Voronoi Diagram and Delaunay Triangulation: Applications and Challenges in Bioinformatics

Binhai Zhu
Computer Science Department
Montana State University
Bozeman, MT 59717, USA
bhz@cs.montana.edu

1. Introduction

In this talk, I first cover the applications of Voronoi Diagram and Delaunay Triangulation based on my personal experience. On the successful part, I will cover the history and development of the Jump-and-Walk algorithm, which is known as the first sublinear geometric algorithm nowadays and has been used in several famous software packages. On the bioinformatics applications, I will focus on three challenges: (1) contour interpolation, (2) protein folding, and (3) protein local structure alignment. I think that Voronoi Diagram and Delaunay Triangulation can play important roles in handling these problems.

2. Jump-and-Walk: History and Applications

Although the concepts of Voronoi Diagram (VD) and Delaunay Triangulation (DT) were brought out long time ago, the automatic construction of both was only started in 1970s. One of the first incremental algorithm to construct Delaunay Triangulation of a set of points $S, |S| = n$, in the plane was due to Green and Sibson [7]. In that algorithm, the most important procedure is to dynamically locate a point in the temporary Delaunay Triangulation (before the update). It looks like the traditional point location problem [9], yet it is not. The reason is that in this case constructing complex data structures for the purpose of point location is impractical, especially when n is large.

What Green and Sibson proposed was a very simply heuristic method: randomly pick a vertex (or the mid-point of a random edge) p and walk in the DT from p to the query point q to locate the triangle containing q. It was empirically observed that for a DT with n vertices, this takes roughly $O(n^{1/2})$ time but no proof was given (and in the worst case it certainly takes $O(n)$ time). In 3D a similar phenomenon was observed by Bowyer [1].

Almost two decades later, this method was revisited and improved (i.e., instead of selecting one random vertex, one can select m vertices and walk from the one closest to q for the point location). When the vertices of DT are quasi-uniformly distributed it can be shown that on average this algorithm takes $O(n^{1/3})$ time when q is slightly away from the boundary of the convex domain where the vertices of the DT are drawn from and when $m = \Theta(n^{1/3})$ [5]. (The boundary condition was thrown away in the recent work by Devroye, Lemaire and Moreau [4].) In 3D there was a similar bound of roughly $O(n^{1/4})$ time [8]. This algorithm was coined as 'Jump-and-Walk' since 1996.

Recently, sublinear algorithms were designed for various problems, including geometric problems. Chazelle, Liu and Magen showed that for arbitrary DT, the point location (without preprocessing) can be done in $\Theta(n^{1/2})$ time [3]. (They also mentioned that Jump-and-Walk was in fact the first sublinear geometric algorithm — of course, with formal proof and in the average case.) But that does not contradict with the bound we obtained for Jump-and-Walk. In practice Jump-and-Walk was mostly used to add Steiner points in a DT to improve the quality of the triangulation (like in finite element analysis). Most of the time, when enough Steiner points are added the resulting triangulation is almost like a DT for quasi-uniform points. This is the reason why Jump-and-Walk has been successfully used in several famous software packages like Triangle, QHULL, CGAL and X3D Grid Generation System, etc.

3. Challenges in Bioinformatics

3.1. Contour Interpolation

Contour interpolation is the problem of reconstructing a 3D solid from two parallel slices of polygons, with Steiner points. This problem is important in medical image processing, medical science, bioinformatics, solid modeling and computational neuroscience. A lot of research has been

conducted since 1970s, but none with a theoretical performance guarantee on the quality of the reconstructed 3D polyhedron.

Recently, by using the Constrained Delaunay Triangulation, we obtained the first non-trivial theoretical bound for this problem [2]. The bound itself is not meaningful in practice. It would be interesting to know whether practically meaningful bound can be obtained, as this might shed light on the existence of better practical algorithms/systems.

3.2. Protein folding

Protein folding, or more precisely, protein structure prediction from sequences of amino acids, is a very important problem in bioinformatics. The problem is in general still open. In the past, most of the results have been obtained based on some toy models (e.g., over a grid). I will cover the potential application of Voronoi (Power) Diagram in this problem. This kind of research has been started very recently, for example, the work by Dupuis, *et al.* [6] and by Soyer, *et al.* [10].

3.3. Protein local structure alignment

Nowadays, the Protein Data Bank (PDB) contains close to 50,000 protein chains (and this number is increasing almost on a daily basis). Given a new protein chain, each vertex corresponding to a residue, it would be very helpful to extract the common structure of it and a group of proteins in the PDB — especially if we know that this common structure contributes to some special properties and pathways for that group of proteins. Due to the size of the problem and the amount of potential computation involved, little progress has been made in the past and this problem remains to be solved. Voronoi Diagram might play an important role in solving this problem.

References

[1] A. Bowyer. Computing Dirichlet tessellations. *The Computer Journal*, 24:162–166, 1981.

[2] S. Bereg, M. Jiang and B. Zhu. Contour interpolation with bounded dihedral angles. In *Proc. 9th ACM Sypm. on Solid Modeling and Applications*, pages 303-308, Genova, Italy, 2004.

[3] B. Chazelle, D. Liu and A. Magen. Sublinear geometric algorithms. In *Proceedings of STOC'03*, pages 531–540, 2003.

[4] L. Devroye, C. Lemaire and J-M. Moreau. Expected time analysis for Delaunay point location. *Comput. Geom.*, 29(2):61–89, 2004.

[5] L. Devroye, E. P. Mücke and B. Zhu. A note on point location in Delaunay triangulations of random points. *Algorithmica*, Special Issue on Average Case Analysis of Algorithms, 22(4):477-482, 1998.

[6] F. Dupuis, J-F. Sadoc, R. Jullien, B. Angelov and J-P. Mornon. Voro3D: 3D Voronoi tessellations applied to protein structures. *Bioinformatics*, 21(8):1715-1716, 2005.

[7] P. J. Green and R. Sibson. Computing Dirichlet tessellations in the plane. *The Computer Journal*, 21:168–173, 1978.

[8] E. P. Mücke, I. Saias and B. Zhu. Fast randomized point location without preprocessing in two and three-dimensional Delaunay triangulations. *Comput. Geom.*, 12(1/2):63-83, 1999 (also in SoCG'96, pages 274-283).

[9] F. P. Preparata and M.I. Shamos. *Computational Geometry: An Introduction*. Springer-Verlag, 1985.

[10] A. Soyer, J. Chomilier, R. Jullien and J. Sadoc. Voronoi tessellation reveals the condensed matter character of folded proteins. *Phys. Rev. Lett.*, 85:3532-3535, 2000.