Sampling and Topological Inference For General Shapes

F. Chazal, D. Cohen-Steiner and A. Lieutier

1 Introduction

In this paper [1], we use the framework of distance functions to study some geometric and topological approximation problems on compact subsets of $\mathbb{R}^n$.

1.1 Motivation.

In many practical situations, the object of study is only known through a finite set of possibly noisy sample points. It is then desirable to try to recover the geometry and the topology of the object from this information. The most obvious example is probably surface reconstruction, where the points are measured on the surface of a real world object. Also, a current research topic in cosmology is to study the large scale structure formed by the galaxies, which seems to be an interconnected network of walls and filaments [13]. In other applications, the shape of interest may live in a higher dimensional space, as for instance in machine learning and in particular in *manifold learning*. This is also the case in time series analysis, when the shape of study is the attractor of a dynamical system sampled by a sequence of observations [12].

In this context, an important question is to find a sampling condition guaranteeing that the object can be reconstructed correctly. Besides providing theoretical guarantees, such a condition may be used to drive the recontruction process itself. Indeed, a possible reconstruction strategy is to look for the shapes that are best sampled by the data points. In what follows, we investigate these questions in a fairly general setting, assuming a very simple reconstruction process.

1.2 Previous work.

The currently most successful framework for dealing with such problems is based on the notion of ε-sample, introduced by Amenta et al [2]. A sampling of a shape K is an ε-sampling if every point p in K has a sample point at distance at most $\varepsilon \,\mathrm{lfs}_K(p)$, where $\mathrm{lfs}_K(p)$ denotes the local feature size of p, that is the distance from p to the medial axis of the complement of K. It has been shown that surfaces smoothly embedded in $\mathbb{R}^3$ can be reconstructed homeomorphically from any 0.06-sampling using the Cocone algorithm [1]. The major limitation of the ε-sample framework is that it cannot handle sharp edges. Indeed, the local feature size vanishes on such edges, implying that any ε-sample must have an infinite number of points in their vicinity. Boissonnat et al [3] recently gave sampling conditions applying to non smooth surfaces in $\mathbb{R}^3$ and guaranteeing that the restricted Delaunay triangulation of the surface is isotopic to the surface. Yet, to the best

[1] already presented in [4]

of our knowledge, no reconstruction algorithm comes with theoretical guarantees in the case of non-smooth shapes, except for curves [8].

One of the simplest methods for reconstructing arbitrary shapes is to output an offset of the sampling for a suitable value α of the offset parameter. Topologically, this is equivalent to taking the α-shape [10] of the data points, which can be computed efficiently in $\mathbb{R}^3$ using the Delaunay triangulation. Recently, Niyogi, Smale and Weinberger [11] proved that this method indeed provides reconstructions having the correct homotopy type for densely enough sampled smooth submanifolds of $\mathbb{R}^n$. The precise sampling condition is that the Hausdorff distance between the sampling and the submanifold should not exceed $\sqrt{9}-\sqrt{8}$ times the minimum of the local feature size function over the submanifold. Note that this condition is similar to the ε-sampling condition, except that noise is allowed and local adaptivity to the variations of the local feature size is lost. It was recently shown that under a similar but locally adaptive condition, the union of balls centered on data points and with radius proportional to the local feature size is homotopy equivalent to the submanifold [5]. Both conditions suffer from the same problem as ε-samplings, namely they do not apply to shapes other than smooth submanifolds. For more general shapes, the homology groups can be estimated correctly from a sufficiently dense sampling using topological persistence techniques [6, 7], but it is not known how to actually build a reconstruction having the correct homology groups.

1.3 Contributions.

In this paper, we introduce a parameterized set of sampling conditions that permit to extend Niyogi, Smale and Weinberger's result to a large class of compact subsets of euclidean space. Our sampling conditions resemble those used by Niyogi, Smale and Weinberger, except that they are not based on the local feature size, but on a parameterized notion of feature size we call the *μ-reach*. The μ-reach of a compact set K is the minimum distance between a point in K and a point in the *μ-medial axis* of $\mathbb{R}^n \setminus K$, which is a filtered version of the medial axis. In particular, the μ-reach interpolates between the minimum of the local feature size (for $\mu = 1$), and the weak feature size (for $\mu \to 0$). A crucial ingredient in our approach is a generalization of a result on the separation of the critical values of distance functions recently obtained by Dey et al [9] in the smooth case. We also introduce the concept of critical function of a compact set, which can be used to choose the offset parameter appropriately by searching for the values that optimize the sampling quality. Along the way, we obtain new stability results for the μ-medial axis, as well as for critical points, critical values, and for the critical function itself.

References

[1] N. Amenta, S. Choi, T. Dey and N. Leekha *A simple algorithm for homeomorphic Surface Reconstruction.* International Journal of Computational Geometry and Applications Vol.12, No. 1,2(2002) pp.125-141

[2] N. Amenta and M. Bern *Surface Reconstruction by Voronoi Filtering* Discrete and Computational Geometry, no 22 pp.481-504(1999)

[3] J.D. Boissonnat and S. Oudot, *Provably Good Sampling and Meshing of Lipschitz Surfaces*, in 22nd ACM Symposium on Computational Geometry 2006.

[4] F. Chazal, D. Cohen-Steiner, A. Lieutier, *A Sampling Theory for Compacts in Euclidean Space*, Proceedings of the 22nd ACM Symposium on Computational Geometry 2006.

[5] F. Chazal, A. Lieutier, *Topology guaranteeing manifold reconstruction using distance function to noisy data*, Proceedings of the 22nd ACM Symposium on Computational Geometry 2006.

[6] F. Chazal, A. Lieutier, *Weak feature size and persistent homology: computing homology of solids in $\mathbb{R}^n$ from noisy data samples* 21st Symposium on Computational geometry (2005) pp. 255-262

[7] D. Cohen-Steiner, H. Edelsbrunner and J. Harer, *Stability of Persistence Diagrams*, Proc. 21th ACM Sympos. Comput. Geom. (2005), 263-271.

[8] T. K. Dey and R. Wenger, *Reconstructing Curves with Sharp Corners*, Proc. 16th ACM Sympos. Comput. Geom. (2000), 233-241.

[9] T. K. Dey, J Giesen, E.A. Ramos and B. Sadri, *Critical points of the distance to an epsilon-sampling of a surface and flow-complex-based surface reconstruction* . 21st symposium on Computational geometry (2005), pp 218-227.

[10] H. Edelsbrunner, E. P. Mücke. *Three-dimensional alpha shapes*, ACM Trans. Graphics 13 (1994), 43-72.

[11] P. Niyogi, S. Smale and S. Weinberger, *Finding the Homology of Submanifolds with High Confidence from Random samples*, to appear in Discrete and Computational Geometry.

[12] V. Robins, J.D. Meiss and L. Bradley, *Computing Connectedness: Disconnectedness and Discreteness*, Physica D 139:276-300, 2000.

[13] J.L. Starck, V.J. Martínez, D.L. Donoho, O. Levi, P. Querre and E. Saar, *Analysis of the Spatial Distribution of Galaxies by Multiscale Methods*, arXiv:astro-ph/0406425 v1 18 Jun 04.