Two Extensions to Manifold Learning Algorithms using α-Complexes

Maia Fraser
Department of Computer Science, University of Chicago
maia@cs.uchicago.edu

October 14, 2010

Abstract

A large body of work has grown up over the past decade [2, 9, 10, 13, 14, 3, 4, 5, 8, 15, 19, 20] dealing with ways of recovering topological invariants of a submanifold $\mathcal{M} \subset \mathbb{R}^N$ from a point cloud Z associated to $\mathcal{M}$ – most of these use α-complexes and assume Z is from $\mathcal{M}$ itself (though some consider base spaces more general than manifolds), while [13, 14] use Čech complexes and consider more general sampling. In addition, most compute exclusively homological invariants, though [13, 14] and [4] do give homotopy equivalence results. The purpose of this report is to briefly point out two simple extensions which can be made to some of the above algorithms. The first consists of a simple pre-processing step (due to [14]) which may be added to many of the other algorithms to extend them to noisy data sampled from a distribution on all of $\mathbb{R}^N$, centred but not necessarily supported on $\mathcal{M}$. This extension also results in statements regarding the homotopy equivalence of the computed simplicial complex and $\mathcal{M}$. The second extension is a brief comment describing how the fundamental group of a simplicial complex may be computed. In particular, using the homotopy equivalence guarantee obtained by our first extension one obtains a similar guarantee on isomorphism of the computed fundamental group with $\pi_1(\mathcal{M})$, the fundamental group of $\mathcal{M}$. We remark, however, that since this group is not in general abelian, its triviality is not decidable.

1 Introduction

Our first, statistical, extension is restricted to methods which assume an underlying smooth submanifold $\mathcal{M} \subset \mathbb{R}^N$ and use α-complexes. Most α-complex- based methods recently proposed [3, 5] make use of the stability of persistence diagrams shown in [6] (Homology Inference Theorem). Although that result in fact applies to a wider class of underlying spaces than just smooth submanifolds of $\mathbb{R}^N$, these methods are often used in practice for the latter. The purpose of this note is to remark that in this restricted (manifold) setting a much stronger result than the stability result of [6] actually applies: namely under similar conditions on denseness of Z in $\mathcal{M}$, the α-complex $A(Z,\alpha)$ computed for sufficiently small α is in fact homotopy equivalent to $\mathcal{M}$ and moreover, if Z is sampled from a suitable distribution on all of $\mathbb{R}^N$ that is centred on $\mathcal{M}$, then the same conclusion holds, with high probability. These extensions are obtained immediately by applying results of [13] and [14]. We remark that the Homology Inference Theorem of [6] shows equality of dimension of the persistent homology groups of $A(Z,\alpha)$ and $\mathcal{M}$. While this is in principle a different claim from ours, in fact it may be seen to follow. We discuss this briefly after defining α- and Čech complexes but postpone a detailed proof for a subsequent paper.

Our second extension, the calculation of fundamental group, applies to any method using a simplicial complex. In particular if it is used with the above methods then one recovers with high probability the fundamental group $\pi_1(\mathcal{M})$.

As mentioned, we have access to $\mathcal{M}$ only via a cloud of points Z that has been *observed*. This sample may be from a distribution μ with support in $\mathcal{M}$, or with support in a tubular neighborhood of $\mathcal{M}$, or from a distribution μ "centred" on $\mathcal{M}$. Based on this discrete set of points one wishes to define a simplicial complex from whose topology one may draw conclusions regarding the topology of $\mathcal{M}$. There are two key steps here.

First, one needs to construct a suitable simplicial complex K. Generally this construction uses small distance between points as an indication that they should be part of a common simplex, though there are different ways of making this precise. Implicit in this step is always a notion of scale, α, defining what constitutes a "small" distance.

The second step is to infer from the constructed simplicial complex something about the topology of $\mathcal{M}$: for example connectedness, or certain Betti numbers, homology groups themselves, or even homotopy invariants. Here, one needs a result proving the relevant *faithfulness* of $K(Z,\alpha)$, i.e. a guarantee that the topological property of $K(Z,\alpha)$ that one is computing is also shared by $\mathcal{M}$, under some assumptions that can be achieved in practice. In particular, some conditions on the density of the sample are clearly required in order to use $K(Z,\alpha)$ to deduce anything about the topology of $\mathcal{M}$ (if for example all points of Z were concentrated in one small piece of $\mathcal{M}$ one could say nothing about $\mathcal{M}$ globally). We briefly look at two choices for $K(Z,\alpha)$ and then discuss the relevant faithfulness results.

One commonly used option for $K(Z,\alpha)$ is the α-complex already mentioned. This was defined by Edelsbrunner in 1995, [11]. Another (related but more classical) option is the Čech complex, used by Niyogi, Smale, Weinberger [13, 14]. In each case, vertices are points of Z. The **Čech complex** $\text{Čech}(Z,\alpha)$ is defined by:

$$\sigma = [z_0 z_1 \dots z_p] \text{ is a } p\text{-simplex iff } \bigcap_{j=0}^{p} B(z_j,\alpha) \neq \emptyset$$

.

The **α-complex** $A(Z,\alpha)$ is defined[1] by:

$$\sigma = [z_0 z_1 \dots z_p] \text{ is a } p\text{-simplex iff } \bigcap_{j=0}^{p} [B(z_j,\alpha) \cap V_j] \neq \emptyset,$$

where $V_j = \{x \in \mathbb{R}^N : \forall z \in Z,\ d(x,z_j) \leq d(x,z)\}$ is the Voronoi cell of $z_j \in Z$.

Let U be the union of balls of radius α around points of Z. Both of the collections of sets – $\{B(z_j,\alpha) : z_j \in Z\}$ and $\{B(z_j,\alpha) \cap V_j : z_j \in Z\}$ – are finite covers of U and the complexes just defined are their **nerves** (see Hatcher [12]). Moreover the sets in these covers are convex. The Nerve Lemma [16] therefore implies the nerves are homotopy equivalent to U, and to each other; Edelsbrunner also gives a direct proof in [11]. $A(Z,R)$ has the advantage that it is a subcomplex of the Delauney triangulation of U and so has size $O(n^{\lceil N/2 \rceil})$ (where $n = |Z|$) however this upper bound is often nearly achieved, which becomes prohibitive in a high dimensional ambient space $\mathbb{R}^N$. More recently de Silva [8] defined the witness complex, where a small subset of the n points of Z are chosen as "landmarks" and used as the vertices of the complex (while all other points of Z are used only as "witnesses" to the existence of simplices). Beyond two and three dimensional manifolds [7, 8] one needs additional techniques to obtain faithfulness [15]. We do not specifically address methods using witness complexes in this report.

[1]Warning: usually the notation $A(Z,\alpha)$ is used to denote the analogous complex obtained using balls of radius $\alpha/2$ intersected with Voronoi cells.

Further work. Note that the finite open cover used for $A(Z,\alpha)$ is a subcover of that used for $\check{\text{C}}\text{ech}(Z,\alpha)$ and so $A(Z,\alpha)$ is a subcomplex of $\check{\text{C}}\text{ech}(Z,\alpha)$ at each α. Since the proof of the Nerve Lemma is functorial, the homotopy equivalences it produces between complexes $A(Z,\alpha)$ and $\check{\text{C}}\text{ech}(Z,\alpha)$ commute with inclusion in a natural way. This allows one to conclude that the image of the k'th homology group of $A(Z,\alpha_1)$ in that of $A(Z,\alpha_2)$ induced by inclusion for $\alpha_1 < \alpha_2$ will be isomorphic to the corresponding image for $\check{\text{C}}\text{ech}(Z,\alpha_1)$ and $\check{\text{C}}\text{ech}(Z,\alpha_2)$. In other words persistent homology groups of the two complexes coincide. Moreover the same argument applies for the manifold $\mathcal{M}$ at suitably small scales and thus one may derive a persistent homology result like that of [6] from the homotopy equivalence result described here. We will address this in a subsequent paper.

2 Extension to Noisy Data

In [14], Niyogi et al consider data that are sampled from a distribution on all of $\mathbb{R}^N$ that satisfies the following:

Definition 2.1 (strong variance condition). *A probability distribution μ on $\mathbb{R}^N$ is said to satisfy the strong variance condition if it is given by the push forward of a probability density function P on the normal bundle $N\mathcal{M}$ for which two conditions hold: 1) the conditional distribution $P(y|x)$ on the normal fiber $N_x\mathcal{M}$ at $x \in \mathcal{M}$ is Normal$(0,\Sigma)$ for some constant $\Sigma = \sigma^2 I_{N-n}$ independent of x, and 2) there exist a, b such that $0 < a \leq P(x) \leq b$, $\forall x \in \mathcal{M}$, where $P(x)$ is the marginal on $\mathcal{M}$.*

This is used to prove the statistical proposition:

Proposition 2.2 ([14] Proposition 5.1)**.** *Let μ be a probability measure on $\mathbb{R}^N$ satisfying the strong variance condition and suppose σ is sufficiently small compared to the reach τ of $\mathcal{M}$. Let $\delta > 0$ and let $Z = \{z_1,\ldots,z_n\}$ be n data points drawn in i.i.d. fashion according to μ from $\mathcal{M}$. Let $\bar{Z} \subseteq Z$ be the data points left after cleaning (by the procedure proposed in [14], reproduced below). Then if n is sufficiently large (an explicit figure is given in [14] in terms of δ, N, d and τ) it follows that with probability greater than $1-\delta$, $\bar{Z}$ is a set of points which*

- *is contained in the R-tubular neighborhood of M, and*
- *is R-dense for M, i.e., $M \subset \bigcup_{z \in Z'} B(z,R)$.*

Here R is assumed to be some number that is small enough to satisfy the hypotheses of Proposition 2.5 below. Moreover, if the co-dimension $N-d$ of $\mathcal{M}$ is high then the sample complexity (i.e. the lower bound on n just mentioned) is independent of N.

The following cleaning procedure uses two pre-chosen thresholds s and m. Appropriate choices for these are given explicitly in [14] in terms of the reach τ and the specific distribution μ. Essentially s is bounded below by a term involving σ and bounded above by τ (with the strong variance condition implying this interval is nonempty). On the other hand m is bounded below by an expression in terms of τ and the measure μ of suitably small balls. Details may be found in Section 3.1 of [14]. Together these result in an R for Proposition 2.2 which is sufficiently small so as to satisfy the hypotheses of Proposition 2.5.

Algorithm 2.3. [Cleaning Procedure]

Input: data points $Z = \{z_1, \dots, z_n\}$ drawn i.i.d. from μ.
Output: data points $\bar{Z}$
Procedure:

1. Construct the nearest neighbor graph with n vertices (each vertex associated to a data point) with adjacency matrix

$$W_{ij} = 1 \Leftrightarrow ||z_i - z_j|| < s.$$

2. Let d_i be the degree of the i'th vertex (associated to z_i) of the graph. Throw away all data points whose degree is smaller than m and put the others in $\bar{Z}$.

3. Output $\bar{Z}$.

This result is combined with a proposition from [13] to establish their main theorem:

Theorem 2.4 ([14] Theorem 5.1)**.** *Let $\bar{Z}$ be as given in the previous Proposition. Then, with probability greater than $1-\delta$, $\check{C}ech(\bar{Z}, \alpha)$ is homotopy equivalent to $\mathcal{M}$.*

The proposition from [13] used is the following:

Proposition 2.5 ([13] Proposition 7.1)**.** *Let Z be a set of points in the tubular neighborhood of radius R around $\mathcal{M}$. Let U be given by $U = \bigcup_{z \in Z} B(z, \varepsilon)$. If Z is R-dense in $\mathcal{M}$ then $\mathcal{M}$ is a deformation retract of U for all sufficiently small R, ε (explicit bounds given in [13] in terms of the reach τ of $\mathcal{M}$) and so $\check{C}ech(Z, \varepsilon)$ is homotopy equivalent to $\mathcal{M}$.*

On the other hand the stability result of [6] states:

Theorem 2.6 (Homology Inference Theorem of [6])**.** *For all real ε with*

$$d_H(\mathcal{M}, Z) < \varepsilon < hfs(\mathcal{M})/4$$

and all sufficiently small $\alpha > 0$, the dimensions of the k'th homology group of $Tub_\alpha(\mathcal{M})$ and the k'th persistent homology group induced by inclusion of $A(Z, \varepsilon/2)$ into $A(Z, 3\varepsilon/2)$ (see footnote [1]) are either both infinite or both finite and equal (for any $k \in \mathbb{N}$).

Here, d_H is Hausdorff distance. The following lemma shows that the lower bound in the hypothesis of the Theorem is equivalent to Z lying in the ε-tubular neighborhood of $\mathcal{M}$ and being ε-dense in $\mathcal{M}$. On the other hand *hfs* stands for *homological feature size*. As commented in [6], this is at least as large as the reach τ.

Lemma 2.7. *Given a compact set K and a compact manifold M possibly with boundary, both embedded in $\mathbb{R}^N$, let $\varepsilon > 0$ then $d_H(Z, S) < \varepsilon$ iff K lies in a ε-neighborhood of M and is ε-dense in M.*

Proof. By definition $d_H(Z, S) < \varepsilon$ if and only if both

$$\sup_{x \in M} \inf_{k \in K} d(x, k) < \varepsilon \quad \text{and} \quad \sup_{k \in K} \inf_{x \in M} d(x, k) < \varepsilon.$$

The first is true if and only if each $x \in M$ is within ε of some $k \in K$, i.e. K is ε-dense in M. The second is true if and only if each $k \in K$ is within ε of some $x \in M$, i.e K lies within the ε-neighborhood of M. □

The hypothesis of compactness is a mild one in practice. Any set that lies in a bounded region of $\mathbb{R}^N$ and has positive reach must automatically be compact (since, positive reach implies the set must contain all its limit points, i.e. be closed). Using the Lemma we may re-state the Homology Inference Theorem (for compact $\mathcal{M}$):

Theorem (re-worded Homology Inference Theorem [6]). *Let Z be a set of points in the tubular neighborhood of radius* ε *around* $\mathcal{M}$. *If Z is* ε*-dense in* $\mathcal{M}$ *then for all sufficiently small* α, ε *(specifically* $\varepsilon < hfs(\mathcal{M})/4$*) the dimensions of the k'th homology group of* $Tub_\alpha(\mathcal{M})$ *and the k'th persistent homology group induced by inclusion of* $A(Z, \varepsilon/2)$ *into* $A(Z, 3\varepsilon/2)$ *(see footnote [1]) are either both infinite or both finite and equal (for any* $k \in \mathbb{N}$*).*

However, using the homotopy equivalence of $A(Z,\varepsilon)$ and Čech(Z,ε), one immediately obtains:

Corollary 2.8 (to Proposition 2.5). *Let Z be a set of points in the tubular neighborhood of radius* ε *around* $\mathcal{M}$. *Let U be given by* $U = \bigcup_{z \in Z} B(z, \alpha)$. *If Z is* ε*-dense in* $\mathcal{M}$ *then for all sufficiently small* α, ε *(with in particular* $\varepsilon < \tau$*)* $A(Z, \alpha)$ *is homotopy equivalent to U which deformation retracts to* $\mathcal{M}$.

In contrast to the Homology Inference Theorem, this result gives homotopy equivalence of $A(Z,\alpha)$ with $\mathcal{M}$ itself instead of just coinciding dimensions of the persistent homology groups for a tubular neighborhood of $\mathcal{M}$ and those for the α-complexes in a similar scale range. This comes at a cost of assuming the underlying space is a well-conditioned submanifold of $\mathcal{M}$ (i.e. with positive reach) and taking possibly smaller ε.

Moreover, using the homotopy equivalence of $A(Z,\varepsilon)$ and Čech(Z,ε), one may also obtain a statistical version of this result, as a Corollary to Theorem 2.4 (via Proposition 2.2):

Corollary 2.9. *Let* μ *be a probability measure on* $\mathbb{R}^N$ *satisfying the strong variance condition and suppose* σ *is sufficiently small compared to the reach* τ *of* $\mathcal{M}$. *Let* $\delta > 0$ *and let* $Z = \{z_1, \ldots, z_n\}$ *be n data points drawn in i.i.d. fashion according to* μ *from* $\mathcal{M}$. *Let* $\bar{Z} \subseteq Z$ *be the data points left after cleaning (by Algorithm 2.3). Then if n is sufficiently large (an explicit figure is given in [14] in terms of* δ*, N, d and* τ*) and* α *sufficiently small it follows that, with probability greater than* $1-\delta$, $A(\bar{Z}, \alpha)$ *is homotopy equivalent to* $\mathcal{M}$.

3 Computing the Fundamental Group

In this section, we give an algorithm for computing the fundamental group of a simplical complex. When used after the procedures mentioned in the previous section, one can thus recover with high probability the fundamental group of a suitably sampled manifold.

The output of the algorithm gives the number of connected components of the simplicial complex and for each one its fundamental group, presented by a set of generators and a set of relations on those generators. However, we remark that by Adian's result of 1957 [1], the triviality of a finitely presented group is not decidable. In fact, Adian showed: if P is a Markov property of finitely presented groups, then P is not decidable. Some examples of Markov properties are: being the trivial group, being finite, being abelian. Even knowing that a group is perfect (as we would if we found the first homology group to be trivial, since it is the abelianization of the fundamental group) will not help: if we could decide that problem, we could decide triviality in general for finitely presented groups.

For some special cases triviality of the fundamental group is decidable by other means: if we know are dealing with the fundamental group of a closed surface then we may simply test if the first homology is trivial; for compact 3 manifolds, since there is an algorithm (Rubinstein [17], Thompson [18]) to decide homeomorphism with the 3-sphere), this also decides triviality of the fundamental group by the Poincaré Conjecture.

We remark that the fundamental group of a simplicial complex depends only on the 2-skeleton of the complex (see Hatcher [12]). We define a special data structure for storing a 2-skeleton. It will moreover allow us to store any cell complex obtained from a 2-skeleton by collapsing 1-simplices. We call this data structure DEGEN2SKEL.

Definition 3.1. [DEGEN2SKEL] *This abstract data structure consists of a set* Σ_0 *(vertices), a multiset* Σ_1 *(edges)consisting of pairs of elements of* Σ_0*, a set* Σ_2 *(faces) consisting of triples of elements of* $\Sigma_1 \cup \{\mathbb{1}\}$ *(where* $\mathbb{1}$ *is a special reserved symbol). This data structure serves iteration through* Σ_0 *or* Σ_1 *or* Σ_2*, modification of the first or second coordinate of an element of* Σ_1*, replacing by* $\mathbb{1}$ *the first, second or third coordinate of an element of* Σ_2*, and deletion of elements of* Σ_0 *which do not occur in any element of* Σ_1*.*

Definition 3.2. [allowable] *A DEGEN2SKEL* $(\Sigma_0, \Sigma_1, \Sigma_2)$ *is* allowable *if every element of* Σ_2 *corresponds to a* 3*-cycle of vertices, i.e.* $\forall \tau \in \Sigma_2$*, if* $\tau = ([u,v],[w,x],[y,z])$ *then* $v = w$*,* $x = y$*,* $z = u$*, assuming any* $\mathbb{1}$ *is interpreted as a repeated vertex of the form* $[v,v]$ *with* v *arbitrary.*

Definition 3.3. [simplicial] *A DEGEN2SKEL* $(\Sigma_0, \Sigma_1, \Sigma_2)$ *is* simplicial *if every element of* Σ_1 *has distinct first and second coordinates, and no element of* Σ_2 *has a coordinate* $\mathbb{1}$*.*

Note there is an obvious bijective correspondence between allowable simplicial DEGEN2SKEL's and 2-dimensional simplicial complexes. Moreover, all allowable DEGEN2SKEL's, even those which are not simplicial, correspond uniquely to cell complexes and so to topological spaces. We will therefore speak of them as topological spaces, meaning implicitly the associated cell complexes.

Algorithm 3.4. [COMPUTING FUNDAMENTAL GROUP]
Input: a simplicial allowable DEGEN2SKEL $(\Sigma_0, \Sigma_1, \Sigma_2)$ which represents the simplicial complex K
Output: an allowable DEGEN2SKEL $(\Sigma_0, \Sigma_1, \Sigma_2)$ which has one vertex (v) per connected component (K_v) of K, whose edges with a given vertex v are generators of $\pi_1(K_v)$ and whose faces are relations.
Procedure:

```
for (σ ∈ Σ1) do
    if (the vertices of σ are distinct) then
        COLLAPSE(σ)
    end if
end for
return (Σ0, Σ1, Σ2)
```

This algorithm calls the following subroutine which modifies the existing DEGEN2SKEL $(\Sigma_0, \Sigma_1, \Sigma_2)$.

Algorithm 3.5. [COLLAPSE]
Input: an allowable DEGEN2SKEL $(\Sigma_0, \Sigma_1, \Sigma_2)$, an element $\sigma = [v_0, v_1]$ in Σ_1
Effect: $(\Sigma_0, \Sigma_1, \Sigma_2)$ has been modified so σ and v_1 no longer appear; the new value of $(\Sigma_0, \Sigma_1, \Sigma_2)$ is homotopy equivalent to the initial one.
Procedure:

```
for (σ′ ∈ Σ1) do
    if (the first or second coordinate of σ′ is v1) then
        replace it by v0
    end if
end for
delete v1 from Σ0
for (τ ∈ Σ2) do
    if (the first, second or third coordinate of τ is σ) then
        replace it by 𝟙
    end if
end for
```

It is easily verified that homotopy type of $(\Sigma_0, \Sigma_1, \Sigma_2)$ is a loop invariant in both algorithms. Moreover when COMPUTING FUNDAMENTAL GROUP stops, there are no edges with distinct vertices left. Viewing $(\Sigma_0, \Sigma_1, \Sigma_2)$ as a cell complex, the edges and faces using a given vertex v define a connected component of the complex. Its fundamental group is easy to compute. All edges are generators and any face (e_1, e_2, e_3), where the e_i are edges or $\mathbb{1}$, corresponds to the relation $e_1 e_2 e_3 = \mathbb{1}$, where $\mathbb{1}$ is the identity of the group. Therefore the algorithm COMPUTING FUNDAMENTAL GROUP indeed computes the fundamental group of each of the connected components of the input simplical complex.

Acknowledgements

The observations made here grew out of discussions with Partha Niyogi and Shmuel Weinberger. The report was not completed until after Partha's untimely passing. He will be greatly missed. All errors and omissions are the author's responsibility and any comments on such are very much appreciated.

References

[1] S. I. Adian, *The unsolvability of certain algorithmic problems in the theory of groups*. Tr. Mosk. Mat. Obshch., **6** (1957), pp. 231298.

[2] E. Carlsson, G. Carlsson, V. de Silva, *An Algebraic Topological Method for Feature Identification*. Int. J. Comput. Geometry Appl. **16** (2006). p 291-314.

[3] F. Chazal, A. Lieutier, *Smooth Manifold Reconstruction from Noisy and Non Uniform Approximation with Guarantees*. Computational Geometry: Theory and Applications, **40** (2008), p.156-170.

[4] F. Chazal, A. Lieutier, *Weak feature size and persistent homology: computing homology of solids in $\mathbb{R}^N$ from noisy data samples*. In Proc. 21st Annual Symposium on Computational Geometry (2005), p. 255-262.

[5] F. Chazal, D. Cohen-Steiner, M. Glisse, L.J. Guibas, and S. Y. Oudot, *Proximity of Persistence Modules and their Diagrams*. In Proc. 25th Annual Symposium on Computational Geometry (2009), p. 237-246.

[6] D. Cohen-Steiner, H. Edelsbrunner and J. Harer, *Stability of persistence diagrams*. Discrete Comput. Geom. 37 (2007), p. 103-120.

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

[8] V. de Silva, G. Carlsson, *Topological estimation using witness complexes*. Eurographics Symposium on Point-Based Graphics (2004).

[9] H. Edelsbrunner, D. Letscher and A. Zomorodian, *Topological persistence and simplification*. Discrete Comput. Geom. 28 (2002), p. 511-533.

[10] H. Edelsbrunner, J. Harer *Persistent homology – a survey*. Surveys on Discrete and Computational Geometry. Twenty Years Later, 257-282, eds. J. E. Goodman, J. Pach and R. Pollack, Contemporary Mathematics **453**, Amer. Math. Soc., Providence, Rhode Island, 2008.

[11] H. Edelsbrunner, *The union of balls and its dual shape*. Discrete & Computational Geometry **13**, 3-4 (1995), p. 415-440.

[12] A. Hatcher, *Algebraic Topology*, Cambridge University Press, 2002.

[13] P. Niyogi , S. Smale , S. Weinberger, *Finding the Homology of Submanifolds with High Confidence from Random Samples*. Discrete & Computational Geometry, 39 (2008), p.419-441.

[14] P. Niyogi , S. Smale , S. Weinberger, *A Topological View of Unsupervised Learning and Clustering*, preprint (2008).

[15] S. Oudot, L. Guibas, J.-D. Boissonat, *Manifold Reconstruction in Arbitrary Dimensions Using Witness Complexes*. SCG-07, June 6-8, Gyeongju, South Korea (2007).

[16] J. J. Rotman, *An Introduction to Algebraic Topology*. Springer-Verlag, 1988.

[17] J. Rubinstein, *An algorithm to recognize the* 3-*sphere*. Proc. of the International Congress of Mathematicians, Vol 1, 2 (Zurich, 1994), Birkhausen, Basel, pp. 601-611.

[18] A.Thompson, *Thin position and the recognition problem for* S^3. Math. Res. Lett. 1 (1994), pp. 613-630.

[19] A. Zomorodian, *Computing and Comprehending Topology: Persistence and Hierarchical Morse Complexes*. PhD thesis, Duke University (2001) (advisor H. Edelsbrunner).

[20] A. Zomorodian, G. Carlsson, *Computing persistent homology*. Symposium on Computational Geometry (2004), p. 347-356.