PhD thesis proposal

Probabilistic analysis of geometric structures

Advisor: Olivier Devillers, INRIA

Location: INRIA, INRIA Nancy Grand Est
615, rue du Jardin Botanique
B.P. 101
54602 Villers-lès-Nancy cedex
FRANCE
Contact: Olivier Devillers <Olivier.Devillers@inria.fr>

1 Context and motivation

Geometric problems are central in many areas of science and engineering. Computational geometry, the study of combinatorial and algorithmic problems in a geometric setting, and in particular triangulations, have tremendous practical applications in areas such as computer graphics, computer vision and imaging, scientific visualization, geographic information systems, astronomy, computational biology... Traditionally, the complexity of computational geometry algorithms is studied in the worst case setting. This kind of analysis is often quite pessimistic compared to real life data.

On the other hand, probability analyzes several geometric problems such as the size of the convex hull or the Delaunay triangulation of a random set of points, but under hypotheses of uniform distribution of the data in some domain. This kind of analysis is often too optimistic compared to real life data.

Two examples of such discrepancy between worst case and random settings are the convex hull and the Delaunay triangulation. For the convex hull, focusing on the case of dimension two, the size of the convex hull is between constant and linear without random hypotheses while the expected size is $\Theta(n^{\frac{1}{3}})$ or $\Theta(\log n)$ if the points are evenly distributed in a disk or a square. The worst case size of the 3D Delaunay triangulation of n points may be as high as $\Omega(n^2)$, but this worst case happens only in quite pathological configurations that have been designed on purpose. If the points are evenly distributed in some region, the expected size is only $\Theta(n)$.

Analysing data structures and algorithms under random hypotheses more complicated that uniform distribution is of a great importance to have a better foot with real applications. We propose the three following research directions.

2 Research directions

Smoothed analysis

The smoothed analysis, introduced to analyze the simplex algorithm [ST04], interpolates between worst case and uniform distribution using a noise parameter. When there is no noise, we have the worst case situation and when the noise goes to infinity we are in a full random situation. Analyzing how the situation evolves in between indicates how the worst case complexities are pathological or realistic.

In geometry, some work has been done to analyze the size of the convex hull under uniform and Gaussian noise [DGGT15], but the current results are still not tight and the more interesting problem of Delaunay triangulation needs to be addressed.

Points on surfaces $\longrightarrow$

Reconstructing a surface from a point set is one of the main use of 3D Delaunay triangulation in real applications. In such applications, by definition the points are not distributed in the whole 3D space but on a surface, thus understanding the expected complexity of the Delaunay triangulation of a random sample of a surface is an important problem with practical implications.

There are several results concerning the size of the Delaunay triangulation when the point set is a sampling of a surface. These results depends of course of the surface definition and the sampling definition. The surface can be smooth, or only piecewise smooth, it can also be *generic* if it does not include pieces of circle (no spheres, cones, cylinders…) or not. The sampling can be a good-sampling (no *holes* nor *concentration* in the sampling) or random. Current known results on the size of the triangulation of a sample of n points are
- $O(n \log n)$ for a *good-sampling* of a generic surface [ABL03],
- $\Omega(n\sqrt{n})$ for a sampling of a general (non-generic) surface satisfying the same good-sampling hypothesis [Eri05], and
- $\Theta(n \log n)$ for a random sampling of a cylinder [DEG08].

Since a random sample is a good sample with high probability, the first result allow to deduce a complexity of $O(n\text{polylog}n)$ for a random sampling of a generic surface.

These results are not tight, and the very reasonable case of a random sampling of the surface is not covered in a good way.

In practice, the measurements may also be noisy, making a link between this subject and the smooth analysis.

Conditional distribution

Another way of going beyond uniform distribution is the use of conditional distributions. To understand such distributions we propose to try to simulate the uniform distribution conditioned by some rare events. An example of such situation is the generation of n uniform random points in some domain (e.g. a disk) under the condition that they are in convex position.

We have in mind two directions to attack this problem: first finding a weaker condition than convexity that would both be simple to sample, and close enough to allow rejection to be efficient, and second adapting the promising random walk techniques in [Vem05]. Such a random generator will give probability estimates for Sylvester's problem (what is the probability that n random points are in convex position), which is still unsolved.

3 Software

The above research directions will have several software outcomes: random generators, new algorithms for building geometric structures optimized for some distributions. These software aspects will be first coded using CGAL, and then submitted to CGAL if appropriate.

CGAL is the Computational Geometry Algorithms Library, an Open Source Library implementing state of the art algorithms in computational geometry.[1] The impact of CGAL extends far beyond computational geometry. Application fields include particle physics, fluid dynamics, shape matching, image processing, geometry processing, computer graphics, computer vision, shape reconstruction, mesh generation, virtual worlds, geophysics, and medical imaging.[2]

CGAL is also commercialized by GeometryFactory, an Inria startup.[3]

4 Knowledge involved:

The topic involves knowledge in both probability and computer science, thus candidate will need good skills in at least one of these two domains and some knowledge in the other. Regarding software coding, CGAL is coded using C++ with an heavy use of the "template" mechanism, and some familiarity with C++ or another similar language is welcome.

[1] http://www.cgal.org/

[2] http://www.cgal.org/projects.html. CGAL 2D and 3D triangulations have also been integrated in Matlab in 2009.

[3] http://geometryfactory.com/

5 Advisor

Thesis will be advised by Olivier Devillers at INRIA. There is a collaboration on that topic with Philippe Chassaing (IECL) on probabilistic aspects.

References

[ABL03] Dominique Attali, Jean-Daniel Boissonnat, and André Lieutier, *Complexity of the Delaunay triangulation of points on surfaces: The smooth case*, Proc. 19th Annual Symposium on Computational Geometry, 2003, pp. 201–210.

[DEG08] Olivier Devillers, Jeff Erickson, and Xavier Goaoc, *Empty-ellipse graphs*, 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'08), 2008, pp. 1249–1256.

[DGGT15] Olivier Devillers, Marc Glisse, Xavier Goaoc, and Rémy Thomasse, *On the smoothed complexity of convex hulls* , 31st International Symposium on Computational Geometry (Eindhoven, Netherlands), Lipics, June 2015.

[Eri05] Jeff Erickson, *Dense point sets have sparse Delaunay triangulations or "... but not too nasty"*, Discrete & Computational Geometry **33** (2005), 83–115.

[ST04] Daniel A. Spielman and Shang-Hua Teng, *Smoothed analysis: Why the simplex algorithm usually takes polynomial time*, Journal of the ACM **51** (2004), 385 – 463.

[Vem05] Santosh Vempala, *Geometric random walks: a survey*, Combinatorial and computational geometry **52** (2005), 573–612.