ik, for clarity.

- Case $k = 0$: every point of L is a vertex of $\mathcal{D}_X(L)$, whether it is witnessed by a point of W or not.

- General case of Ik: assume that the result holds up to some $k \geq 0$. Assume further that $\varepsilon \leq \frac{1}{4^{k+2}}$. Let $\sigma = [p_0,\cdots,p_{k+1}]$ be a simplex of $\mathcal{C}_X^W(L)$, and let $w_0 \in W$ be a witness of σ. Consider without loss of generality that the p_i are ordered such that $\mathrm{d}_X(w_0,p_0) \geq \cdots \geq \mathrm{d}_X(w_0,p_{k+1})$. Then, the closed geodesic ball $B_0 = \overline{B_X}(w_0, \mathrm{d}_X(w_0,p_0))$ contains the p_i and no other point of L. Moreover, p_0 belongs to ∂B_0. We will prove by induction that B_0 can be shrunk to some closed geodesic ball B_{k+1} such that all the p_i lie on ∂B_{k+1}, while B_{k+1} still contains no other point of L. The center of B_{k+1} will then be equidistant to all the vertices of σ, and the latter will therefore be proved to be in $\mathcal{D}_X(L)$. The induction, named Ir for clarity, states that there is a closed geodesic ball B_r that contains the p_i and no other point of L, and such that $p_0,\cdots,p_r$ lie on ∂B_r.

• Case $r = 0$: initially, we have $p_0 \in \partial B_0$, and B_0 contains the p_i and no other point of L.

• General case of Ir ($0 \leq r < k$): assume that we have found a closed geodesic ball B_r that satisfies the requirements. In particular, we have $p_0,\cdots,p_r \in \partial B_r$. This means that the center w_r of B_r belongs to $\mathcal{V}'_X(p_0)\cap\cdots\cap\mathcal{V}'_X(p_r)$, where $\mathcal{V}'_X(p_i)$ denotes the cell of p_i ($i \leq r$) in the geodesic Voronoi diagram of $L\setminus\{p_{r+1},\cdots,p_{k+1}\}$. Moreover, since $[p_0,\cdots,p_{k+1}]$ belongs to $\mathcal{C}_X^W(L)$, so does its subsimplex $[p_0,\cdots,p_r]$, which therefore belongs also to $\mathcal{D}_X(L)$, by the induction hypothesis of Ik. Hence, $\mathcal{V}_X(p_0)\cap\cdots\cap\mathcal{V}_X(p_r)$ is not empty. Let $\tilde{w}_r \in \mathcal{V}_X(p_0)\cap\cdots\cap\mathcal{V}_X(p_r)$. Since the cell of any p_i in $\mathcal{V}_X(L\setminus\{p_{r+1},\cdots,p_{k+1}\})$ contains the cell of p_i in $\mathcal{V}_X(L)$, $\tilde{w}_r$ also belongs to $\mathcal{V}'_X(p_0)\cap\cdots\cap\mathcal{V}'_X(p_r)$.

We claim that $\mathcal{V}'_X(p_0)\cap\cdots\cap\mathcal{V}'_X(p_r)$ is path-connected. Indeed, for any point $x \in X$, the geodesic distance from x to L is finite, because L is a geodesic εsfs-

sample of X. And since w_0 witnesses $[p_0, \cdots, p_{k+1}]$, all the p_i lie in the same path-connected component of X as w_0, therefore the geodesic distance between x and $L \setminus \{p_{r+1}, \cdots, p_{k+1}\}$ is still finite, and by Lemma 4.13 it is bounded by $\left(\frac{3+\varepsilon}{1-\varepsilon}\right)^{k+1} \varepsilon \operatorname{sfs}(x)$. This quantity is less than $4^{k+1}\varepsilon \operatorname{sfs}(x) \leq \frac{1}{4}\operatorname{sfs}(x)$, since by the induction hypothesis of Ik we have $\varepsilon \leq \frac{1}{4^{k+2}} < \frac{1}{5}$. Hence, $L \setminus \{p_{r+1}, \cdots, p_{k+1}\}$ is a geodesic ε'sfs-sample of X, for some $\varepsilon' \leq \frac{1}{3}$. As a consequence, $\mathcal{V}'_X(p_0) \cap \cdots \cap \mathcal{V}'_X(p_r)$ is path-connected, by Proposition 4.11.

Since w_r and $\tilde{w}_r$ both belong to $\mathcal{V}'_X(p_0) \cap \cdots \cap \mathcal{V}'_X(p_r)$, which is path-connected, there exists a path $\gamma : I \to \mathcal{V}'_X(p_0) \cap \cdots \cap \mathcal{V}'_X(p_r)$ such that $\gamma(0) = w_r$ and $\gamma(1) = \tilde{w}_r$. For all $s \in I$, $\gamma(s)$ is equidistant to $p_0, \cdots, p_r$, and closer to these points than to any other point of $L \setminus \{p_{r+1}, \cdots, p_{k+1}\}$, in the intrinsic metric. Moreover, for all $j = r+1, \cdots, k+1$, the map $f_j : s \mapsto \mathrm{d}_X(\gamma(s), p_0) - \mathrm{d}_X(\gamma(s), p_j)$ is continuous, and we have $f_j(0) = \mathrm{d}_X(w_r, p_0) - \mathrm{d}_X(w_r, p_j) \geq 0$ since B_r contains p_j and has p_0 on its boundary, whereas $f_j(1) = \mathrm{d}_X(\tilde{w}_r, p_0) - \mathrm{d}_X(\tilde{w}_r, p_j) \leq 0$ since $\tilde{w}_r$ is a witness of $[p_0, \cdots, p_r]$. Thus, $f_j(s) = 0$ for at least one value $s \in I$. Let s_j be the smallest such s.

Consider now $\tilde{j} = \operatorname{argmin}_{j=r+1,\cdots,k+1} s_j$, and assume without loss of generality that $\tilde{j} = r+1$. We then have $f_{r+1}(s_{r+1}) = 0$ and $f_j(s_{r+1}) \geq 0$ for all $j = r+2, \cdots, k+1$. This means that the point $w_{r+1} = \gamma(s_{r+1})$ is equidistant to $p_0, \cdots, p_{r+1}$, and farther from these points than from $p_{r+2}, \cdots, p_{k+1}$. In addition, w_{r+1} is closer to $p_0, \cdots, p_{r+1}$ than to any other point of $L \setminus \{p_{r+2}, \cdots, p_{k+1}\}$, since $w_{r+1} \in \gamma(I) \subseteq \mathcal{V}'_X(p_0)$. It follows that the closed geodesic ball $B_{r+1} = \overline{B_X}(w_{r+1}, \mathrm{d}_X(w_{r+1}, p_0))$ contains $p_0, \cdots, p_{k+1}$ and no other point of L, and that $p_0, \cdots, p_{r+1}$ lie on ∂B_{r+1}. This concludes the induction Ir, and hereby also the induction Ik. □

Note that, for the conclusion of Theorem 4.14 to hold, it is mandatory to make an assumption on the density of the landmarks set L, since otherwise some boundary effects could occur. As an example, take for X an annulus and for L a set of three landmarks evenly distributed around the hole of the annulus: $\mathcal{D}_X(L)$ is then reduced to the boundary of the triangle formed by the three landmarks, whereas since L has only three points, the triangle is witnessed and therefore it belongs to $\mathcal{C}^W_X(L)$.

Our next result (Theorem 4.17) is an analog of Theorem 3.2 of [Guibas and Oudot 2007]. It involves a relaxed version of the witness complex, defined as follows:

Definition 4.15. *Given a subset X of $\mathbb{R}^2$, two subsets W, L of X such that L is finite, and an integer $\nu \geq 0$, a simplex σ with vertices in L is* ν-witnessed *by $w \in W$ if the vertices of σ belong to the path-connected component of X that contains w and to the $\nu+1$ landmarks closest to w in the intrinsic metric. The* geodesic ν-witness complex *of L relative to W, or $\mathcal{C}^W_{X,\nu}(L)$ for short, is the maximum abstract simplicial complex made of ν-witnessed simplices. Its dimension is at most ν.*

Theorem 4.17 assumes that L is a $\frac{\varepsilon}{1+\varepsilon}$sfs-sparse sample, which means by Definition 4.1 that every pair of landmarks $p \neq q$ must satisfy $\mathrm{d}_X(p,q) \geq \frac{\varepsilon}{1+\varepsilon} \min\{\operatorname{sfs}(p), \operatorname{sfs}(q)\}$. The bound on ε depends on the *doubling dimension* of (X, d_X), defined as the smallest integer d such that every open (resp. closed) geodesic ball can be covered by

a union of 2^d open (resp. closed) geodesic balls of half its radius. The doubling dimension measures the shape complexity of X, and it can be arbitrarily large. As an example, take for X a comb-shaped domain made of a rectangle of dimensions 1×2, to which are glued k branches of length 1 and width $\frac{2}{2k-1}$ as shown in Figure 2 (left). The geodesic distance from any point of X to the center point p is at most 2, so that X is covered by the closed geodesic ball $B_X(p,2)$. Consider now the closed geodesic balls $B_X(q_i,1)$, $1\leq i\leq k$, where points q_i are located at the tips of the k branches of X. Every ball $B_X(q_i,1)$ is included in the branch of q_i, therefore the balls $B_X(q_i,1)$ are pairwise disjoint. Thus, at least k closed geodesic balls of radius 1 can be packed inside a closed geodesic ball of radius 2, which implies that the doubling dimension of (X,d_X) is at least $\log_2 k$, according to the following result by Kolmogorov and Tikhomirov:

Lemma 4.16 from [Kolmogorov and Tikhomirov 1961]. *Given any subset Y of X, and any real number $r>0$, the maximum number of pairwise-disjoint open (resp. closed) geodesic balls of radius r that can be packed inside Y is at most the minimum number of open (resp. closed) geodesic balls of radius r that are necessary to cover Y.*

Theorem 4.17. *Let X be a Lipschitz domain in the plane, of doubling dimension d. Let W be a geodesic δ*sfs*-sample of X, and L a geodesic ε*sfs*-sample of X that is also $\frac{\varepsilon}{1+\varepsilon}$*sfs*-sparse. If $\varepsilon+2\delta<1$, then, for any integer $\nu\geq 2^{ld}-1$, where $l=\left\lceil \log_2 \frac{3+\varepsilon+2\delta/\varepsilon}{1-\varepsilon-2\delta}\right\rceil$, $\mathcal{D}_X(L)$ is included in $\mathcal{C}^W_{X,\nu}(L)$.*

Proof. Let σ be a simplex of $\mathcal{D}_X(L)$, and let c be a point of its dual geodesic Voronoi cell $\mathrm{V}_X(\sigma)$. Since W is a geodesic δsfs-sample of X, there is a point $w\in W$ at geodesic distance at most δ $\mathrm{sfs}(c)$ from c. Moreover, since L is a geodesic εsfs-sample of X, every vertex v of σ is at geodesic distance less than ε $\mathrm{sfs}(c)$ from c. It follows that $\mathrm{d}_X(w,v)<(\delta+\varepsilon)$ $\mathrm{sfs}(c)$. Now, since L is $\frac{\varepsilon}{1+\varepsilon}$sfs-sparse, every two landmarks v,v' located in the open geodesic ball $B_X(w,(\varepsilon+\delta)\ \mathrm{sfs}(c))$ satisfy: $\mathrm{d}_X(v,v')\geq \frac{\varepsilon}{1+\varepsilon}$ $\mathrm{sfs}(v)$, assuming without loss of generality that $\mathrm{sfs}(v)\leq \mathrm{sfs}(v')$. Since sfs is 1-Lipschitz in the intrinsic metric (Lemma 3.3), we have: $\mathrm{sfs}(v)\geq \mathrm{sfs}(c)-\mathrm{d}_X(v,c)\geq \mathrm{sfs}(c)-(\varepsilon+2\delta)\ \mathrm{sfs}(c)=(1-\varepsilon-2\delta)\ \mathrm{sfs}(c)$. Thus, the landmarks inside $B_X(w,(\varepsilon+\delta)\ \mathrm{sfs}(c))$ are at least $\frac{\varepsilon(1-\varepsilon-2\delta)}{1+\varepsilon}$ $\mathrm{sfs}(c)$ away from one another in the intrinsic metric. Hence, they are centers of pairwise-disjoint open geodesic balls of same radius $\frac{\varepsilon(1-\varepsilon-2\delta)}{2(1+\varepsilon)}$ $\mathrm{sfs}(c)$, packed inside the open geodesic ball of center c and radius $(\varepsilon+\delta+\frac{\varepsilon(1-\varepsilon-2\delta)}{2(1+\varepsilon)})\mathrm{sfs}(c)=\frac{3\varepsilon+\varepsilon^2+2\delta}{2(1+\varepsilon)}$ $\mathrm{sfs}(c)$. According to Lemma 4.16, there are at most 2^{ld} such balls, where $l=\left\lceil \log_2 \frac{3\varepsilon+\varepsilon^2+2\delta}{\varepsilon(1-\varepsilon-2\delta)}\right\rceil=\left\lceil \log_2 \frac{3+\varepsilon+2\delta/\varepsilon}{1-\varepsilon-2\delta}\right\rceil$. It follows that σ is ν-witnessed by w whenever $\nu\geq 2^{ld}-1$. Since this is true for every simplex σ of $\mathcal{D}_X(L)$, the latter is included in $\mathcal{C}^W_{X,\nu}(L)$ whenever $\nu\geq 2^{ld}-1$. □

It follows from Theorems 4.14 and 4.17 that, whenever L and W are dense enough, $\mathcal{D}_X(L)$ is sandwiched between $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$, provided that ν is chosen sufficiently large. The simulation results presented in Section 7 suggest that even small values of ν are sufficient in practice. Note however that, in some cases, neither $\mathcal{C}^W_X(L)$ nor $\mathcal{C}^W_{X,\nu}(L)$ coincides exactly with $\mathcal{D}_X(L)$. This fact, already observed in [Guibas and Oudot 2007] in a Euclidean setting, motivates the use of

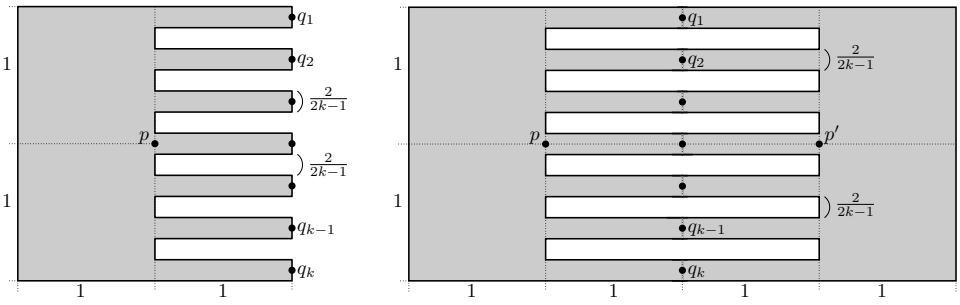


Fig. 2. Left: a Lipschitz domain with doubling dimension at least k. Right: the size of $\{q_1, \cdots, q_k\}$ is $\frac{k}{2}$ times that of $\{p, p'\}$, although both point sets are geodesic sfs-samples of the domain and $q_1, \cdots, q_k$ are centers of pairwise-disjoint geodesic balls of same radius $\frac{\text{sfs}}{2}$.

persistent homology between $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$ for computing the homology of $\mathcal{D}_X(L)$ without building the latter complex explicitly.

5. UNIONS OF GEODESIC BALLS AND THEIR NERVES

Given a Lipschitz domain X in the plane, and two finite subsets $W \subseteq L \subset X$, we saw in the previous section (Theorems 4.14 and 4.17) that the following sequence of inclusions holds provided that W, L are dense with respect to the systolic feature size of X and that the relaxation parameter ν is large enough:

$$\mathcal{C}^W_X(L) \subseteq \mathcal{D}_X(L) \subseteq \mathcal{C}^W_{X,\nu}(L).$$

In the conference version of this paper [Gao et al. 2008] we showed how the above sequence of inclusions can be used to infer the homology of the domain X. Specifically, considering singular homology with coefficients in an arbitrary field, we showed that the inclusion $\mathcal{C}^W_X(L) \hookrightarrow \mathcal{D}_X(L)$ induces surjective homomorphisms at homology level, while the inclusion $\mathcal{D}_X(L) \hookrightarrow \mathcal{C}^W_{X,\nu}(L)$ induces injective homomorphisms. Intuitively, this means that the homology classes of cycles of $\mathcal{D}_X(L)$ already exist in $\mathcal{C}^W_X(L)$ and do not die in $\mathcal{C}^W_{X,\nu}(L)$. As a result, the inclusion $\mathcal{C}^W_X(L) \hookrightarrow \mathcal{C}^W_{X,\nu}(L)$ encodes the same homological information as $\mathcal{D}_X(L)$, and therefore as X itself, by Theorem 4.3. More formally, for all $k \in \mathbb{N}$, the rank of the linear map $H_k(\mathcal{C}^W_X(L)) \to H_k(\mathcal{C}^W_{X,\nu}(L))$ induced by the inclusion between the witness complexes is equal to the kth Betti number of $\mathcal{D}_X(L)$, which by Theorem 4.3 coincides with the kth Betti number of X.

In this section we want to proceed further and study the ranks of the linear maps induced at homology level by inclusions of type $\mathcal{C}^W_{X,\nu}(L) \hookrightarrow \mathcal{C}^W_{X,\nu'}(L)$, where $0 \leq \nu \leq \nu'$ are arbitrary values of the relaxation parameter. Moreover, we want to study other families of simplicial complexes that are also easy to build in practice. In particular, we are interested in *Rips complexes* in the geodesic distance:

DEFINITION 5.1. *Given a finite point set $L \subset X$ and a real parameter $\alpha > 0$, the* (Vietoris-)Rips complex $\mathcal{R}_\alpha(L)$ *is the abstract simplicial complex of vertex set L whose simplices correspond to non-empty subsets of L of diameter less than α in the geodesic distance* d_X.

Our analysis uses the approach of [Chazal and Oudot 2008], which we will now describe briefly and adapt to our context. The main idea of [Chazal and Oudot 2008] is to relate Rips and witness complexes to the so-called *Čech complexes*, defined below:

DEFINITION 5.2. *Given a finite point set $L \subset X$ and a real parameter $\alpha > 0$, the* Čech complex $\mathcal{C}_\alpha(L)$ *is the nerve of the union of open geodesic balls of same radius α about the points of L.*

Since Čech complexes can be potentially difficult to compute, they are not meant to be constructed in practice. However, they can be used as an intermediate algebraic construction for the analysis of the topological structures of Rips or witness complexes. Indeed, on the one hand, the topology of the Čech complex is tied to the one of its dual union of balls via the Nerve Theorem 4.5, provided that the balls form a good cover of the union, as per Definition 4.4. On the other hand, as proved *e.g.* in [Chazal and Oudot 2008], the one-parameter family of Čech complexes is interleaved with the one-parameter family of Rips complexes in the following sense:

$$\forall \alpha > 0,\ \mathcal{C}_{\frac{\alpha}{2}}(L) \subseteq \mathcal{R}_\alpha(L) \subseteq \mathcal{C}_\alpha(L). \tag{3}$$

The analysis of [Chazal and Oudot 2008] uses the above interleaving property to derive relations between the ranks of the linear maps induced at homology level by inclusions between Rips complexes and the ranks of linear maps induced by inclusions between Čech complexes. More precisely, from Eq. (3) one deduces the following sequence of inclusions for all $\beta \geq 2\alpha$:

$$\mathcal{C}_{\alpha/2}(L) \subseteq \mathcal{R}_\alpha(L) \subseteq \mathcal{C}_\alpha(L) \subseteq \mathcal{R}_\beta(L) \subseteq \mathcal{C}_\beta(L). \tag{4}$$

By simple algebraic arguments, this sequence of inclusions implies the following inequalities between the ranks of the homomorphisms induced at homology level by inclusions: $\forall \beta \geq 2\alpha$, $\forall k \in \mathbb{N}$,

$$\text{rank } H_k(\mathcal{C}_{\alpha/2}(L)) \to H_k(\mathcal{C}_\beta(L)) \leq \text{rank } H_k(\mathcal{R}_\alpha(L)) \to H_k(\mathcal{R}_\beta(L)) \leq \dim H_k(\mathcal{C}_\alpha(L)). \tag{5}$$

These inequalities provide upper and lower bounds on the ranks of the linear maps induced at homology level by inclusions of type $\mathcal{R}_\alpha(L) \hookrightarrow \mathcal{R}_\beta(L)$. The rest of the analysis of [Chazal and Oudot 2008] consists in working out sufficient conditions under which the upper and lower bounds coincide with the Betti numbers of X. To do so, it relates the one-parameter family of Čech complexes to its dual one-parameter family of unions of balls. Recall indeed from Definition 5.2 that $\mathcal{C}_\alpha(L)$ is the nerve of the union of open geodesic balls of same radius α about the points of L. Let us call L^α this union, and $\{L^\alpha\}$ the associated collection of open geodesic balls. The analysis of [Chazal and Oudot 2008] provides the following key result, which can be viewed as a persistent variant of the Nerve Theorem 4.5:

LEMMA 5.3. *For any parameters $\alpha \leq \beta$, if $\{L^\alpha\}$ forms a good open cover of L^α and $\{L^\beta\}$ forms a good open cover of L^β, then there exist homotopy equivalences $L^\alpha \to \mathcal{C}_\alpha(L)$ and $L^\beta \to \mathcal{C}_\beta(L)$ that commute with the canonical inclusions $L^\alpha \hookrightarrow L^\beta$ and $\mathcal{C}_\alpha(L) \hookrightarrow \mathcal{C}_\beta(L)$ at homology level.*

In other words, the inclusions $L^\alpha \hookrightarrow L^\beta$ and $\mathcal{C}_\alpha(L) \hookrightarrow \mathcal{C}_\beta(L)$ carry the same homological information, that is: for all $k \in \mathbb{N}$, the linear maps $H_k(L^\alpha) \to H_k(L^\beta)$

and $H_k(\mathcal{C}_\alpha(L)) \to H_k(\mathcal{C}_\beta(L))$ induced by inclusions have the same rank.

Now, if we assume that L is a geodesic ε-sample of some length space X, then L^α coincides with X as soon as $\alpha > \varepsilon$, and for all $\beta \geq \alpha > \varepsilon$, the canonical inclusion $L^\alpha \hookrightarrow L^\beta$ is the identity of X, which implies that the rank of $H_k(L^\alpha) \to H_k(L^\beta)$ coincides with the kth Betti number of X. Combined with Lemma 5.3, this fact implies that, for all $\alpha > 2\varepsilon$ and $\beta \geq 2\alpha$ such that $\{L^{\alpha/2}\}$, $\{L^\alpha\}$, $\{L^\beta\}$ form good open covers of $L^{\alpha/2}$, L^α, L^β respectively, the rank of $H_k(\mathcal{C}_{\alpha/2}(L)) \to H_k(\mathcal{C}_\beta(L))$ and the dimension of $H_k(\mathcal{C}_\alpha(L))$ coincide with the kth Betti number of X. Thus, the upper and lower bounds in Eq. (5) coincide with the kth Betti number of X, which implies the following:

THEOREM 5.4. *Let X be a length space that admits a finite geodesic ε-sample L. Then, for all $k \in \mathbb{N}$, for all $\alpha > 2\varepsilon$ and $\beta \geq 2\alpha$ such that $\{L^{\alpha/2}\}$, $\{L^\alpha\}$, $\{L^\beta\}$ form good open covers of $L^{\alpha/2}$, L^α, L^β respectively, the rank of the homomorphism $H_k(\mathcal{R}_\alpha(L)) \to H_k(\mathcal{R}_\beta(L))$ induced by inclusion coincides with the kth Betti number of X.*

In [Chazal and Oudot 2008], the analysis takes place in Euclidean space $\mathbb{R}^d$, where balls are convex and their intersections contractible (if not empty). In [Chazal et al. 2009], the analysis is extended to the case of compact Riemannian manifolds, with or without boundary, where geodesic balls are convex and their intersections contractible (if not empty) up to the so-called convexity radius of the manifold. Thus, the assumption of having good covers in Theorem 5.4 holds as long as β is smaller than the convexity radius. In the present context, the domain X is not a Riemannian manifold since its boundary can be non-smooth. Yet, the above properties of geodesic balls still hold provided that the radii are not more than a fraction of the systolic feature size of X:

LEMMA 5.5. *If X is a Lipschitz planar domain, then any finite collection of open geodesic balls of radii at most $\frac{1}{3}\mathrm{sfs}(X)$ forms a good open cover of its union in X.*

Combined with Theorem 5.4, this result implies that, if L is a geodesic ε-sample of a Lipschitz planar domain X, for some $\varepsilon < \frac{1}{12}\mathrm{sfs}(X)$, then, for any choice of parameters $\alpha \in (2\varepsilon, \frac{1}{6}\mathrm{sfs}(X)]$ and $\beta \in [2\alpha, \frac{1}{3}\mathrm{sfs}(X)]$, the Betti numbers of X can be obtained as the ranks of the homomorphisms induced at homology level by the inclusion $\mathcal{R}_\alpha(L) \hookrightarrow \mathcal{R}_\beta(L)$.

Lemma 5.5 is the main new result of this section. Its proof turns out to be rather elaborate, and in fact it draws some interesting connections between the systolic feature size and the distance to the cut locus on the one hand (see Lemma 5.6 in Section 5.1), as well as between Lipschitz planar domains and a class of length spaces called *Alexandrov spaces* on the other hand (see Theorem 5.10 in Section 5.2). The proof is detailed in Sections 5.1 and 5.2, while Section 5.3 adapts the above analysis to the case of witness complexes.

5.1 Systolic feature size and cut locus

A noticeable feature of the systolic feature size is its close relationship with the so-called cut-locus. For any given path $\gamma : I \to X$, we call *support* of γ the set $\gamma(I)$. If γ is a shortest path between $x = \gamma(0)$ and $y = \gamma(1)$, then $\gamma(I)$ is called a *shortest path support* between x and y. Note that different paths may have identical

supports. In particular, a shortest path support may be shared by shortest paths as well as non-shortest paths (think of the latter as moving back and forth along the support). 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 different shortest paths supports to x in X. In other words, a point $y \in X$ belongs to $\mathrm{CL}_X(x)$ iff there exist two paths $\gamma, \gamma' : I \to X$ such that $\gamma(0) = \gamma'(0) = x$, $\gamma(1) = \gamma'(1) = y$, $|\gamma| = |\gamma'| = \mathrm{d}_X(x, y)$, and $\gamma(I) \neq \gamma'(I)$. The geodesic distance from x to its cut-locus is denoted by $\mathrm{d}_X(x, \mathrm{CL}_X(x))$.

LEMMA 5.6. *If X is a Lipschitz domain in the plane, then $\forall x \in X$,* $\mathrm{sfs}(x) = \mathrm{d}_X(x, \mathrm{CL}_X(x))$.

PROOF. We first show that $\mathrm{sfs}(x) \geq \mathrm{d}_X(x, \mathrm{CL}_X(x))$. This is clearly true if the path-connected component X_i of X that contains x is simply connected, since in such a case we have $\mathrm{sfs}(x) = +\infty$. Assume now that X_i is not simply connected, and let $\gamma : (S^1, 1) \to (X, 1)$ be a non null-homotopic loop through x in X, of length $2\mathrm{sfs}(x) < +\infty$. Such a loop exists, by Lemma 3.2. Moreover, according to [Burago et al. 2001, Proposition 2.5.9], we can assume without loss of generality that γ is parameterized with constant speed, that is: $\forall s \in I$, $|\gamma_{[0,s]}| = s|\gamma|$. We then have $|\gamma_{[0,1/2]}| = |\gamma_{[1/2,1]}| = \frac{1}{2}|\gamma| = \mathrm{sfs}(x)$. Call respectively γ' and γ'' the paths $\gamma_{[0,1/2]}$ and $\gamma_{[1/2,1]}$. These are two paths between x and $y = \gamma(1/2)$ in X, hence their lengths are at least $\mathrm{d}_X(x, y)$. We claim that $|\gamma'| = |\gamma''| = \mathrm{d}_X(x, y)$. Indeed, let ζ be a shortest path from x to y in X. Since γ is not null-homotopic in X, γ' and $\bar{\gamma}''$ are not homotopic relative to ∂I in X, and therefore $\gamma' \cdot \bar{\zeta}$ or $\bar{\gamma}'' \cdot \bar{\zeta}$ (say $\gamma' \cdot \bar{\zeta}$) is not null-homotopic in X. It follows that $|\gamma' \cdot \bar{\zeta}| \geq 2\mathrm{sfs}(x)$. Now, if $|\zeta| < |\gamma''|$, then we have $|\gamma' \cdot \bar{\zeta}| = |\gamma'| + |\zeta| < |\gamma'| + |\gamma''| = |\gamma| = 2\mathrm{sfs}(x)$, which raises a contradiction with the previous sentence. Therefore, $|\gamma'| = |\gamma''| = |\zeta| = \mathrm{d}_X(x, y)$. Finally, we claim that the supports of γ' and γ'' are distinct. Assume for a contradiction that $\gamma'(I) = \gamma''(I)$. Then, for all $s' \in [0, 1/2]$, there exists $s'' \in [1/2, 1]$ such that $\gamma(s') = \gamma(s'')$. This implies that $\mathrm{d}_X(x, \gamma(s')) = \mathrm{d}_X(x, \gamma(s''))$. But since γ' and $\bar{\gamma}''$ are shortest paths from x to y in X, we have $\mathrm{d}_X(x, \gamma(s')) = |\gamma_{[0,s']}|$ and $\mathrm{d}_X(x, \gamma(s'')) = |\gamma_{[s'',1]}|$. It follows that $s' = 1 - s''$, because γ is parameterized with constant speed. This means that $\gamma' = \bar{\gamma}''$, which implies that $\gamma = \gamma' \cdot \gamma''$ is null-homotopic in X, which contradicts our assumption. Thus, we have $\gamma'(I) \neq \gamma''(I)$, as well as $|\gamma'| = |\gamma''| = \mathrm{d}_X(x, y)$, which means that y belongs to $\mathrm{CL}_X(x)$. Therefore, $\mathrm{sfs}(x) = |\gamma'| = |\gamma''| = \mathrm{d}_X(x, y) \geq \mathrm{d}_X(x, \mathrm{CL}_X(x))$.

Let us now show that $\mathrm{sfs}(x) \leq \mathrm{d}_X(x, \mathrm{CL}_X(x))$. Assume for a contradiction that there is a point $y \in \mathrm{CL}_X(x)$ such that $\mathrm{d}_X(x, y) < \mathrm{sfs}(x)$. Point y has at least two shortest paths γ, γ' from x whose supports differ. Assume without loss of generality that γ, γ' are parameterized with constant speed. Then, for all $0 \leq s < s' \leq 1$, we have $\gamma(s) \neq \gamma(s')$, since otherwise the path $\gamma_{[0,s]} \cdot \gamma_{[s',1]}$ would connect x to y and be strictly shorter than γ, hereby contradicting the fact that the latter is a shortest path from x to y. Thus, γ is an injection from I to X. Given any points $u, v \in \gamma(I)$, with $\gamma^{-1}(u) \leq \gamma^{-1}(v)$, we call γ_{uv} the path $\gamma_{[\gamma^{-1}(u), \gamma^{-1}(v)]}$. By the same argument, γ' is also an injection from I to X, and we use the same notation for subpaths.

Since the supports of γ and γ' differ, we have $\gamma(I) \setminus \gamma'(I) \neq \emptyset$ or $\gamma'(I) \setminus \gamma(I) \neq \emptyset$ –

say $\gamma(I) \setminus \gamma'(I) \neq \emptyset$. Let γ_{uv} be a maximal subarc of γ satisfying $\gamma_{uv}(]0,1[) \cap \gamma'(I) = \emptyset$. Here, u and v are the two endpoints of γ_{uv}, and by maximality we have $u \neq v$ and $u, v \in \gamma(I) \cap \gamma'(I)$. Since γ_{uv} and γ'_{uv} are injective, and since their images in X have common endpoints but disjoint relative interiors, the path $\gamma'_{uv} \cdot \bar{\gamma}_{vu}$ is a simple loop, and therefore it divides the plane into two connected components, one of which (called C) is bounded, by the Jordan curve theorem. Moreover, we have $\partial C = (\gamma'_{uv} \cdot \bar{\gamma}_{vu})(I)$, and the degree of the loop with respect to any point of C is non-zero. Now, since γ, γ' are shortest paths from x to y, with $\mathrm{d}_X(x,y) < \mathrm{sfs}(x)$, the image of the loop $\gamma'_{uv} \cdot \bar{\gamma}_{vu}$ lies in the open geodesic ball $B_X(x, \mathrm{sfs}(x))$. Hence, by Lemma 3.4, the loop is null-homotopic in X, and since its degree with respect to the points of C is non-zero, any homotopy with a constant map in X passes through the points of C, which therefore belong to X. Thus, between points u and v, γ and γ' sandwich a region C that is included in X. We will show that there exist shortcuts to γ, γ' in C, hereby contradicting the fact that γ and γ' are shortest paths from x to y in X.

Consider the line segment $[u,v]$, and choose a positively-oriented orthonormal frame such that point u is at the origin, line (u,v) is vertical, and point v lies above u. Let λ_{uv} denote the path $s \mapsto (1-s)u + sv$.

- If $[u,v]$ is included in C, then the paths $\gamma_{xu} \cdot \lambda_{uv} \cdot \gamma_{vy}$ and $\gamma'_{xu} \cdot \lambda_{uv} \cdot \gamma'_{vy}$ connect x to y in X. And since $\gamma_{uv}(I)$ and $\gamma'_{uv}(I)$ differ, one of them at least (say $\gamma_{uv}(I)$) differs from $[u,v]$, which implies that $|\gamma_{uv}| > \mathrm{d}_E(u,v) = |\lambda_{uv}|$ and hence that $|\gamma| > |\gamma_{xu} \cdot \lambda_{uv} \cdot \gamma_{vy}|$, which contradicts the fact that γ is a shortest path from x to y in X.

- If now $[u,v]$ is not included in C, then there is a point $p \in [u,v]$ that does not belong to C. On the horizontal line passing through p, C lies on the right or on the left of (u,v), say on the right. Let c be a rightmost point of $\overline{C}$. We have $c \notin \{u,v\}$ because c lies on the right of line (u,v). Note that $c \in \partial C$, and assume without loss of generality that $c \in \gamma_{uv}(I)$, which implies that $c \notin \gamma'_{uv}(I)$ since $c \notin \{u,v\}$. Let α be the connected component of $\gamma_{uv}(I) \setminus (u,v)$ that contains c. Since γ_{uv} is a simple arc, α is a subarc of γ_{uv}, starting and ending on (u,v), and passing through c. Let l be the vertical line passing through c. Note that $\overline{C}$ does not intersect the right half-plane bounded by l. Nevertheless, other components of $\partial C \setminus (u,v)$ may touch l, including some subarcs of γ'_{uv}. However, by paring C infinitesimally in their vicinity, one can easily ensure that α is the only arc of ∂C that touches l. Hence, from now on, we assume without loss of generality that $l \cap \overline{C} \subseteq \alpha$. This implies that $\gamma'_{uv}(I)$ does not touch l, since $\alpha \subseteq \gamma_{uv}(]0,1[)$, which does not intersect $\gamma'_{uv}(I)$. Therefore, the rightmost vertical line l' touching $\gamma'_{uv}(I)$ lies on the left of l. Let $\delta > 0$ denote the Euclidean distance between l and l'.

Consider the open Euclidean ball $B_E(c,\delta)$. Since $c \in \overline{C}$, there exists a point c'' lying in $C \cap B_E(c,\delta)$. Since C is open in $\mathbb{R}^2$, we have $c'' \notin \partial C$. Let l'' be the vertical line passing through c''. Note that l'' is located on the right of l'. Let u'' and v'' be the first points of intersection of l'' with ∂C above and below c''. We have $[u'', v''] \subseteq \overline{C}$. Moreover, $u'' \neq v''$ because $c'' \notin \partial C$. In addition, u'' and v'' belong to $\gamma_{uv}(I)$, since they lie on l'' and hence on the right of l'. Finally, $[u'', v'']$ differs from $\gamma_{u''v''}(I)$ because $[u'', v'']$ passes through $c'' \notin \partial C$. As a result, the path $\lambda_{u''v''}$, defined by $s \mapsto (1-s)u'' + sv''$, is included in $\overline{C} \subseteq X$, it connects points u'', v''

of $\gamma_{uv}(I)$, and it is shorter than $\gamma_{u''v''}$. It follows that the path $\gamma_{xu''} \cdot \lambda_{u''v''} \cdot \gamma_{v''y}$ connects x to y in X, and is strictly shorter than γ, which contradicts the fact that γ is a shortest path from x to y in X. This shows that every point y inside the open geodesic ball $B_X(x, \mathrm{sfs}(x))$ has only one shortest path support to x. It follows that $\mathrm{sfs}(x) \leq \mathrm{d}_X(x, \mathrm{CL}_X(x))$, which concludes the proof of Lemma 5.6. □

The fact that the geodesic distance of a point $x \in X$ to its cut-locus is equal to half the length of the shortest non null-homotopic loop through x was already known in the case of planar domains with polygonal boundaries [Mitchell 1991]. Lemma 5.6 above extends this result to the case of planar domains with Lipschitz boundaries.

5.2 Lipschitz planar domains are Alexandrov spaces

The background material used in this section comes from Chapters 4 and 9 of [Burago et al. 2001], to which we refer the reader for further details.

We call *geodesic triangle* any collection of three distinct points $a, b, c \in X$ connected by three shortest paths supports $\tau_{ab}, \tau_{bc}, \tau_{ca}$ in X. Note that the three vertices alone may not define a geodesic triangle uniquely since there may be several different shortest paths supports connecting a same pair of vertices.

Definition 5.7. *Given a geodesic triangle of vertices $a, b, c \in X$, a* comparison triangle *is a triangle $(\bar{a}, \bar{b}, \bar{c})$ in the Euclidean plane such that $\mathrm{d}_E(\bar{a}, \bar{b}) = \mathrm{d}_X(a, b)$, $\mathrm{d}_E(\bar{b}, \bar{c}) = \mathrm{d}_X(b, c)$, and $\mathrm{d}_E(\bar{c}, \bar{a}) = \mathrm{d}_X(c, a)$.*

Although three distinct points in X may not define a unique geodesic triangle, they always define a unique comparison triangle up to an isometry of the Euclidean plane.

Definitions 5.8 and 5.9 below consider the *shapes* of small enough geodesic triangles as a criterion for a length space to have bounded curvature. This criterion is inspired from results in Riemannian geometry, where manifolds of negative curvature tend to have *skinny* triangles, whereas manifolds of positive curvature have rather *fat* triangles. Here, the skinniness of a geodesic triangle is measured with respect to a comparison triangle in the Euclidean plane.

Definition 5.8 Angle condition. *A geodesic triangle $(a, b, c, \tau_{ab}, \tau_{bc}, \tau_{ca})$ satisfies the* angle condition *if the angles formed by $\tau_{ab}, \tau_{bc}, \tau_{ca}$ at the vertices a, b, c are well-defined and at most the corresponding angles in a comparison triangle.*

In the above definition, by angle between two paths $\alpha, \beta : I \to X$ emanating from a same point $p = \alpha(0) = \beta(0)$ is meant the limit quantity $\lim_{s,t \to 0} \widetilde{\angle}(\alpha(s), p, \beta(t))$, if it exists, where $\widetilde{\angle}(\alpha(s), p, \beta(t))$ denotes the inner angle[8] at the vertex corresponding to p in a comparison triangle of $(p, \alpha(s), \beta(t))$. This limit may not always exist in general. Below we prove that, in the special case of Lipschitz planar domains, small enough geodesic triangles have concave edges (Claim 5.10.3) whose tangents at the vertices of the triangles are well-defined, which implies that angles between edges are also well-defined.

[8]This angle is defined uniquely because the comparison triangle is defined uniquely up to an isometry of the Euclidean plane.

DEFINITION 5.9. *A length space X is an* Alexandrov space with non-positive curvature *if around each point of X there is a neighborhood such that every geodesic triangle within this neighborhood satisfies the angle condition of Definition 5.8.*

Alexandrov spaces of non-positive curvature are sometimes called *CAT(0)-spaces* in the literature, where CAT stands for Cartan-Alexandrov-Toponogov, and where (0) indicates the upper bound on the curvature. Note also that curvature bounds are usually derived from distance conditions, not angle conditions. As proved in [Burago et al. 2001, Theorem. 4.3.5], distance and angle conditions are in fact equivalent.

The main result of this section is that Lipschitz planar domains are CAT(0)-spaces:

THEOREM 5.10. *Every Lipschitz domain X in the plane, endowed with the length structure inherited from $(\mathbb{R}^2, \mathrm{d}_E)$, is an Alexandrov space of non-positive curvature. More precisely, for any open geodesic ball $B \subset X$ of radius at most $\frac{1}{3}\mathrm{sfs}(X)$, and for any distinct points $a, b, c \in B$, the geodesic triangle formed by a, b, c and their (unique) shortest paths supports satisfies the angle condition of Definition 5.8.*

The proof of the theorem uses four intermediate results, stated as Claims 5.10.1 through 5.10.4 and proved on the fly.

Proof of Theorem 5.10. Observe first that, since the diameter of B is less than $\mathrm{sfs}(X)$, the shortest paths supports between a, b, c are defined uniquely, by Lemma 5.6. For more clarity, we call τ_{ab}, τ_{bc}, and τ_{ca} these paths supports — dashed in Figure 3 (left).

CLAIM 5.10.1. *The paths supports τ_{ab}, τ_{bc} and τ_{ca} are simple planar curves that pairwise intersect along connected subarcs incident to their common endpoints.*

PROOF. Since τ_{ab}, τ_{bc} and τ_{ca} are shortest paths supports, they have to be simple, since otherwise they could be shortened. Consider now τ_{ab} and τ_{bc}. These paths supports intersect at their common endpoint b. Assume that they have another point b' of intersection. Then, the arc of τ_{ab} that connects b to b' is a shortest path support between the two points in X. Idem for the arc of τ_{bc} that connects b to b'. Therefore, these two arcs coincide, by Lemma 5.6. It follows that τ_{ab} and τ_{bc} must intersect along a common subarc incident to their common endpoint b. The same is true for τ_{bc} and τ_{ca} on the one hand, and for τ_{ca} and τ_{ab} on the other hand. □

Let $\tau_{bb'}$ be the common subarc of τ_{ab} and τ_{bc}, $\tau_{cc'}$ the common subarc of τ_{bc} and τ_{ca}, and $\tau_{aa'}$ the common subarc of τ_{ca} and τ_{ab}. The shortest paths supports $\tau_{a'b'}$, $\tau_{b'c'}$ and $\tau_{c'a'}$ are then uniquely defined as subarcs of τ_{ab}, τ_{bc} and τ_{ca} respectively. Note that if $a' = b'$ or $a' = c'$ or $b' = c'$, then it must be the case that $a' = b' = c'$, by Claim 5.10.1.

CLAIM 5.10.2. *If $a' = b' = c'$, then the curve $\tau = \tau_{a'b'} \cup \tau_{b'c'} \cup \tau_{c'a'}$ is reduced to a point. Else, τ is a simple closed curve whose complement in $\mathbb{R}^2$ has two path-connected components, one of which (called Ω) is bounded and contained in X.*

PROOF. Since $\tau_{a'b'}, \tau_{b'c'}, \tau_{c'a'}$ are shortest paths supports, they are reduced to a same point if $a' = b' = c'$. Else, we have $a' \neq b'$, $b' \neq c'$ and $a' \neq c'$, and the

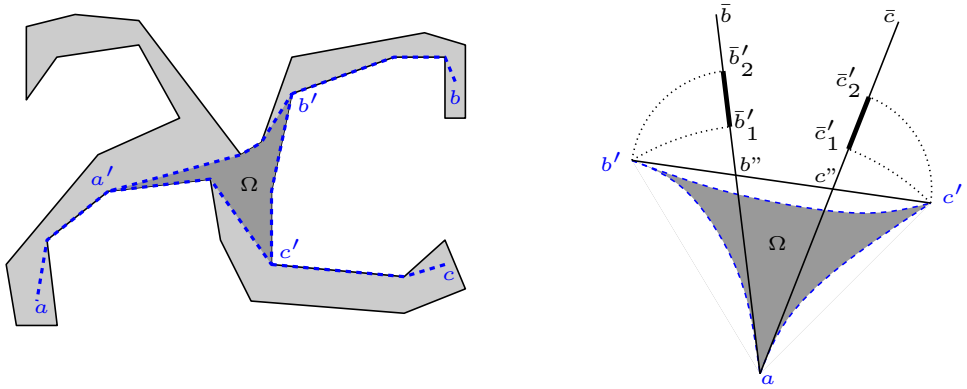


Fig. 3. Left: a geodesic triangle (a, b, c) in a Lipschitz planar domain (light grey). Shortest paths supports are dashed. The non-singular part of the triangle, of vertices a', b', c' and interior Ω, is shown in dark grey. Right: bold segments show the possible locations of $\bar{b}'$ and $\bar{c}'$ on rays $[a, \bar{b})$ and $[a, \bar{c})$.

definition of a', b', c' derived from Claim 5.10.1 ensures that τ is a simple closed curve. Then, the Jordan curve theorem guarantees that τ divides $\mathbb{R}^2$ into two distinct connected components, one of which (called Ω) is bounded. Let $\gamma_{a'b'} : I \to X$ be a shortest path between a' and b', $\gamma_{b'c'} : I \to X$ a shortest path between b' and c', and $\gamma_{c'a'} : I \to X$ a shortest path between c' and a'. Let now $\gamma = \gamma_{a'b'} \cdot \gamma_{b'c'} \cdot \gamma_{c'a'}$. We have $\gamma(I) = \tau$. Furthermore,

$$|\gamma| = \mathrm{d}_X(a', b') + \mathrm{d}_X(b', c') + \mathrm{d}_X(c', a') \leq \mathrm{d}_X(a, b) + \mathrm{d}_X(b, c) + \mathrm{d}_X(c, a) < 2\mathrm{sfs}(X),$$

which implies that γ is null-homotopic in X, by definition of $\mathrm{sfs}(X)$. Let $\Gamma : S^1 \times I \to X$ be a homotopy between γ and a constant map in X. For any point $x \in \Omega$, we have $\deg_x \gamma = \pm 1$ since the loop γ winds once around Ω. If x did not belong to $\Gamma(S^1 \times I)$, then Γ would be a homotopy between γ and a constant map in $\mathbb{R}^2 \setminus \{x\}$, thus by Corollary 2.1 we would have $\deg_x \gamma = 0$, thereby raising a contradiction. It follows that $\Gamma(S^1 \times I)$ contains all the points of Ω, which is therefore included in X. □

It follows from Claim 5.10.2 that the geodesic triangle formed by $\tau_{a'b'}$, $\tau_{b'c}$ and $\tau_{c'a'}$ is either reduced to a point, or an embedded triangle in the plane, whose interior Ω is included in X. From now on, we denote the triangle by (a', b', c') for simplicity.

Claim 5.10.3. *The triangle (a', b', c') has concave edges.*

Proof. Consider an edge of the triangle, say for instance $\tau_{a',b'}$. For any pair of points x, y on this edge, consider the Euclidean segment $[x, y]$. We will show that $[x, y] \cap \Omega = \emptyset$. Assume for a contradiction that this is not the case, and let (x', y') be a connected component of $[x, y] \cap \Omega$. This component is an open subsegment of $[x, y]$, and its endpoints lie on $\tau_{a'b'}$. Call $\tau_{a'x'}$ the subarc of $\tau_{a'b'}$ that connects a' to x' and $\tau_{y'b'}$ the subarc of $\tau_{a'b'}$ that connects y' to b'. Replacing $\tau_{a'b'}$ by $\tau_{a'x'} \cup (x', y') \cup \tau_{y'b}$ yields a path support between a' and b' that is strictly shorter than $\tau_{a'b'}$, yet still included in X (since $\Omega \subset X$). This contradicts the fact that

$\tau_{a'b'}$ is a shortest path support between a' and b' in X. $\square$

The fact that the edges of (a', b', c') are concave implies that their tangents at the three vertices are well-defined when a', b', c' are distinct, as shown in Figure 3 (right).

We can now prove that the inner angles of the geodesic triangle (a, b, c) are well-defined, taking for instance the case of vertex a: if $a \neq a'$, then τ_{ab} and τ_{ca} coincide in the vicinity of a (as in Figure 3 (left) for instance), and therefore the inner angle $\hat{a}$ is zero; if $a = a' = b' = c'$, then a lies on the shortest path support τ_{bc}, and therefore $\hat{a} = \pi$, since a, b, c are assumed to be distinct; else, $a = a'$ and a', b', c' are distinct, and $\hat{a}$ coincides with the angle formed by the two rays emanating from a and tangent to τ_{ab} and τ_{ca} respectively[9]. In every case, the inner angle $\hat{a}$ is well-defined. The same is true for $\hat{b}$ and $\hat{c}$.

CLAIM 5.10.4. *The angles $\hat{a}, \hat{b}, \hat{c}$ are not larger than the corresponding angles in a comparison triangle.*

PROOF. Take for instance vertex a. If $a \neq a'$, then we have $\hat{a} = 0$, which cannot be more than the value of the corresponding angle in a comparison triangle. If $a = a' = b' = c'$, then we have $\hat{a} = \pi$. But since a belongs to the shortest path support τ_{bc}, we have $\mathrm{d}_X(b, c) = \mathrm{d}_X(b, a) + \mathrm{d}_X(a, c)$, which implies that a comparison triangle must be flat, with an inner angle at a equal to π. Consider finally the case where $a = a'$ and a', b', c' are distinct. Let $[a, \bar{b})$ and $[a, \bar{c})$ be the rays emanating from a and tangent to τ_{ab} and τ_{ca} respectively. On $[a, \bar{b})$, the point $\bar{b}$ is placed such that its Euclidean distance to a is equal to $\mathrm{d}_X(a, b)$. Similarly, we place point $\bar{c}$ on $[a, \bar{c})$ such that $\mathrm{d}_E(a, \bar{c}) = \mathrm{d}_X(a, c)$. Assume that the following inequality holds:

$$\mathrm{d}_E(\bar{b}, \bar{c}) \leq \mathrm{d}_X(b, c). \tag{6}$$

Then, any comparison triangle of (a, b, c) must have an inner angle at a that is at least the angle $\hat{a}$ between $[a, \bar{b})$ and $[a, \bar{c})$, which proves the claim.

Let us now prove Eq. (6). Since the triangle (a, b', c') is embedded in the plane with concave edges, b' and c' must lie outside the wedge formed by rays $[a, \bar{b})$ and $[a, \bar{c})$, and the edge $\tau_{b'c'}$ (as well as the Euclidean segment $[b', c']$) must intersect the wedge. Let b'' be the unique intersection point between $[b', c']$ and $[a, \bar{b})$, and c'' the unique intersection point between $[b', c']$ and $[a, \bar{c})$. We place a point $\bar{b}'$ on $[a, \bar{b})$ such that $\mathrm{d}_E(a, \bar{b}') = \mathrm{d}_X(a, b')$. We also let $\bar{b}'_1, \bar{b}'_2 \in [a, \bar{b})$ be such that $\mathrm{d}_E(\bar{b}'_1, a) = \mathrm{d}_E(b', a)$ and $\mathrm{d}_E(\bar{b}'_2, b'') = \mathrm{d}_E(b', b'')$. Since the edge $\tau_{a,b'}$ is concave, it coincides with the graph of some convex real-valued function in an appropriate orthogonal frame of abcissa line (a, b'). Observe that the Euclidean line segments $[a, b'']$ and $[b'', b']$, once concatenated, also form a concave triangle edge, therefore $[a, b''] \cup [b'', b']$ coincides with the graph of some convex function in the same frame as above. And since $[a, b''] \cup [b'', b']$ lies below $\tau_{a,b'}$ in that frame, its length must

[9]This is an easy consequence of the concavity of the edges of (a, b', c'). Indeed, when two points $p \in \tau_{ab'}$ and $q \in \tau_{c'a}$ converge to a, the ratios $\frac{\mathrm{d}_X(p,a)}{\mathrm{d}_E(p,a)}$ and $\frac{\mathrm{d}_X(q,a)}{\mathrm{d}_E(q,a)}$ converge to 1. Furthermore, for p, q close enough to a, the Euclidean line segment $[p, q]$ is included in (a', b', c') and therefore in X, which implies that $\frac{\mathrm{d}_X(p,q)}{\mathrm{d}_E(p,q)} = 1$. Thus, as p, q converge to a, the angle of a comparison triangle tends to $\widehat{paq}$, which converges to the angle between the tangents to τ_{ab} and τ_{ca}.

be greater. As a result, we have:

$$\mathrm{d}_E(a,b') \leq \mathrm{d}_X(a,b') \leq \mathrm{d}_E(a,b'') + \mathrm{d}_E(b'',b').$$

This implies that point $\bar{b}'$ lies in-between $\bar{b}'_1$ and $\bar{b}'_2$ along the ray $[a,\bar{b})$. Similarly, placing $\bar{c}'$ on $[a,\bar{c})$ such that $\mathrm{d}_E(a,\bar{c}') = \mathrm{d}_X(a,c')$, and letting $\bar{c}'_1, \bar{c}'_2 \in [a,\bar{c})$ be such that $\mathrm{d}_E(\bar{c}'_1,a) = \mathrm{d}_E(c',a)$ and $\mathrm{d}_E(\bar{c}'_2,c'') = \mathrm{d}_E(c',c'')$, we have that $\bar{c}'$ lies in-between $\bar{c}'_1$ and $\bar{c}'_2$ along $[a,\bar{c})$. Assuming without loss of generality that $\mathrm{d}_X(a,b') \geq \mathrm{d}_X(a,c')$, we then have

$$\mathrm{d}_E(\bar{b}',\bar{c}') \leq \mathrm{d}_E(\bar{b}'_2,\bar{c}'_1). \tag{7}$$

In addition, since $[b',c']$ crosses the wedge bounded by $[a,\bar{b}) = [a,b'')$ and $[a,\bar{c}) = [a,\bar{c}'_1)$, point $\bar{c}'_1$ lies inside the wedge bounded by $[a,b'')$ and $[a,c')$. Now, inside this wedge, the arc of circle $\partial B_E(a,\mathrm{d}_E(a,c'))$ is included in the closed Euclidean ball $B_E(b'',\mathrm{d}_E(b'',c'))$. It follows that $\mathrm{d}_E(b'',\bar{c}'_1) \leq \mathrm{d}_E(b'',c')$, and by the triangle inequality,

$$\mathrm{d}_E(\bar{b}'_2,\bar{c}'_1) \leq \mathrm{d}_E(\bar{b}'_2,b'')+\mathrm{d}_E(b'',\bar{c}'_1) = \mathrm{d}_E(b',b'')+\mathrm{d}_E(b'',\bar{c}'_1) \leq \mathrm{d}_E(b',b'')+\mathrm{d}_E(b'',c'),$$

which is equal to $\mathrm{d}_E(b',c')$. It follows that $\mathrm{d}_E(\bar{b}'_2,\bar{c}'_1) \leq \mathrm{d}_X(b',c')$. Combined with Eq. (7), this inequality yields $\mathrm{d}_E(\bar{b}',\bar{c}') \leq \mathrm{d}_X(b',c')$. Now, recall that $\mathrm{d}_X(b,b') = \mathrm{d}_X(a,b) - \mathrm{d}_X(a,b')$ since b' lies on the shortest path support τ_{ab}. Idem, $\mathrm{d}_X(c,c') = \mathrm{d}_X(a,c) - \mathrm{d}_X(a,c')$ since c' lies on the shortest path support τ_{ca}. Therefore, we have $\mathrm{d}_E(\bar{b},\bar{b}') = \mathrm{d}_X(b,b')$ and $\mathrm{d}_E(\bar{c},\bar{c}') = \mathrm{d}_X(c,c')$. Combining these relations with the triangle inequality, we obtain:

$$\mathrm{d}_E(\bar{b},\bar{c}) \leq \mathrm{d}_E(\bar{b},\bar{b}') + \mathrm{d}_E(\bar{b}',\bar{c}') + \mathrm{d}_E(\bar{c}',\bar{c}) \leq \mathrm{d}_X(b,b') + \mathrm{d}_X(b',c') + \mathrm{d}_X(c',c),$$

which is equal to $\mathrm{d}_X(b,c)$ since b',c' lie on the shortest path support τ_{bc} between b and c. This proves Eq. (6), and thus also the claim. □

Claim 5.10.4 concludes the proof of Theorem 5.10. □

Open geodesic balls of X in which the angle condition of Definition 5.8 is satisfied by all geodesic triangles are often called *normal balls* in the literature. They enjoy many interesting properties, among which the most important ones to us are the fact that normal balls are convex (*i.e.* any two points in a normal ball B have a unique shortest path support, which is also included in B), and the fact that for any point $p \in X$ the map $q \mapsto \gamma_{pq}$, where γ_{pq} is a shortest path from p to q parametrized with constant speed, is uniquely defined and continuous within any normal ball that contains p. As a result, intersections of normal balls are either empty, or convex and contractible — see Propositions 9.1.16 and 9.1.17 as well as Remark 9.1.18 of [Burago et al. 2001]. Combined with Theorem 5.10, this fact proves Lemma 5.5.

5.3 The case of witness complexes

The one-parameter families of Čech and witness complexes can be interleaved in a same way as in Eq. (3), modulo some additional conditions on the landmarks and witnesses densities:

LEMMA 5.11. *Let X be a Lipschitz domain in the plane, of doubling dimension d. Let W be a geodesic δ-sample of X, and L an ε-sparse geodesic ε-sample of X. For any parameter $\alpha > 0$, we have $\mathcal{C}_\alpha(L) \subseteq \mathcal{C}^W_{X,\nu}(L)$ as soon as $\nu \geq 2^{ld}$, where $l = \lceil \log_2 \frac{2\alpha+\varepsilon+2\delta}{\varepsilon} \rceil$. Conversely, for any parameter ν, we have $\mathcal{C}^W_{X,\nu}(L) \subseteq \mathcal{C}_\alpha(L)$ for all values $\alpha \geq (2\nu+3)\varepsilon$.*

PROOF. Let $\alpha > 0$ be a parameter, and $\sigma = [p_0, \cdots, p_k]$ a simplex of $\mathcal{C}_\alpha(L)$. The open geodesic balls $B_X(p_i, \alpha)$ have a non-empty common intersection. Let c be a point in the intersection, and let $w \in W$ be a point of W closest to c in the intrinsic metric. We then have $\mathrm{d}_X(w, c) \leq \delta$, which implies that $\mathrm{d}_X(w, p_i) < \alpha + \delta$ for all $i = 0, \cdots, k$. Now, since L is ε-sparse, the points of L that lie within geodesic distance $\alpha + \delta$ of w are centers of pairwise-disjoint open geodesic balls of same radius $\frac{\varepsilon}{2}$, packed inside the open ball $B_X(w, \alpha + \delta + \frac{\varepsilon}{2})$. Since the doubling dimension of X is d, the maximum possible number of such balls is at most 2^{ld}, where $l = \lceil \log_2 \frac{2\alpha+\varepsilon+2\delta}{\varepsilon} \rceil$. This implies that the vertices of σ are among the 2^{ld} points of L nearest to w in the intrinsic metric. As a result, w is a ν-witness of σ as soon as $\nu \geq 2^{ld}$. Since this is true for any simplex $\sigma \in \mathcal{C}_\alpha(L)$, we conclude that $\mathcal{C}_\alpha(L) \subseteq \mathcal{C}^W_{X,\nu}(L)$ for all $\nu \geq 2^{ld}$.

Let now $\nu \in \mathbb{N}$ be a parameter, and σ a simplex of $\mathcal{C}^W_{X,\nu}(L)$. Consider any ν-witness w of σ. The vertices of σ are among the $\nu+1$ points of L closest to w in the geodesic distance, and they all lie in the same path-connected component of X as w. Therefore, their geodesic distances to w are less than $(2\nu+3)\varepsilon$, according to Lemma 4.13 and its subsequent comment. Thus, for all $\alpha \geq (2\nu+3)\varepsilon$, w belongs to the open geodesic balls of same radius α centered at the vertices of σ, whose common intersection is therefore non-empty. It follows that $\mathcal{C}^W_{X,\nu}(L) \subseteq \mathcal{C}_\alpha(L)$. □

Letting $l(\alpha) = \lceil \log_2 \frac{2\alpha+\varepsilon+2\delta}{\varepsilon} \rceil$ and $\nu(\alpha) = 2^{l(\alpha)d}$, we deduce from Lemma 5.11 the following inclusions, which correspond to the ones of Eq. (3) for witness complexes: $\forall \alpha > 0$,

$$\mathcal{C}_\alpha(L) \subseteq \mathcal{C}^W_{X,\nu(\alpha)}(L) \subseteq \mathcal{C}_{(2\nu(\alpha)+3)\varepsilon}(L).$$

The above inclusions induce a sequence similar to the one of Eq. (4): $\forall \beta \geq (2\nu(\alpha)+3)\varepsilon$,

$$\mathcal{C}_\alpha(L) \subseteq \mathcal{C}^W_{X,\nu(\alpha)}(L) \subseteq \mathcal{C}_\beta(L) \subseteq \mathcal{C}^W_{X,\nu(\beta)}(L) \subseteq \mathcal{C}_{(2\nu(\beta)+3)\varepsilon}(L).$$

This sequence provides upper and lower bounds on the ranks of the homomorphisms induced at homology level by the inclusion $\mathcal{C}^W_{X,\nu(\alpha)}(L) \hookrightarrow \mathcal{C}^W_{X,\nu(\beta)}(L)$, as in Eq. (5): $\forall \beta \geq (2\nu(\alpha)+3)\varepsilon$, $\forall k \in \mathbb{N}$,

$$\operatorname{rank} H_k(\mathcal{C}_\alpha(L)) \to H_k(\mathcal{C}_{(2\nu(\beta)+3)\varepsilon}(L)) \leq \operatorname{rank} H_k(\mathcal{C}^W_{X,\nu(\alpha)}(L)) \to H_k(\mathcal{C}^W_{X,\nu(\beta)}(L)) \leq \dim H_k(\mathcal{C}_\beta(L)).$$

Equality between the upper and lower bounds is guaranteed by Lemma 5.5, using the same analysis as in the introduction of Section 5 and assuming that $\alpha > \varepsilon$ and $(2\nu(\beta)+3)\varepsilon \leq \frac{1}{3}\mathrm{sfs}(X)$. We thus obtain:

THEOREM 5.12. *Let X be a Lipschitz planar domain, W a geodesic δ-sample of X, and L a finite geodesic ε-sample of X. Then, for any choice of parameters α and $\beta \geq (2\nu(\alpha)+3)\varepsilon$ such that $\alpha > \varepsilon$ and $(2\nu(\beta)+3)\varepsilon \leq \frac{1}{3}\mathrm{sfs}(X)$, the Betti numbers of X can be obtained as the ranks of the homomorphisms induced at homology level by the inclusion $\mathcal{C}^W_{X,\nu(\alpha)}(L) \hookrightarrow \mathcal{C}^W_{X,\nu(\beta)}(L)$, provided that δ, ε are small enough.*

6. ALGORITHMS

In this section, we describe high-level procedures for estimating sfs, for generating geodesic εsfs-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. Section 7 will be devoted to the implementation of such oracles on a sensor network.

6.1 Computing the systolic feature size

Lemma 5.6 suggests a simple procedure for computing the systolic feature size: given a Lipschitz domain X in the plane, and a point $x \in X$, grow a geodesic ball B about x at constant speed, 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 ∂B as the radius of B increases – this wavefront evolves as the iso-level sets of the geodesic distance to x:

- if at some stage the wavefront *self-intersects*, meaning that there is a point $y \in \partial B$ with at least two different shortest paths supports to x, then interrupt the growing 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 5.6 is equal to $\mathrm{sfs}(x)$. The procedure relies on two oracles: the first one detects whether B covers X_x entirely; the second one detects whether the wavefront self-intersects at a given value r of the radius of B, or rather, between two given values $r_1 < r_2$ of the radius of B.

6.2 Generating geodesic εsfs-samples

Given a Lipschitz domain X in the plane, and a real number $\varepsilon > 0$, we can use the procedure of Section 6.1 to generate geodesic εsfs-samples of X. Our algorithm relies on a greedy packing strategy that builds a point set L iteratively by inserting at each iteration a point of X that is far away from the current point set L in the intrinsic metric.

In the initialization phase, the algorithm selects an arbitrary point $p \in X$ and sets $L = \{p\}$. It also assigns to p the open geodesic ball B_p of center p and radius $\frac{\varepsilon}{1+\varepsilon}\,\mathrm{sfs}(p)$, where $\mathrm{sfs}(p)$ is estimated using the procedure of Section 6.1. If $\mathrm{sfs}(p) = +\infty$, then B_p coincides with the path-connected component of X containing p. The main loop of the algorithm proceeds in a similar fashion. At each iteration, an arbitrary point $q \in X \setminus \bigcup_{p\in L} B_p$ is selected and inserted in L. Point q is then assigned the open geodesic ball B_q of center q and radius $\frac{\varepsilon}{1+\varepsilon}\,\mathrm{sfs}(q)$. The process stops when $X \setminus \bigcup_{p\in L} B_p = \emptyset$.

The algorithm uses a variant of an oracle of Section 6.1, 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 open ball B_p, and we have $\mathrm{d}_X(x, L) \leq \mathrm{d}_X(x, p) < \frac{\varepsilon}{1+\varepsilon}\,\mathrm{sfs}(p)$, which is at most $\varepsilon\,\mathrm{sfs}(x)$ since sfs is 1-Lipschitz in the intrinsic metric (Lemma 3.3). Moreover, $\mathrm{d}_X(x, p)$ is finite because B_p is included in the path-connected component of X containing p. Therefore, upon termination, L is a geodesic εsfs-sample of X. Let us show that the algorithm

indeed terminates:

LEMMA 6.1. *For all $\varepsilon > 0$, the algorithm terminates.*

PROOF. Our approach is to bound the pairwise Euclidean distances between the points of L from below by some positive value, and then to apply a packing argument. Let $h = \min\{1, \mathrm{sfs}(X)\}$. Note that we do not use $\mathrm{sfs}(X)$ directly, since the latter might be infinite. In contrast, $0 < h < +\infty$.

Consider any two points p, q inserted in L by the algorithm, and assume without loss of generality that q was inserted after p. If $\mathrm{sfs}(p) = +\infty$, then the ball B_p coincides with X_p, the path-connected component of X that contains p. Therefore, q does not belong to X_p, and we have $\mathrm{d}_X(p,q) = +\infty > \frac{h\varepsilon}{1+\varepsilon}$. If $\mathrm{sfs}(p) < +\infty$, then $\mathrm{d}_X(p,q)$ is at least the radius of B_p, which is equal to $\frac{\varepsilon}{1+\varepsilon}\mathrm{sfs}(p) \geq \frac{\varepsilon}{1+\varepsilon}\mathrm{sfs}(X) \geq \frac{h\varepsilon}{1+\varepsilon}$. In any case, we have $\mathrm{d}_X(p,q) \geq \frac{h\varepsilon}{1+\varepsilon}$ for all points $p, q \in L$. We will now bound this quantity from below by another quantity depending on $\mathrm{d}_E(p,q)$, which will then enable us to use a packing argument.

Consider the set K of all pairs of points x, y of X such that $\mathrm{d}_X(x,y) \geq \frac{h\varepsilon}{1+\varepsilon}$. K is a closed subset of $X \times X$, which is compact since X is, hence K itself is also compact. It follows that the map[10] $g(x,y) = \frac{\mathrm{d}_E(x,y)}{\mathrm{d}_X(x,y)}$ reaches its minimum m over K. This minimum is positive since $\forall (x,y) \in K$, we have $\mathrm{d}_X(x,y) > 0$, which implies that $x \neq y$ and hence that $\mathrm{d}_E(x,y) > 0$.

From the previous paragraphs, we deduce that, for all points $p, q \in L$, $\mathrm{d}_E(p,q)$ is at least $m\ \mathrm{d}_X(p,q) \geq \frac{mh\varepsilon}{1+\varepsilon}$. Hence, the points of L are centers of pairwise-disjoint open Euclidean balls of same radius $\frac{mh\varepsilon}{2(1+\varepsilon)} > 0$, packed inside $X \oplus B_E\left(0, \frac{mh\varepsilon}{2(1+\varepsilon)}\right)$, where $\oplus$ stands for the Minkowski sum. Since X is compact, so is $X \oplus B_E\left(0, \frac{mh\varepsilon}{2(1+\varepsilon)}\right)$, which therefore contains only finitely many disjoint open Euclidean balls of same positive radius. It follows that L is finite. And since the algorithm inserts one point in L per iteration, the process terminates. □

We will now show that the size of the output of the algorithm lies within a constant factor of the optimal, the constant depending on the doubling dimension of (X, d_X).

LEMMA 6.2. *For any $\varepsilon \in]0,1[$, the output landmarks set is $\frac{\varepsilon}{1+\varepsilon}$*sfs*-sparse, and its size is within 2^{ld} times the size of any geodesic ε*sfs*-sample of X, where $l = \left\lceil \log_2 \frac{3+3\varepsilon+2\varepsilon^2}{1-\varepsilon} \right\rceil$ and where d is the doubling dimension of (X, d_X).*

The influence of the doubling dimension d of X is illustrated in Figure 2 (right), where the domain consists of two copies of the domain of Figure 2 (left), glued together along the tips of their branches. The systolic feature size at any point of X is at least half the perimeter of a hole, which is equal to $2 + \frac{2}{2k-1}$. Consider the sets $P = \{p, p'\}$ and $Q = \{q_1, \cdots, q_k\}$. For any point $x \in X$, the geodesic distance from x to P is at most 2, as in the case of Figure 2 (left). As for the geodesic distance from x to Q, it is at most $2 + \frac{2}{2k-1}$. Both distances are bounded from above by $\mathrm{sfs}(x)$, so that P and Q are geodesic sfs-samples of X. Now, for any

[10] This map is well-defined since $\mathrm{d}_X(x,y) \geq \frac{h\varepsilon}{1+\varepsilon} > 0$ for all $(x,y) \in K$.

$q_i \in Q$, the geodesic distance from q_i to any other q_j is greater than half the length of the shortest loop through q_i that winds around a hole. Therefore, the geodesic distance from q_i to $Q \setminus \{q_i\}$ is greater than $\mathrm{sfs}(q_i)$. It follows that Q is sfs-sparse. However, the size of Q is $\frac{k}{2}$ times the size of P, where k is of the order of 2^d, as observed before Theorem 4.17.

Proof of Lemma 6.2. Let L be the output landmarks set. Given any two points $p \neq q \in L$, assume without loss of generality that p was inserted in L before q. Then, q does not belong to the open geodesic ball of center p and radius $\frac{\varepsilon}{1+\varepsilon}\ \mathrm{sfs}(p)$. Hence, $\mathrm{d}_X(p,p) \geq \frac{\varepsilon}{1+\varepsilon}\ \mathrm{sfs}(p)$, which is at least $\frac{\varepsilon}{1+\varepsilon}\ \min\{\mathrm{sfs}(p), \mathrm{sfs}(q)\}$. Therefore, L is $\frac{\varepsilon}{1+\varepsilon}$sfs-sparse.

Let now $L' \subset X$ be any geodesic εsfs-sample of X. Consider the function $\pi : L \to L'$ that maps each point of L to its nearest neighbor in L' in the intrinsic metric, breaking ties arbitrarily. We then have $|L| = \sum_{q \in L'} |\pi^{-1}(\{q\})|$. Therefore, to bound the size of L, it is enough to bound the size of each set $\pi^{-1}(\{q\})$.

Let $q \in L'$, and let $p_1, \cdots, p_k$ be the points of $\pi^{-1}(\{q\})$. All the points p_i belong to the path-connected component X_q of X that contains q, since L' is a geodesic εsfs-sample of X. If $\mathrm{sfs}(q) = +\infty$, then X_q is simply connected, and therefore the algorithm picks only one point from X_q. It follows that $|\pi^{-1}(q)| = 1$. Assume from now on that $\mathrm{sfs}(q) < +\infty$, which means that X_q is not simply connected and hence that the $\mathrm{sfs}(p_i)$ are finite.

Since L' is a geodesic εsfs-sample of L, for all $i = 1, \cdots, k$ we have $\mathrm{d}_X(p_i, q) < \varepsilon\ \mathrm{sfs}(p_i)$, which is at most $\frac{\varepsilon}{1-\varepsilon}\ \mathrm{sfs}(q)$ since sfs is 1-Lipschitz in the intrinsic metric (Lemma 3.3). Hence, the p_i belong to the open geodesic ball of center q and radius $\frac{\varepsilon}{1-\varepsilon}\ \mathrm{sfs}(q)$. Moreover, assuming without loss of generality that $p_1, \cdots, p_k$ were inserted in L in this order, we have that, for all $1 \leq i < j \leq k$, p_j does not belong to the open geodesic ball of center p_i and radius $\frac{\varepsilon}{1+\varepsilon}\ \mathrm{sfs}(p_i)$. Hence, $\mathrm{d}_X(p_i, p_j) \geq \frac{\varepsilon}{1+\varepsilon}\ \mathrm{sfs}(p_i)$, which is at least $\frac{\varepsilon}{(1+\varepsilon)^2}\ \mathrm{sfs}(q)$ since $\mathrm{d}_X(p_i, q) \leq \varepsilon\ \mathrm{sfs}(p_i)$ and since sfs is 1-Lipschitz in the intrinsic metric. Therefore, the p_i are centers of pairwise-disjoint open geodesic balls of radius $\frac{\varepsilon}{2(1+\varepsilon)^2}\ \mathrm{sfs}(q)$, packed inside the open geodesic ball of center q and radius $\left(\frac{1}{1-\varepsilon} + \frac{1}{2(1+\varepsilon)^2}\right) \varepsilon\ \mathrm{sfs}(q) = \frac{3+3\varepsilon+2\varepsilon^2}{2(1-\varepsilon)(1+\varepsilon)^2}\ \varepsilon\ \mathrm{sfs}(q)$.

It follows from the previous paragraph that the size of $\pi^{-1}(\{q\})$ is bounded by the maximum number of open geodesic balls of radius $\frac{\varepsilon}{2(1+\varepsilon)^2}\ \mathrm{sfs}(q)$ that can be packed inside an open geodesic ball of radius $\frac{3+3\varepsilon+2\varepsilon^2}{2(1-\varepsilon)(1+\varepsilon)^2}\ \varepsilon\ \mathrm{sfs}(q)$. By Lemma 4.16, this number is at most the minimum number n of geodesic balls of radius $\frac{\varepsilon}{2(1+\varepsilon)^2}\ \mathrm{sfs}(q)$ that are necessary to cover a geodesic ball of radius $\frac{3+3\varepsilon+2\varepsilon^2}{2(1-\varepsilon)(1+\varepsilon)^2}\ \varepsilon\ \mathrm{sfs}(q)$. The ratio between the two radii is $\frac{3+3\varepsilon+2\varepsilon^2}{1-\varepsilon}$, therefore n is at most $\left(2^d\right)^l = 2^{ld}$, where $l = \left\lceil \log_2 \frac{3+3\varepsilon+2\varepsilon^2}{1-\varepsilon} \right\rceil$ and d is the doubling dimension of (X, d_X). Thus, for all point q of L', the size of $\pi^{-1}(q)$ is at most 2^{ld}, which implies that $|L| \leq 2^{ld}\ |L'|$. □

Note that the algorithm introduced in this section can also be used to generate (uniform) ε-sparse geodesic ε-samples of X, for any input $\varepsilon > 0$. It suffices indeed to remove the estimation of sfs from the algorithm, which is no longer needed, and to consider open geodesic balls of radius ε instead of radius $\frac{\varepsilon}{1+\varepsilon}$sfs. The arguments

of the proofs of Lemmas 6.1 and 6.2 still hold in this context, and the technical details are slightly simpler.

6.3 Computing the homology of Lipschitz domains in the plane

Given a finite sampling L of some Lipschitz planar domain X, a variant of the procedure of Section 6.1 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, provided that L is dense enough, by Theorem 4.3. However, 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 type of discrete setting, it is relevant to replace the construction of $\mathcal{D}_X(L)$ by the ones of $\mathcal{C}^W_{X,\nu}(L)$ or $\mathcal{R}_\alpha(L)$, for some subset $L \subseteq W$ of landmarks, since these constructions only require to compare geodesic distances at the points of L or W. The Betti numbers of $\mathcal{D}_X(L)$ (and hence the one of X) can then be obtained as the ranks of the homomorphisms induced at homology level by the inclusions $\mathcal{C}^W_{X,\nu}(L) \hookrightarrow \mathcal{C}^W_{X,\nu'}(L)$ or $\mathcal{R}_\alpha(L) \hookrightarrow \mathcal{R}_{\alpha'}(L)$, for well-chosen parameters ν, ν' or α, α', thanks to the results of Section 5.

More precisely, if we choose for instance to use witness complexes, then we can select two integer parameters $\nu \leq \nu'$ and build $\mathcal{C}^W_{X,\nu}(L)$ and $\mathcal{C}^W_{X,\nu'}(L)$ by means of comparisons between the geodesic distances from the points of W to the points of L. Then, using simplicial homology with coefficients in a field, which in practice will be $\mathbb{Z}/2$ – omitted in our notations, we have that for all $k \in \mathbb{N}$ the inclusion map $i : \mathcal{C}^W_{X,\nu}(L) \hookrightarrow \mathcal{C}^W_{X,\nu'}(L)$ induces a homomorphism $i_k^* : H_k^\Delta(\mathcal{C}^W_{X,\nu}(L)) \to H_k^\Delta(\mathcal{C}^W_{X,\nu'}(L))$. By applying the persistence algorithm [Zomorodian and Carlsson 2005] to the filtration $\mathcal{C}^W_{X,\nu}(L) \hookrightarrow \mathcal{C}^W_{X,\nu'}(L)$, we can compute the rank of i_k^*. Now, thanks to Theorem 5.12, for any given choice of parameters $\nu' > \nu' > 0$, the rank of i_k^* coincides with the kth Betti number of X provided that W, L are dense enough (*i.e.* that δ, ε are small enough). Thus, the homology of the domain can be inferred using witness complexes, under sufficient sampling density.

7. APPLICATION TO SENSOR NETWORKS

We have implemented the algorithms of Section 6 in the context of sensor networks, where the nodes do not have geographic locations, and where the intrinsic metric is approximated by the shortest path length in the connectivity graph $G = (W, E)$, which is assumed to comply with the *geodesic* unit disk graph model. This means that each node has a geodesic communication range of μ, so that two nodes $w, w' \in W$ are connected in the graph iff $\mathrm{d}_X(w, w') \leq \mu$. All edges have a unit weight, and we denote by d_G the associated graph distance – also called hop-count distance. This geodesic unit disk graph model is the analog of the standard Euclidean unit disk graph model in the intrinsic metric.

LEMMA 7.1. *Assume that W is a geodesic δ-sample of X, with $\delta < \frac{\mu}{2}$. Then, for all nodes $w, w' \in W$, we have:*

$$\frac{\mathrm{d}_X(w,w')}{\mu} \leq \mathrm{d}_G(w,w') \leq \frac{\mathrm{d}_X(w,w')}{\mu}\left(1 + \frac{4\delta}{\mu} + \frac{\mu}{\mathrm{d}_X(w,w')}\right)$$

PROOF. Let $w, w' \in W$ be two nodes of the graph. We first give an upper

bound on d_G. Consider a shortest path ζ from w to w' inside X. We have $|\zeta| = \mathrm{d}_X(w, w')$. Let $0 = t_0 \leq t_1 \leq \cdots \leq t_{m-1} \leq t_m = 1$ be distributed along I in such a way that $\mathrm{d}_X(\zeta(t_i), \zeta(t_{i+1})) = \mu - 2\delta$ for all $i = 0, \cdots, m-2$, while $\mathrm{d}_X(\zeta(t_{m-1}), \zeta(t_m)) \leq \mu - 2\delta$. Clearly, we have $m = \left\lceil \frac{\mathrm{d}_X(w,w')}{\mu - 2\delta} \right\rceil$. For all i, let w_i be a point of W closest to $\zeta(t_i)$ in the intrinsic metric. Since W is a geodesic δ-sample of X, we have $w_0 = \zeta(t_0) = w$, $w_m = \zeta(t_m) = w'$, and $\mathrm{d}_X(w_i, \zeta(t_i)) \leq \delta$ for any other i. It follows from the triangle inequality that: $\mathrm{d}_X(w_i, w_{i+1}) \leq \mathrm{d}_X(w_i, \zeta(t_i)) + \mathrm{d}_X(\zeta(t_i), \zeta(t_{i+1})) + \mathrm{d}_X(\zeta(t_{i+1}), w_{i+1}) \leq \mu$. Therefore, $[w_i, w_{i+1}]$ is an edge of the communication graph G, and thus to ζ corresponds a path γ in G. Both ζ and γ connect w to w' and are made of m pieces stitched together. Hence, $\mathrm{d}_G(w, w') \leq m = \left\lceil \frac{\mathrm{d}_X(w,w')}{\mu - 2\delta} \right\rceil$, which is bounded from above by:

$$\left\lceil \frac{\mathrm{d}_X(w,w')}{\mu} \left(1 + \frac{4\delta}{\mu}\right) \right\rceil \leq \frac{\mathrm{d}_X(w,w')}{\mu} \left(1 + \frac{4\delta}{\mu}\right) + 1 = \frac{\mathrm{d}_X(w,w')}{\mu} \left(1 + \frac{4\delta}{\mu} + \frac{\mu}{\mathrm{d}_X(w,w')}\right).$$

Let us now give a lower bound on d_G. Let γ be any path from w to w' in the communication graph G. For any consecutive nodes w_i, w_{i+1} along the path, we have $\mathrm{d}_X(w_i, w_{i+1}) \leq \mu$ since $[w_i, w_{i+1}]$ is an edge of G. Therefore, by the triangle inequality, γ must have at least $\left\lceil \frac{\mathrm{d}_X(w,w')}{\mu} \right\rceil$ edges. Since this is true for any path γ from w to w' in G, $\mathrm{d}_G(w, w') \geq \left\lceil \frac{\mathrm{d}_X(w,w')}{\mu} \right\rceil \geq \frac{\mathrm{d}_X(w,w')}{\mu}$. □

Assume now that L is a $\frac{\varepsilon}{1+\varepsilon}$sfs-sparse geodesic εsfs-sample[11] of X. Suppose that $\delta \ll \mu \ll \varepsilon \ll 1$. Given a witness $w \in W$, every landmark $p \in L$ that is not its closest landmark satisfies: $\mathrm{d}_X(w, p) = \Omega(\varepsilon) \gg \mu$, which implies that $\mathrm{d}_G(w, p)$ is an accurate approximation to $\frac{\mathrm{d}_X(w,p)}{\mu}$, by Lemma 7.1. If now p is the landmark closest to w, then we may as well have $\mathrm{d}_X(w, p) \ll \mu$, but in this case we also have $\mathrm{d}_X(w, p) \ll \mathrm{d}_X(w, q)$ for all $q \in L \setminus \{p\}$, which implies that $\mathrm{d}_G(w, p) < \mathrm{d}_G(w, q)$. As a result, d_G may change the order of the distances between the landmarks and w, but interverted distances must have similar values. In this respect, we can say that d_X is a faithful approximation to d_X, as it is known that the persistent homology of the family of ν-witness complexes is stable under such small perturbations [Chazal et al. 2009].

Systolic feature size computation. Given a node x, we estimate the geodesic distance of x to its cut-locus, which by Lemma 5.6 is equal to $\mathrm{sfs}(x)$. Wang *et al.* [Wang et al. 2006] 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 [Wang et al. 2006], every hole of perimeter greater

[11] One may as well assume that L is an ε-sparse geodesic ε-sample of X, in a uniform version of the setting.

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 systolic feature size as the minimum hop value.

Landmarks selection and witness complex computation. The landmarks selection procedure implements the incremental algorithm of Section 6.2 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 systolic 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 geodesic witness complex is computed in a similar way as in [Fang et al. 2005]. 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 centralized manner. In a planar setting, where only the Betti numbers β_0 and β_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. This reduces the message complexity drastically.

As for ν-witness complexes, they are computed with the exact same procedure, except that each node stores its geodesic distances to its $\nu + 1$ nearest landmarks.

Simulation results and discussion. Figures 4 through 8 present our simulation results. We consider n sensor nodes randomly distributed in a Lipschitz planar domain. Two nodes within unit Euclidean distance of each other are connected, so that the resulting communication network is a unit disk graph. The average node degree in this graph is denoted by 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 graph distance). Our aim is to evaluate the dependency of the landmarks selection and homology computation on various parameters. For the homology computation we use the pair of complexes $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$, where L is the landmarks set and ν is an integer parameter that ranges typically between 2 and 11. The inclusion $\mathcal{C}^W_X(L) \subseteq \mathcal{C}^W_{X,\nu}(L)$ holds because we restrict our construction to the 2-skeleta of the complexes. Figure 4 shows a typical example, with $\varepsilon = 0.5$ (a) and $\varepsilon = 0.25$ (b). In both cases, only the genuine 3 holes persist and are therefore identified as non-trivial 1-cycles in the geodesic Delaunay triangulation.

- *Nodes density.* We vary the number of nodes from 217 to 355. The average degree remains the same. The result is shown in Figure 5. Again, the persistent homology between the witness complex $\mathcal{C}^W_X(L)$ and the ν-witness complex $\mathcal{C}^W_{X,\nu}(L)$ 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 dense enough.

- *Landmarks density.* Figure 6 shows our results on the same setup as above, with $\varepsilon = 0.85$ (a) and $\varepsilon = 0.15$ (b). In the first case, only two holes are captured, because of the low landmarks density. In the second case, three non-genuine holes are not destroyed in the ν-witness complex, because the value of the relaxation parameter ν is too small given the relatively low nodes density. Increasing ν from 2 to 4 produces the correct answer (c). But setting ν 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 ν ($\nu \leq 4$), 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 4.14 and 4.17.
- *Weighted graph distance vs. hop-count distance.* Since the hop-count distance is a poor approximation to the geodesic distance, the range of values of ε that work fine with it is reduced. In Figure 7 for instance, the scheme works well with $\varepsilon = 0.5$, but not with $\varepsilon = 0.25$, in contrast with the results of Figure 4.
- *Packing strategy.* Figure 8 shows some of our sampling results. It appears that different packing strategies can produce samples of very different sizes, as predicted by Lemma 6.1. Maximizing the ratio $\frac{\mathrm{d}_X(q,L)}{\mathrm{sfs}(q)}$ at each iteration seems to be a very effective strategy in practice, but it is also time-consuming, and it tends to choose landmarks near the boundaries of the domain, which can be a quality or a defect, depending on the application considered.

8. CONCLUSION

We have introduced a new quantity, called the systolic feature size, and showed that it is well-suited for the sampling and analysis of Lipschitz domains in the plane. In particular, given a domain X and a landmarks set L that is sufficiently densely sampled from X, the bound on the density depending on the systolic feature size of X, we have proved that the geodesic Delaunay triangulation of L is homotopy equivalent to X. The systolic feature size depends essentially on the global topology of X, and it is rather insensitive to the local geometry. As a result, it enables to have very sparse sets of landmarks, which makes it a convenient theoretical tool for geometric data analysis. In this context, we have devised generic procedures for estimating the systolic feature size and for generating geodesic εsfs-samples of Lipschitz planar domains.

With more practical applications in mind, we have focused on the geodesic witness complex and its relaxed version, proving that these two complexes sandwich the geodesic Delaunay triangulation under some conditions. As an application, we have shown that it is possible to estimate the homology of a Lipschitz planar domain X from a finite set of landmarks L without actually building $\mathcal{D}_X(L)$ explicitly, by constructing $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$ and computing their persistent homology. To give theoretical guarantees to this approach, we proved in the conference version of the paper that the persistent homology between $\mathcal{C}^W_X(L)$ and $\mathcal{C}^W_{X,\nu}(L)$ coincides with the homology of $\mathcal{D}_X(L)$, yet under some fairly stringent sampling conditions. Our practical experiments in the context of sensor networks suggest that milder conditions should be sufficient. Taking a different approach in the present paper, we have uncovered some sufficient conditions that depend solely on the systolic

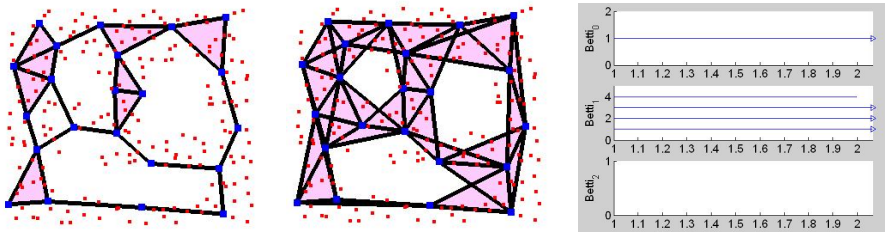
feature size.

This work can be improved in several ways:

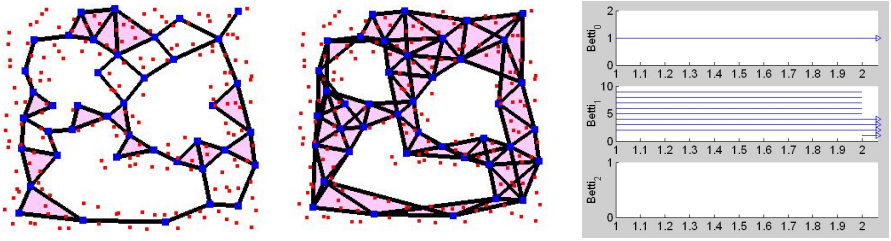
– One may look at bounded domains in higher-dimensional Euclidean spaces, with applications in robotics and geometric data analysis. Note that our approach relies on cycles bounding holes being non-contractible. In higher dimensions, voids cannot stop cycles from contracting to a point, so our framework does not apply as it is, and higher-dimensional homotopy groups need to be considered.
– Another possible improvement would be 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 [Freedman and Chen 2007].
– Finally, in order to make the approach fully practical, it would be necessary to devise distributed variants of the procedures that build the simplicial complexes and compute the persistent homology. Whether such variants exist is still an open question at this time.

ACKNOWLEDGMENTS

The authors wish to thank Jean-Daniel Boissonnat for pointing out the connection that exists between the systolic feature size (originally called *homotopy feature size* in the conference version of the paper) and the concept of systole. They also thank the anonymous referees for their careful reading of the paper and for their constructive criticism, which really helped improve the paper.

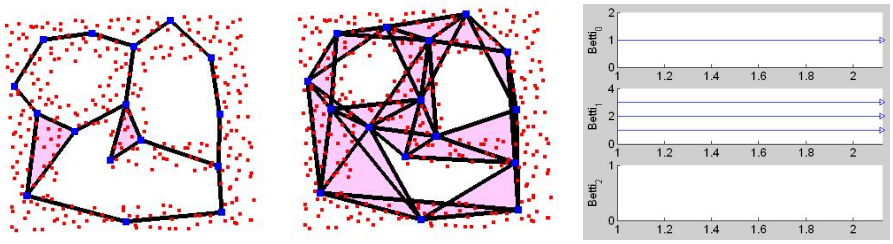


(a) $n = 217$, $d \approx 7.66$, $\varepsilon = 0.5$, $\nu = 2$, weighted graph distance.

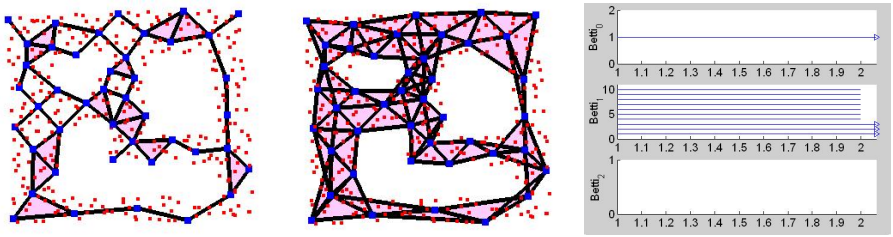


(b) $n = 217$, $d \approx 7.66$, $\varepsilon = 0.25$, $\nu = 2$, weighted graph distance.

Fig. 4. From left to right: witness complex, relaxed witness complex, persistence barcode of the filtration.



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



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

Fig. 5. Same setting as above, with a higher nodes density.

REFERENCES

Amenta, N. and Bern, M. 1999. Surface reconstruction by Voronoi filtering. *Discrete Comput. Geom. 22,* 4, 481–504.

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

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

Axelsson, A. and M^{c}Intosh, A. 2004. Hodge decompositions on weakly Lipschitz domains. *Advances in Analysis and Geometry: New Developments Using Clifford Algebras 30,* 3.

Bobenko, A. I. and Springborn, B. A. 2007. A discrete laplace-beltrami operator for simplicial surfaces. *Discrete and Computational Geometry 38,* 4 (December), 740–756.

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

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

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

Borsuk, K. 1967. *Theory of Retracts*. Monografie Matematyczne, vol. 44. Polish Scientific Publishers, Warsaw, Poland.

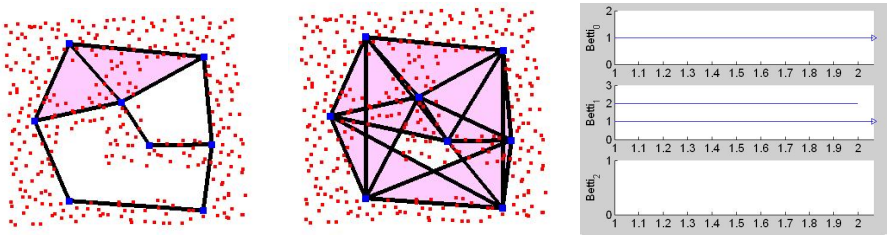
Burago, D., Burago, Y., and Ivanov, S. 2001. *A Course in Metric Geometry*. Graduate Studies in Mathematics, vol. 33. American Mathematical Society, Providence, RI.

Cannon, J., Conner, G., and Zastrow, A. 2002. One-dimensional and planar sets are aspherical. *Topology Appl. 120,* 1-2, 23–45.

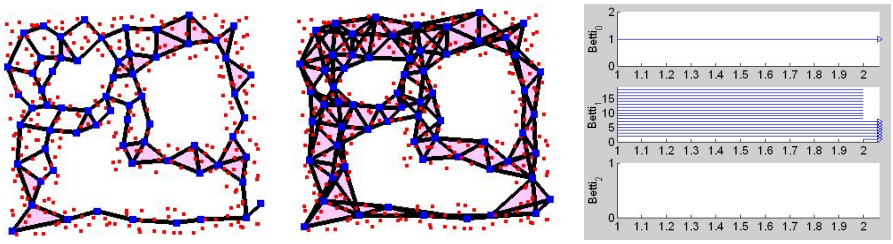
Chazal, F., Cohen-Steiner, D., Glisse, M., Guibas, L. J., and Oudot, S. Y. 2009. Proximity of persistence modules and their diagrams. In *Proc. 25th Annual Symposium on Computational Geometry*. 237–246.

Chazal, F., Guibas, L. J., Oudot, S. Y., and Skraba, P. 2009. Analysis of scalar fields over point cloud data. In *Proc. 20th ACM-SIAM Sympos. Discrete Algorithms*. 1021–1030.

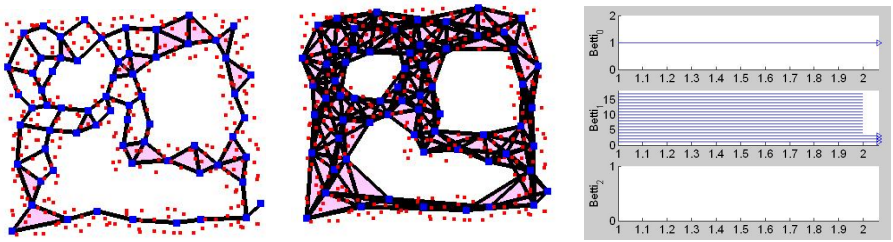
Chazal, F. and Lieutier, A. 2007. Stability and computation of topological invariants of solids in $\mathbb{R}^n$. *Discrete Comput. Geom. 37,* 4, 601–617.



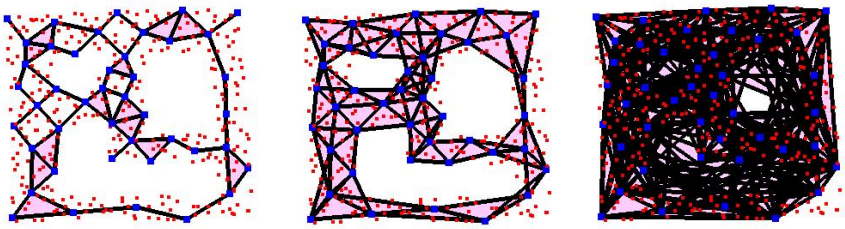
(a) $n = 353$, $d \approx 7.66$, $\varepsilon = 0.85$, $\nu = 2$, weighted graph distance.



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



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



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

Fig. 6. Effect of varying parameter ν, versus landmarks density.

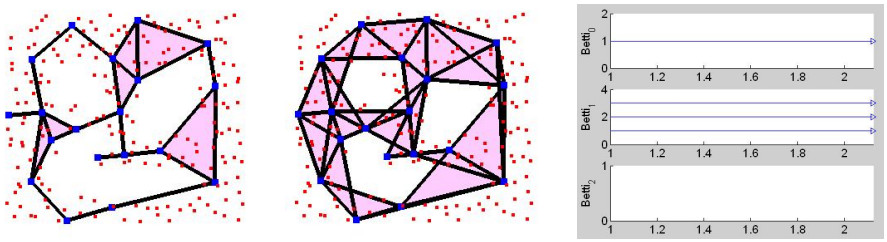
Chazal, F. and Oudot, S. Y. 2008. Towards persistence-based reconstruction in Euclidean spaces. In *Proc. 24th ACM Sympos. Comput. Geom.* 232–241.

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

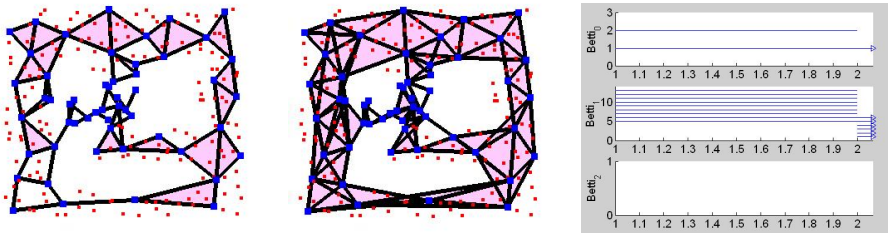
Daverman, R. J. 2007. *Decompositions of Manifolds.* American Mathematical Society, Providence, RI.

de Silva, V. 2008. A weak characterisation of the Delaunay triangulation. *Geometriae Dedicata 135,* 1 (August), 39–64.

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

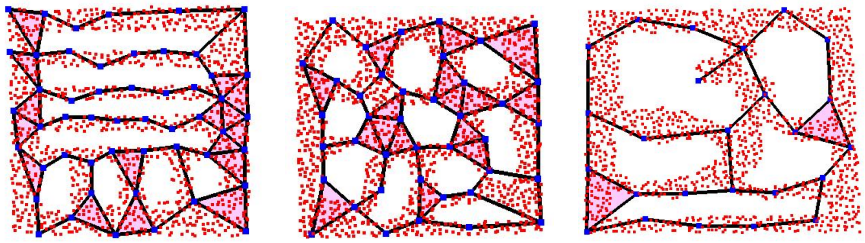


(a) $n = 217$, $d \approx 7.66$, $\varepsilon = 0.5$, $\nu = 2$, hop-count distance.

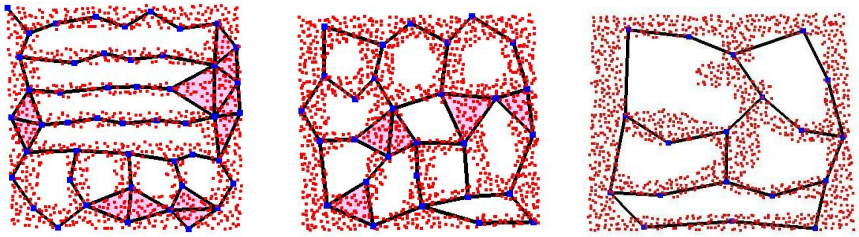


(b) $n = 217$, $d \approx 7.66$, $\varepsilon = 0.25$, $\nu = 2$, hop-count distance.

Fig. 7. Same setting as above, with the weighted graph distance replaced by the hop-count distance.



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



(b) $\varepsilon = \frac{1}{3}$, insertion of node q that maximizes $\frac{\mathrm{d}_X(q,L)}{\mathrm{sfs}(q)}$ outside $\bigcup_{p \in L} \overline{B}_p$.

Fig. 8. Landmarks sets obtained by two different packing strategies, and their geodesic witness complexes.

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

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

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

Fang, Q., Gao, J., Guibas, L. J., de Silva, V., and Zhang, L. 2005. GLIDER: gradient landmark-based distributed routing for sensor networks. In *Proc. of the 24th Conference of the IEEE Communication Society (INFOCOM)*. Vol. 1. 339–350.

Federer, H. 1996. *Geometric Measure Theory*. Classics in Mathematics. Springer-Verlag. Reprint of the 1969 ed.

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

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

Gao, J., Guibas, L. J., Oudot, S. Y., and Wang, Y. 2008. Geodesic delaunay triangulation and witness complex in the plane. In *Proc. 18th ACM-SIAM Sympos. on Discrete Algorithms*. 571–580.

Gromov, M. 1996. Systoles and intersystolic inequalities. In *Actes de la Table Ronde de Géométrie Différentielle (Luminy, 1992)*. Sémin. Congr. Soc. Math. France, 291–362.

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

Hatcher, A. 2001. *Algebraic Topology*. Cambridge Univ. Press.

James, I. M. 1995. *Handbook of Algebraic Topology*. North Holland.

Kolmogorov, A. and Tikhomirov, V. 1961. ϵ-entropy and ϵ-capacity of sets of functions. *Translations of the AMS 17*, 277–364.

Lee, J. M. 2002. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer-Verlag.

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

Mitchell, J. S. B. 1991. A new algorithm for shortest paths among obstacles in the plane. *Ann. Math. Artif. Intell. 3*, 83–106.

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

Wang, Y., Gao, J., and Mitchell, J. S. B. 2006. Boundary recognition in sensor networks by topological methods. In *Proc. of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*. 122–133.

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

Zomorodian, A. and Carlsson, G. 2005. Computing persistent homology. *Discrete Comput. Geom. 33,* 2, 249–274.

Received ...; accepted ...

A. APPENDIX – PROOF OF PROPOSITION 2.2

We use singular homology with real coefficients, so that our homology groups are vector spaces over the field $\mathbb{R}$ – omitted in our notations. Please refer to [Hatcher 2001, Chapter 2] for an introduction to homology theory.

Proof of (i). The proof is by induction on k. The case $k = 1$ is trivially true. Assume now that the result is true up to some $k \geq 1$, and consider $k + 1$

planar sets $X_1, \cdots, X_{k+1}$ satisfying the hypotheses of Proposition 2.2 (i). Notice that each path-connected component of $X_1 \cap \cdots \cap X_{k+1}$ is the intersection of some path-connected component Y of $\bigcap_{i=1}^{k} X_i$ with $Z = X_{k+1}$, which by the induction hypothesis are simply connected. Intuitively, the presence of a hole in the intersection $Y \cap Z$ would automatically imply the presence of a hole in Y or in Z. Thus, the path-connected components of $Y \cap Z$ must be simply connected, since Y and Z are.

Formally, since Y, Z and $Y \cap Z$ are ANR's, the triad $(Y \cup Z, Y, Z)$ is excisive and the Mayer-Vietoris long exact sequence holds:

$$\cdots \to H_2(Y \cup Z) \xrightarrow{\partial_2} H_1(Y \cap Z) \xrightarrow{\phi} H_1(Y) \oplus H_1(Z) \to \cdots$$

Since Y and Z are simply connected, we have $H_1(Y) = H_1(Z) = 0$, therefore $\ker \phi = H_1(Y \cap Z)$. By exactness, $\ker \phi$ is also equal to $\operatorname{im} \partial_2$, which is trivial since we have $H_2(Y \cup Z) = 0$, Y and Z being subsets of $\mathbb{R}^2$. As a result, we have $H_1(Y \cap Z) = 0$. Since $H_1(Y \cap Z)$ is the direct sum of the $H_1(C)$, for C ranging over all the path-connected components of $Y \cap Z$, we have $H_1(C) = 0$ for each path-connected component C of $Y \cap Z$. This implies that the fundamental group of C is trivial: indeed, since C is a path-connected planar set, its fundamental group is either free or uncountable, and therefore it is trivial if and only if its abelianization (which is precisely $H_1(C)$) is. As a conclusion, C is simply connected, which proves the result for $k+1$ and thereby concludes the induction. □

To prove (ii), we need an easy intermediate result:

LEMMA A.1. *If X, Y are path-connected planar sets such that $X \cap Y \neq \emptyset$, then $X \cup Y$ is path-connected.*

PROOF. Let $p \in X \cap Y$, and let q be any other point of $X \cup Y$. If $q \in X$, then there exists a path between p and q in X, which is path-connected. Otherwise, q lies in Y, and there exists a path between p and q in Y, which is also path-connected. Therefore, every point of $X \cup Y$ is path-connected to p in $X \cup Y$, which is therefore path-connected. □

We can now prove (ii):

Proof of (ii). Assume that $X \cap Y$ is not empty. Intuitively, the topological type of $X \cup Y$ partially determines the topological type of $X \cap Y$, in the sense that $X \cup Y$ would have a hole if ever $X \cap Y$ were not path-connected, since X, Y themselves are path-connected. Formally, since X, Y and $X \cap Y$ are ANR's, the triad $(X \cup Y, X, Y)$ is excisive and the Mayer-Vietoris long exact sequence holds:

$$\cdots \to H_1(X \cup Y) \xrightarrow{\partial_1} H_0(X \cap Y) \xrightarrow{\phi} H_0(X) \oplus H_0(Y) \xrightarrow{\psi} H_0(X \cup Y) \xrightarrow{\partial_0} 0.$$

Since $X \cap Y \neq \emptyset$, Lemma A.1 tells us that $X \cup Y$ is path-connected, therefore $\dim H_0(X \cup Y) = 1$. This implies that $\dim \ker \partial_0 = 1$, and hence that $\operatorname{rank} \psi = \dim \ker \partial_0 = 1$, by exactness. By the homomorphism theorem, we have $\dim \ker \psi = \dim(H_0(X) \oplus H_0(Y)) - \operatorname{rank} \psi$, which is equal to 1 since X and Y are path-connected. Hence, by exactness, $\operatorname{rank} \phi = \dim \ker \psi = 1$. Moreover, since by assumption $X \cup Y$ is simply connected, we have $\dim H_1(X \cup Y) = 0$, which implies

that rank $\partial_1 = 0$. By exactness, we have $\dim \ker \phi = \text{rank } \partial_1 = 0$. Hence, by the homomorphism theorem, $\dim H_0(X \cap Y) = \dim \ker \phi + \text{rank } \phi = 1$, which means that $X \cap Y$ is path-connected. $\square$