1 Research proposal (15p.)

1.1 State-of-the-art and objectives

The research and development pipeline in Computational Geometry. Computational geometry emerged as a discipline in the seventies and has met with considerable success in providing foundations to solve basic geometric problems including data structures, convex hulls, triangulations, Voronoi diagrams, geometric arrangements and geometric optimisation [?]. This initial development of the discipline has been followed, in the mid-nineties, by a vigorous effort to make computational geometry more effective. Situated in-between basic theoretical research and the development of robust software is the emerging study of effective methods (algorithms and data structures) of geometric computing, namely methods that are not only theoretically proved but also work well in practice. Several EC projects (CGAL, GALIA) established an outstanding research momentum and gave a leading role to Europe in this context. They led to successful techniques and tools, most notably the CGAL library [?], a unique tool that provides a well-organised, robust and efficient software environment for developing geometric applications. CGAL is considered as one of the main achievements of the field and is by now the standard in Geometric Computing, with a large diffusion worldwide and varied applications in both academia and industry. CGAL has no equivalent counterpart in the world.

3D Geometric Modeling. Approximating complex shapes through meshing is a fundamental problem on the agenda of several communities like Numerical Analysis, Computer Graphics, Geometry Processing and Computer-Aided Design. Emblematic problems are mesh generation that aims at sampling and meshing a given domain, and surface reconstruction that constructs an approximation of a surface which is only known through a set of points. Although these problems have received considerable attention in the past, it is only during the last 10 years that the Computational Geometry community established solid theoretical foundations to the problem, most notably in the emerging new area of Computational Topology. This approach has shown to be very successful and led to recent breakthroughs in *mesh generation* [?] and *surface reconstruction* [?]. The Geometrica group took a leading role in this research and contributed major theoretical advances as well as practical developments in the form of fast, safe and quality-guaranteed CGAL components for mesh generation and shape reconstruction [?]. Those components are now used worldwide in academia and in industry for various applications in Geometric Modeling, Medical Imaging and Geology.

Noisy data. When dealing with approximation and samples, one needs stability results to ensure that the quantities that are computed, geometric or topological invariants, are good approximations of the real ones. *Topological persistence* was recently introduced as a powerful tool for the study of the topological invariants of sampled spaces [?, ?]. Given a point cloud in Euclidean space, the approach consists in building a simplicial complex whose elements are filtered by some user-defined function. This filter basically gives an order of insertion of the simplices in the complex. The persistence algorithm, first introduced by Edelsbrunner, Letscher and Zomorodian [?], is able to track down the topological invariants of the filtered complex as the latter is being built. As proved by Cohen-Steiner et al. [?], under reasonable conditions on the input point cloud, and modulo a right choice of filter, the most persistent invariants in the filtration correspond to invariants of the space underlying the data. Thus, the information extracted by the persistence algorithm is global, as opposed to the locality relationships used by the dimensionality reduction techniques.

In this respect, topological persistence appears as a complementary tool to dimensionality reduction. In particular, it enables to determine whether the input data is sampled from a manifold with trivial topology, a mandatory condition for dimensionality reduction to work properly. Note however that it does not tell how and where to cut the data to remove unwanted topological features.

Multiscale reconstruction is a novel approach [?]. Taking advantage of the ideas of persistence, the approach consists in building a one-parameter family of simplicial complexes approximating the input at various scales. Differently from above, the family may not necessarily form a filtration, but it has other nice properties. In particular, for a sufficiently dense input data set, the family contains a long sequence of complexes that approximate the underlying space provably well, both in a topological and in a geometric sense. In fact, there can be several such sequences, each one corresponding to a plausible reconstruction at a certain scale. Thus, determining the topology and shape of the original object reduces to finding the stable sequences in the one-parameter family of complexes. However, multiscale reconstruction, at least in its current form, still has a complexity that scales up exponentially with the dimension of the ambient space. Hence, it can only be applied to low-dimensional data sets in practice.

High dimensional spaces. Dimensionality reduction is certainly one of the most popular approaches to high-dimensional data analysis. It consists in projecting the data points down to a linear subspace, whose dimension supposedly coincides with the intrinsic dimension of the data. This approach is elegant in that it helps detect the intrinsic parameters of the data, and by doing so it also reduces the complexity of the problem. Dimensionality reduction techniques fall into two classes: linear methods, e.g. principal component analysis (PCA) or multi-dimensional scaling (MDS), and non-linear methods, e.g. isomap or locally-linear embedding (LLE). The second class of algorithms is more powerful in that it computes more general (in fact, non-linear) projections. On the whole, dimensionality reduction works well on data sets sampled from manifolds with low curvature and trivial topology. Although the condition on the curvature is mainly a sampling issue, the condition on the topology is mandatory for the projection onto a linear subspace to make sense.

Many of the results of Computational Geometry have been extended to arbitrary dimension. In particular, worst-case optimal algorithms are known for computing convex hulls, Voronoi diagrams and Delaunay triangulations in any dimension. Hence, in principle, the methods developed for 3D applications should be extendable to higher dimensions. However, the size of these structures depends exponentially on the dimension of the embedding space, which makes them only useful in moderate dimensions [?]. Many ideas have been suggested to bypass this curse of dimensionality. A first approach looks for more realistic combinatorial analyses such as smoothed analysis that bounds the expected complexity under some small random perturbation of the data. This led to the celebrated analysis of Linear Programming of Spielmann and Teng [?]. Another route is to trade exact for approximate algorithms [?]. An important example is the search for approximate nearest neighbours. Another example is the theory of core sets that was shown to provide good approximate solution to some optimization problems like computing the smallest enclosing ball, or computing an optimal separating hyperplane (SVM). These tools are both extremely useful but limited to basic operations and have not been yet applied in the context of geometric modeling. A third approach assumes that the intrinsic dimension of the object of interest has a much lower intrinsic dimension than the dimension of the embedding space. This is the usual assumption in Manifold Learning. It is then possible to resort to techniques derived from the 3D case and to approximate complex shapes by simplicial complexes (the analogue of triangulations in higher-dimesnional spaces). Various types of simplicial complexes have been proposed such as the Czech and the RIPS complexes, and the more recent Delaunay-like complexes such as the α-complex [?, ?], the witness complex [?, ?] and the Delaunay tangential complex [?]. They differ by their combinatorial and algorithmic complexities, and their power to approximate a shape. Under appropriate sampling conditions, we have shown that one can reconstruct a provably correct approximation of a smooth k-dimensional manifold M embedded in $\mathbb{R}^d$ in a time that depends only linearly on d [?]. Researchers have also turned their focus to the somewhat easier problem of inferring topological invariants of the shape without explicitly reconstructing it with the hope that more lightweight data structures and weaker sampling conditions would be appropriate for this simpler task [?, ?].

1.2 Methodology

Our overall goal is to settle the foundations for Geometric Modeling in higher dimensions by developing

- Sounded approaches providing guarantees even in the presence of noise or outliers,
- Algorithms that are of both theoretical and practical interest, that are amenable to theoretical analysis and fully validated experimentally,
- A software platform that is generic, robust and efficient.

The proposal is structured into the following four workpackages:

- WP 1: Computational geometry in non euclidean spaces.
- WP 2: Dimension-sensitive algorithms and data structures.
- WP 3: Robust geometric models.
- WP 4: Platform for geometric modelling in higher dimensions.

WP 2: Dimension-sensitive algorithms and data structures

Central to the techniques we intend to develop is the construction of simplicial complexes. A graph is an example of a 1-dimensional simplicial complex but simplicial complexes are much more powerful than graphs and allow to approximate complicated shapes of arbitrary dimension and topology. They offer a flexible data structure to represent and process higher-dimensional shapes and recent developments have shown that simplicial complexes computed on top of point clouds are primary tools to capture the topology of the underlying space of the data.

In 3-dimensions, 2 and 3-dimensional simplicial complexes of surfaces and volumes are widely used in graphics, scientific computing and manufacturing. Because of its numerous interesting properties and of the existence of extremely efficient algorithms to compute it, the Delaunay triangulation has become one of the most famous and widely used geometric data structures that spread out accross all sciences. The algorithms we have implemented in CGAL are among the most reliable and fast algorithms. They have been included in the heart of MATLAB.

The algorithms used in 3 dimensions extend to any dimension but their complexity grows exponentially with the dimension which makes them useless for real applications beyong say dimension 6 [?, ?]. In order for algorithms and implementations to scale with the dimension, we need to exploit (hidden) structure of the data and to design dimension-sensitive algorithms and data structures.

Design of small yet faithfull simplicial complexes. Given a set of points V in $\mathbb{R}^d$, a number of simplicial complexes with vertex set V have been proposed. A first class of simplicial complexes uses a parameter α which can be used to order the simplices of the complex.

The Čech complex is the nerve of the set B_α of balls of radius α centered at the points of V. The nerve of B_α associates a i-simplex to any subset of $i+1$ balls that have a common intersection. This is a simplicial complex that is in general not embeddable in $\mathbb{R}^d$. Moreover, it is usually very big and difficult to compute since it requires to detect whether a subset of balls of $\mathbb{R}^d$ intersect.

A simpler to compute simplicial complex is the Rips complex whose edges are the same as for the Čech complex. The higher dimensional simplices of the Rips complex are obtained by computing the cliques of sizes 3, 4 etc. in the graph of the edges. This simplicial complex is much easier to compute than the Čech complex and it has the vremarkable property that it can be constructed in a purely combinatorial way from its 1-skeleton. Such a simplicial complex is called a *flag complex*. Nevertheless, the Rips complex is not embedded in $\mathbb{R}^d$ and may have a dimension much higher than the dimension of the underlying structure of the data.

Various simplicial complexes have been derived from the Delaunay triangulation of the vertices. The α-complex is the nerve of the restriction of the Delaunay triangulation to the union of the balls of B_α. This complex is embedded in $\mathbb{R}^d$ (provided that the vertices are in general position) but very difficult to compute in high dimension for the same reason as the Čech complex.

Other simplicial complexes derived from the Delaunay triangulation do not involve any parameter, most notably the restricted Delaunay triangulation, the tangential Delaunay complex and the witness complex. Those complexes are especially designed for the case where V samples a topological space of small dimension k, the central hypothesis in Machine Learning. Both the restricted and the tangential Delaunay complexes are embedded in $\mathbb{R}^d$, have dimension k (under a mild general position assumption). Still, these simplicial complexes are limited to small k. The witness complex is another variant of the Delaunay triangulation introduced by Vin de Silva and Carlsson. The witness complex is embedded in $\mathbb{R}^d$ and is remarkably easy to compute in any dimension since the only numerical operations involves in its construction are comparisons of distances.

It should be noted that the Rips complex and the witness complex can both be computed from the knowledge of the distances between the vertices. Hence these complexes can be computed in any discrete metric space.

We identify four research directions :

1. Classifying simplicial complexes
2. Combinatorial and algorithmic complexity
3. Compact representation
4. New types of simplicial complexes

Classifying simplicial complexes. Some equivalences between the various types of simplicial complexes are known. For example, the Rips and the Čech complexes are identical for the L_1 norm and for the Euclidean norm, we have

$$\mathrm{Rips}() \subset \mathrm{Cech}() \subset \mathrm{Rips}().$$

Recently, we have established conditions under which the witness complex, the restricted Delaunay triangulation and the tangential complex are identical []. A more complete classification is required to better understand these structures and their properties.

Combinatorial and algorithmic complexity. A main limitation of using simplicial complexes is their combinatorial and algorithmic complexity. Differently from polytopes, very few results are known. The flag random complex is a noticeable exception []. Other types of random abstract complexes have to be studied from a combinatorial point of view. Geometric simplicial complexes should also be considered. We intend to develop probabilistic analyses.

Compact representation of simplicial complexes. *state of the art* Currently no code allows to manipulate simplicial complexes of arbitrary dimension in a routine way as is possible for 2 and 3-dimensional triangulations of $\mathbb{R}^3$ [**?**, **?**, **?**]. We are aware of only a few works on the design of data structures for general simplicial complexes. Brisson [**?**] and Lienhardt [**?**] have introduced data structures to represent d-dimensional cell complexes, most notably subdivided manifolds. While those data structures have nice algebraic properties, they are very redundant and do not scale to large data sets or high dimensions. Another representation for simplicial complexes has been proposed by de Floriani et al. [**?**]. This data structure encodes all the simplices of a simplicial complex and a subset of the incidence relations between the simplices. Experimental results are only reported for 3-dimensional objects and the data structure does not seem well adapted to much higher dimensions. More recently, Attali et al. [**?**] have proposed an efficient data structure to represent and simplify flag complexes, a special family of simplicial complexes including the Rips complex. This work is close in spirit to our work but is limited to flag complexes and simplicial complexes that are almost flag complexes. Experimental results are only reported in dimension 3.

full representation Needed to store info at the faces. Recently, we have experimented with a tree representation for simplicial complexes that seems to perform very well. Simplicial complexes of 500 millions of simplices have been constructed and stored on a laptop. The nodes of the tree are in bijection with the simplices (of all dimensions) of the simplicial complex, and each node in the tree requires $O(1)$ space. Hence, our data structure explicitly stores all the simplices of the complex but does not represent explicitly the adjacency relations between the simplices. Storing all the simplices provides generality, and the tree structure of our representation enables us to implement basic operations on simplicial complexes efficiently. In particular to retrieve of incidence relations. In addition, more compact storage could be further obtained by using well-known succinct representations of trees [**?**, **?**, **?**, **?**].

Our data structure is purely combinatorial and can represent any abstract simplicial complex. Geometric simplicial complexes can be represented on top of such a representation by adding the appropriate geometric operations. We validate our approach on several types of simplicial complexes.

compact representation. Flag complexes. A second line of research is to design compact representations of the simplicial complexes men- tioned above. A lot of work has been done for triangulated surfaces (2-dimensional simplicial complexes). Much less has been done in three dimension [16] and almost nothing in higher dimen- sions. One approach is to use semi-implicit representations where we only store the 1-skeleton of the data structures (i.e. the vertices and the edges but not the faces of higher dimensions). The full data structure is then reconstructed locally on the fly when needed. Preliminary work on Delaunay triangulations has shown that this approach dramatically reduces the space requirement while it increases the computational time by a mild factor only [18]. Further work on this topic will involve the adaptation to this context of techniques to represent graphs in a compact way while allowing fast queries of the data structure. We also intend to explore variants of this approach to efficiently code k-skeletons of simpicial complexes.

other known results on compact data structures (graphs, trees)

new computational paradigms. Parallelism, out of core.

New types of simplicial complexes. ==========

Various simplicial complexes have been proposed. The Čech complex is important in the mathematical literature because it is known to capture the homotopy type of the underlying object under general conditions (nerve theorem). However, the Čech complex The Rips complex

including the Cech, the Rips and the α-complex **??**. These complexes realize various compromises between their power to capture the topology of the underlying topological space and their complexity. One of the main bottlenecks that limit their use resides in the fact that the size of those data structures increases exponentially with the dimension of the ambient space. Restricted Delaunay triangulation.

Recently, new simplicial complexes have been defined whose complexity depend on the intrinsic dimension of the data. Witness, tangential.

Distinguish abstract/geometric, the nature of predicates involved. Connections : Čech = Rips for L1 and well known inclusion (equation). Very recently (unpublished work), we have established conditions under which RDT=TDC=intrinsic DT= witness. Such results involve a better understanding and simpler algorithms.

Our goal is to provide a fine analysis of the *complexity* of those complexes and to develop various strategies to compute those cell complexes, aiming to by-pass the curse of dimensionality. Such a study will lead to efficient code for cell complexes in higher dimensions that will be integrated in the platform.

It is important to have a better understanding of the complexity of the cell complexes mentioned above. Some results are known for simple cases. The Delaunay triangulation of point sets that belong to lower dimensional manifolds has a complexity that is lower than in the general case where the points span the entire space [].

Random complexes.

Smoothed analysis. We intend to measure the effect of perturbations (either noise or computed perturbations) on the mathematical properties and combinatorial complexity of those structures.

Complexity bounds for simplicial complexes of well sampled substructures (e.g. submanifolds).

Lastly, we intend to *implement* those structures, experiment with them and see how they be- have under realistic conditions. We intend to deliver CGAL code for the most efficient data struc- tures. Some implementation of Delaunay triangulations in higher dimensions exists, the only fully robust being the one developed by one the partners [18]. Prototype codes for other simplicial com- plexes have been developed in the context of the PLEX library http://comptop.stanford.edu/ programs/plex/ developed at Stanford university. However, the code is primarily intended for pedagogical and research purposes and does not address efficiency and robustness issues.

Validation. 1. Compute the combinatorial structure of the flow complex on an image data sets where the different dimensions are given by the images pixels, e.g., for the Semeion Handwritten Digit Data Set from the UCI repository [8]. This is a data set of 1593 points in 256 dimensions.

2. Deliver a CGAL package for computing Delaunay triangulations, witness and Rips complexes for data set with dimensions up to 10-15, as required e.g. by the analysis of astronomic data.

3. Provide benchmarks

Dimension-sensitive algorithms and data structures Work Package 2.3: Output sensitive detection of substructure

Objectives.

Typically geometric algorithms are analyzed with respect to worst-case complexity, which is often an unrealistic measure, especially when dimension is a parameter. We focus on output-sensitive and expected-case complexities. We plan to develop algorithms and data structures that balance complexity and expressiveness for representing objects of intrinsically low dimension in high-dimensional ambient spaces, and for extracting important substructures, with methods whose complexity depends on the size of the computed object. Our implementations support methods developed in Work Package 3, and enhance Cgals d-dimensional kernel in conjunction with Work Package 2.1. In particular, we examine simplicial complexes based on restricted Delaunay triangu- lations that scale well with dimension.

Work plan. Restricted Delaunay triangulations and variants.

Our first goal is to compute simplicial complexes that are good approximations of a given submanifold. In small dimensions, the approach consists in computing a subdivision of the whole space and then to extract a complex that approximates the manifold. Two main such approaches have been considered. The first one uses grids and the marching cube algorithm. Extending this approach in higher dimensions is limited due to the fact that the complexity of the grid depends exponentially on the ambient dimension. A second approach is to use Delaunay triangulations and approach the manifold by extracting the so-called restricted Delaunay triangulation (RDT). Though more adaptive than the grid approach, this method suffers also from the curse of dimensionality. Furthermore, the presence of badly shaped simplices (so-called slivers) leads to bad approximations of the manifold. To overcome these diffi- culties, we propose to construct instead variants of the RDT that scale better with the dimension and avoid badly shaped simplices. Several directions will be pursued: restricting the Delaunay triangulations to tangent spaces (tangential complex), and relaxing the definition of Delaunay tri- angulations (witness and Rips complexes). We intend to exhibit conditions on the input data under which these complexes are a good approximation of the manifold, and to provide algorithms whose complexity depends on the intrinsic dimension of the manifold.

Non smooth : homology or homotopy type via α-shapes. easier constructions ? weaker sampling conditions.

======================WITNESS

Various subcomplexes of the Delaunay triangulation have been used with success for approximating k-submanifolds of $\mathbb{R}^d$ from finite collections of sample points. Perhaps the most popular one in small dimensions ($k \in \{1, 2\}$ and $d \in \{2, 3\}$) is the so-called *restricted Delaunay complex*, defined as the subcomplex spanned by those Delaunay simplices whose dual Voronoi faces intersect the manifold. Its main selling point is to be a faithful approximation to the manifold underlying the data points, in terms of topology (ambient isotopy), of geometry (Hausdorff proximity), and of differential quantities (normals, curvatures, etc), and this under very mild conditions on the sampling density. These qualities explain its success in the context of curve and surface meshing or reconstruction, where it is used either as a data structure for the algorithms, or as a mathematical tool for their analysis, or both.

The story becomes quite different when the data is sitting in higher dimensions, where two major bottlenecks appear:

(i) The nice structural properties mentioned above no longer hold when the dimension k of the submanifold is 3 or more. In particular, normals may become arbitrarily wrong [?], and more importantly the topological type of the complex may deviate significantly from the one of the manifold [?]. These shortcomings question the usefulness of the complex as a theoretical tool.

(ii) It is not known how to compute the restricted Delaunay complex without computing the full-dimensional Delaunay triangulation or at least its restriction to some local d-dimensional neighborhood. The resulting construction time incurs an exponential dependence on the ambient dimension d, which makes the complex a prohibitively costly data structure in practice.

To address problem (i), Cheng *et al.* [?] suggested to use weighted Delaunay triangulations. The intuition underlying their approach is simple: when the restricted Delaunay complex contains badly shaped simplices, called *slivers*, its behavior in their vicinity may be arbitrarily bad: wrong normals, wrong local homology, and so on. By carefully assigning weights to the data points, one can remove all the slivers from the restricted Delaunay complex and thus have it recover its good structural properties. This idea was carried on in subsequent work [?, ?], and it is now considered a fairly common technique in Delaunay-based manifold reconstruction.

Yet, the question of computing a good set of weights given the input point cloud remains. This question is closely connected to problem (ii) above, since determining which simplices are the slivers to be removed requires first to compute the restricted Delaunay complex. To address this issue, it has been proposed to build some superset of the restricted Delaunay complex, from which the slivers are removed [?, ?]. After the operation the superset becomes equal to the restricted Delaunay complex, and thus it shares its nice properties. Unfortunately, the supersets proposed so far were pretty crude, and their construction times depended exponentially on the ambient dimension d, which made the approach quickly intractable in practice.

To circumvent the building time issue, Boissonnat and Ghosh [?] proposed to use a different subcomplex of the Delaunay triangulation, called the *tangential complex*, whose construction reduces to computing local Delaunay triangulations in (approximations of) the k-dimensional tangent spaces of the manifold at the sample points. Once these local triangulations have been computed, the tangential complex is assembled by gluing them together. Consistency issues between the local triangulations may appear, which are solved once again by a careful weight assignment over the set of data points to remove slivers. The catch is that the slivers to be removed are determined directly from the complex, not from some superset, so the complexity of the sliver removal phase reduces to a linear dependence on the ambient dimension d, while keeping an exponential dependence on the intrinsic dimension k. This makes the approach tractable under the common assumption that the data points live on a manifold with small intrinsic dimension, embedded in some potentially very high-dimensional space. Yet, the obtained complex is not the restricted Delaunay complex, and the question of whether the latter can be effectively retrieved remains open.

Enter the witness complex. In light of the apparent hardness of manifold reconstruction, researchers have turned their focus to the somewhat easier problem of inferring some topological invariants of the manifold without explicitly reconstructing it. Their belief was that more lightweight data structures would be appropriate for this simpler task, and it is in this context that Vin de Silva introduced the *witness complex* [?]. Given a point cloud W, his idea was to carefully select a subset L of landmarks on top of which the complex would be built, and to use the remaining data points to drive the complex construction. More precisely, a point $w \in W$ is called a *witness* for a simplex $\sigma \in 2^L$ if no point of L is closer to w than the vertices of σ are, i.e. if there is a ball centered at w that excludes the vertices of σ from the rest of the points of L. The witness complex is then the largest abstract simplicial complex that can be assembled using only witnessed simplices. The geometric test for being a witness can be viewed as a simplified version of the classical Delaunay predicate, and its great advantage is to require a mere comparison of distances. As a result, witness complexes can be built in arbitrary metric spaces, and the construction time is bound to the size of the input point cloud rather than to the dimension d of the ambient space (save for distance computations, which depend linearly on d).

Since its introduction, the witness complex has met a real success [?, ?, ?, ?, ?, ?], which can be explained by its close connection to the Delaunay triangulation and restricted Delaunay complex. In his seminal paper [?], de Silva showed that the witness complex is always a subcomplex of the Delaunay triangulation $\mathrm{Del}(L)$, provided that the data points lie in some Euclidean space or more generally in some Riemannian manifold of constant sectional curvature. With applications in reconstruction in mind, Attali *et al.* [?] and Guibas and Oudot [?] considered the case where the data points lie on or close to some k-submanifold of $\mathbb{R}^d$, and they showed that the witness complex is equal to the restricted Delaunay complex when $k = 1$, and a subset of it when $k = 2$. Unfortunately, the case of 3-manifolds is once again problematic, and it is now a well-known fact that the restricted Delaunay and witness complexes may differ significantly (no respective inclusion, different topological types, etc) when $k \geq 3$ [?]. To overcome this issue, Boissonnat, Guibas and Oudot [?] resorted to the sliver removal technique described above on some superset of the witness complex, whose construction incurs an exponential dependence on d. The state of affairs as of now is that the complexity of witness complex based manifold reconstruction is exponential in d, and whether it can be made only polynomial in d (while still exponential in k) remains an open question.

Our contributions. In this paper we analyze carefully the reasons why the restricted Delaunay and witness complexes fail to include each other, and from there we derive a new set of conditions under which the two complexes are equal. In addition to being very natural, our conditions involve a small superset of the witness complex, whose construction time depends only linearly on d while exponentially on k. Furthermore, they can be satisfied with constant probability using a random weight assignment process, which reduces the complexity of witness complex based manifold reconstruction to a polynomial dependence on d and thereby makes it practical.

Another aspect of the problem that is addressed in our analysis is the following: Suppose the data points have been sampled randomly from the manifold according to some known probability distribution. Then, what is the probability that the restricted Delaunay and witness complexes include each other? Following previous work in manifold learning from randomly sampled data [?], we consider the case where the underlying probability distribution is uniform, and we show that in this model the restricted Delaunay complex is included in the witness complex with high probability. We also argue that the reverse inclusion may be a rare event, and to bridge the gap we show how the classical deterministic sliver removal technique can succeed at obtaining the reverse inclusion with high probability.

============SIMPLEX TREE

================

random simplicial complexes, smoothed analysis

References

[1] P. Alliez, L. Rineau, S. Tayeb, J. Tournois, and M. Yvinec. 3D mesh generation. In *CGAL User and Reference Manual*. CGAL Editorial Board, 3.8 edition, 2011.

[2] J.-D. Boissonnat, D. Cohen-Steiner, B. Mourrain, G. Rote, and G. Vegter. Meshing of surfaces. In J.-D. Boissonnat and M. Teillaud, editors, *Effective Computational Geometry for Curves and Surfaces*, pages 181–229. Springer-Verlag, Mathematics and Visualization, 2006.

[3] J.-D. Boissonnat, O. Devillers, and S. Hornus. Incremental construction of the Delaunay graph in medium dimension. In *Proc. 25th Annual Symposium on Computational Geometry*, pages 208–216, 2009.

[4] J.-D. Boissonnat and A. Ghosh. Manifold reconstruction using tangential Delaunay complexes. In *Proc. 26th Annual Symposium on Computational Geometry*, 2010.

[5] J.-D. Boissonnat, L. J. Guibas, and S. Oudot. Manifold reconstruction in arbitrary dimensions using witness complexes. *Discrete and Computational Geometry*, 42(1):37–70, 2009.

[6] J.-D. Boissonnat and M. Yvinec. *Algorithmic geometry.* Cambridge University Press, 1998.

[7] G. Carlsson and V. de Silva. Topological estimation using witness complexes. In *Symposium on Point-Based Graphics*, 2004.

[8] F. Chazal and S. Y. Oudot. Towards Persistence-Based Reconstruction in Euclidean Spaces. In *Proc. ACM Symp. on Computational Geometry*, pages 232–241, 2008.

[9] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete Comput. Geom.*, 37(1):103–120, 2007.

[10] V. de Silva. A weak characterisation of the Delaunay triangulation. *Geometriae Dedicata*, 135:39–64, 2008.

[11] T. Dey. *Curve and Surface Reconstruction : Algorithms with Mathematical Analysis.* Cambridge University Press, 2007.

[12] H. Edelsbrunner. The union of balls and its dual shape. *Discrete Comput. Geom.*, 13:415–440, 1995.

[13] H. Edelsbrunner and J. Harer. Persistent homology — a survey. In J. P. J. E. Goodman and R. Pollack, editors, *Surveys on Discrete and Computational Geometry. Twenty Years Later*, Contemporary Mathematics 453, pages 257–282. Amer. Math. Soc., 2008.

[14] H. Edelsbrunner, D. G. Kirkpatrick, and R. Seidel. On the shape of a set of points in the plane. *IEEE Transactions on Information Theory*, 29:551–559, 1983.

[15] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28:511–533, 2002.

[16] R. Ghrist. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc., 45(1)*, pages 61–75, 2008.

[17] S. Har-Peled. *Geometric Approximation Algorithms.* Amer. Math. Soc., 2011.

[18] P. Niyogi, S. Smale, and S. Weinberger. Finding the Homology of Submanifolds with High Confidence from Random Samples. *Discrete and Computational Geometry*, 39(1):419–441, 2008.

[19] D. Spielmann and S.-H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM*, 51(3):385–463, 2004.

[20] Cgal. Computational Geometry Algorithms Library. http://www.cgal.org.

2 Ressources