Sampling and Topological Inference For General Shapes

Frédéric Chazal
INRIA Futurs

David Cohen-Steiner
INRIA Sophia-Antipolis

André Lieutier
Dassault Systèmes

Introduction

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. 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. In this context, an important question is to find a sampling condition guaranteeing that the shape 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. We investigate these questions in a fairly general setting assuming a very simple reconstruction process.

The distance functions framework

Topological and geometric features of a shape cannot be directly extracted from an approximating data: for example, the number of connected components of a shape is obviously not the same as the one of a point cloud approximating it. Worse, the occurence of some features may depend on a "scale" at which both the data and the shape are viewed: for example, viewed with human eyes, the surface of a real world object may look very regular but at a microscopic scale it appears as a much more complicated surface with many holes, tunnels, *etc*...

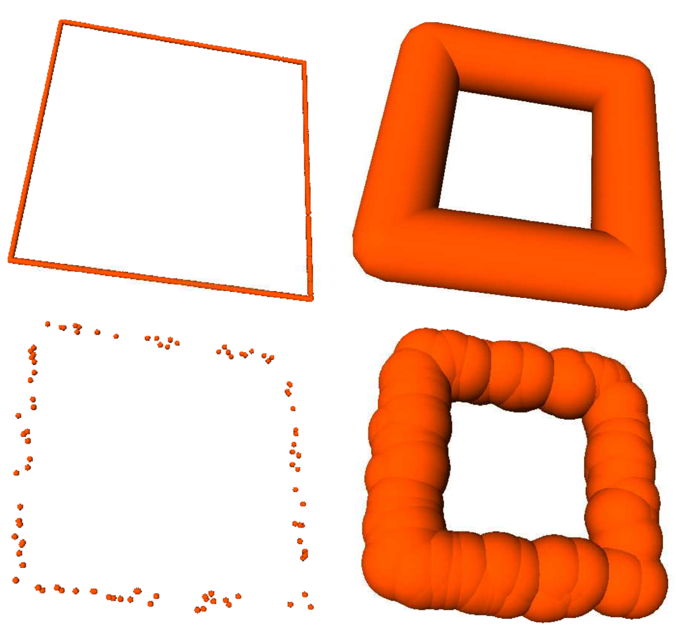
Idea: study the topology of the level sets of the distance function to the data and compare it to the one of the level sets of the distance function to the shape.

The *distance function* d_K to a shape $K \subset \mathbb{R}^n$ (i.e. a compact set) is defined by

$$d_K(x) = \inf_{y \in K} d(x,y) \quad \text{for all } x \in \mathbb{R}^n$$

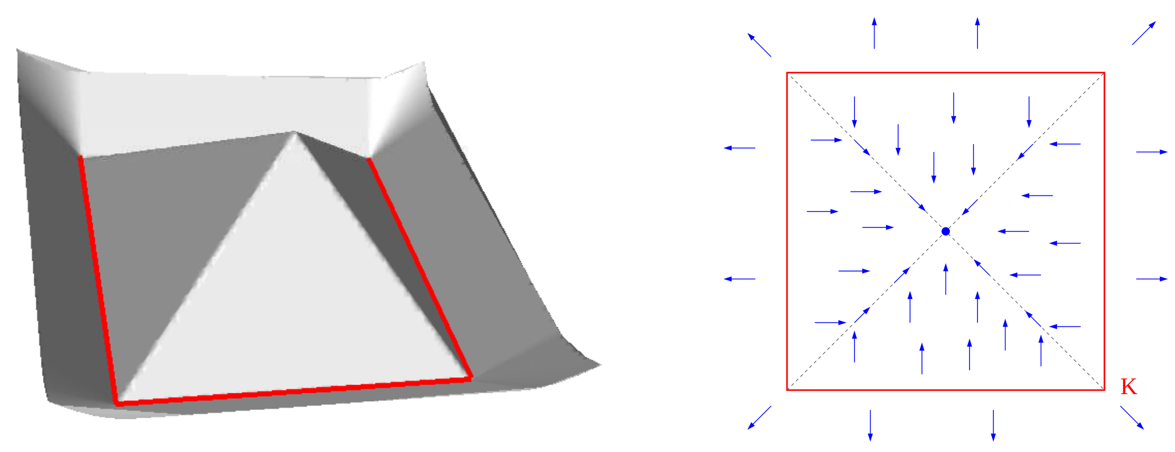
where $d(x,y)$ denotes the usual euclidean distance. The *offsets* K^r of K are the sub-level sets of the distance function: $K^r = d_K^{-1}([0,r])$.

Closeness between two shapes K and K' is measured by the mean of the *Hausdorff distance* $d_H(K,K')$ which is the smallest $r \geq 0$ such that $K \subset K'^r$ and $K' \subset K^r$.



Gradient and offset topology

Although distance functions are not differentiable everywhere, they behave almost like differentiable functions. In particular, one can define the gradient of d_K: intuitively, its direction is the one in which the "slope"' of the graph $\{(y, d_K(y)) : y \in \mathbb{R}^n\}$ of d_K is the biggest at $(x, d_K(x))$ (see figure below). The norm of the gradient has then to be the "slope" of the graph of d_K in this direction.



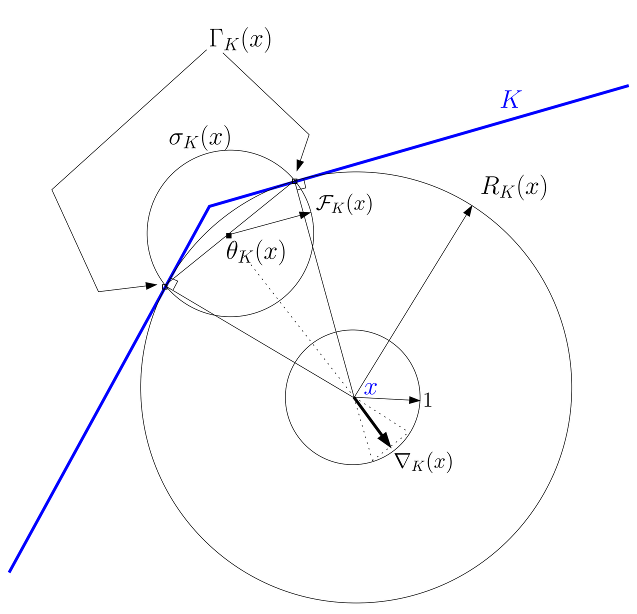
More formally, for $x \in \mathbb{R}^n$, $\Gamma_K(x)$ is the set of points in K closest to x:

$$\Gamma_K(x) = \{y \in K \mid d(x,y) = d(x,K)\}$$

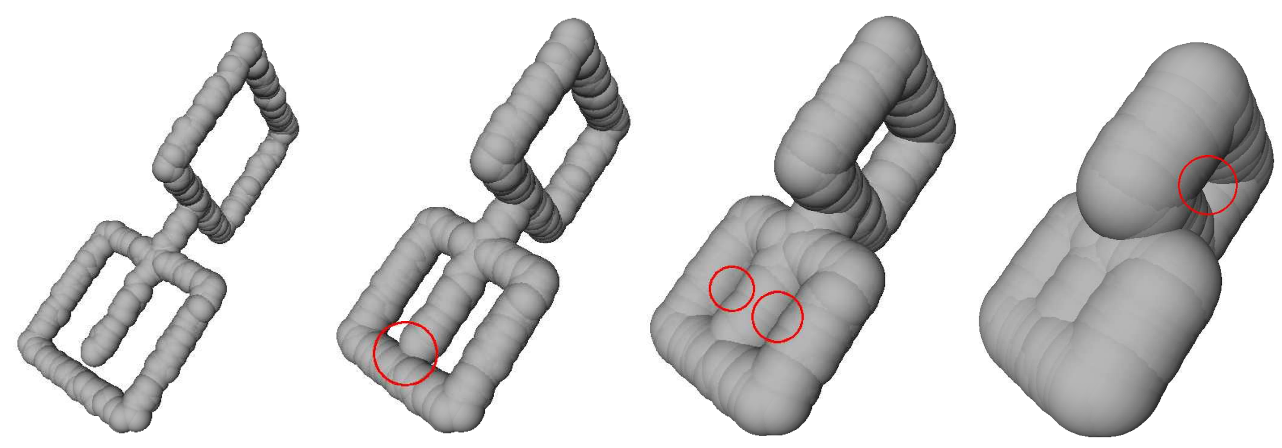
There is a unique smallest closed ball $\sigma_K(x)$ enclosing $\Gamma_K(x)$. We denote by $\theta_K(x)$ its center and by $F_K(x)$ its radius. For $x \in \mathbb{R}^n \setminus K$, *the generalized gradient* $\nabla_K(x)$ is defined by:

$$\nabla_K(x) = \frac{x - \theta_K(x)}{d_K(x)}$$

It is natural to set $\nabla_K(x) = 0$ for $x \in K$.



A *critical point* is a point x such that $\nabla_K(x) = 0$. The topology of the offsets of K can only change at critical points: if $0 < r_1 < r_2$ are such that $K^{r_2} \setminus K^{r_1}$ does not contain any critical point of d_K, then all the level sets $d_K^{-1}(r)$, $r \in [r_1, r_2]$, are homeomorphic and even isotopic topological manifolds and the offsets K^{r_1} and K^{r_2} are also isotopic (Isotopy lemma).



The *weak feature size* $\mathit{wfs}(K)$ of K is the infimum of the critical values of d_K. From the Isotopy lemma, it may be viewed as the "minimum size of the topological features" of K.

Stability of critical points

Once we know how the topology of the offsets of a given shape changes, it is necessary to compare the topology of the offsets of two close shapes K and K'. This is closely related to the study of the behavior of the critical points under perturbation of the shape. Unfortunately, it is easy to see that the critical points of a shape K are not stable when one replaces it by a close shape K'. To overcome this difficulty, one introduces a parametrized notion of critical point. Given $0 \leq \mu < 1$, a point $x \in \mathbb{R}^n$ is a *μ-critical point* of K if $\|\nabla_K(x)\| \leq \mu$.

Critical point stability theorem. *Let K and K' be two compact subsets of $\mathbb{R}^n$ and $d_H(K,K') \leq \varepsilon$. For any μ-critical point x of K, there is a $(2\sqrt{\varepsilon/d_K(x)} + \mu)$-critical point of K' at distance at most $2\sqrt{\varepsilon d_K(x)}$ from x.*

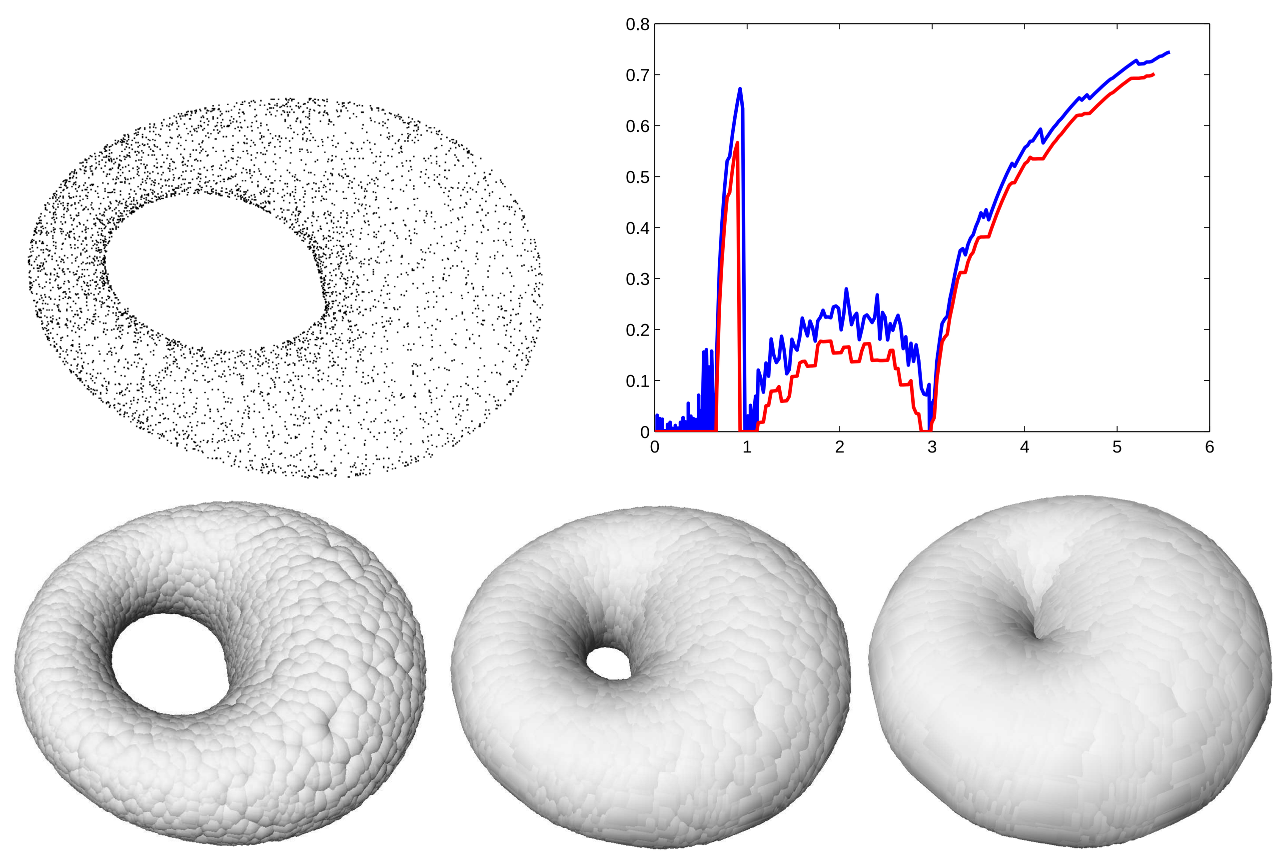
It is possible to "encode" all the μ-critical values of d_K in a real-valued one variable function. The *critical function* $\chi_K : (0, +\infty) \to \mathbb{R}_+$ of K is defined by:

$$\chi_K(d) = \inf_{d_K^{-1}(d)} ||\nabla_K||$$

For a point cloud the critical function can be easily computed from the Delaunay triangulation of the point set. Knowing the critical function of K' and the Hausdorff distance between K and K', the stability of μ-critical points provides a lower bound for χ_K.

Critical function stability theorem. *If $d_H(K,K') \leq \varepsilon$ then for all $d \geq 0$, we have $\inf\{\chi_{K'}(u) \mid u \in I(d,\varepsilon)\} \leq \chi_K(d) + 2\sqrt{\frac{\varepsilon}{d}}$ where $I(d,\varepsilon) = [d - \varepsilon, d + 2\chi_K(d)\sqrt{\varepsilon d} + 3\varepsilon]$*

It is thus possible to locate intervals on which the critical function of K does not vanish and the topology of the corresponding offsets of K does not change. The figure below illustrates this in the case of a point cloud sampled around a torus shape in $\mathbb{R}^3$ (which is not a torus of revolution). The critical function of the point cloud is in blue. The red curve is the lower bound for the critical function of any shape K at distance less than some fixed threshold (here 0.001 - the diameter of the torus being 10) from the point cloud. We distinguishes three intervals with stable topology for K: the first one corresponds to offsets having the topology of a torus (bottom left), the second one corresponds to solid torus with a void homeomorphic to a ball inside (bottom middle - not visible from outside) and the third one is unbounded and corresponds to offsets that have the topology of a ball (bottom right).



Sampling conditions and reconstruction

Although we are able to detect, from the critical function, common intervals of stable topology for all the shapes K located at some given distance from K', it remains to relate the topology of the offsets of K' to the ones of K. Fortunately, it happens that if the length of an interval where the critical function of K' does not vanish is sufficiently large with respect to the infimum of $\chi_{K'}$ then the offsets (in this interval) of K and the offsets of K' are homemorphic and even isotopic.

Level sets isotopy theorem *Let $K, K' \subset \mathbb{R}^n$ be two compact sets such that $d_H(K,K') < \varepsilon$ for some $\varepsilon > 0$. If $a > 0$ is such that $\chi_{K'} > 2\sqrt{\frac{2\varepsilon}{a-\varepsilon}}$ on the interval $[a - \varepsilon - 2\sqrt{2\varepsilon(a+\varepsilon)}, a + \varepsilon + 2\sqrt{2\varepsilon(a+\varepsilon)}]$ then $R_K^{-1}(a)$ and $R_{K'}^{-1}(a)$ are isotopic hypersurfaces and K^a and K'^a are also isotopic.*

It is important to notice that the previous theorem does not require any knowledge on K (except an upperbound on its distance to K'). This is particularly useful in practical applications where the approximation K' is usually the only information we know about K. In particular, from K' we are able to decide if there exists at some given distance ε some shape with given "intervals of topological stability" for its offsets. From this it is then possible to exhibit sampling conditions insuring that the topology of the offsets of K can be reliably recovered from the offsets of K'.

The *μ-reach, $r_\mu(K)$* of K is the infimum of the values d such that $\chi_K(d) < \mu$. Given two non-negative real numbers κ and μ, we say that $K' \subset \mathbb{R}^n$ is a *(κ, μ)-approximation* of $K \subset \mathbb{R}^n$ if the Hausdorff distance between K and K' does not exceed κ times the μ-reach of K.

Isotopic reconstruction theorem. *Let $K \subset \mathbb{R}^n$ be a compact set such that $r_\mu(K) > 0$ for some $\mu > 0$. Let K' be a (κ, μ)-approximation of K where*

$$\kappa < \min\left(\frac{4\sqrt{2}-5}{14}, \frac{\mu^2}{16 + 2\mu^2}\right)$$

and let d, d' be such that $0 < d < wfs(K)$ and $\frac{4\kappa r_\mu}{\mu^2} \leq d' < r_\mu(K) - 3\kappa r_\mu$ Then the level set $R_{K'}^{-1}(d')$ is isotopic to the level set $R_K^{-1}(d)$ (the same holds for $K'^{d'}$ and K^d).

Reference

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 / to appear in Discrete and Computational Geometry.

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE