PhD thesis proposal

Simulation of random geometric structures

Advisor: Olivier Devillers

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

1 Context and motivation

The complexity of geometric algorithms is often express in terms of the input size and sometimes of the output size, but the behavior of these algorithms may depend drastically on the geometric distribution of the input. The difference between the best-case and the worst-case can be important. An alternative is to analyze these algorithms under the hypothesis of a probabilistic distribution of the data. When the input is a point set, an easy probabilistic hypothesis is to consider that points are independently distributed under some law (Poisson distribution, uniform distribution...), unfortunately this hypothesis of independance may be unsuitable for several applications.

The use of non independent distribution is very difficult to analyze theoretically, thus having access to simulations of practical instances would be of tremendous help to guide the intuition and to lead to reasonable conjectures. Unfortunately, the generation of useful random instances is a difficult question in itself. We plan to attack three specific questions, as described below: simulation of geometric structures, of conditioned structures, and of dependent pointsets.

2 Objectives

The goal of this research is to design and realize tools for the generation of several random geometric objects. These tools will be used to benchmark geometric algorithms and to establish new probability conjectures.

3 Knowledge involved:

- mathematical aspects (probability)
- algorithmic aspects
- C++ (templates, etc)

4 Collaborations and funding

The thesis will be done in the framework of the ANR Aspag (https://members.loria.fr/Olivier.Devillers/aspag/).

The PhD candidate will be member of the Gamble team at LORIA and INRIA Nancy – Grand Est. The Gamble team is focused on computational geometry but, through the APSAG project, the candidate will have to work with specialists in probability and in analysis of algorithms.

5 Working program

The thesis will address the following questions:

5.1 Generation of geometric structures.

Random generation of geometric structures has so far been the subject of significantly less research than for combinatorial structures. The well-established approaches for combinatorial structures (bijections, recursive or analytic techniques, to name a few) all fall apart when dealing with geometric objects. We intend to address the problem of random generation of geometric structures and to develop general techniques. Some specific questions of interest concern the simulation of *sparse or partial structures*, and the relationship between the distributions induced by *geometric and combinatorial sampling.* We also intend to work at *making geometric simulation tools more accessible* to the potential users.

Sparse structures. One of the general frameworks we will consider is that of sparse structures, for which typically only a vanishing proportion of the pointset is part of the output. Problems of this kind abound: in convex hulls, in power diagrams, or when computing the boundary of a union of objects, it is very likely that the vast majority of the input data has no influence on the computed structure. Our goal is to design algorithms of random generation with a near-linear running time in the output size. For the convex hull of n random points in a disk, we reached an $O(n^{\frac{1}{3}} \log^2 n)$ complexity, far below the $O(n \log n)$ naive complexity [5]. We plan to extend this kind of approach to other settings. The algorithms will initially be problem dependent, but we expect to obtain general techniques that can adapt to a variety of problems and situations.

Partial structures. In some other situations, it may be the case that even if all the input participates in the definition of the output, only a part of the output is relevant. Examples of this kind of situation include paths in Delaunay triangulations: either intrinsic (shortest paths) or defined through a pathfinding algorithm.

Access to simulations. Finally, one of the objectives will be to provide a more user-oriented access to simulations. Implementation in CGAL is a significant step, but we will also provide a more accessible, less programming-intensive, interface through *ipelets* [2]. For instance, the simulator we have developed for Delaunay triangulations, in which we can generate samples with millions of points in a few seconds and extract relevant statistics, has already played an essential role in obtaining novel theoretical asymptotic estimates about extremal cells [1].

5.2 Conditioned structures.

A closely related —but fundamentally different— open question is the design of algorithms that simulate random objects conditional on the occurrence of some rare event. From a probabilistic point of view, this relates to large deviations estimates, the precise quantifications of extreme values of various parameters, and the concept of quasi-stationary distributions.

Such questions arise, e,.g., for *points in convex position*, portions of *tessellations in the neighborhood of exceptional locations*, and *conditioned geometric networks*.

In the thesis we will focus on *points in convex position.* The problem, closely related to a famous Sylvester's problem, is to devise an efficient algorithm to generate a set of n random points in some domain, under the condition that they are in convex position. Other than the cases of the triangle and square domains, for which a solution is provided by Valtr [11, 12], the probability of convex position vanishes so quickly that naive algorithms have trouble dealing with more than 15 points. We plan to explore two possible approaches to solve this problem: either find a weaker condition than convexity that is easy to simulate yet close enough that rejection becomes feasible, or adapting the random walk techniques in [13]. We strongly believe that the simultaneous focus on simulation algorithms and on probability estimates for Sylvester's problem will yield beneficial mutual strengthening.

5.3 Simulation of dependent distributions.

Efficiently generating a determinantal point process (DPP) with just a few hundred points is currently out of reach. Yet, this question is of fundamental importance judging from the practical applications of DPP in physics or networks. We plan to take on this challenge, and we believe that our combined expertise will yield significant improvements.

The simulation algorithm proposed in [7] is subtle and elegant and is simply implementable for the simulation of a DPP configuration of a few dozens of points [9]. As the number of points grows, numerical instability problems compounded with the algorithm's complexity greatly increase the practical difficulty and computation time. Moreover, this algorithm assumes that the eigenvalues and eigenvectors of an infinite dimensional map are known. It is thus necessary to develop new simulation techniques not relying on such a decomposition. The specific case of the Ginibre point process has been treated by an ASPAG member [3] using a very interesting truncation technique of the kernel; generating techniques based on coupling from the past have been explored in [14]. However both approaches are limited: one is only approximate because of the truncation, and the other assumes that the spectral representation is available. A promising solution is to consider perfect simulation involving a dominating Poisson process as in [8].

Moreover, the algorithm in [7], as well as many others, raises questions about the asymptotics of some functionals of DPP: such estimates are indeed crucial for the running times of the algorithms. For instance, there are few known examples [10, 6] of central limit theorems concerning DPP. They exclusively focus on the asymptotic properties of the number of points in an increasing domain, and the proofs are specific for each kernel. This leaves wide open the problem of generalizing these proofs to a larger class of processes and to more general functionals. We believe that the new techniques set up in [4] will be of tremendous help to reach a positive conclusion.

References

[1] N. Chenavier and R. Hemsley. Extremes for the inradius in the Poisson line tessellation. *Advances in Applied Probability*, 2016. URL: http://arxiv.org/abs/1502.00135, doi:10.1017/apr.2016.14.

[2] Otfried Cheong. The ipe extensible drawing editor.

[3] L. Decreusefond, I. Flint, and A. Vergne. Efficient simulation of the Ginibre point process. *Journal of Applied Probability*, 52(4), 2015. URL: http://arxiv.org/abs/1310.0800.

[4] L. Decreusefond, M. Schulte, and C. Thäle. Functional Poisson approximation in Kantorovich-Rubinstein distance with applications to U-statistics and stochastic geometry. *Annals of Probability*, 44(3):2147–2197, 2015. URL: http://de.arxiv.org/abs/1406.5484.

[5] O. Devillers, P. Duchon, and R. Thomasse. A generator of random convex polygons in a disc. AofA - 25th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, 2014. URL: https://hal.inria.fr/hal-01015603.

[6] Peter J. Forrester and Joel L. Lebowitz. Local central limit theorem for determinantal point processes. *Journal of Statistical Physics*, 157(1):60–69, 2014.

[7] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág. Determinantal processes and independence. *Probab. Surv.*, 3:206–229 (electronic), 2006.

[8] Wilfrid S. Kendall and Jesper Møller. Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability*, 32(3):844–865, 2000.

[9] Frédéric Lavancier, Jesper Møller, and Ege Rubak. Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2014.

[10] AlexanderB. Soshnikov. Gaussian fluctuation for the number of particles in Airy, Bessel, Sine, and other determinantal random point fields. *Journal of Statistical Physics*, 100(3-4):491–522, 2000.

[11] Pavel Valtr. Probability that n random points are in convex position. *Discrete & Computational Geometry*, 13:637–643, 1995.

[12] Pavel Valtr. The probability that n random points in a triangle are in convex position. *Combinatorica*, 16(4):567–573, 1996.

[13] Santosh Vempala. Geometric random walks: a survey. *MSRI volume on Combinatorial and Computational Geometry*, 2005.

[14] Anaïs Vergne, Ian Flint, Laurent Decreusefond, and Philippe Martins. Disaster Recovery in Wireless Networks: A Homology-Based Algorithm. In *ICT 2014*, Lisbon, Portugal, May 2014. URL: https://hal.archives-ouvertes.fr/hal-00800520, doi:10.1109/ICT.2014.6845113.