



HAL
open science

Optimal Reconstruction Might be Hard

Dominique Attali, André Lieutier

► **To cite this version:**

Dominique Attali, André Lieutier. Optimal Reconstruction Might be Hard. *Discrete and Computational Geometry*, 2013, 49 (2), pp.133-156. 10.1007/s00454-012-9475-8 . hal-00785089

HAL Id: hal-00785089

<https://hal.science/hal-00785089>

Submitted on 5 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal Reconstruction Might be Hard*

Dominique Attali[†]

André Lieutier[‡]

September 7, 2012

Abstract

Given as input a point set \mathcal{S} that samples a shape \mathcal{A} , the condition required for inferring Betti numbers of \mathcal{A} from \mathcal{S} in polynomial time is much weaker than the conditions required by any known polynomial time algorithm for producing a topologically correct approximation of \mathcal{A} from \mathcal{S} . Under the former condition which we call the *weak precondition*, we investigate the question whether a polynomial time algorithm for reconstruction exists. As a first step, we provide an algorithm which outputs an approximation of the shape with the correct Betti numbers under a slightly stronger condition than the weak precondition. Unfortunately, even though our algorithm terminates, its time complexity is unbounded. We then identify at the heart of our algorithm a test which requires answering the following question: given two 2-dimensional simplicial complexes $L \subset K$, does there exist a simplicial complex containing L and contained in K which realizes the persistent homology of L into K ? We call this problem the *homological simplification* of the pair (K, L) and prove that this problem is NP-complete, using a reduction from 3SAT.

1 Introduction

Previous works. The problem of reconstructing shapes from point clouds has been well studied in computer graphics, computational geometry, machine learning, and other areas. Reconstruction methods aim at building an approximation of a shape from a set of data points that sample it. The resulting object may then be measured or used for certain tasks such as rendering, storing, searching in a data base and so on. In this context, it is desirable that the result of the reconstruction reflects the topology of the original shape. During the past two decades, a lot of research went into finding sampling conditions which guarantee a topologically correct reconstruction. First sampling conditions were assuming shapes to be compact smooth surfaces embedded in the Euclidean three-dimensional space and data points to be noise-free [5, 1, 3, 9, 19, 4]. Since then, much effort has been put into weakening sampling conditions so that a wider class of shapes can be reconstructed from sparser and less accurate samples.

An important step has been to allow noise in the sample. Maybe one of the simplest noise model supposes that each point of the sample lies within some distance of the sampled shape (the sample is *accurate*) and each point of the sampled shape lies within some distance of a sample point (the sample is *dense*). When both distances are bounded by the same value ε , we say that the *Hausdorff distance* between the shape and the sample is upper bounded by ε . First sampling conditions were assuming the Hausdorff distance to be less than a fraction the *reach* of the shape [27]. The reach of \mathcal{A} is the infimum of distances

*This work is partially supported by ANR Project *GIGA* ANR-09-BLAN-0331-01.

[†]Gipsa-lab, Grenoble, France. Dominique.Attali@gipsa-lab.grenoble-inp.fr

[‡]Dassault Systèmes, France. Andre.Lieutier@3ds.com

between points in \mathcal{A} and points in its medial axis. Unfortunately, the reach vanishes for shapes with sharp concave edges and therefore is not suitable for expressing sampling conditions for non-smooth manifolds or stratified objects. To deal with this problem, Boissonnat and Oudot in [10] considered Lipschitz manifolds while Chazal, Cohen-Steiner and Lieutier in [11] considered a large class of non-smooth compact sets called *sets with a positive μ -reach*. The μ -reach of \mathcal{A} is the infimum of distances between points in \mathcal{A} and points in its μ -medial axis which for $\mu \in (0, 1]$ is a stable subset of the medial axis defined in [11]. More recently, authors in [8] proved under sampling conditions weaker than the one in [11] that an r -offset of the sample provides a topologically correct reconstruction of any shape with a positive μ -reach for some suitable value of the parameter r . It should be noted that this reconstruction can be computed efficiently (*i.e.* in polynomial time). In this paper, we ask the following question: can we weaken further these sampling conditions and still be able to construct a topologically correct reconstruction of a shape from a sample of it?

The starting point of this work was the observation made in [16] that Betti numbers of a shape \mathcal{A} can be derived efficiently from the point set \mathcal{S} , as long as its Hausdorff distance to \mathcal{A} remains smaller than a fourth the *weak feature size* of \mathcal{A} . The weak feature size is another notion of feature size equal to the infimum of distances between points in \mathcal{A} and critical points of the distance function to \mathcal{A} . As critical points form a subset of the μ -medial axis, the weak feature size is larger than the μ -reach for all values of $\mu \in (0, 1]$. Hence, conditions for computing efficiently the Betti numbers of a shape \mathcal{A} are significantly milder than the conditions known for building efficiently a topologically correct approximation of \mathcal{A} . We refer to the mild sampling condition sufficient for inferring Betti numbers as the *weak sampling condition*. This condition is tight.

Optimal reconstruction. We call any algorithm that would be able to produce a topologically correct reconstruction under the weak sampling condition an *optimal reconstruction algorithm*. We explain in Section 3 that, even though no realistic version of an optimal reconstruction algorithm is known today, the weak sampling condition ensures that the sample contains in principle enough information on the sampled shape to produce without ambiguity a topologically correct reconstruction of it. Starting from this observation, we give in Section 4 a “naive” algorithm which, at the expense of not being efficient, produces a reconstruction with the correct Betti numbers under conditions slightly stronger than the weak sampling condition.

The main question we pursue is: can we do better? More precisely, does there exist a polynomial time optimal reconstruction algorithm? This problem is closely related to the persistence-sensitive simplification of real-valued functions, whose goal is to filter out topological noise in sub-level sets. Indeed, reconstruction can be thought of as the simplification of distance functions to the samples. For functions defined on triangulated 2-manifolds, polynomial algorithms have been devised [22, 6, 25]. Still, persistence-sensitive simplification of functions in higher dimension remains elusive.

Homological simplification. In Section 5, we focus on the test at the heart of our naive algorithm. This test requires to answer the following question: given two 2-dimensional simplicial complexes $L \subset K$, does there exist a simplicial complex X containing L and contained in K such that the maps induced by the inclusions $L \hookrightarrow X$ and $X \hookrightarrow K$ on all modulo 2 homology groups are respectively surjective and injective. We call this problem the *homological simplification* of the pair (K, L) and prove that it is NP-complete. Although this result is negative, we believe that it casts new light on the problem of finding a topologically correct reconstruction under weak sampling conditions and opens further research tracks as mentioned in Section 6.

Outline. Section 2 presents the necessary background. Section 3 defines what we mean by an optimal reconstruction algorithm. Section 4 presents an algorithm which can be regarded as an approximation of an optimal reconstruction algorithm. This algorithm requires to be able to solve a problem, which we prove is NP-hard in Section 5. Section 6 concludes the paper.

2 Background

The goal of this section is to recall three closely related concepts useful for expressing sampling conditions in shape reconstruction. Given a shape \mathcal{A} , we define the reach $r_1(\mathcal{A})$, the μ -reach $r_\mu(\mathcal{A})$ for any $\mu \in (0, 1]$ and the weak feature size $\text{wfs}(\mathcal{A})$. As we shall see, these quantities are related by the following inequality: $r_1(\mathcal{A}) \leq r_\mu(\mathcal{A}) \leq \text{wfs}(\mathcal{A})$. All three concepts can be derived from the critical function of the shape. This leads us to introduce the critical function, which requires first to define the norm of the gradient to the distance function.

The distance function to a compact set plays a central role in several recent works related to topologically guaranteed reconstruction [24, 20, 11]. For a compact set $\mathcal{A} \subset \mathbb{R}^N$, the distance function $d_{\mathcal{A}} : \mathbb{R}^N \rightarrow \mathbb{R}^+$ maps every point $q \in \mathbb{R}^N$ to

$$d_{\mathcal{A}}(q) = \min_{a \in \mathcal{A}} \|a - q\|.$$

Although not differentiable, $d_{\mathcal{A}}$ admits several notions of extended gradient [14, 24]. For our purpose, we shall introduce a real valued function $\Psi_{\mathcal{A}} : \mathbb{R}^N \setminus \mathcal{A} \rightarrow [0, 1]$ which coincides with the norm of the gradient defined in [24]. Let $\frac{d}{dt^+}(\cdot)|_{t=0}$ denote the right derivative with respect to the variable t at $t = 0$. For $q \in \mathbb{R}^N \setminus \mathcal{A}$ and $v \in \mathbb{S}^{N-1}$, one can check [24] that the quantity $\frac{d}{dt^+}d_{\mathcal{A}}(q + tv)|_{t=0}$ is well-defined and belongs to $[-1, 1]$. We define $\Psi_{\mathcal{A}}$ as:

$$\Psi_{\mathcal{A}}(q) = \max \left\{ 0, \sup_{v \in \mathbb{S}^{N-1}} \frac{d}{dt^+}d_{\mathcal{A}}(q + tv)|_{t=0} \right\}.$$

Roughly speaking, $\Psi_{\mathcal{A}}(q)$ quantifies at which maximal speed the distance function to \mathcal{A} can increase in a neighborhood of q . We are now ready to recall the definition of the critical function $\chi_{\mathcal{A}}$ introduced in [11]. The critical function maps every positive real number $\rho > 0$ to the infimum of $\Psi_{\mathcal{A}}$ over the set of points at distance ρ from \mathcal{A} :

$$\chi_{\mathcal{A}}(\rho) = \inf_{d_{\mathcal{A}}(q)=\rho} \Psi_{\mathcal{A}}(q).$$

The critical function is lower semi-continuous [11] and allows to define two quantities, the μ -reach and the weak feature size of \mathcal{A} denoted respectively $r_\mu(\mathcal{A})$ and $\text{wfs}(\mathcal{A})$:

$$\begin{aligned} r_\mu(\mathcal{A}) &= \inf \{ \rho > 0, \chi_{\mathcal{A}}(\rho) < \mu \}, \\ \text{wfs}(\mathcal{A}) &= \inf \{ \rho > 0, \chi_{\mathcal{A}}(\rho) = 0 \}. \end{aligned}$$

The reach of \mathcal{A} is equal to $r_1(\mathcal{A})$. From the definition, it is clear that $r_1(\mathcal{A}) \leq r_\mu(\mathcal{A}) \leq \text{wfs}(\mathcal{A})$ for any $\mu \in (0, 1]$. Figure 1 shows the critical function $\chi_{\mathcal{A}}$ for a simple shape \mathcal{A} in the Euclidean plane, which consists of the points at distance R from a full rectangle of width ℓ and length L .

To shed light on these notions, it is useful to make some connections with the medial axis. The *medial axis* of \mathcal{A} is the set of points $q \notin \mathcal{A}$ which have at least two closest points in \mathcal{A} . Alternatively, it is the

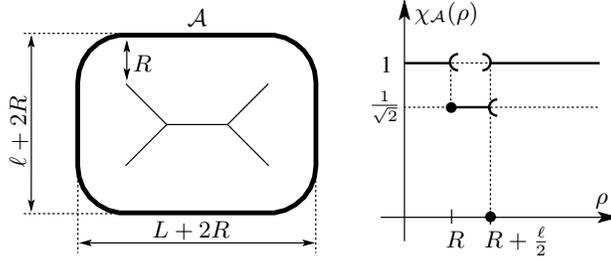


Figure 1: Left: the shape \mathcal{A} is the outer closed thick curve and its medial axis consists of the five thin inner segments. Right: critical function $\chi_{\mathcal{A}}$. We have $r_{\mu}(\mathcal{A}) = R$ for $\mu > \frac{1}{\sqrt{2}}$ and $r_{\mu}(\mathcal{A}) = R + \frac{l}{2} = \text{wfs}(\mathcal{A})$ for $\mu \leq \frac{1}{\sqrt{2}}$.

locus of points q for which $\Psi_{\mathcal{A}}(q) < 1$. Any point q for which $\Psi_{\mathcal{A}}(q) = 0$ is called a *critical point* of the distance function and lies on the medial axis. The reach is the minimum of distances between points in \mathcal{A} and points in its medial axis. The weak feature size is the minimum of distances between points in \mathcal{A} and critical points.

For instance, the function $\Psi_{\mathcal{A}}$ of the shape \mathcal{A} depicted in Figure 1 evaluates to 0 on the horizontal line of the medial axis which constitutes the only critical points in this case, evaluates to $\frac{1}{\sqrt{2}}$ on the other points of the medial axis and evaluates to 1 on all points of the plane that neither belong to \mathcal{A} nor to its medial axis.

For completeness, we also recall the related notion of local feature size, introduced by Amenta [2] for reconstructing smooth shapes. The *local feature size* is a real-valued function which maps every point of \mathcal{A} to its distance to the medial axis. Notice that the local feature size and its infimum, the reach, vanish on non-smooth objects as soon as they contain a sharp concave corner or edge. For this reason, we will focus in Section 3.2 on sampling conditions based on the weak feature size and μ -reaches which apply to a large class of non-smooth shapes.

Given $\eta > 0$, the η -offset of \mathcal{A} is the set of points at distance η or less from \mathcal{A} , $\mathcal{A}^{\eta} = d_{\mathcal{A}}^{-1}([0, \eta])$. As in Morse theory, topological changes in offsets occur only at critical values. More precisely, as stated in [23, 13]:

Lemma 1 (Topological Stability of Offsets). *If $0 < x < y < \text{wfs}(\mathcal{A})$, then the inclusion map $\mathcal{A}^x \hookrightarrow \mathcal{A}^y$ is a homotopy equivalence.*

3 The quest for an optimal reconstruction algorithm

Section 3.1 contains our definition of a (homological) faithful reconstruction, which formalizes what we mean by a “topologically correct reconstruction”. Section 3.2 compares two algorithms for inferring information on a shape \mathcal{A} known through a finite sample \mathcal{S} . The first algorithm outputs the Betti numbers of \mathcal{A} and the second algorithm outputs a faithful approximation of \mathcal{A} . We then define in Section 3.3 an *optimal reconstruction algorithm* as one that would produce the output of the second algorithm with the input and precondition of the first algorithm.

3.1 Faithful reconstructions

To prepare our definition of an optimal reconstruction algorithm, we first introduce in this section the notions of faithful reconstruction and faithful homological reconstruction. For the second notion, we shall consider

a fixed field F and take coefficients in F for homology [26, Chapter 1]. Hence, the property of being a faithful homological reconstruction will depend on the choice of F .

Definition 1 (Faithful (homological) reconstruction). *We say that a subset $\mathcal{R} \subset \mathbb{R}^N$ is a faithful reconstruction of the compact set $\mathcal{A} \subset \mathbb{R}^N$ if there exist real numbers x, y such that $0 < x < y < \text{wfs}(\mathcal{A})$ and the following two properties hold:*

- $\mathcal{A}^x \subset \mathcal{R} \subset \mathcal{A}^y$
- the inclusion maps $\mathcal{A}^x \hookrightarrow \mathcal{R}$ and $\mathcal{R} \hookrightarrow \mathcal{A}^y$ are homotopy equivalences.

We say that \mathcal{R} is a faithful homological reconstruction when the last condition is relaxed to:

- the inclusion maps $\mathcal{A}^x \hookrightarrow \mathcal{R}$ and $\mathcal{R} \hookrightarrow \mathcal{A}^y$ induce isomorphisms on all homology groups.

For any $0 < \eta < \text{wfs}(\mathcal{A})$, the η -offset of \mathcal{A} is clearly a faithful reconstruction of \mathcal{A} . A faithful reconstruction is always a faithful homological reconstruction. As expected, the converse is not true: a punctured Poincaré sphere nested between a point and a ball is an example where inclusions are not homotopy equivalences but yet induce isomorphisms on homology groups [15]. Interestingly, this example does not embed in \mathbb{R}^3 .

Note that in the above definition, if one of the two inclusion maps $\mathcal{A}^x \hookrightarrow \mathcal{R}$ or $\mathcal{R} \hookrightarrow \mathcal{A}^y$ is a homotopy equivalence, so is the other one. Indeed, by Lemma 1, $\mathcal{A}^x \hookrightarrow \mathcal{A}^y$ is a homotopy equivalence and we can conclude by applying Lemma 2 below. A similar statement can be made for the second part of the definition.

Lemma 2. *Consider three nested spaces $\mathcal{A} \subset \mathcal{B} \subset \mathcal{C}$. If two of the three inclusions $i : \mathcal{A} \hookrightarrow \mathcal{B}$, $j : \mathcal{B} \hookrightarrow \mathcal{C}$ and $k = j \circ i : \mathcal{A} \hookrightarrow \mathcal{C}$ are homotopy equivalences, then the third one is a homotopy equivalence also.*

Proof. Let us write $f \simeq g$ if the two maps f and g are homotopic. We consider the three cases in turn. If i and j are homotopy equivalences with homotopy inverses i' and j' respectively, then $i' \circ j'$ is clearly a homotopy inverse of $k = j \circ i$. If j and k are homotopy equivalences with homotopy inverses j' and k' respectively, then using $k = j \circ i$ we get that $j' \circ k = j' \circ j \circ i \simeq i$ and $k' \circ j$ is a homotopy inverse of $i \simeq j' \circ k$. Similarly, if i and k are homotopy equivalences with homotopy inverses i' and k' respectively, then using $k = j \circ i$ we get that $k \circ i' = j \circ i \circ i' \simeq j$ and $i \circ k'$ is a homotopic inverse of $j \simeq k \circ i'$. \square

The following observation will be useful: If x and y are two real numbers such that $\mathcal{A}^x \subset \mathcal{R} \subset \mathcal{A}^y$ with $0 < x < y < \text{wfs}(\mathcal{A})$, then \mathcal{R} is a faithful homological reconstruction of \mathcal{A} if and only if $\mathcal{A}^x \hookrightarrow \mathcal{R}$ induces isomorphisms on all homology groups. This is a direct consequence of Lemmas 1 and 2.

3.2 Comparing existing algorithms

In this section, we present two algorithms and compare their inputs, preconditions and outputs. Specifically, given as input a finite sample \mathcal{S} of an unknown shape \mathcal{A} , the first algorithm recovers the Betti numbers of \mathcal{A} and the second algorithm constructs a faithful approximation of \mathcal{A} . Each algorithm relies on a key theorem that states sampling conditions ensuring correctness. Both algorithms are polynomial in the size of the sample. Recall that the *Hausdorff distance* between two compact sets \mathcal{A} and \mathcal{B} of \mathbb{R}^N is defined by:

$$d_H(\mathcal{A}, \mathcal{B}) = \|d_{\mathcal{B}} - d_{\mathcal{A}}\|_{\infty} = \sup_{q \in \mathbb{R}^N} |d_{\mathcal{B}}(q) - d_{\mathcal{A}}(q)|.$$

Algorithm for computing Betti numbers. A powerful tool for inferring Betti numbers from geometric approximations is topological persistence [21]. Theorem 3 below is a corollary of the Persistence Stability Theorem [16] and can also be derived by flow based arguments [13]. Before stating it, we need the following definition.

Definition 2 (Persistent Betti numbers). *Let $\mathcal{A} \subset \mathbb{R}^N$ be a compact set and let $0 \leq x \leq y$. The p -th (x, y) -persistent Betti number of \mathcal{A} is the rank of the homomorphism induced by inclusion $\mathcal{A}^x \hookrightarrow \mathcal{A}^y$:*

$$\beta_p^{x,y}(\mathcal{A}) = \text{rank}(\mathbf{H}_p(\mathcal{A}^x) \hookrightarrow \mathbf{H}_p(\mathcal{A}^y))$$

It is worth noting that the (x, y) -persistent Betti numbers are finite whenever $x < y$ [18].

Theorem 3 (Homology Inference [16, 13]). *Let \mathcal{A} and \mathcal{S} be two compact subsets of \mathbb{R}^N and suppose there exists a real number $\alpha > 0$ such that*

$$d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4} \text{wfs}(\mathcal{A})$$

Then, $\beta_p(\mathcal{A}) = \beta_p^{\alpha, 3\alpha}(\mathcal{S})$.

The above theorem leads immediately to a polynomial time algorithm for inferring Betti numbers of a shape \mathcal{A} when the sample \mathcal{S} of \mathcal{A} is finite. Indeed, writing $K_\alpha(\mathcal{S})$ for the α -complex of \mathcal{S} , the persistent Betti numbers can be expressed as

$$\beta_p^{\alpha, 3\alpha}(\mathcal{S}) = \text{rank}(\mathbf{H}_p(K_\alpha(\mathcal{S})) \hookrightarrow \mathbf{H}_p(K_{3\alpha}(\mathcal{S}))).$$

In particular, they can be computed in time cubic the size of $K_{3\alpha}(\mathcal{S})$. Since for a fixed dimension, the size of α -complexes is polynomial in the number of vertices, it follows that $\beta_p(\mathcal{A})$ can also be computed in polynomial time the size of the sample.

Algorithm for computing a faithful reconstruction. Suppose we want to reconstruct a shape \mathcal{A} from a sample \mathcal{S} . A standard way to do this is to output an r -offset of \mathcal{S} . In practice, this computation can be replaced by the computation of $K_r(\mathcal{S})$, which shares the same homotopy type. Both computations can be done in polynomial time if the sample is finite. Assuming the shape has a positive μ -reach, it has been proved in [11, 8] that if the Hausdorff distance between \mathcal{A} and \mathcal{S} is less than a fraction the μ -reach of \mathcal{A} , then this method provides indeed a faithful reconstruction of \mathcal{A} for some suitable value of the parameter r . Precisely, setting $\lambda(\mu) = \frac{\mu^2}{5\mu^2+12}$ and $I_\mu(\alpha) = \{\frac{4\alpha}{\mu^2}\}$, we have:

Theorem 4 (Reconstruction Theorem [11, 8]). *Let \mathcal{A} and \mathcal{S} be two compact subsets of \mathbb{R}^N and suppose there exists two real numbers $\alpha > 0$ and $\mu \in (0, 1]$ such that*

$$d_H(\mathcal{S}, \mathcal{A}) < \alpha < \lambda(\mu)r_\mu(\mathcal{A})$$

Then, \mathcal{S}^r is a faithful reconstruction of \mathcal{A} for all $r \in I_\mu(\alpha)$.

The same theorem has been established in [8] but with a larger constant $\lambda(\mu)$, a different interval of admissible values $I_\mu(\alpha)$ and different proof techniques. In both cases, $\lambda(\mu) < \frac{1}{4}$ and $\lim_{\mu \rightarrow 0} \lambda(\mu) = 0$.

Table 1: Input, precondition and output of two polynomial time algorithms derived from Theorem 3 and Theorem 4. The notation “ $\exists \mathcal{A}$ ” stands for “there exists a compact set $\mathcal{A} \subset \mathbb{R}^N$ ”. \mathcal{S} designates a finite set of \mathbb{R}^N . α and μ designate two real numbers with $\alpha > 0$ and $\mu \in (0, 1]$.

Input	Precondition	Output
(\mathcal{S}, α)	$\exists \mathcal{A}, d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4} \text{wfs}(\mathcal{A})$	Betti numbers of \mathcal{A}
$(\mathcal{S}, \alpha, \mu)$	$\exists \mathcal{A}, d_H(\mathcal{S}, \mathcal{A}) < \alpha < \lambda(\mu)r_\mu(\mathcal{A})$	a faithful reconstruction of \mathcal{A}

Comparing sampling conditions. Table 1 summarizes inputs, preconditions and outputs of the two polynomial time algorithms described above and inspired by Theorems 3 and 4. Note that the precondition of the first algorithm is significantly weaker than the precondition of the second one especially when μ is small because $r_\mu(\mathcal{A}) \leq \text{wfs}(\mathcal{A})$, $\lambda(\mu) < \frac{1}{4}$ and $\lambda(\mu)$ tends to zero as $\mu \rightarrow 0$. The gap between the two preconditions leads to the question whether the precondition of the second algorithm can be weakened and replaced by the precondition of the first algorithm. This question motivates our definition of an optimal reconstruction algorithm in the next section.

3.3 Optimal reconstruction algorithms

Note that the precondition required by the first algorithm which we call the *weak precondition* is equivalent to saying that the following set is non-empty:

$$W(\mathcal{S}, \alpha) = \left\{ \mathcal{A} \text{ compact set of } \mathbb{R}^N \mid d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4} \text{wfs}(\mathcal{A}) \right\} \neq \emptyset.$$

By Theorem 3, all shapes in $W(\mathcal{S}, \alpha)$ share the same Betti numbers and the first algorithm returns the Betti numbers of any $\mathcal{A} \in W(\mathcal{S}, \alpha)$. We claim that if the input (\mathcal{S}, α) of the first algorithm satisfies the weak precondition, that is, if $W(\mathcal{S}, \alpha) \neq \emptyset$, then the output of the second algorithm is completely determined as well. To explain this, let us first recall the following theorem from [13]:

Theorem 5 ([13]). *Let \mathcal{A} and \mathcal{B} be two compact subsets of \mathbb{R}^N and $\alpha > 0$ a real number such that*

$$d_H(\mathcal{A}, \mathcal{B}) < 2\alpha < \frac{1}{2} \min \{ \text{wfs}(\mathcal{A}), \text{wfs}(\mathcal{B}) \}$$

Then, $\mathcal{B}^{2\alpha}$ is a faithful reconstruction of \mathcal{A} .

We provide below a quick proof.

Proof. Consider the following diagram in which arrows represent inclusion maps between spaces:

$$\begin{array}{ccccc}
 A_0 & \longrightarrow & A_1 & \longrightarrow & A_2 \\
 & \searrow & \nearrow & & \nearrow \\
 & & & & \\
 & \nearrow & \searrow & & \searrow \\
 B_0 & \longrightarrow & B_1 & \longrightarrow & B_2
 \end{array}$$

It is not too hard to prove that if the horizontal arrows are homotopy equivalences, then the other arrows are homotopy equivalences also. Note that we can always find $\delta > 0$ such that $d_H(\mathcal{A}, \mathcal{B}) < 2\alpha - \delta$. Setting $X_0 = \mathcal{X}^\delta$, $X_1 = \mathcal{X}^{2\alpha}$ and $X_2 = \mathcal{X}^{4\alpha - \delta}$ for $X \in \{A, B\}$, we get immediately that $\mathcal{B}^{2\alpha}$ is a faithful reconstruction of \mathcal{A} . \square

Suppose \mathcal{A} and \mathcal{B} both belong to $W(\mathcal{S}, \alpha)$. Applying a triangular inequality, we get that \mathcal{A} , \mathcal{B} and α fulfill conditions of Theorem 5 and therefore, $\mathcal{B}^{2\alpha}$ is a faithful reconstruction of \mathcal{A} . Hence, any 2α -offset of a shape $\mathcal{B} \in W(\mathcal{S}, \alpha)$ is a faithful reconstruction of any shape $\mathcal{A} \in W(\mathcal{S}, \alpha)$. For this reason, we say that if the pair (\mathcal{S}, α) satisfies the weak precondition, then it carries in principle enough information about the unknown shape \mathcal{A} to determine without ambiguity a faithful reconstruction of it.

Furthermore, the weak precondition is tight. To explain what this means, let us introduce the set

$$W(\mathcal{S}, \alpha, \eta) = \left\{ \mathcal{A} \text{ compact set of } \mathbb{R}^N \mid d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4}(\text{wfs}(\mathcal{A}) - \eta) \right\}.$$

For negative values of η , the set $W(\mathcal{S}, \alpha, \eta)$ is a superset of $W(\mathcal{S}, \alpha)$. We claim that for any $\eta < 0$, $W(\mathcal{S}, \alpha, \eta)$ may contain objects that do not have the same homology. To construct such an example, consider the two shapes \mathcal{O} and \mathcal{U} described in [13] and the sample \mathcal{S} pictured on Figure 2. By construction, we have $\text{wfs}(\mathcal{O}) = 2$, $\text{wfs}(\mathcal{U}) = +\infty$ and adjust the angle between the two bars in shape \mathcal{U} so that $d_H(\mathcal{S}, \mathcal{O}) = d_H(\mathcal{S}, \mathcal{U}) < \frac{1}{2} - \frac{\eta}{8}$ for some $\eta < 0$. Both \mathcal{O} and \mathcal{U} belong to $W(\mathcal{S}, \frac{1}{2} - \frac{\eta}{4}, \eta)$ but $\beta_1(\mathcal{O}) \neq \beta_1(\mathcal{U})$. Therefore, the weak precondition is the weakest amongst the preconditions expressed in terms of Hausdorff distance and critical functions that allows to retrieve a faithful reconstruction without ambiguity.

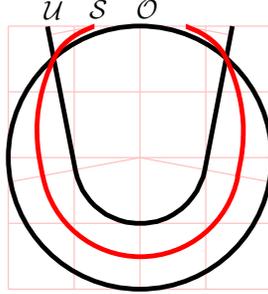


Figure 2: The angle between the two bars in shape \mathcal{U} is adjusted such that $d_H(\mathcal{S}, \mathcal{O}) = d_H(\mathcal{S}, \mathcal{U}) < \frac{1}{2} - \frac{\eta}{8}$.

We are now ready to define what we mean by an optimal reconstruction algorithm.

Definition 3 (Optimal reconstruction algorithm). *We call optimal reconstruction algorithm any algorithm that takes as input a pair (\mathcal{S}, α) satisfying the weak precondition, that is, such that $W(\mathcal{S}, \alpha) \neq \emptyset$, and outputs a set which is a faithful reconstruction of all the shapes in $W(\mathcal{S}, \alpha)$.*

The main question motivating this work is whether there exists a polynomial time optimal reconstruction algorithm. Given as input a pair (\mathcal{S}, α) satisfying the weak precondition, the previous discussion suggests the following strategy for computing a faithful reconstruction: enumerate all compact sets in \mathbb{R}^N and return the 2α -offset of the first compact set \mathcal{B} that belongs to $W(\mathcal{S}, \alpha)$. Indeed, by Theorem 5, we know that the output $\mathcal{B}^{2\alpha}$ is a faithful reconstruction of every shape $\mathcal{A} \in W(\mathcal{S}, \alpha)$. Of course, this procedure is unrealistic and the goal of the next section is to present an effective version of it. To achieve this goal, we will replace the search of a faithful reconstruction by the search of a faithful homological reconstruction and strengthen slightly the weak precondition, assuming instead that $W(\mathcal{S}, \alpha, \eta) \neq \emptyset$ for some arbitrarily small $\eta > 0$.

4 Naive algorithms for homological reconstruction

In this section, we describe a naive algorithm that outputs a faithful homological reconstruction under conditions slightly stronger than the weak precondition. We call it “naive” since it has an unbounded time complexity. The idea is to explore a set of cubical sets, refining the size of the cubes until we find a solution.

We proceed in four steps. Given a set \mathcal{S} that samples a shape \mathcal{A} , we first prove the existence of a cubical set that can be derived from the sample \mathcal{S} and which is a faithful reconstruction of the shape \mathcal{A} . Second, we discuss a simple test to decide whether a cubical set is a faithful homological reconstruction of a shape. Based on this test, we then give a reconstruction algorithm for shapes with a positive μ -reach (NAIVE_1) and finally, derive a reconstruction algorithm for shapes with a lower bounded weak feature size (NAIVE_2).

4.1 Cubical sets

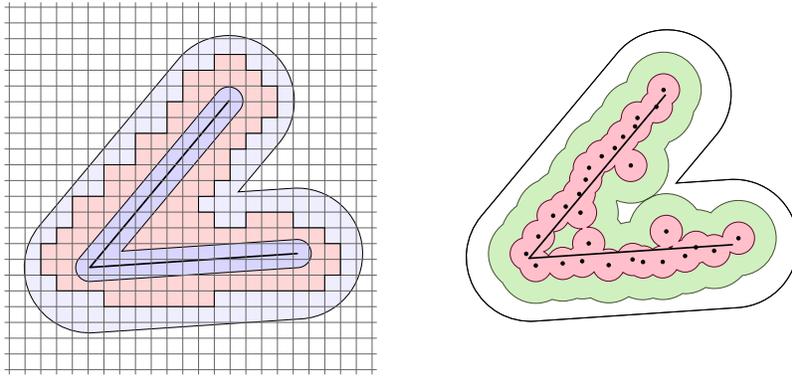


Figure 3: Left: the cubical set (in pink) is nested between two offsets (in light purple) of the V-shaped black curve and is a faithful reconstruction of it. Right: two offsets \mathcal{S}^l and \mathcal{S}^k of the sample.

Let us start with some definitions. An ε -voxel is a closed cube with edge length ε and whose vertices belong to the lattice $\varepsilon\mathbb{Z}^N$. We call any finite union of ε -voxels an ε -cubical set. Let $\mathcal{S} \subset \mathbb{R}^N$ be a compact set and consider three real numbers $\alpha > 0$, $\mu \in (0, 1]$ and $\eta > 0$. The goal of this section is to prove the existence of cubical sets that are faithful reconstructions of all shapes in the set

$$W(\mathcal{S}, \alpha, \eta, \mu) = \left\{ \mathcal{A} \text{ compact set of } \mathbb{R}^N \mid d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4}(r_\mu(\mathcal{A}) - \eta) \right\}.$$

We proceed in three phases. First, we recall a result from [7] that states the existence of cubical sets that are faithful reconstructions of shapes with a positive reach (Lemma 6). We then deduce the existence of cubical sets that are faithful reconstructions of shapes with a positive μ -reach (Lemma 7) before proving the same for shapes in $W(\mathcal{S}, \alpha, \eta, \mu)$ (Lemma 8).

Lemma 6 (Corollary 3 in [7]). *For $d_N = \frac{1}{40N^3\lceil\sqrt{N}\rceil}$ and for all compact sets $\mathcal{A} \subset \mathbb{R}^N$ with reach greater than $\rho > 0$, there exists a $(d_N\rho)$ -cubical set \mathcal{X} such that $\mathcal{A} \subset \mathcal{X} \subset \mathcal{A}^\rho$ and the inclusion maps $\mathcal{A} \hookrightarrow \mathcal{X}$ and $\mathcal{X} \hookrightarrow \mathcal{A}^\rho$ are homotopy equivalences.*

Using the above lemma, we derive the existence of cubical sets which are faithful reconstructions of shapes with a positive μ -reach.

Lemma 7. *There exists a positive constant c_N depending only upon the ambient dimension N such that the following property holds: for all real numbers x, y and $\mu \in (0, 1]$ and all compact sets $\mathcal{A} \subset \mathbb{R}^N$ satisfying $r_\mu(\mathcal{A}) > y > x > 0$, there exists a $(c_N \mu(y - x))$ -cubical set \mathcal{X} such that $\mathcal{A}^x \subset \mathcal{X} \subset \mathcal{A}^y$ and the inclusion maps $\mathcal{A}^x \hookrightarrow \mathcal{X}$ and $\mathcal{X} \hookrightarrow \mathcal{A}^y$ are homotopy equivalences. In particular, the cubical set \mathcal{X} is a faithful reconstruction of \mathcal{A} (see Figure 3, left).*

Proof. The proof consists in extending Lemma 6 to the situation where compact sets have a positive μ -reach with the constant $c_N = \frac{d_N}{2}$. The key ingredient in the proof is a result in [12] which says that if we dilate a shape with a positive μ -reach and then erode it again, we can adjust the parameters of the dilation and erosion in such a way that the resulting shape has a positive reach. Precisely, given a set $\mathcal{Y} \subset \mathbb{R}^N$, we denote respectively by $\overline{\mathcal{Y}}$ and \mathcal{Y}^c the closure and the complement of \mathcal{Y} . For any compact set $\mathcal{Y} \subset \mathbb{R}^N$ and any real number $\rho > 0$, let $\mathcal{Y}^{-\rho} = \overline{((\mathcal{Y}^c)^\rho)^c}$ and consider the set $\mathcal{B} = (\mathcal{A}^y)^{-\mu(y-x)}$. We know from [12] that the reach of \mathcal{B} is greater than or equal to $\mu(y - x)$ and the inclusion maps corresponding to the sequence

$$\mathcal{A}^x \subset \mathcal{B} \subset \mathcal{A}^y$$

are homotopy equivalences. We can now apply Lemma 6 to the set \mathcal{B} whose reach is greater than $\rho = \frac{\mu(y-x)}{2}$. This gives the existence of a $(c_N \mu(y - x))$ -cubical set \mathcal{X} such that:

$$\mathcal{B} \subset \mathcal{X} \subset \mathcal{B}^\rho$$

and the maps corresponding to inclusions are homotopy equivalences. Using $\mathcal{B}^\rho = ((\mathcal{A}^y)^{-2\rho})^\rho \subset \mathcal{A}^y$, we get the sequence of inclusions

$$\mathcal{A}^x \subset \mathcal{X} \subset \mathcal{A}^y.$$

in which the inclusion map $\mathcal{A}^x \hookrightarrow \mathcal{X}$ is a homotopy equivalence. By Lemma 2, \mathcal{X} is a faithful reconstruction of \mathcal{A} . \square

Unfortunately, the shape is only known through a finite set of points that sample it. Nonetheless, next lemma states that we can deduce from the mere knowledge of the sample a faithful reconstruction of the underlying shape which is a cubical set. Recall that $V_\varepsilon(\mathcal{Y})$ denotes the union of ε -voxels that intersect the set $\mathcal{Y} \subset \mathbb{R}^N$.

Lemma 8. *Let $\alpha, \eta > 0$ and $\mu \in (0, 1]$ be real numbers and let \mathcal{A} and \mathcal{S} be compact subsets of \mathbb{R}^N such that $d_H(\mathcal{S}, \mathcal{A}) < \alpha$. Then, for:*

$$\varepsilon = \frac{\eta}{4\sqrt{N} + \frac{2}{c_N \mu}}, \quad l = \frac{\eta}{2} + \alpha, \quad k = \eta + 3\alpha - \varepsilon\sqrt{N},$$

we have the sequence of inclusions:

$$\mathcal{A}^{\frac{\eta}{2}} \subset V_\varepsilon(\mathcal{S}^l) \subset \mathcal{A}^{\frac{3\eta}{4} + 2\alpha} \subset V_\varepsilon(\mathcal{S}^k) \subset \mathcal{A}^{\eta + 4\alpha}.$$

Furthermore, if we assume $d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4}(r_\mu(\mathcal{A}) - \eta)$, there exists an ε -cubical set \mathcal{X} such that:

$$\mathcal{A}^{\frac{\eta}{2}} \subset V_\varepsilon(\mathcal{S}^l) \subset \mathcal{X} \subset V_\varepsilon(\mathcal{S}^k) \subset \mathcal{A}^{\eta + 4\alpha}$$

and the inclusion maps $\mathcal{A}^{\frac{\eta}{2}} \hookrightarrow \mathcal{X}$ and $\mathcal{X} \hookrightarrow \mathcal{A}^{\eta + 4\alpha}$ are homotopy equivalences. In particular, \mathcal{X} is a faithful reconstruction of \mathcal{A} .

Proof. Note that for all compact sets $\mathcal{Y} \subset \mathbb{R}^N$, we have $\mathcal{Y} \subset V_\varepsilon(\mathcal{Y}) \subset \mathcal{Y}^{\varepsilon\sqrt{N}}$. It follows that for all $t \geq 0$, we have the following sequence of inclusions:

$$\mathcal{A}^t \subset \mathcal{S}^{t+\alpha} \subset V_\varepsilon(\mathcal{S}^{t+\alpha}) \subset \mathcal{S}^{t+\alpha+\varepsilon\sqrt{N}} \subset \mathcal{A}^{t+2\alpha+\varepsilon\sqrt{N}}.$$

Applying this sequence twice, once for $t = \frac{\eta}{2}$ and once for $t = \eta + 2\alpha - \varepsilon\sqrt{N}$, we get that

$$\mathcal{A}^{\frac{\eta}{2}} \subset V_\varepsilon(\mathcal{S}^{\frac{\eta}{2}+\alpha}) \subset \mathcal{A}^{\frac{\eta}{2}+2\alpha+\varepsilon\sqrt{N}} \subset \mathcal{A}^{\eta+2\alpha-\varepsilon\sqrt{N}} \subset V_\varepsilon(\mathcal{S}^{\eta+3\alpha-\varepsilon\sqrt{N}}) \subset \mathcal{A}^{\eta+4\alpha}.$$

The value ε has been chosen precisely such that the parameters of the two offsets of \mathcal{A} in the middle differ by $\frac{\varepsilon}{c_N\mu}$. Specifically, writing $x = \frac{\eta}{2} + 2\alpha + \varepsilon\sqrt{N}$ and $y = \eta + 2\alpha - \varepsilon\sqrt{N}$, we have $y - x = \frac{\varepsilon}{c_N\mu}$. Hence, applying Lemma 7 to \mathcal{A} , we get the existence of an ε -cubical set \mathcal{X} such that $\mathcal{A}^x \subset \mathcal{X} \subset \mathcal{A}^y$ and the maps corresponding to the inclusions are homotopy equivalences. The lemma follows. \square

4.2 Homological simplification

Almost all pieces are in place to write a reconstruction algorithm. Given as input a sample \mathcal{S} of a shape \mathcal{A} , Lemma 8 suggests to enumerate all cubical sets nested between the two cubical sets $\mathcal{L} = V_\varepsilon(\mathcal{S}^l)$ and $\mathcal{K} = V_\varepsilon(\mathcal{S}^k)$ and stop as soon as we find a faithful homological reconstruction (see Figure 4). Yet, we still need to discuss how to recognize that a cubical set \mathcal{X} nested between cubical sets \mathcal{L} and \mathcal{K} is actually a faithful homological reconstruction of shape \mathcal{A} . For this, we will suppose that simplicial complexes $L \subset X \subset K$ triangulate cubical sets $\mathcal{L} \subset \mathcal{X} \subset \mathcal{K}$ and characterizes $\mathcal{X} = |\mathcal{X}|$ using the notion of homological simplification introduced below.

Definition 4 (Homological simplification). *Let $L \subset K$ be two simplicial complexes. The simplicial complex X is said to be a homological simplification of the pair (K, L) if $L \subset X \subset K$ and the maps $j_* : \mathbf{H}_p(L) \rightarrow \mathbf{H}_p(X)$ and $i_* : \mathbf{H}_p(X) \rightarrow \mathbf{H}_p(K)$ induced by inclusions are respectively surjective and injective for all integers $p \geq 0$.*

A useful observation is that since we are working with coefficients in F and homology groups are finite-dimensional vector spaces, X is a homological simplification of the pair (K, L) if and only if X realizes the persistent homology of L into K . This is a consequence of the following lemma:

Lemma 9. *For any sequence of finite-dimensional vector spaces $U \rightarrow V \rightarrow W$, the map $U \rightarrow V$ is surjective and the map $V \rightarrow W$ is injective if and only if $\text{Rank}(U \rightarrow W) = \dim(V)$.*

Proof. Indeed, if $j : U \rightarrow V$ is surjective and $i : V \rightarrow W$ is injective then

$$\text{Rank}(i \circ j) = \dim(i \circ j(U)) = \dim(i(V)) = \dim(V).$$

Conversely, note that

$$\begin{aligned} \dim(V) &\geq \text{Rank}(i) \geq \text{Rank}(i \circ j), \\ \dim(V) &\geq \text{Rank}(j) \geq \text{Rank}(i \circ j). \end{aligned}$$

Thus, if $\dim(V) = \text{Rank}(i \circ j)$, then $\dim(V) = \text{Rank}(i) = \text{Rank}(j)$ and so i is injective and j surjective. \square

The next lemma follows immediately.

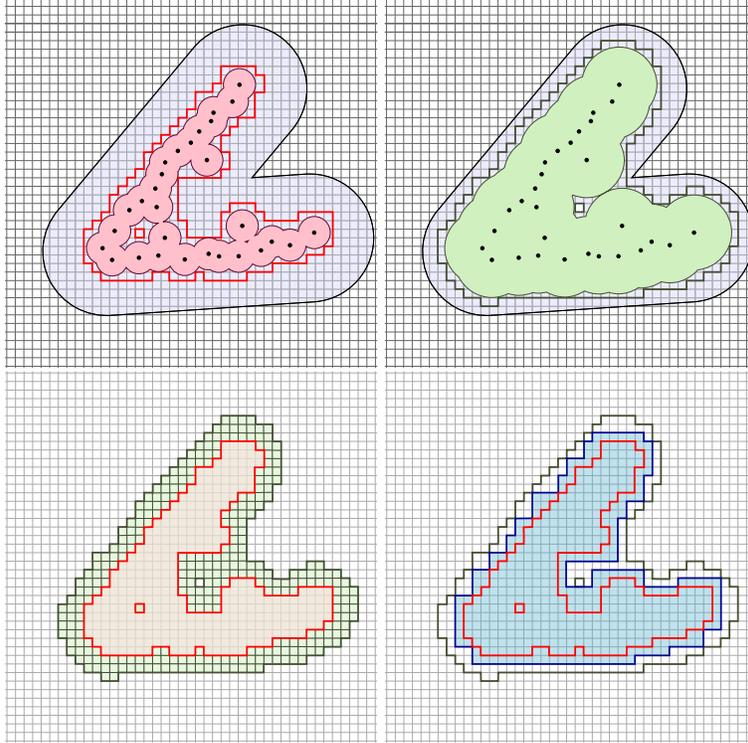


Figure 4: Top: boundaries of $\mathcal{L} = V_\varepsilon(S^l)$ and $\mathcal{K} = V_\varepsilon(S^k)$ are depicted in red and dark green. Bottom: the cubical set \mathcal{X} in blue is nested between \mathcal{L} and \mathcal{K} and is a faithful homological reconstruction of \mathcal{A} .

Lemma 10. Consider a sequence of simplicial complexes $L \subset X \subset K$. The simplicial complex X is a homological simplification of the pair (K, L) if and only if $\mathbf{H}_p(X)$ is isomorphic to the image of the homomorphism $\mathbf{H}_p(L) \rightarrow \mathbf{H}_p(K)$ induced by the inclusion $L \subset K$, for all integers $p \geq 0$.

Lemma 11. Let x_1, x_2, x_3 be three real numbers and \mathcal{A} a compact subset of \mathbb{R}^N such that $0 < x_1 < x_2 < x_3 < \text{wfs}(\mathcal{A})$. Let $L \subset K$ be two simplicial complexes such that:

$$\mathcal{A}^{x_1} \subset |L| \subset \mathcal{A}^{x_2} \subset |K| \subset \mathcal{A}^{x_3}.$$

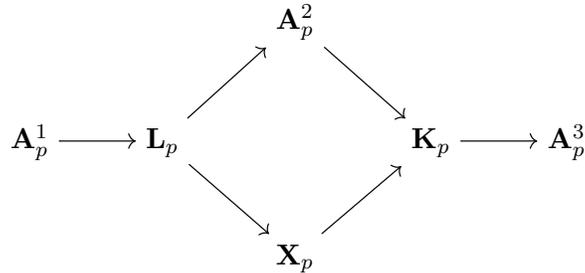
Suppose X is a simplicial complex such that $L \subset X \subset K$. Then, X is a homological simplification of the pair (K, L) if and only if $|X|$ is a faithful homological reconstruction of \mathcal{A} .

Proof. First observe that for any diagram of vector spaces $\mathbf{A}^1 \rightarrow \mathbf{L} \rightarrow \mathbf{A}^2 \rightarrow \mathbf{K} \rightarrow \mathbf{A}^3$ where the maps $\mathbf{A}^1 \rightarrow \mathbf{A}^2 \rightarrow \mathbf{A}^3$ are isomorphisms, we have $\text{Rank}(\mathbf{L} \rightarrow \mathbf{K}) = \dim(\mathbf{A}^i)$ for all $i \in \{1, 2, 3\}$. Indeed, $\mathbf{A}^1 \rightarrow \mathbf{A}^2$ bijective implies that $\mathbf{L} \rightarrow \mathbf{A}^2$ surjective and $\mathbf{A}^2 \rightarrow \mathbf{A}^3$ bijective implies that $\mathbf{A}^2 \rightarrow \mathbf{K}$ injective. We conclude by applying Lemma 9 to the diagram $\mathbf{L} \rightarrow \mathbf{A}^2 \rightarrow \mathbf{K}$. Consider now the diagram of vector

Table 2: Naive reconstruction algorithms. $|X|$ denotes the underlying space of the simplicial complex X . $sc(\mathcal{X})$ denotes a triangulation of the cubical set \mathcal{X} compatible with inclusion.

<pre> NAIVE_1 ($\mathcal{S}, \alpha, \eta, \mu$) PRECONDITION: $\exists \mathcal{A}, d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4}(\text{wfs}(\mathcal{A}) - \eta)$ OUTPUT: either \emptyset or a faithful homological reconstruction of all $\mathcal{A} \in W(\mathcal{S}, \alpha, \eta)$. $\varepsilon \leftarrow \frac{\eta}{4\sqrt{N}+2/(c_N\mu)}$ $l \leftarrow \frac{\eta}{2} + \alpha; \quad k \leftarrow \eta + 3\alpha - \varepsilon\sqrt{N}$ $L \leftarrow sc(V_\varepsilon(\mathcal{S}^l)); \quad K \leftarrow sc(V_\varepsilon(\mathcal{S}^k))$ for all (X such that $L \subset X \subset K$) if (X is a homological simplification of (K, L)) return X end for return \emptyset </pre>	<pre> NAIVE_2 ($\mathcal{S}, \alpha, \eta$) PRECONDITION: $\exists \mathcal{A}, d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4}(\text{wfs}(\mathcal{A}) - \eta)$ OUTPUT: a faithful homological reconstruction of all $\mathcal{A} \in W(\mathcal{S}, \alpha, \eta)$. $\mu \leftarrow 1$ $\alpha' \leftarrow \alpha + \frac{\eta}{8}$ $\eta' \leftarrow \frac{\eta}{4}$ while (TRUE) $\mathcal{X} \leftarrow \text{NAIVE_1}(\mathcal{S}, \alpha', \eta', \mu)$ if ($\mathcal{X} \neq \emptyset$) return \mathcal{X} $\mu \leftarrow \frac{\mu}{2}$ end while </pre>
--	---

spaces



in which $\mathbf{A}_p^i = \mathbf{H}_p(\mathcal{A}^{x_i})$, $\mathbf{L}_p = \mathbf{H}_p(|L|)$, $\mathbf{K}_p = \mathbf{H}_p(|K|)$, $\mathbf{X}_p = \mathbf{H}_p(|X|)$, and the arrows represent inclusion maps. We prove the lemma by establishing equivalences between the following five statements:

- (i) X is a homological simplification of the pair (K, L) ;
- (ii) $\mathbf{L}_p \rightarrow \mathbf{X}_p$ is surjective and $\mathbf{X}_p \rightarrow \mathbf{K}_p$ is injective for all $p \geq 0$;
- (iii) $\dim(\mathbf{X}_p) = \text{Rank}(\mathbf{L}_p \rightarrow \mathbf{K}_p) = \dim(\mathbf{A}_p^i)$ for all $i \in \{1, 2, 3\}$ and all $p \geq 0$;
- (iv) all maps $\mathbf{A}_p^1 \rightarrow \mathbf{X}_p \rightarrow \mathbf{A}_p^3$ are isomorphisms for all $p \geq 0$;
- (v) $|X|$ is a faithful homological reconstruction.

By definition of a homological simplification, (i) is equivalent to (ii). By Lemma 9, (ii) is equivalent to (iii). To prove (iii) \implies (iv), we just need to observe that $\mathbf{A}_p^1 \rightarrow \mathbf{A}_p^3$ is a bijection. The reverse implication is obvious. By definition of a faithful homological reconstruction and using the observation at the end of Section 3.1, (iv) is equivalent to (v). \square

4.3 First naive reconstruction algorithm

We are now ready to describe our first reconstruction algorithm `NAIVE_1`. Its pseudocode is given in Table 2, left. Recall that $V_\varepsilon(\mathcal{Y})$ designates the union of ε -voxels that intersect the subset $\mathcal{Y} \subset \mathbb{R}^N$. Given as input

the 4-tuple $(\mathcal{S}, \alpha, \eta, \mu)$, the algorithm proceeds as follows. It chooses a voxel size ε , two offset parameters l and k (see Table 2 for the exact values of ε , l and k) and derives from the sample \mathcal{S} two ε -cubical sets $\mathcal{L} = V_\varepsilon(\mathcal{S}^l)$ and $\mathcal{K} = V_\varepsilon(\mathcal{S}^k)$, obtained by collecting all ε -voxels intersecting respectively \mathcal{S}^l and \mathcal{S}^k (see Figures 3 and 4). For all cubical sets \mathcal{X} containing \mathcal{L} and contained in \mathcal{K} , the algorithm then considers three nested simplicial complexes $L \subset X \subset K$ triangulating the three cubical sets $\mathcal{L} \subset \mathcal{X} \subset \mathcal{K}$ in a way that is consistent with the grid. It then returns the underlying space \mathcal{X} of X if the simplicial complex X is a homological simplification of the pair (K, L) (see definition above). If no homological simplification X is found between L and K , the algorithm returns the empty set.

Theorem 12. *Let $\mathcal{S} \subset \mathbb{R}^N$, $\alpha > 0$, $\mu \in (0, 1]$, and $\eta > 0$. Assuming the precondition $W(\mathcal{S}, \alpha, \eta) \neq \emptyset$ on the input, the algorithm `NAIVE_1` outputs either the empty set or a faithful homological reconstruction of all shapes in $W(\mathcal{S}, \alpha, \eta)$. If furthermore we assume the stronger precondition $W(\mathcal{S}, \alpha, \eta, \mu) \neq \emptyset$ on the input, the algorithm `NAIVE_1` does not return the empty set.*

Proof. The correctness of the algorithm `NAIVE_1` relies on the lemmas stated in the previous sections. Let $\mathcal{A} \in W(\mathcal{S}, \alpha, \eta)$. Equivalently, $d_H(\mathcal{S}, \mathcal{A}) < \alpha$ and $\eta + 4\alpha < \text{wfs}(\mathcal{A})$. By Lemma 8, we thus have the sequence of inclusions:

$$\mathcal{A}^{\frac{\eta}{2}} \subset |L| \subset \mathcal{A}^{\frac{3\eta}{4} + 2\alpha} \subset |K| \subset \mathcal{A}^{\eta + 4\alpha},$$

with $\eta + 4\alpha < \text{wfs}(\mathcal{A})$. Lemma 11 then implies that if X is a homological simplification of the pair (K, L) , its underlying space $\mathcal{X} = |X|$ is a faithful homological reconstruction of \mathcal{A} . Furthermore, if $\mathcal{A} \in W(\mathcal{S}, \alpha, \eta, \mu)$ or equivalently if $d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4}(r_\mu(\mathcal{A}) - \eta)$, Lemmas 8 and 11 guarantee that the algorithm returns a faithful homological simplification of \mathcal{A} (and not \emptyset). \square

Let us bound the time complexity of a more efficient version of the algorithm in which voxels are not decomposed into simplices. Let D be the diameter of \mathcal{S} and set $D' = D + 2(\eta + 3\alpha)$. It is not difficult to check that this simpler algorithm has time complexity $O(2^{|K|}|K|^3) = O\left(2^{\left(\frac{D'}{\varepsilon}\right)^N} \left(\frac{D'}{\varepsilon}\right)^{3N}\right)$. Indeed, the size of K is $O((D'/\varepsilon)^N)$. Checking if X is a homological simplification of (K, L) takes cubic time the size of K and the number of cubical sets \mathcal{X} between \mathcal{L} and \mathcal{K} is $O(2^{|K|})$. If the voxels are decomposed into simplices, the running time increases but remains finite.

4.4 Second naive reconstruction algorithm

We now describe our second reconstruction algorithm `NAIVE_2`. Its pseudocode is given in Table 2, right. The algorithm takes as input a triplet $(\mathcal{S}, \alpha, \eta)$. Starting with $\mu = 1$, it calls `NAIVE_1` with decreasing values of μ until `NAIVE_1` returns a non-empty set.

Theorem 13. *Let $\mathcal{S} \subset \mathbb{R}^N$, $\alpha > 0$, and $\eta > 0$. Assuming the precondition $W(\mathcal{S}, \alpha, \eta) \neq \emptyset$ on the input, the algorithm `NAIVE_2` outputs a faithful homological reconstruction of all shapes in $W(\mathcal{S}, \alpha, \eta)$ after a finite number of iterations.*

Proof. The algorithm terminates thanks to the lower semi-continuity of the critical function $\chi_{\mathcal{A}}$. Suppose $W(\mathcal{S}, \alpha, \eta) \neq \emptyset$ and let $\mathcal{A} \in W(\mathcal{S}, \alpha, \eta)$, i.e. such that $d_H(\mathcal{S}, \mathcal{A}) < \alpha < \frac{1}{4}(\text{wfs}(\mathcal{A}) - \eta)$. Because $\chi_{\mathcal{A}}$ is lower semi-continuous, it attains its minimum $\mu' > 0$ over the interval $[\frac{\eta}{8}, 4\alpha + \frac{7\eta}{8}]$. Setting $\mathcal{A}' = \mathcal{A}^{\frac{\eta}{8}}$,

$\alpha' = \alpha + \frac{\eta}{8}$ and $\eta' = \frac{\eta}{4}$, we have $r_{\mu'}(\mathcal{A}') > 4\alpha + \frac{3\eta}{4} = 4\alpha' + \eta'$ (see Figure 5 for an explanation) and therefore

$$d_H(\mathcal{S}, \mathcal{A}') \leq d_H(\mathcal{S}, \mathcal{A}) + d_H(\mathcal{A}, \mathcal{A}') < \alpha + \frac{\eta}{8} = \alpha' < \frac{1}{4} (r_{\mu'}(\mathcal{A}') - \eta').$$

It follows that for all $0 < \mu \leq \mu'$ we have $\mathcal{A}' \in W(\mathcal{S}, \alpha', \eta', \mu) \neq \emptyset$ because $r_{\mu'}(\mathcal{A}') \leq r_{\mu}(\mathcal{A}')$. Thus, at some point, the algorithm `NAIVE_1` will be called with input $(\mathcal{S}, \alpha', \eta', \mu)$ satisfying $W(\mathcal{S}, \alpha', \eta', \mu) \neq \emptyset$ and by Theorem 12 will return a non-empty set to the algorithm `NAIVE_2`. When this happens, the result is a faithful reconstruction of every shape in $W(\mathcal{S}, \alpha', \eta')$ and in particular of \mathcal{A} since $d_H(\mathcal{S}, \mathcal{A}) < \alpha' < \frac{1}{4}(\text{wfs}(\mathcal{A}) - \eta')$ as can be easily checked. \square

Note that algorithm `NAIVE_1` can be considered as an approximation of an optimal reconstruction algorithm. Even though the algorithm terminates, its time complexity is unbounded.

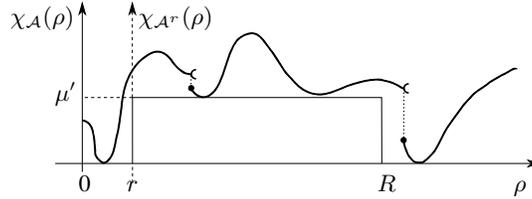


Figure 5: Performing an r -offset translates the critical function to the left by r [11]. Thus, $\chi_{\mathcal{A}}(\rho) \geq \mu'$ on $[r, R]$ implies $r_{\mu'}(\mathcal{A}^r) > R - r$.

5 Homological simplification is NP-complete

In this section, we focus on the problem of computing a homological simplification and prove that this problem is NP-complete, at least when $F = \mathbb{Z}_2$. We denote the p -th homology group of K by $\mathbf{H}_p(K)$ and work with coefficients in the field \mathbb{Z}_2 of integers modulo 2. A *simplicial pair* (K, L) consists of a (finite) simplicial complex K and a subcomplex $L \subset K$. When clear from the context, we will simply speak of the pair (K, L) and omit “simplicial”. We say that the pair (K, L) is p -dimensional if the simplicial complex K has dimension p .

Definition 5 (Homological simplification problem). *The homological simplification problem takes as input a simplicial pair (K, L) and asks whether there exists a simplicial complex X which is a homological simplification of the pair (K, L) .*

The size of the problem is the number of simplices in K . We are now ready to state our main theorem:

Theorem 14. *The homological simplification problem of 2-dimensional simplicial pairs is NP-complete.*

Proof. To check that a candidate X is a homological simplification of the p -dimensional pair (K, L) , it is enough to compute the dimension of the p -th homology group of X and compare it to the rank of the persistent p -th homology group of K into L , for all p . Since all computations can be done in time cubic in the number of simplices in K , we deduce that the homological simplification problem of p -dimensional simplicial pairs is in NP. In Section 5.1, we prove that this problem is NP-hard for $p = 2$ by reducing 3SAT to it in polynomial time. Figure 6 summarizes the reduction. \square

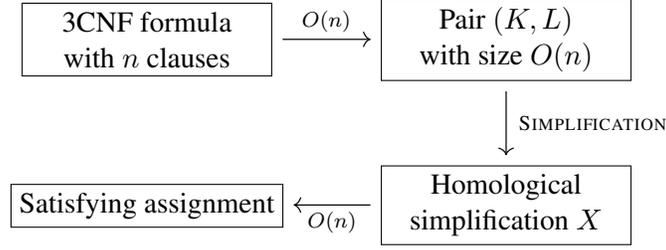


Figure 6: Diagram of the reduction.

5.1 Reduction from 3SAT

3SAT. A Boolean formula E is in *3-conjunctive normal form*, or 3CNF, if it is a conjunction (AND) of n clauses c_1, c_2, \dots, c_n , each of which is a disjunction (OR) of three literals, each literal being a Boolean variable or its negation [17]. Specifically, $E = \bigwedge_{1 \leq i \leq n} c_i$ and each clause c_i has the form

$$c_i = \left(e_i^1 v_{j_i^1} \right) \vee \left(e_i^2 v_{j_i^2} \right) \vee \left(e_i^3 v_{j_i^3} \right),$$

where $j_i^k \in \{1, \dots, m\}$, $v_{j_i^k}$ is a Boolean variable and $e_i^k \in \{\mathbf{1}, \neg\}$ is either the identity symbol $\mathbf{1}$ or the negation symbol \neg , for $1 \leq k \leq 3$. The 3SAT problem takes as input a 3CNF formula E and determines whether one can assign a value TRUE or FALSE to each variable of E such that E evaluates to TRUE. An assignment of variables which makes E evaluates to TRUE is called a *satisfying assignment*. Since the number m of variables used in formula E is at most three times the number n of clauses, *i.e.* $m \leq 3n$, we let n be the size of the 3SAT problem. 3SAT is known to be NP-complete.

Reduction algorithm. We describe a reduction algorithm that transforms in linear time any instance E of the 3SAT problem into an instance (K, L) of the homological simplification problem in such a way that (K, L) has a homological simplification if and only if E has a satisfying assignment. Given a 3CNF formula E of n clauses c_1, \dots, c_n and m variables v_1, \dots, v_m , we construct a 2-dimensional simplicial pair (K, L) as follows; see Figure 7. The simplicial complex L consists of

- a vertex A ;
- two vertices B_i and C_i and three edges AB_i , B_iC_i and C_iA for each clause c_i ;
- two vertices V_j and W_j and the edge V_jW_j for each variable v_j .

Besides simplices in L , the simplicial complex K contains three triangles per literal and two edges per variable. Specifically, if $e_i^k = \mathbf{1}$, we add the three triangles $AB_iV_{j_i^k}$, $B_iC_iV_{j_i^k}$ and $C_iAV_{j_i^k}$ and their edges. If $e_i^k = \neg$, we add the three triangles $AB_iW_{j_i^k}$, $B_iC_iW_{j_i^k}$ and $C_iAW_{j_i^k}$ and their edges. Moreover, we add edges AV_j and AW_j for all $j \in \{1, \dots, m\}$. Observe that the size of K is only a constant factor larger than the size of E and its construction requires linear time in n .

Let $f_* : \mathbf{H}_p(L) \rightarrow \mathbf{H}_p(K)$ be the homomorphism induced by the inclusion $L \subset K$. Since K is connected, we have $f_*(\mathbf{H}_0(L)) = \mathbb{Z}_2$. Furthermore, $f_*(\mathbf{H}_1(L)) = 0$ since a base of the 1-cycles in L is given by the n cycles $\sigma_i = AB_i + B_iC_i + C_iA$ and σ_i is homologous to 0 in K for each $i \in \{1, \dots, n\}$. Last, $f_*(\mathbf{H}_2(L)) = 0$ since L contains no 2-simplices. By Lemma 10, we obtain that X is a homological

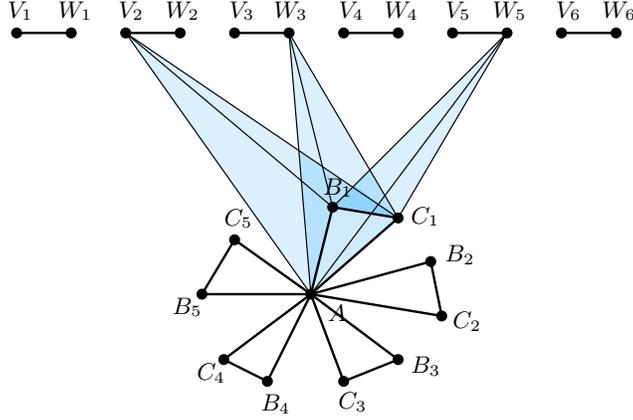


Figure 7: Simplicial complex L output by the reduction of a formula with five clauses and six variables and triangles in K created by clause $c_1 = v_2 \vee \neg v_3 \vee \neg v_5$.

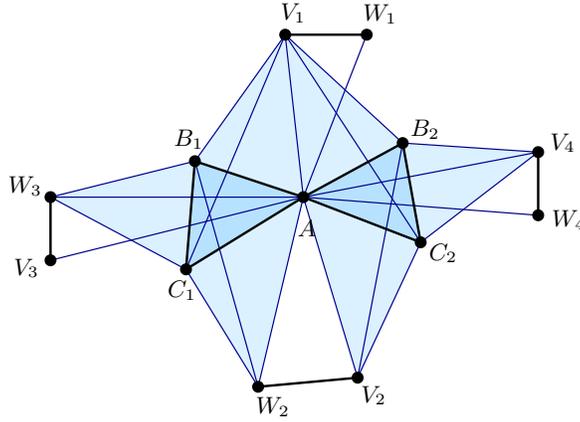


Figure 8: Pair (K, L) produced by the reduction of formula $(v_1 \vee \neg v_2 \vee \neg v_3) \wedge (v_1 \vee v_2 \vee v_4)$. L consists of the vertices and bold edges.

simplification of the pair (K, L) if and only if $\mathbf{H}_0(X) = \mathbb{Z}_2$, $\mathbf{H}_1(X) = 0$ and $\mathbf{H}_2(X) = 0$. Keeping this in mind, we establish the following lemma, in which (K, L) designates the pair output by our reduction algorithm when applied to formula E .

Lemma 15. *The pair (K, L) has a homological simplification if and only if the formula E has a satisfying assignment. Furthermore, given a homological simplification of the pair (K, L) , computing a satisfying assignment for E takes linear time.*

Proof. Suppose the pair (K, L) has a homological simplification X and let us prove that E has a satisfying assignment. First, we claim that X cannot contain both edges AV_j and AW_j , for $1 \leq j \leq m$. Indeed, if both edges AV_j and AW_j were in X , we could consider the cycle $\tau = AV_j + V_jW_j + W_jA$. Since the edge V_jW_j bounds no triangle in K , the cycle τ cannot be homologous to 0 in X , contradicting $\mathbf{H}_1(X) = 0$.

The claim allows us to assign to each variable v_j either the value TRUE if the edge AV_j belongs to X or the value FALSE if the edge AW_j belongs to X . If none of the edges AV_j and AW_j belong to X , then we assign to v_j an arbitrary value in $\{\text{TRUE}, \text{FALSE}\}$; see Figure 9. Note that the computation of this assignment

can be done in linear time. We now check that this assignment of boolean values to the variables v_j is a satisfying assignment, in other words we show that all clauses c_i are satisfied for $1 \leq i \leq n$.

Since $\mathbf{H}_1(X) = 0$, the 1-cycle $AB_i + B_iC_i + C_iA$ is a boundary in X . This implies that at least one triangle of X contains AB_i on its boundary. By construction, AB_i belongs to exactly three triangles in K , namely the triangles $AB_iY_i^k$ for $1 \leq k \leq 3$ where Y_i^k designates $V_{j_i^k}$ if $e_i^k = \mathbf{1}$ and $W_{j_i^k}$ if $e_i^k = \neg$. It follows that one of the three triangles $AB_iY_i^k$ must belong to X and, in turn, at least one of the three edges AY_i^k for $1 \leq k \leq 3$ is in X . This implies that one of the three literals $e_i^k v_{j_i^k}$ in clause c_i evaluates to TRUE and hence c_i is satisfied.

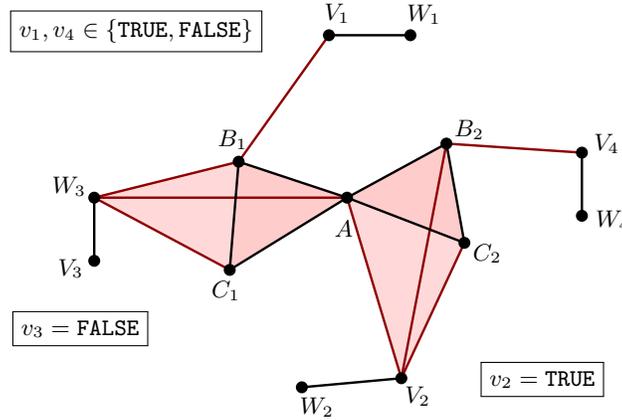


Figure 9: A homological simplification of the pair (K, L) drawn in Figure 8 and output by the reduction of formula $E = (v_1 \vee \neg v_2 \vee \neg v_3) \wedge (v_1 \vee v_2 \vee v_4)$. Corresponding satisfying assignments for E .

Conversely, suppose variables v_1, \dots, v_m have been assigned values that cause E to evaluate to TRUE and let us prove that the pair (K, L) has a homological simplification X . We construct X starting from L and adding some simplices of K as follows; see Figure 10. We begin by adding the edge AV_j if $v_j = \text{TRUE}$ and the edge AW_j if $v_j = \text{FALSE}$, for all $j \in \{1, \dots, m\}$. Since values of v_1, \dots, v_m are a satisfying assignment, we can choose one literal $e_i v_{j_i}$ in each clause c_i that is true. Let $Y_i = V_{j_i}$ if $e_i = \mathbf{1}$ and $Y_i = W_{j_i}$ if $e_i = \neg$. Note that by construction, the edge AY_i is already in X . We then add the three triangles AB_iY_i , $B_iC_iY_i$ and C_iAY_i to X , for all $i \in \{1, \dots, n\}$.

Let us check that X is indeed a solution to the homological simplification problem, *i.e.* $\mathbf{H}_0(X) = \mathbb{Z}_2$, $\mathbf{H}_1(X) = 0$ and $\mathbf{H}_2(X) = 0$. For this, we check that X is contractible by collapsing X to A , using a sequence of elementary collapses. First, observe that exactly one of the two vertices V_j or W_j belongs to no other simplices than the edge V_jW_j . For instance, if $v_j = \text{TRUE}$, then by construction $AV_j \in X$ and $AW_j \notin X$. Thus, W_j belongs to no other simplices than V_jW_j and we can collapse the edge V_jW_j to the vertex V_j by removing the pair of simplices (W_j, V_jW_j) . Similarly, if $v_j = \text{FALSE}$, we collapse the edge V_jW_j to the vertex W_j . For all $i \in \{1, \dots, n\}$, we apply five elementary collapses, first removing the three triangles AB_iY_i , $B_iC_iY_i$ and C_iAY_i and their edges AB_i , B_iC_i and C_iA , then removing the edges B_iY_i and C_iY_i and their vertices B_i and C_i . Last, we collapse every edge AY_i for $1 \leq i \leq n$ to the vertex A . \square

- [2] N. Amenta, M. Bern, and D. Eppstein. The crust and the β -skeleton: Combinatorial curve reconstruction. *Graphical Models and Image Processing*, 60(2):125–135, 1998.
- [3] N. Amenta, S. Choi, T. Dey, and N. Leekha. A simple algorithm for homeomorphic surface reconstruction. In *Proceedings of the sixteenth annual symposium on Computational Geometry*, pages 213–222. ACM New York, NY, USA, 2000.
- [4] N. Amenta, S. Choi, and R. Kolluri. The power crust, unions of balls, and the medial axis transform. *Computational Geometry: Theory and Applications*, 19(2-3):127–153, 2001.
- [5] D. Attali. r -regular shape reconstruction from unorganized points. *Computational Geometry: Theory and Applications*, 10:239–247, 1998.
- [6] D. Attali, M. Glisse, S. Hornus, F. Lazarus, and D. Morozov. Persistence-sensitive simplification of functions on surfaces in linear time. *Manuscript, INRIA*, 2008.
- [7] D. Attali and A. Lieutier. Reconstructing shapes with guarantees by unions of convex sets. Submitted to *Discrete and Computational Geometry*. <http://hal.archives-ouvertes.fr/hal-00427035/en/>.
- [8] D. Attali, A. Lieutier, and D. Salinas. Vietoris-Rips complexes also provide topologically correct reconstructions of sampled shapes. *Computational Geometry: Theory and Applications (CGTA)*, 2012. Accepted.
- [9] J. Boissonnat and F. Cazals. Smooth surface reconstruction via natural neighbour interpolation of distance functions. *Computational Geometry: Theory and Applications*, 22(1-3):185–203, 2002.
- [10] J. Boissonnat and S. Oudot. Provably good sampling and meshing of Lipschitz surfaces. In *Proceedings of the twenty-second annual symposium on Computational geometry*, page 346. ACM, 2006.
- [11] F. Chazal, D. Cohen-Steiner, and A. Lieutier. A sampling theory for compact sets in Euclidean space. *Discrete and Computational Geometry*, 41(3):461–479, 2009.
- [12] F. Chazal, D. Cohen-Steiner, A. Lieutier, and B. Thibert. Shape smoothing using double offsets. In *Proc. of the ACM symposium on Solid and physical modeling*, pages 183–192. ACM New York, NY, USA, 2007.
- [13] F. Chazal and A. Lieutier. Stability and computation of topological invariants of solids in \mathbb{R}^n . *Discrete and Computational Geometry*, 37(4):601–617, 2007.
- [14] F. Clarke. *Optimization and nonsmooth analysis*. Society for Industrial Mathematics, 1990.
- [15] D. Cohen-Steiner. Personal communication. 2008.
- [16] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37(1):103–120, 2007.
- [17] T. Cormen, C. Leiserson, R. Rivest, and C. Stein. *Introduction to algorithms*, 2001.
- [18] V. de Silva. Personal communication. 2009.
- [19] T. Dey, S. Funke, and E. Ramos. Surface reconstruction in almost linear time under locally uniform sampling. In *Abstracts 17th European Workshop Comput. Geom.*, pages 129–132. Citeseer, 2001.

- [20] T. Dey, J. Giesen, E. Ramos, and B. Sadri. Critical points of the distance to an epsilon-sampling of a surface and flow-complex-based surface reconstruction. In *Proc. of the twenty-first annual symposium on Computational geometry*, page 227. ACM, 2005.
- [21] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28(4):511–533, 2002.
- [22] H. Edelsbrunner, D. Morozov, and V. Pascucci. Persistence-sensitive simplification functions on 2-manifolds. In *Proceedings of the twenty-second annual symposium on Computational geometry*, page 134. ACM, 2006.
- [23] K. Grove. Critical point theory for distance functions. In *Proc. of Symposia in Pure Mathematics*, volume 54, pages 357–386, 1993.
- [24] A. Lieutier. Any open bounded subset of \mathbb{R}^n has the same homotopy type as its medial axis. *Computer-Aided Design*, 36(11):1029–1046, 2004.
- [25] D. Morozov. Homological Illusions of Persistence and Stability. *Ph.D. Dissertation, Duke University*, 2008.
- [26] J. Munkres. *Elements of algebraic topology*. Perseus Books, 1993.
- [27] P. Niyogi, S. Smale, and S. Weinberger. Finding the Homology of Submanifolds with High Confidence from Random Samples. *Discrete Computational Geometry*, 39(1-3):419–441, 2008.