EXPECTED TIME ANALYSIS FOR DELAUNAY POINT LOCATION

Luc Devroye, Christophe Lemaire and Jean-Michel Moreau

February 16, 2004

ABSTRACT. We consider point location in Delaunay triangulations with the aid of simple data structures. In particular, we analyze methods in which a simple data structure is used to first locate a point close to the query point. For points uniformly distributed on the unit square, we show that the expected point location complexities are $\Theta(\sqrt{n})$ for the Green-Sibson rectilinear search, $\Theta(n^{1/3})$ for Jump and Walk, $\Theta(n^{1/4})$ for BinSearch and Walk (which uses a 1-dimensional search tree), $\Theta(n^{0.056\ldots})$ for search based on a random 2-d tree, and $\Theta(\log n)$ for search aided by a 2-d median tree.

KEYWORDS AND PHRASES. Delaunay triangulation, Voronoi diagram, point location, computational geometry, expected time analysis, random tree, probabilistic analysis.

CR CATEGORIES: 3.74, 5.25, 5.5.

Research of the first author was sponsored by NSERC Grant A3456 and by FCAR Grant 90-ER-0291. Address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. Email: luc@cs.mcgill.ca.

Introduction

Delaunay triangulations are fundamental geometric structures (Borouchaki and George, 1997; Okabe, Boots and Sugihara, 1992; Boissonnat and Yvinec, (1995, 1998)). Guibas and Stolfi (1985) described the basic data structure for storing such triangulations. Fundamental operations include insertion, deletion and point location. In this paper, we reconsider the point location problem, motivated by the numerical problems encountered by many of the fancy solutions seen in some recent papers, and by the need to keep memory use to a minimum. In particular, we analyze the expected time of some simple algorithms that require elementary data structures. The present paper should be considered as a companion paper of Devroye, Lemaire and Moreau (1999) where several of the algorithms of this paper are experimentally compared with the major algorithms for data sets up to 12 million in size. We present all the probabilistic results that serve to support and explain these simulations.

The randomized algorithm of Guibas, Knuth and Sharir (1992), and the partially or fully dynamic randomized algorithms of Devillers (2002), Boissonnat and Teillaud (1993), and Devillers, Meiser and Teillaud (1992) all use a hierarchy of triangulations or space partitions, with a point location search started at the root of the hierarchy. The bottom of the hierarchy typically represents the full Delaunay triangulation. While these methods yield optimal randomized query times of $\Theta(\log n)$ for any data sets in the plane, they are relatively complicated to implement, and use quite a bit of memory, Devillers (2002) excepted. The present paper, in contrast, does not require any advanced data structure beyond the Guibas-Stolfi implementation of the triangulation.

We assume throughout this paper that the data $X_1, \ldots, X_n$ and the query point X are independent and uniformly distributed on $[0,1]^2$. We will use the abbreviation i.i.d. to denote independent and identically distributed random variables. The Delaunay triangulation for $X_1, \ldots, X_n$ will simply be called the random Delaunay triangulation. In 1978, Green and Sibson proposed rectilinear search, based on ideas of Lawson (1977). Here one draws a data point at random, and walks in the Delaunay triangulation to X to determine the triangle for X. It was conjectured there that the expected time is $O(\sqrt{n})$. We will prove this.

The jump and walk method of Devroye, Mücke and Zhu (1998) offers good prospects without any additional data structure: take a random sample of size $n^{1/3}$, find the closest point Y to the query point X, and walk along the Delaunay structure, following the segment (Y, X) to determine the triangle for X. We will reprove that this takes $O(n^{1/3})$ expected time.

Minimal one-dimensional organization with the help of a one-dimensional search tree reduces the expected time even further. For example, the binary search and walk method developed by Lemaire (1997) will be shown to have expected time $O(n^{1/4})$. Dubbed BinSearch and Walk, the walk is started from a point Y that is found at a cost of $n^{1/4}$ using a 1-d search tree based on x-coordinates only.

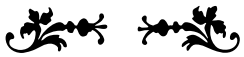
Further improvements are possible if the data are organized by means of a random 2-d tree (Bentley, 1975). An algorithm is presented below whose expected point location complexity is $O(n^{0.056\cdots})$, with further improvements unlikely with such a structure in view of recent results of Chanzy, Devroye and Zamora-Cura (2001).

Finally, if we organize the data by means of a perfectly balanced 2d-tree, using medians for splitting, and if we locate the leaf region for X, and let Y be the parent of that leaf region, then the expected time for this search and the subsequent Delaunay triangulation walk from Y to X is shown here to be $O(\log n)$.

Previous attempts at studying Delaunay edge crossings were sometimes restricted by the model: for example, in infinite Poisson process models, the irregular behavior of Delaunay triangulations near the perimeter of the support is avoided. Others consider a finite model for the data, but assume that the query point is at least ϵ away from the rim of the support (see, e.g., Bern, Eppstein and Yao, 1991, or Devroye, Lemaire and Moreau, 1999). The present analysis provides a global attack on the problem, which could serve as an example for other similar analyses. The paper is rather technical, as the boundary effect can be devastating for certain complexities. For example, we know that the convex hull of n uniformly distributed points in the unit square has expected size $\Theta(\log n)$. In fact, the longest edge on that convex hull has expected value $\Theta(1)$. For point location queries near the border of the square, walks from Y to X as described above will very likely cut many of these long edges that are prevalent near the borders. Thus, in particular, there is a difference between the expected worst case query time (when X is chosen as a function of the data so as to make the point location time worst; this will be studied elsewhere) and the expected uniform query time (when X is uniformly distributed on the square and independent of the data). It is the latter situation we are dealing with here.

The first section sets up some useful auxiliary results. We will base several inequalities on the relationship between the Delaunay triangulation and the halfmoon graph (Lemma 1). Lemma 2 states that Delaunay edges that have some point at least $\sqrt{\log n/n}$ away from the border cannot exceed about $\sqrt{\log n/n}$ in length. So, this threshold defines a rim beyond which the boundary effect is considerable. Lemma 4 for example shows that a segment of length L away from the rim gets cut roughly by $L\sqrt{n}$ Delaunay edges. That Lemma suffices to establish the Green-Sibson result. Lemmas 5 through 7 give very

useful properties on the probability that all cells in a rectangular grid are occupied, and on the maximal occupancy. These make the proof of Lemma 8 shorter, where we show that the expected maximal degree is unlikely to be more than $O(\log^2 n)$. Contrast this with the observation that the expected degree of a node given that it is not in the rim is $O(1)$ (Lemma 9). The results on the degrees are essential for Theorems 2, 3 and 4.



Properties of the random Delaunay triangulation

In this section, we give a short proof of some results related to the shape of a random uniform Delaunay triangulation. Throughout this section, $X_1, \ldots, X_n$ are i.i.d. uniform $[0,1]^2$ random variables. We define a halfcircle supported by a segment (x, y) as one of the two half circles of the circle with x and y on the perimeter at opposite ends, and divided by the segment (x, y).

LEMMA 1. *Let $x_1, \ldots, x_n$ be points in the plane. If (x_i, x_j) is a Delaunay edge, then one of two halfcircles supported by (x_i, x_j) cannot contain any other data point.*

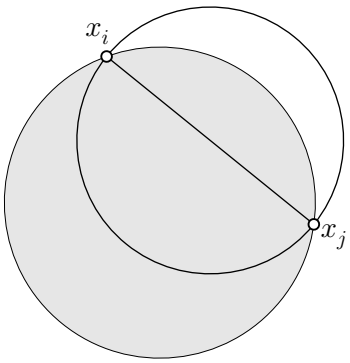


Figure 1. Proof of Lemma 1: if (x_i, x_j) is a Delaunay edge, then there exists a circle (shaded) with those two points on its perimeter that does not contain any other data point in its interior. But this circle contains one of the two halfcircles supported by (x_i, x_j), and thus one of these has no data point in its interior.

We call the halfmoon graph for $x_1, \ldots, x_n$ the graph obtained by connecting x_i with x_j when one of the two halfcircles supported by (x_i, x_j) contains no x_k, $k \neq i, j$. By Lemma 1, the Delaunay triangulation is a subgraph of the halfmoon graph. It is more convenient to bound edge lengths in the halfmoon graph. Let 1_A denote the indicator function of an event A, i.e., a function that is one if A is true and 0 otherwise.

LEMMA 2. *Consider a random Delaunay triangulation. Let $H_{n,u}$ be the maximal edge length among those edges of the halfmoon graph for which some point on the edge is at distance at least u from the square's perimeter. Then*

$$\mathbb{P}\{H_{n,u} \geq t\} \leq (ne)^2 e^{-\frac{\pi n(\min(u,t))^2}{8}}.$$

Thus, for $\epsilon > 0$, with $t_n = ((16+\epsilon)\log n/\pi n)^{1/2}$,

$$\limsup_{n\to\infty} \mathbb{P}\left\{H_{n,t_n} \geq t_n\right\} = 0 \ .$$

Let $L_{n,u}$ be the length of the longest Delaunay edge emanating from a data point that is at least distance u from the perimeter of the unit square. Then, for $\ell \geq 2u$:

$$\mathbb{P}\{L_{n,u} \geq \ell\} \leq n^2 e^{-(n-2)u\ell\sqrt{3/16}} \ .$$

PROOF. Let $B = [u, 1-u]^2$, $0 < u < 1/2$. Let $X_1, \ldots, X_n$ be the data points. For fixed X_i, X_j, let $X_{ij} = (X_i + X_j)/2$, and note that if (X_i, X_j) intersects B (an event denoted by E_{ij}), then X_{ij} is at least at distance $\min(u, \|X_i - X_j\|/2)$ from the perimeter. By the union bound, if we denote by A_{ij} and B_{ij} the two halfcircles supported by (X_i, X_j), we see that

$$\begin{aligned}
\mathbb{P}\{H_{n,u} \geq t\} &\leq \sum_{i\neq j} \mathbb{P}\left\{E_{ij} \cap [\|X_i - X_j\| \geq t] \cap [A_{ij} \text{ is empty, or } B_{ij} \text{ is empty }]\right\} \\
&\leq n(n-1)\mathbb{P}\left\{E_{12} \cap [\|X_1 - X_2\| \geq t] \cap [A_{12} \text{ is empty }]\right\} \\
&= n(n-1)\mathbb{E}\left\{1_{E_{12}} 1_{[\|X_1 - X_2\| \geq t]} \left(1 - \mathbb{P}\{X_3 \in A_{12} | E_{12}, \|X_1 - X_2\| \geq t\}\right)^{n-2}\right\}.
\end{aligned}$$

We note that $\mathbb{P}\{X_3 \in A_{12} | E_{12}, \|X_1 - X_2\| \geq t\} \geq \pi \min(u,t)^2/8$, as the circle centered at X_{12} of radius $\min(u,t)/2$ is completely contained in $[0,1]^2$. Thus,

$$\mathbb{P}\{H_{n,u} \geq t\} \leq (ne)^2 e^{-\frac{n\pi \min(t,u)^2}{8}} \ .$$

Consider the last part of the lemma. Fix a data pair X_i, X_j, with X_j at least u away from the perimeter of the unit square, and with $\|X_i - X_j\| \geq \ell$. Then the probability that there is a Delaunay edge from X_i to X_j is maximized if $\|X_i - X_j\| = \ell$. If $\ell \geq u$, then the probability is further maximized by placing X_i on the perimeter of the unit square.

Let R be the rectangle with X_i and X_j at opposite ends, and let T, T' be two triangles of R obtained by splitting it along the diagonal (X_i, X_j). If (X_i, X_j) is to be a Delaunay edge, then one of these triangles must not have any other data points. Given X_i, X_j, the probability of this is bounded from above by $2(1 - (1/2)u\sqrt{\ell^2 - u^2})^{n-2}$. For $\ell \geq 2u$, this is further bounded by

$$2(1 - u\ell\sqrt{3/16})^{n-2} \leq 2e^{-(n-2)u\ell\sqrt{3/16}} .$$

We conclude by the union bound, as there are not more than n^2 pairs. □

LEMMA 3. *Consider n points uniformly distributed on the unit square. Let M_1 be the maximal edge length of any Delaunay triangulation edge emanating from the first point, X_1. Then, for $0 < r < 1/2$,*

$$\sup_{x \in [r, 1-r]^2} \mathbb{P}\{M_1 \geq r | X_1 = x\} \leq 24e^{-3\pi(n-1)r^2/96} .$$

PROOF. Consider the circle C of radius $r\sqrt{3}/2$ centered at $x \in [r, 1-r]^2$. Partition C into 24 equal cones of angle $\pi/12$ each. Let A be the event that one of the cones does not receive one of the $n-1$ other data points. Clearly, as $C \subseteq [0,1]^2$,

$$\mathbb{P}\{A\} \leq 24(1 - 3\pi r^2/96)^{n-1} \leq 24e^{-\frac{3\pi(n-1)r^2}{96}}.$$

If $M_1 \geq r$, then by Lemma 1, there exists a halfcircle H on a diameter r with one end at x that has no data point. We claim that H must necessarily contain one of the 24 cones, and thus one of the 24 cones must be empty. Therefore, $\mathbb{P}\{M_1 \geq r | X_1 = x\} \leq \mathbb{P}\{A\}$, and we are done.

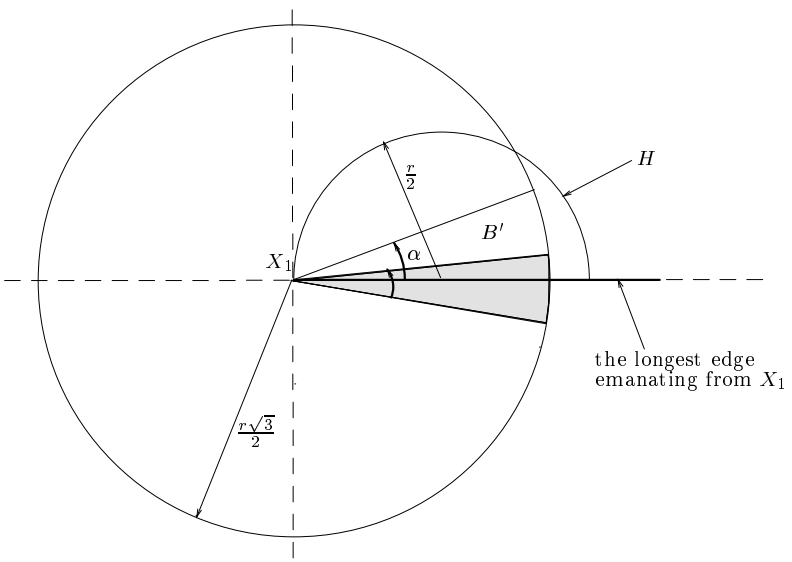


Figure 2.

Assume without loss of generality that x has coordinates $(0,0)$, and that H is supported by $((0,0),(r,0))$, and faces towards the positive y-axis. Let B be the cone containing $(r\sqrt{3}/2, 0)$. Let B' be the next cone in counterclockwise order. To show that $B' \subseteq H$, it suffices to show that its topmost vertex is in H. This vertex has coordinates $r(\sqrt{3}/2)(\cos\alpha, \sin\alpha)$, where $\pi/12 \leq \alpha \leq \pi/6$. The square of the distance from this vertex to the center of H, $(r/2, 0)$, is

$$\begin{aligned}
&r^2\left((3/4)\sin^2\alpha + (3/4)\cos^2\alpha + 1/4 - (\sqrt{3}/2)\cos\alpha\right) \\
&= r^2\left(1 - (\sqrt{3}/2)\cos\alpha\right) \\
&\leq r^2\left(1 - (\sqrt{3}/2)\cos(\pi/6)\right) \\
&= r^2(1 - 3/4) \\
&= (r/2)^2 .
\end{aligned}$$

This concludes the proof of Lemma 3. □

To see this Lemma at work, consider the following Lemma, which complements results of Bose and Devroye (1998):

LEMMA 4. *Let Δ_n denote a Delaunay triangulation for n points distributed uniformly in the unit square, and let L be a line segment independent of Δ_n, where L is completely contained in $B = [2r, 1-2r]^2$ and $r = \sqrt{25 \log n/\pi n}$. Let N denote the number of edges of Δ_n crossed by L. Then there exist constants c_i such that*

$$\mathbb{E}\{N\} \leq c_1 + c_2 \mathbb{E}\{\|X - Y\|\} \sqrt{n} \ .$$

PROOF. We condition on L (i.e., we hold L fixed) and show that

$$\mathbb{E}\{N|L\} \leq c_1 + c_2 \|L\| \sqrt{n}$$

for some universal constants c_i. Let M_i be the maximal length of any Delaunay edge emanating from X_i, and let L_i denote its distance to L. Let S be the collection of indices of data points that are such that a Delaunay edge from it reaches L. Clearly, any Delaunay edge (X_i, X_j) cut by L must have $i, j \in S$. If $G = [H_{n,r} \leq r]$ holds (where $H_{n,r}$ is as in Lemma 2), then L can only be reached by edges starting from nodes in $A = [r, 1-r]^2$. Thus, if G holds, then $S = S \cap T$, where T is the set of indices of X_i's in A. Let G^c be the complement of G. By the planarity of the Delaunay triangulation, there are at most $3n$ Delaunay edges. We thus have, for n large enough,

$$\begin{aligned}
\mathbb{E}\{N\} &\leq 3n\mathbb{P}\{G^c\} + 3\mathbb{E}\{|S|\mathbb{1}_G\} \\
&= 3n\mathbb{P}\{G^c\} + 3\mathbb{E}\{|S \cap T|\mathbb{1}_G\} \\
&\leq 3e^2 n^{-1/8} + 3n\mathbb{P}\{G, X_1 \in A, M_1 > L_1\} \\
&\quad \text{(by Lemma 2)} \\
&\leq o(1) + 3n\mathbb{P}\{X_1 \in A, r \geq M_1 > L_1\} \\
&= o(1) + 3n\mathbb{E}\left\{\mathbb{1}_{[X_1 \in A]}\mathbb{P}\{r \geq M_1 > L_1 | X_1\}\right\} \\
&\leq o(1) + 3n\mathbb{E}\left\{\mathbb{1}_{[X_1 \in A]} 24 e^{-3\pi(n-1)L_1^2/96}\right\} \\
&\quad \text{(by Lemma 3)} \\
&\leq o(1) + 72n\mathbb{E}\left\{e^{-nL_1^2/22}\right\} \\
&= o(1) + 72n \int_0^1 \mathbb{P}\left\{L_1 \leq \sqrt{22\log(1/t)/n}\right\} dt
\end{aligned}$$

$$\leq o(1) + 72n \int_0^1 \left(2\|L\| \sqrt{22 \log(1/t)/n} + 22\pi \log(1/t)/n \right) dt$$

(because we have a uniform distribution, see Figure 3)

$$= o(1) + c_1 + c_2 \|L\| \sqrt{n}$$

since both $\sqrt{\log(1/t)}$ and $\log(1/t)$ are integrable on $[0,1]$. Note that the $o(1)$ term and the constants c_i do not depend upon L. This concludes the proof of Lemma 4. □

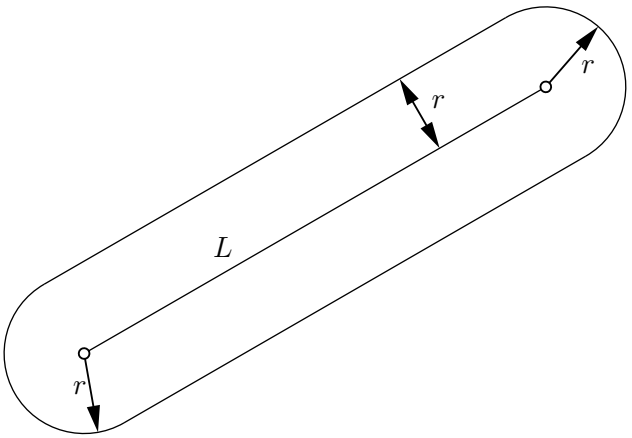


Figure 3. The area of the set of points within distance r of a line of length ℓ is $2\ell r + \pi r^2$.

Throughout this section, $X_1, \ldots, X_n$ are i.i.d. uniform $[0,1]^2$ random variables. For a set $A \subseteq [0,1]^2$, we let $\mu(A)$ denote its probability or area, and we denote by $\mu_n(A)$ its empirical probability:

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[X_i \in A]} \ .$$

For what follows, a few key results on $\mu_n(A)$ are needed. These will make the analysis of various quantities for the Delaunay triangulation particularly straightforward. We need the notion of the Vapnik-Chervonenkis dimension V of a class $\mathcal{A}$ of sets: V is the smallest positive integer such that there exists a set S of V points such that

$$|\{S \cap A : A \in \mathcal{A}\}| < 2^V \ .$$

If this is not possible for any finite set S, then $V = \infty$. For example, if $\mathcal{A}$ is the class of all Borel sets in a Euclidean space, then $V = \infty$. The class of all triangles in the plane has $V = 4$ as the points $(0,0)$, $(3,0)$, $(0,3)$ and $(1,1)$ cannot be broken up into 16 sets by intersecting with triangles: the subset $\{(0,0),(3,0),(0,3)\}$ for example cannot be obtained in this manner.

LEMMA 5. *Fix $p \in (0,1)$. Let $\mathcal{A}$ be a class of subsets of $[0,1]^2$ of equal area $\mu(A) = p$, and let V denote the Vapnik-Chervonenkis dimension of $\mathcal{A}$. Then*

$$\begin{aligned}
&\mathbb{P}\{\text{every set } A \text{ of } \mathcal{A} \text{ receives at least one point}\} \\
&= \mathbb{P}\left\{\inf_{A\in\mathcal{A}} \mu_n(A) > 0\right\} \\
&\geq 1 - 4(2en/V)^V e^{-np/4} \ , n \geq V/2.
\end{aligned}$$

PROOF. The proof is based on an inequality of Vapnik and Chervonenkis (1971) which states that $t > 0$,

$$\mathbb{P}\left\{\sup_{A\in\mathcal{A}} \frac{\mu(A) - \mu_n(A)}{\sqrt{\mu(A)}} \geq t\right\} \leq 4\mathcal{S}_{2n} e^{-nt^2/4} \ ,$$

where $\mathcal{S}_n$ is the maximal number of sets obtainable by intersecting a set of size n with sets from $\mathcal{A}$. Thus, $\mathcal{S}_n \leq 2^n$, but one can do better. In fact,

$$\mathcal{S}_n \leq (en/V)^V \ , n \geq V.$$

As $\mu(A) = p$ in our case, we get, upon setting $t = \sqrt{p}$,

$$\begin{aligned}
\mathbb{P}\left\{\inf_{A\in\mathcal{A}} \mu_n(A) \leq 0\right\} &= \mathbb{P}\left\{\sup_{A\in\mathcal{A}} p - \mu_n(A) \geq p\right\} \\
&= \mathbb{P}\left\{\sup_{A\in\mathcal{A}} \frac{p - \mu_n(A)}{\sqrt{p}} \geq \sqrt{p}\right\} \\
&\leq 4(2en/V)^V e^{-np/4} \ . \ \square
\end{aligned}$$

The previous Lemma is in fact valid for any distribution in any Euclidean space, not just the uniform distribution on $[0,1]^2$. Other generalizations of that inequality may be found in Panchenko (2002) and the references cited there. The next auxiliary results are particular for the uniform distribution though.

LEMMA 6. *Let $\mathcal{A}$ be the $k \times k$ square cells of a regular grid that partitions $[0,1]^2$. Let $X_1, \ldots, X_n$ be i.i.d. and uniformly distributed on the unit square. Then*

$$\mathbb{P}\left\{\min_{A\in\mathcal{A}} \mu_n(A) = 0\right\} \leq k^2 e^{-n/k^2} \leq 1/n^3$$

when $k \leq \sqrt{n/4\log n}$, $n \geq 2$.

PROOF. The first inequality follows from $\mathbb{P}\{\min_{A\in\mathcal{A}} \mu_n(A) = 0\} \le k^2(1-1/k^2)^n$. Replace k formally in this bound by $\sqrt{n/4\log n}$ and obtain the further upper bound $1/4n^3 \log n \le 1/n^3$ for $n \ge 2$. □

We define the *star set* $S(x,t)$ for $x = (x_1, x_2) \in [0,1]^2$, $1 > t > 0$ as the set of all $z = (z_1, z_2) \in [0,1]^2$ with the property that

$$|(z_1 - x_1)(z_2 - x_2)| \le t \ .$$

It is the set of points contained within four hyperbolae centered at x. Note that the probability (area) of any star set is not more than

$$4(t + t\log(1/t)).$$

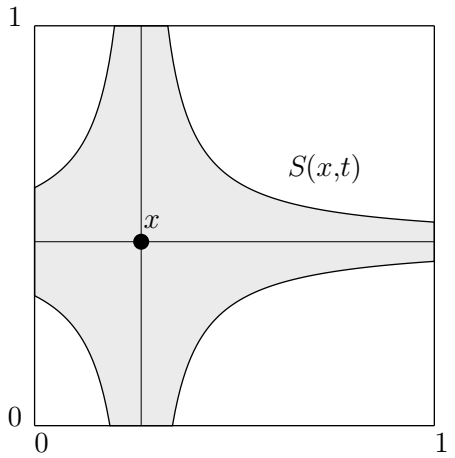


Figure 4. The area of $S(x,t)$ does not exceed $4\int_0^1 \min(1, t/x)\,dx$.

LEMMA 7. *Let $\mathcal{A}_t$, $1 > t > 0$ be the class of all star sets $S(x,t)$, $x \in [0,1]^2$. Then*

$$\mathbb{P}\left\{\sup_{A\in\mathcal{A}_t} \mu_n(A) \ge 2(4t + 5/n)(1 + \log(1/t))\right\} \le n^2(e/4)^{4nt\log(1/t)}.$$

PROOF. Partition $[0,1]$ into a regular $1/n \times 1/n$ grid, and let z denote the center of a grid cell. Let x be an arbitary point in that cell. We claim that

$$S(x,t) \subseteq S(z, t + 5/(4n)).$$

Let $y \in S(x,t)$. Then

$$\begin{aligned}
|z_1 - y_1||z_2 - y_2| &\leq (|y_1 - x_1| + 1/(2n))(|y_2 - x_2| + (1/2n)) \\
&\leq |y_1 - x_1||y_2 - x_2| + (1/2n)(|y_1 - x_1| + |y_2 - x_2|) + 1/(4n^2) \\
&\leq t + 1/n + 1/(4n^2) \\
&\leq t + 5/(4n).
\end{aligned}$$

Thus,

$$\sup_{A \in \mathcal{A}_t} \mu_n(A) \leq \sup_z \mu_n(S(z, t + 5/(4n))) \ .$$

Also, note that the maximal probability of any set $S(z, t + 5/(4n))$ is nor more than

$$4(t + 5/(4n) + (t + 5/(4n)) \log(1/(t + 5/(4n)))) \leq (4t + 5/n)(1 + \log(1/t)) \stackrel{\text{def}}{=} p.$$

If Z denotes a binomial (n,p) random variable, we thus have by the union bound,

$$\begin{aligned}
\mathbb{P}\left\{ \sup_z \mu_n(S(z, t + 5/(4n))) \geq 2p \right\} &\leq \sum_z \mathbb{P}\left\{ \mu_n(S(z, t + 5/(4n))) \geq 2p \right\} \\
&\leq \sum_z \mathbb{P}\left\{ Z - np \geq np \right\} \\
&\leq n^2 (e/4)^{np} \ ,
\end{aligned}$$

where we used an inequality due to Okamoto (1958): for $u > p$,

$$\mathbb{P}\{Z \geq un\} \leq \left(\frac{(1-p)^{1-u} p^u}{(1-u)^{1-u} u^u} \right)^n \ .$$

Set $u = 2p$, and use $(1 + p/(1-2p))^{1-2p} \leq e^p$ to obtain the desired result. Finally, note that $np > 4t \log(1/t)$. □

We now have our first rough bound on the maximal degree of any node in the Delaunay triangulation for $X_1, \ldots, X_n$, with constants explicitly shown.

LEMMA 8. *Let $n \geq 1000$. Let D_x be the degree of x in the Delaunay triangulation for $x, X_1, \ldots, X_n$. Then*

$$\mathbb{P}\left\{\sup_{x\in[0,1]^2} D_x \geq 384\log^2 n\right\} \leq \frac{13.66}{n^2}.$$

Furthermore, $\mathbb{E}\{\sup_x D_x\} \leq 384\log^2 n + 13.66/n$. Let D_i be the degree of X_i in the Delaunay triangulation for $X_1, \ldots, X_n$. Then

$$\mathbb{P}\left\{\sup_{1\leq i\leq n} D_i \geq 384\log^2 n\right\} \leq \frac{13.66}{n}.$$

Finally, $\mathbb{E}\{\sup_i D_i\} \leq 384\log^2 n + 13.66$.

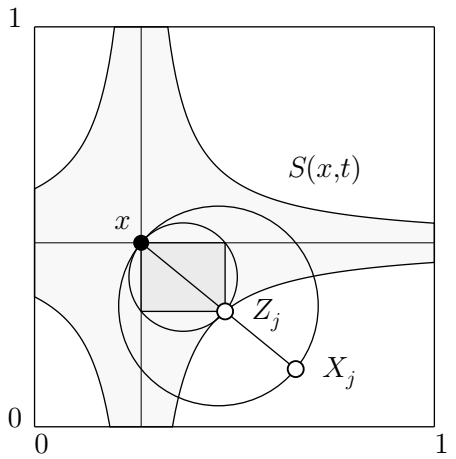


Figure 5. Illustration of the proof of Lemma 8. The shaded triangles each have area $t/2$ and each receive at least one data point when F is true. Therefore, the halfcircles supported by (x, X_j) are each occupied, and (x, X_j) cannnot be a Delaunay edge.

PROOF. Set $t = 48\log n/n$. Let $\mathcal{A}$ be the class of all triangles in $[0,1]^2$ with vertices of the form (a,b), (a,z) and (w,b), of area $t/2$. Let F be the event that each of these triangles receives a data point (from $X_1, \ldots, X_n$). Fix any j. For X_j, we either have that F is false, or that $X_j \in S(x,t)$ (with S as in Lemma 7), or that $F \cap [X_j \notin S(x,t)]$ is true. The last case cannot happen though, because when we consider the two triangles in the rectangle with x and X_j as opposite vertices and diagonal (x, X_j), then each of these triangles contains a further triangle contained in $\mathcal{A}$: indeed, let Z_j be the place on the

boundary of $S(x,t)$ that is crossed by the edge (x, X_j). Then define the triangles obtained by splitting the rectangle defined by x and Z_j about its diagonal (x, Z_j). By definition of $S(x,t)$, the area of the rectangle is t, and thus both triangles are of area $t/2$. But both of these triangles contain a data point when F is true, and both are contained in the halfcircle supported by (x, Z_j). Thus, both halfcircles supported by (x, X_j) contain a data point and thus, no circle with x and X_j on its perimeter can be empty. That implies that (x, X_j) cannot be a Delaunay edge. Thus,

$$D_x \le n\mu_n(S(x,t)) + n\mathbb{1}_{[F^c]} \ .$$

We can now apply Lemmas 5 and 7. Note that for $n \ge 1000$,

$$\begin{aligned} 2n(4t + 5/n)(1 + \log(1/t)) &= (8 \times 48 \log n + 10) \log(en/(48 \log n)) \\ &\le 384(\log n + 10/384)(\log n - 4.8) \\ &\le 384 \log^2 n. \end{aligned}$$

We have

$$\begin{aligned} &\mathbb{P}\left\{\sup_x D_x \ge 384 \log^2 n\right\} \\ &\le \mathbb{P}\left\{\sup_x D_x \ge 2n(4t + 5/n)(1 + \log(1/t))\right\} \\ &\le \mathbb{P}\{F^c\} + \mathbb{P}\left\{\sup_x \mu_n(S(x,t)) \ge 2(4t + 5/n)(1 + \log(1/t))\right\} \\ &\le 4(2en/4)^4 e^{-nt/8} + n^2(e/4)^{4nt \log(1/t)} \\ &\qquad \text{(since } V = 4 \text{ for triangles)} \\ &\le 4(2e/4)^4 n^{-2} + n^2(e/4)^{192 \log n \log(n/48 \log n)} \\ &\le 13.65 n^{-2} + n^2(e/4)^{192 \log n} \\ &\le 13.65 n^{-2} + n^{-72} \\ &\le 13.66 n^{-2}. \end{aligned}$$

This proves the first part. Let D_i be the degree of X_i in the Delaunay triangulation for $X_1, \ldots, X_n$. Then, by the first part,

$$\mathbb{P}\{D_i \ge 384 \log^2(n-1)\} \le \frac{13.66}{n^2} \ , \ n - 1 \ge 1000.$$

The last part then follows by the union bound. □

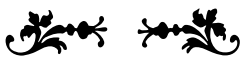
We also need a good estimate of the expected degree far from the border.

LEMMA 9. *Consider $x \in [t, 1-t]^2$ with $t = \sqrt{8 \log n / \pi n}$. If D_1 is the degree of X_1 in the triangulation for $X_1, \ldots, X_n$, then*

$$\mathbb{E}\{D_1 | X_1 = x\} \leq 18 + o(1).$$

PROOF. Using Lemma 1,

$$\begin{aligned}
&\mathbb{E}\{D_1 | X_1 = x\} \\
&\quad \leq n\mathbb{P}\{(X_2, X_1) \text{is a Delaunay edge, and } \|X_2 - X_1\| > t | X_1 = x\} \\
&\qquad + n\mathbb{P}\{(X_2, X_1) \text{is a Delaunay edge, and } \|X_2 - X_1\| \leq t | X_1 = x\} \\
&\quad \leq 2n\left(1 - \pi t^2/8\right)^{n-2} + 2n\mathbb{E}\left\{\left(1 - \pi\|X_1 - X_2\|^2/8\right)^{n-2} \mathbb{1}_{[\|X_2 - X_1\| \leq t]} | X_1 = x\right\} \\
&\qquad \text{(the first term argues by the emptiness of one halfcircle of diameter } t) \\
&\qquad \text{(the second term argues by the emptiness of one halfcircle supported on } (X_1, X_2)) \\
&\quad \leq 2ne^{-(n-2)\log n/n} + 2n \int_0^1 \mathbb{P}\left\{\left(1 - \pi\|X_1 - X_2\|^2/8\right)^{n-2} > u | X_1 = x\right\} du \\
&\quad \leq 2ne^{-(n-2)\log n/n} + 2n \int_0^1 \mathbb{P}\left\{e^{-(n-2)\pi\|X_1 - X_2\|^2/8} > u | X_1 = x\right\} du \\
&\quad = 2 + o(1) + 2n \int_0^1 \mathbb{P}\left\{e^{-(n-2)\pi\|X_1 - X_2\|^2/8} > u | X_1 = x\right\} du \\
&\quad \leq 2 + o(1) + 2n \int_0^1 \frac{8\log(1/u)}{(n-2)} du \\
&\quad \leq 2 + o(1) + (16 + o(1)) \int_0^1 \log(1/u)\, du \\
&\quad = 18 + o(1)\ . \square
\end{aligned}$$



Rectilinear search: the Green-Sibson method

Green and Sibson (1978) propose taking a random point, say X_1, and performing a walk to the random query point X. The complexity of Green-Sibson method is bounded by D_1, the degree of X_1 in the Delaunay triangulation Δ_n for $X_1, \ldots, X_n$ (to find the starting triangle in the Delaunay triangulation) plus N^*, the number of triangles visited by (X, X_1) in Δ_n.

THEOREM 1. *The expected complexity of the Green-Sibson method for a random Delaunay triangulation and an independent and uniformly distributed query point on* $[0,1]^2$ *is* $\Theta(\sqrt{n})$.

PROOF. We will only show the $O(\sqrt{n})$ upper bound. Let N be the number of triangles visited by (X, X_1) in the Delaunay triangulation for $X_2, \ldots, X_n$. Clearly, $|N - N^*| \leq D_1$. Thus,

$$|\mathbb{E}\{N\} - \mathbb{E}\{N^*\}| \leq 6 \ ,$$

because the expected degree of a randomly picked node in any Delaunay triangulation on n nodes is less than 6 (the sums of the degrees being less than $6n$). Therefore, the expected complexity is bounded by $6 + \mathbb{E}\{N^*\} \leq 12 + \mathbb{E}\{N\}$. We finish the proof by showing that $\mathbb{E}\{N\} = O(\sqrt{n})$.

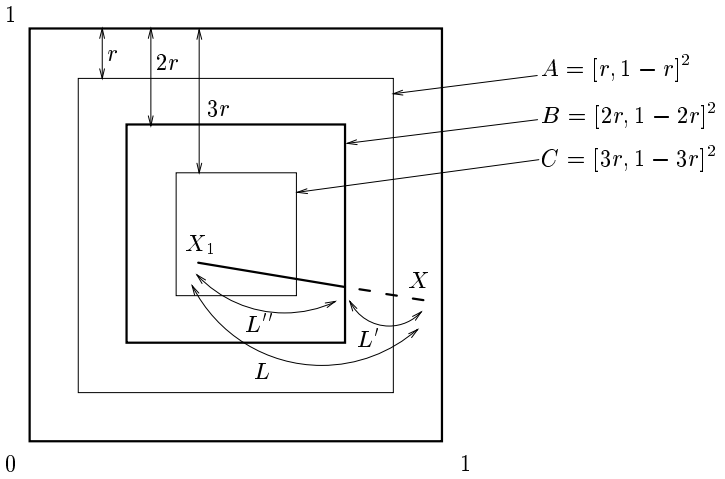


Figure 6.

To do so, define $r = \sqrt{25\log(n-1)/(\pi(n-1))}$, $A = [r, 1-r]^2$, $B = [2r, 1-2r]^2$ and $C = [3r, 1-3r]^2$. Split $L = (X, X_1)$ into two parts, $L'' = L \cap B$ and $L' = L \cap B^c$. Let N''

and N' denote the number of intersections of L'' and L' with the Delaunay triangulation for $X_2, \ldots, X_n$. By Lemma 4,

$$\mathbb{E}\{N''\} \leq c_1 + c_2 \mathbb{E}\{\|L''\|\}\sqrt{n} \leq c_1 + c_2\sqrt{2n} \ .$$

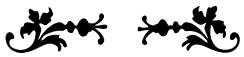
Define $G = [H_{n-1,r} \leq r]$. Let G^c denote its complement. If G holds, then L' can only be reached by edges starting in points of C^c. Thus, under G, N' is bounded from above by three times the cardinality of C^c (because it suffices to consider just the Delaunay triangulation for the points in C^c and to note that this triangulation is planar). Therefore, conditional on L, if $N(C^c)$ denotes the number of data points among $X_2, \ldots, X_n$ falling in C^c,

$$\begin{aligned}
\mathbb{E}\{N'|L\} &\leq \mathbb{E}\left\{3N(C^c)\mathbb{1}_G\right\} + \mathbb{E}\left\{3n\mathbb{1}_{G^c}\right\} \\
&\leq 3\mathbb{E}\left\{N(C^c)\right\} + O(n^{-1/8}) \\
&\leq 36rn + O(n^{-1/8}) \\
&= O(\sqrt{n\log n}) \ .
\end{aligned}$$

But then, as L' is nonempty only if X or X_1 fall outside B,

$$\mathbb{E}\{N'\} \leq \mathbb{P}\{[X \in B^c] \cup [X_1 \in B^c]\} O(\sqrt{n\log n}) \leq 16r \times O(\sqrt{n\log n}) = O(\log n) \ .$$

This concludes the proof of Theorem 1. □



Jump and Walk

Jump and Walk (Devroye, Mücke and Zhu, 1998) proceeds by picking Y as the nearest neighbor of X among $X_1, \ldots, X_k$, where $k = \lfloor n^{1/3} \rfloor$, and walking in the triangulation from Y to X. Devroye, Mücke and Zhu showed that if X is at least distance $\epsilon > 0$ away from the border of the unit square, then the expected time of this method is $\Theta(n^{1/3})$. The purpose of the Theorem below is to bring their results in line with the other ones in this paper, assuming that X is uniformly distributed on $[0,1]^2$. Note that the result is not true if X is allowed to depend upon the data or X is allowed to be deterministic and placed anywhere in the unit square.

A note on the data structure for the Delaunay triangulation is in order here. At each vertex, a list of incident triangles in clockwise order is maintained. One could organize this list as a balanced search tree, but that is in fact not necessary for the results of the previous, present and next sections. To start walking from Y to X in the Jump and Walk algorithm takes an initial cost proportional to the degree $d(Y)$ of Y for the list structure and $O(\log d(Y))$ for the search tree structure. As the maximal degree is $O(\log^2 n)$

in expected value (Lemma 8), this contribution is asymptotically negligible with either implementation.

The distance from X to Y is roughly $1/\sqrt{k} = \Theta(n^{-1/6})$. Then, the number of triangles cut by (X, Y) should be about $\sqrt{n}$ times that, or $n^{1/3}$, by Lemma 4, provided that (X, Y) does not approach the border of the unit square. When (X, Y) is near the border, a special additional argument is needed to show that the expected number of triangles met remains nevertheless $O(n^{1/3})$.

THEOREM 2. *Let Δ_n be a random Delaunay triangulation. If the query point is independent of the data and also uniformly distributed on the unit square, then the expected number of triangles visited by Jump and Walk and the expected complexity of the Jump and Walk algorithm are both $O(n^{1/3})$.*

PROOF. We will first show that $\mathbb{E}\{D\} = O(n^{1/3})$, where D is the degree of Y, and then that $\mathbb{E}\{N^*\} = O(n^{1/3})$, where N^* is the number of triangles visited by (X, Y). While in fact, $\mathbb{E}\{D\} \leq 42$ (see Chazelle et al. (2001)), we opt for a self-contained paper. Clearly, denoting by D_i the degree of X_i,

$$D \leq \max_{1 \leq i \leq k} D_i \leq \sum_{i=1}^{k} D_i$$

so that $\mathbb{E}\{D\} \leq 6k = O(n^{1/3})$. If we consider N, the number of triangles visited by (X, Y) in the triangulation formed by $X_{k+1}, \ldots, X_n$, then clearly,

$$N^* \leq N + \sum_{i=1}^{k} D_i \ ,$$

as the last term bounds the number of triangles added if we enlarge the triangulation to include $X_1, \ldots, X_k$ as well. Here we used the property that adding a point to a Delaunay triangulation can only remove existing edges, and add edges incident to the new point. Hence,

$$\mathbb{E}\{N^*\} \leq \mathbb{E}\{N\} + 6k = \mathbb{E}\{N\} + O(n^{1/3}) \ .$$

Observe that $L = (X, Y)$ and the triangulation considered in N are independent.

We first compute $\mathbb{E}\{\|L\|\}$. If $t < 1/2$, then at least 1/4-th of the circle of radius t centered at any $x \in [0, 1]^2$ is inside $[0, 1]^2$. Thus,

$$\mathbb{E}\{\|L\|\} = \int_0^\infty \mathbb{P}\{\|L\| > t\}\, dt$$

$$
\begin{aligned}
&\leq \int_0^{1/2} \left(1-\pi t^2/4\right)^k \, dt + \int_{1/2}^{\sqrt{2}} (1-\pi/16)^k \, dt \\
&\leq \int_0^{\infty} e^{-\pi t^2 k/4} \, dt + \sqrt{2}(1-\pi/16)^k \\
&= \frac{1+o(1)}{\sqrt{k}} \\
&\sim n^{-1/6} \; .
\end{aligned}
$$

We proceed as in the proof of Theorem 1. Define $r = \sqrt{25\log(n-k)/\pi(n-k)}$, $B = [2r, 1-2r]^2$ and $C = [3r, 1-3r]^2$. Split $L = (X, Y)$ into two parts, $L'' = L \cap B$ and $L' = L \cap B^c$. Let N'' and N' denote the number of intersections of L'' and L' with the Delaunay triangulation for $X_{k+1}, \ldots, X_n$. By Lemma 4,

$$
\mathbb{E}\{N''\} \leq c_1 + c_2 \mathbb{E}\{\|L''\|\}\sqrt{n} \leq c_1 + (c_2 + o(1))n^{1/3} \; .
$$

Define $G = [H_{n-k,r} \leq r]$. Let G^c denotes its complement. If G holds, then L' can only be reached by edges starting in points of C^c. Let $N(C^c)$ denotes the number of data points among $X_{k+1}, \ldots, X_n$ falling in C^c, Thus, under G, N' is bounded from above by $3N(C^c)$ because it suffices to consider just the Delaunay triangulation for the points in C^c and to note that this triangulation is planar. Therefore, conditional on L,

$$
\begin{aligned}
\mathbb{E}\{N'|L\} &\leq \mathbb{E}\left\{3N(C^c)1_G\right\} + \mathbb{E}\left\{3n1_{G^c}\right\} \\
&\leq 3\mathbb{E}\left\{N(C^c)\right\} + O(n^{-1/8}) \\
&\leq 36rn + O(n^{-1/8}) \\
&= O(\sqrt{n\log n}) \; .
\end{aligned}
$$

But then, as L' is nonempty only if X or Y fall outside B,

$$
\begin{aligned}
\mathbb{E}\{N'\} &\leq \mathbb{P}\{[X \in B^c] \cup [Y \in B^c]\}O(\sqrt{n\log n}) \\
&\leq 8r \times O(\sqrt{n\log n}) + \mathbb{P}\{Y \in B^c\}O(\sqrt{n\log n}) \\
&= O(\log n) + \mathbb{P}\{Y \in B^c\}O(\sqrt{n\log n}) \; .
\end{aligned}
$$

We conclude the proof by showing that $\mathbb{P}\{Y \in B^c\} = O\left(n^{-29/105}(\log n)^{1/4}\right)$, so that $\mathbb{E}\{N'\} \leq O\left(n^{47/210}(\log n)^{3/4}\right) = o(n^{1/3})$. For the proof of this fact, let $A = [d, 1-d]^2$, where d will be picked so that $d > 4r$ for n large enough. Let $0 < a = o(n)$ be picked further on. Consider the following:

$$
\begin{aligned}
\mathbb{P}\{Y \in B^c\} &\leq \mathbb{P}\{X \in A, Y \in B^c\} + \mathbb{P}\{X \in A^c, \|L\| < a\} + \mathbb{P}\{X \in A^c, \|L\| > 1/2\} \\
&\quad + \mathbb{P}\{X \in A^c, Y \in B^c, ||L|| \in [a, 1/2]\} \\
&\leq \mathbb{P}\{\|L\| \geq d - 2r\} + 4dk\pi a^2 + 4d\left(1 - \pi(1/2)^2/4\right)^k \\
&\quad + 4d \sup_{x \in [0,1]^2} \mathbb{P}\{Y \in B^c, ||L|| \in [a, 1/2] | X = x\}
\end{aligned}
$$

$$
\begin{aligned}
&\le \left(1-\pi(d-2r)^2/4\right)^k + 4dk\pi a^2 + 4d\exp\left(-\pi k(1/2)^2/4\right) \\
&\qquad + 4d \sup_{x\in[0,1]^2,\ell\in[a,1/2]} \mathbb{P}\{Y\in B^c | X=x, \|L\|=\ell\} \\
&\le \exp\left(-\pi k d^2/16\right) + 4dk\pi a^2 + 4d\exp\left(-\pi k(1/2)^2/4\right) \\
&\qquad + 4d \sup_{x\in[0,1]^2,\ell\in[a,1/2]} \mathbb{P}\{Y\in B^c | X=x, \|L\|=\ell\}
\end{aligned}
$$

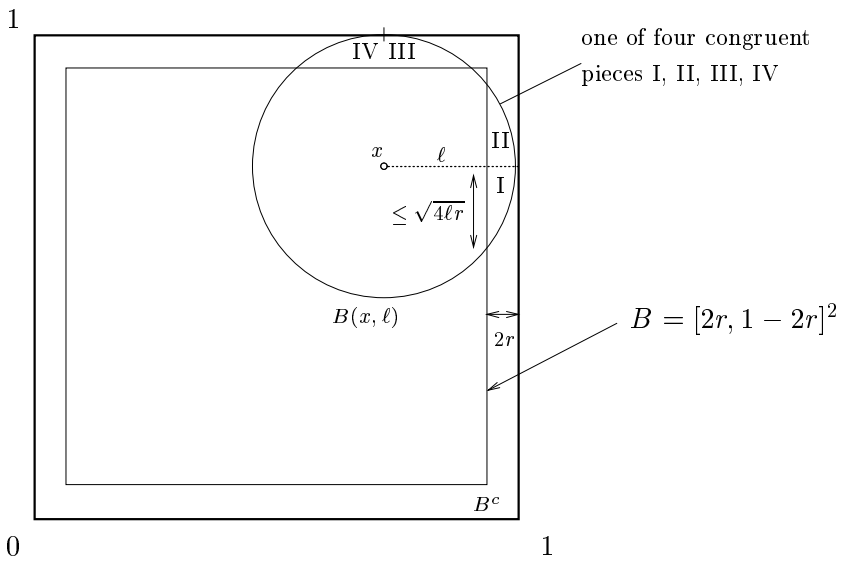


Figure 7.

We look at the last term and refer to Figure 7. Let $B(x,\ell)$ be the circle centered at x of radius ℓ. Given $||L|| = \ell$ and $X = x$, we note that Y is uniformly distributed on the perimeter of $B(x,\ell)$ constrained to the unit square. At least one quarter of this circle is inside the square, so the probability in the last bound is bounded by the length of the perimeter of $B(x,\ell)$ constrained to B^c divided by $\pi\ell/2$. The part of the perimeter in B^c is maximal when $x = (\ell,\ell)$. Decompose the perimeter into two parts, one touching the horizontal side of the unit square, and one touching the vertical side. Cut each part again exactly in half. Each of the four pieces has length not exceeding $2r$ (for the perpendicular distance to the side) plus $\sqrt{\ell^2-(\ell-2r)^2} \le \sqrt{4\ell r}$ (for the distance parallel to the side of the square), which for n large enough is not more than $(8+o(1))\sqrt{\ell r} < 9\sqrt{\ell r}$, provided that $r = o(\ell)$. Therefore, if $\ell \in [a, 1/2]$,

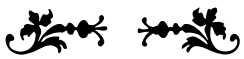
$$
\mathbb{P}\{Y\in B^c | X=x, \|L\|=\ell\} \le \frac{9\sqrt{\ell r}}{\pi\ell/2} = \frac{18}{\pi}\sqrt{\frac{r}{\ell}} \le 6\sqrt{\frac{r}{a}}\ .
$$

We conclude that

$$\mathbb{P}\{Y \in B^c\} \leq \exp\left(-\pi k d^2/16\right) + 4dk\pi a^2 + 4d\exp\left(-\pi k(1/2)^2/4\right) + 24d\sqrt{\frac{r}{a}}\ .$$

We now choose $d = 1/n^{1/7}$ and $a = 1/n^{7/30}$. Note that $r = o(a)$ and $r = o(d)$ as required. We have for n large enough,

$$\mathbb{P}\{Y \in B^c\} \leq \exp\left(-n^{1/21}/6\right) + 25n^{-11/420}\sqrt{r} = O\left(n^{-29/105}(\log n)^{1/4}\right)\ . \square$$



Point location using 1-d binary search trees

Given are $x_1, \ldots, x_n$, points distributed in $[0,1]^2$ all of whose coordinates are different. On top of a standard data structure for Delaunay triangulations, we maintain a simple binary search tree for the x-coordinates, as proposed by Lemaire (1997) in his thesis and subsequent work (Devroye, Lemaire and Moreau, 1999). The purpose of this section is to analyze how the binary search tree helps in Delaunay point location in an expected time sense. Lemaire's method, BinSearch and Walk, requires a parameter k which we shall set to $\lfloor n^{1/4} \rfloor$, but we shall describe intermediate complexities in terms of a general k.

In the algorithm, we first perform BinSearch: locate the unique leaf for the query point X in the tree, and find the order statistics whose ranks differ by the rank of the leaf by at most k. The number of these order statistics returned is thus between k and $2k$. Among those points, keep only the k nearest (what is said below will also be true if this selection step is omitted). In this subset of size k, find the point Y whose y-coordinate is closest to the y-coordinate of the query point X. As this involves finding the $O(k)$ neighbors in a binary search tree, the time cost is $O(k + h)$, where h is the height of the tree (see Cormen, Leiserson and Rivest, 1990). Recall that for most brands of binary search trees, $h = O(\log n)$ in the worst case. For the random binary search tree and the treap, the expected height is $O(\log n)$ (Robson, 1979).

The bounded balance search trees of Nievergelt and Reingold (1973) showed fine experimental results on large data sets in the study of Devroye, Lemaire and Moreau (1999). First, we locate Y in the tree in logarithmic time. Then we determine the location of Y in the Delaunay triangulation in constant time. Finally, we perform Walk: we walk across the segment (Y, X) in the Delaunay triangulation to determine the triangle for X. The extra time needed for this walk is $O(1)$ plus the number of Delaunay edges

crossed by the walk, plus the degree of Y in the Delaunay triangulation (to start off, we must perform binary search among the incident triangles). The latter contribution has expected value $O(\log^2 n)$ by Lemma 8. Thus, the expected time is $O(n^{1/4})$ plus the expected time number of triangles visited by (Y, X).

THEOREM 3. *Consider a random Delaunay triangulation. The expected number of Delaunay edges cut by (X, Y) (found by the procedure above after taking $k = \lfloor n^{1/4} \rfloor$) does not exceed $O(n^{1/4})$. If one organizes the point location by using a binary search tree with expected height $O(n^{1/4})$, then the expected time for point location is $O(n^{1/4})$.*

PROOF. We will make repeated use of the following fact. If (x, y) is a line segment contained in a convex set $C \subseteq [0,1]^2$, then the number of Delaunay edges cut by (x, y) for a Delaunay triangulation Δ_n (which may or may not include x and/or y as vertices) is bounded by $3(N + N')$, where N is the number of points of the Delaunay vertex set in C, and N' is the number of Delaunay edges crossing C. To see this, mark all points in C and all points outside C that have a Delaunay edge that crosses into C. The number of these points does not exceed $N + N'$. Form the Delaunay triangulation for these points. By planarity, the number of edges of this triangulation crossing (x, y) (or any other line segment contained in C) is not more than $3(N + N')$.

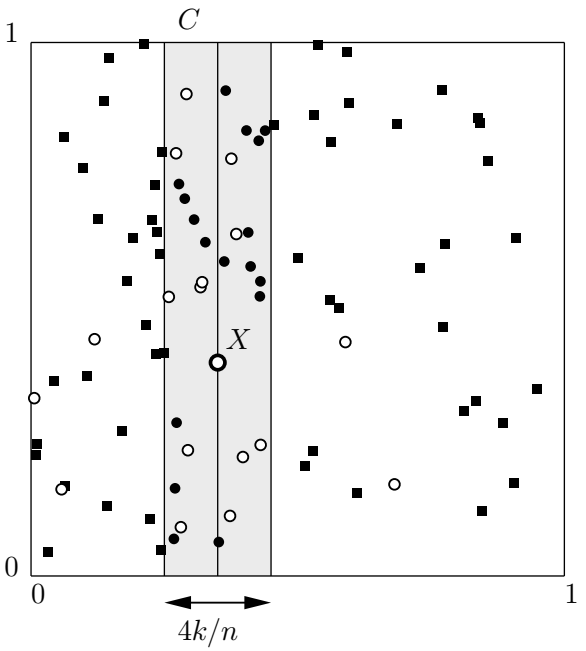


Figure 8. Illustration of the construction of two coupled samples. The square points (outside C) are common to both samples. The white and black round points are particular to one sample. On the strip C, the samples differ, and off C, they are largely identical. However, both are i.i.d. uniform samples when considered separately.

Let $X = (X_x, X_y)$ be the query point. Let $C = [X_x - 2k/n, X_x + 2k/n] \times [0,1]$. Let $Y = X_i = (Y_x, Y_y)$ be the point among $X_1, \ldots, X_n$ whose x-coordinate rank differs from that of X_x by at most k and whose y-coordinate value is nearest to X_y. Let $\mathcal{D}_n$

denote the data $(X_1, \ldots, X_n)$. We need a closely related sample $\mathcal{D}'_n$ (or $X'_1, \ldots, X'_n$). The construction of both samples is now described (see Figure 8). Let p be the area of C, a random variable between 0 and $4k/n$. First generate two independent binomial (n, p) random variables, N and N', denoting the number of data points in C in each sample (before they are generated!). Let $Z_1, \ldots, Z_n$ be an i.i.d. sample drawn uniformly from $C^c \cap [0,1]^2$, and let $W_1, \ldots, W_n, W'_1, \ldots, W'_n$ be an i.i.d. sample drawn uniformly from C. We start by setting $X_i = Z_i$, $i \leq n - N$, $X'_i = Z_i$, $i \leq n - N'$ (so that $X_i = X'_i$ for $i \leq n - \max(N, N')$). Then, set $X_{n-N+i} = W_i$, $i \leq N$, and $X'_{n-N'+i} = W'_i$, $i \leq N'$. This completes the samples. Apply the same random permutation to both samples. We claim that both samples are uniformly distributed on $[0,1]^2$. Also, for any i (after the random permutation), $\mathbb{P}\{X_i \neq X'_i\} = \mathbb{E}\{\max(N, N')/n\}$. This is less than $\mathbb{E}\{(N+N')/n\} \leq 8k/n$. We also have $\mathbb{P}\{X_i \neq X'_i | X\} \leq 8k/n$. If $N(X, Y, \mathcal{D}_n)$ denotes the number of edges of the Delaunay triangulation for $\mathcal{D}_n$ that meet (X, Y), and $N(X, Y, \mathcal{D}'_n)$ denotes the quantity for $\mathcal{D}'_n$, then

$$N(X, Y, \mathcal{D}_n) \leq N(X, Y, \mathcal{D}'_n) + \sum_{i=1}^{n} D_i \mathbb{1}_{\left[X_i \neq X'_i\right]} + \sum_{i=1}^{n} D'_i \mathbb{1}_{\left[X'_i \neq X_i\right]}$$

where D_i is the degree of X_i in the Delaunay triangulation $\mathcal{D}_n$, and D'_i is the degree of X'_i in the Delaunay triangulation $\mathcal{D}'_n$. To see this, note that if we remove the points X_i from $\mathcal{D}_n$ for which $X_i \neq X'_i$, and adjust the Delaunay triangulation after every point removal, then the number of edges changed when removing X_i is equal to D_i, the degree of X_i just before its removal. This may increase the degrees of some other points, but the edges causing such degree increases are already counted. Arguing in this way, the total number of edges changed is not more than $\sum_{i=1}^{n} D_i \mathbb{1}_{\left[X_i \neq X'_i\right]}$. Now, start adding points X'_i (with $X_i \neq X'_i$) one by one to get $\mathcal{D}'_n$ and note that these additions cause at most $\sum_{i=1}^{n} D'_i \mathbb{1}_{\left[X'_i \neq X_i\right]}$ new edge creations. Taking expected values, we note that

$$\begin{aligned}
\mathbb{E}\left\{N(X, Y, \mathcal{D}_n)\right\} &\leq \mathbb{E}\left\{N(X, Y, \mathcal{D}'_n)\right\} + 2\mathbb{E}\left\{\sum_{i=1}^{n} D_i \mathbb{1}_{\left[X_i \neq X'_i\right]}\right\} \\
&= \mathbb{E}\left\{N(X, Y, \mathcal{D}'_n)\right\} + 2n\mathbb{E}\left\{D_1 \mathbb{1}_{[X_1 \neq X'_1]}\right\} \ .
\end{aligned}$$

We consider both terms in the bound separately. Turning to the last term, we condition on the position of X_1: $X_1 = x$. Consider $x \in [t, 1-t]^2$ with $t = \sqrt{8 \log n / \pi n}$, and note that by Lemma 9, $\mathbb{E}\{D_1 | X_1 = x\} \leq 18 + o(1)$. Using this and Lemma 5, we have

$$\begin{aligned}
&2n\mathbb{E}\left\{D_1 \mathbb{1}_{[X_1 \neq X'_1]}\right\} \\
&\quad = 2n\mathbb{E}\left\{D_1 \mathbb{1}_{[X_1 \neq X'_1]} \mathbb{1}_{[X_1 \in [t,1-t]^2]}\right\} + 2n\mathbb{E}\left\{D_1 \mathbb{1}_{[X_1 \neq X'_1]} \mathbb{1}_{[X_1 \notin [t,1-t]^2]}\right\} \\
&\quad \leq 2n\mathbb{E}\left\{\mathbb{1}_{[X_1 \in [t,1-t]^2]} \mathbb{E}\left\{D_1 \mathbb{1}_{[X_1 \neq X'_1]} | X_1\right\}\right\} + 2n\mathbb{E}\left\{\mathbb{1}_{[X_1 \notin [t,1-t]^2]} \mathbb{1}_{[X_1 \neq X'_1]} \sup_x D_x\right\}
\end{aligned}$$

$$
\begin{aligned}
&\text{(where } D_x \text{ is the Delaunay degree of } x \text{ for } x, X_2, \ldots, X_n) \\
&\leq 16k\mathbb{E}\left\{\mathbb{1}_{[X_1\in[t,1-t]^2]}\mathbb{E}\left\{D_1|X_1\right\}\right\} + 2n^2\mathbb{P}\{\sup_x D_x \geq 384\log^2 n\} \\
&\qquad\qquad +2n\times 384\log^2 n\mathbb{P}\left\{X_1\notin[t,1-t]^2, X_1\neq X_1'\right\} \\
&\leq 16k\mathbb{E}\left\{\mathbb{1}_{[X_1\in[t,1-t]^2]}(18+o(1))\right\} + 27.32 + 16k\times 384\log^2 n\,\mathbb{P}\left\{X_1\notin[t,1-t]^2\right\} \\
&\qquad \text{(by Lemma 8, for } n\geq 1000) \\
&\leq 16k(18+o(1)) + 27.32 + 16k\times 4t\times(384\log^2 n) \\
&\sim 288k\ .
\end{aligned}
$$

This leaves us with $\mathbb{E}\left\{N(X,Y,\mathcal{D}_n')\right\}$. To deal with this, we introduce the event $G=[N\geq k]$ (which insures that the point Y is in C). Given G, (X,Y) and $\mathcal{D}_n'$ are independent, and $\mathcal{D}_n'$ is a uniform sample on the unit square. Introduce the event H that the vertical distance between X and Y is less than or equal to $2\log n/k$, and the event $E=[X,Y\in[2r,1-2r]^2]$, where $r=\sqrt{25\log n/\pi n}$. By Lemma 4, we have

$$
\begin{aligned}
\mathbb{E}\left\{N(X,Y,\mathcal{D}_n')\mathbb{1}_{[E]}\mathbb{1}_{[G]}\right\} &\leq c_1 + c_2\mathbb{E}\{\|X-Y\|\mathbb{1}_{[G]}\}\sqrt{n} \\
&\leq c_1 + c_2\mathbb{E}\{(4k/n+1/(k+1))\mathbb{1}_{[G]}\}\sqrt{n} \\
&\leq c_1 + c_2(4k/n+1/k)\sqrt{n} \\
&= O(n^{1/4}),
\end{aligned}
$$

where we used the fact that the expected vertical distance between X and Y is at most $1/(k+1)$. Next,

$$
\begin{aligned}
\mathbb{E}\left\{N(X,Y,\mathcal{D}_n')\mathbb{1}_{[G^c]}\right\} &\leq 3n\mathbb{P}\{G^c\} \\
&\leq 3n\mathbb{P}\{\text{binomial}(n,4k/n)<k\} \\
&\leq 3ne^{-cn^{1/4}}
\end{aligned}
$$

by an exponential tail inequality for binomials (see, e.g., Angluin and Valiant, 1979). Furthermore,

$$
\begin{aligned}
\mathbb{E}\left\{N(X,Y,\mathcal{D}_n')\mathbb{1}_{[H^c\cap G]}\right\} &\leq 3n\mathbb{P}\{H^c\cap G\} \\
&\leq 3n\left(1-2\log n/k\right)^k \\
&\leq 3ne^{-2\log n} \\
&= \frac{3}{n}\ .
\end{aligned}
$$

Finally, note that if $E^c\cap G\cap H$ hold, then X must have a y-coordinate that falls within $2r+2\log n/k$ of either 0 or 1, an event that has probability bounded by $4r+4\log n/k < 5\log n/n^{1/4}$ for n large enough. Let R be the rectangle centered at X of dimensions $4k/n$ times $4\log n/k$, intersected with the unit square. Under $E^c\cap G\cap H$, the segment (X,Y) is completely contained in R. Let N^* be the number of X_i's in $\mathcal{D}_n'$ that have a Delaunay edge

emanating from it that has a nonempty intersection with R. By conditional independence (given G) of

$$\mathbb{E}\left\{N(X,Y,\mathcal{D}'_n)\mathbb{1}_{[E^c\cap H\cap G]}\right\} \leq \frac{5\log n}{n^{1/4}}\mathbb{E}\{3N^*\} \ .$$

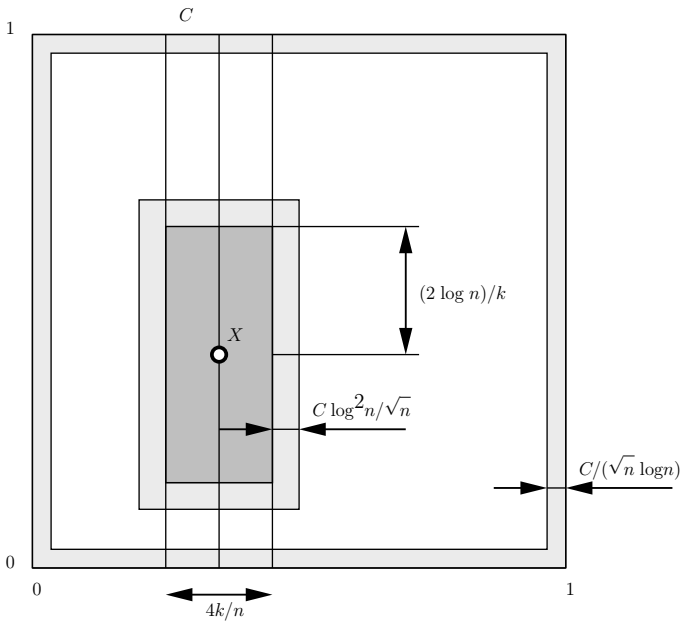


Figure 9. iIllustration of the construction of R and T. All shaded areas together form T. The darkest shading is reserved for the recatngle R centered at X.

We are done if we can show that $\mathbb{E}\{N^*\} = O(\sqrt{n}/\log n)$. Let T be the set consisting of all points of the unit square that are within distance $C/\sqrt{n}\log n$ from the perimeter of the unit square, or at distance less than $C\log^2 n/\sqrt{n}$ from R (thus, T includes R) (see Figure 9). The area of T is not more than $4C/\sqrt{n}\log n + 4C^2\log^4 n/n + (8k/n + 8\log n/k)C\log^2 n/\sqrt{n}$, which for n large enough is not more than $5C/\sqrt{n}\log n$. For such large n, the expected number of points of $\mathcal{D}'_n$ in T is not more than $5C\sqrt{n}/\log n$. Let L be the length of the largest Delaunay edge emanating from a point not in T. If $L < C\log^2 n/\sqrt{n}$, then any edge visiting R must start from a data point in T. So,

$$\mathbb{E}\{N^*\} \leq \frac{5C\sqrt{n}}{\log n} + n \times \mathbb{P}\{L \geq C\log^2 n/\sqrt{n}\} \ .$$

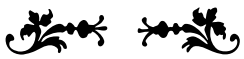
By the last part of Lemma 2 and the definition of T, we bound the last term in the upper bound by

$$n \times n^2 e^{-\frac{(n-2)C^2\log n\sqrt{3/16}}{n}}$$

for all n large enough. The last term is $O(1)$ if we set $C = 48^{1/4}$. Putting everything together, we have shown that

$$\mathbb{E}\left\{N(X, Y, \mathcal{D}'_n)\right\} = O(n^{1/4}) \ . \ \square$$

Using a proof method as in the present paper, it is possible to show that BinSearch and Walk in 3d with parameter $k \approx n^{2/9}$ takes expected time $\Theta(n^{2/9})$ for data points and a query point i.i.d. and uniformly distributed in the unit cube. The analysis could be undertaken by combining and adapting the arguments from the present paper and from Devillers, Pion and Teillaud (2001).



Expected complexity of Delaunay point location with balanced 2d trees

The unit square is recursively partitioned by a perfectly balanced 2-d tree (Bentley, 1975), that is, the partitioning alternates directions between x and y, and member sets in the partition are rectangles. Every node in the 2-d tree receives one data point, about which the remaining points are split. Leaves correspond to empty rectangles. It is clear, therefore, that there are n non-leaf nodes, and $n+1$ leaves. For a rectangle that properly contains a collection of data points, if a vertical split is made, it is made at the x-median of these points, where the x-median is uniquely defined if the number of points is odd, and is the leftmost of the two candidate medians when n is even. Horizontal splits are made about y-medians. A rectangle with one point is thus split about this point, and this results in two empty leaf rectangles.

When splits alternate between horizontal and vertical, we end up with a rather balanced 2-d tree. This structure is used in a static manner then for point location in a Delaunay triangulation for $x_1, \ldots, x_n$. We assume that the triangles incident to each vertex are stored by using balanced search trees. Let x be a given query point. First, we determine in $O(\log_2 n)$ worst-case time the leaf region for x. Then we let y be the node at the parent of this leaf. This node's position in the Delaunay triangulation is of course known (in $O(1)$ time). Finally, we walk across the segment (y, x) in the Delaunay triangulation to determine the triangle for x. The extra time needed for this walk is $O(1)$, plus $O(\log n)$ (for the binary search at y among the triangles incident to y), plus N, the number of Delaunay edges crossed by the walk. Note that Lemma 8 implies that without the search tree organization of the triangles, we would get an expected complexity $O(\log^2 n)$. However, that is not good enough here. In fact, the expected maximal degree is $O(\log n)$, but that takes a much more careful argument not presented in this paper. It

would allow us to conclude that the next theorem is true even without the search trees at each vertex.

It suffices thus to study N. To do so, we assume that the data $X_1, \ldots, X_n$ are drawn uniformly at random from the unit square, and that X, the query point, is independent of the data and also drawn uniformly from the unit square. The point at which the walk is started is denoted by Y. We will prove that the expected complexity of a random query is $O(\log n)$. The walk from Y to X crosses on the average $O(1)$ Delaunay edges, so that the expected complexity of the walk from Y to X takes $O(1)$ expected time. However, this result is rather delicate to prove, and it is much easier to show that the walk takes $O(\log n)$ expected time. It is this proof we will present here. Note also that we do not claim that the expected time from Y to a given $X = x$ is $O(\log n)$ on the average. In fact, this is not true, due to boundary effects in Delaunay triangulations.

THEOREM 4. *The expected value of the number of Delaunay edges crossed by (X, Y), and the expected complexity of point location by the method described above, when $X_1, \ldots, X_n$ and X are i.i.d. and uniformly distributed in the unit square is $O(\log n)$.*

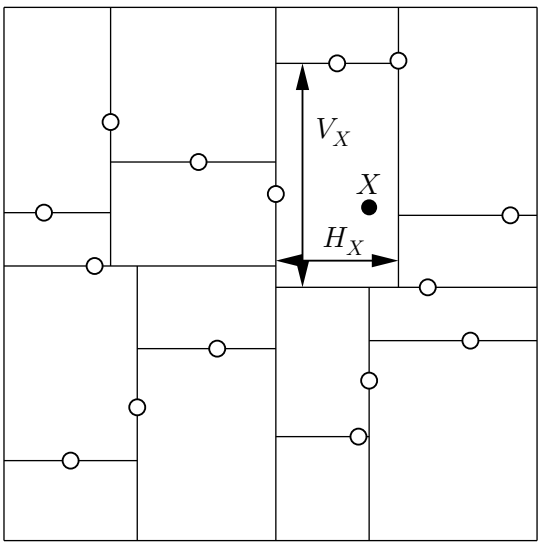


Figure 10. The partition of space by a balanced 2d tree. The dimensions of the unique (leaf) rectangle to which X belongs are called H_X and V_X.

PROOF. Set $C = 16$, $D = 17$. Let X and Y be as in the algorithm: X is the query point, and Y is the data point at which the walk is started. It is also the parent node of the leaf region for X. Let V_X and H_X be the vertical and horizontal dimensions of the leaf rectangle to which X belongs. We will need the following lemma about V_x, proved in the appendix. As this rectangle has no other data points in its interior, the only Delaunay edges that can possibly be crossed by the segment (X, Y) are those that emanate from points outside the leaf rectangle. Let N_i denote the number of points from among $X_1, \ldots, X_n$ that fall in the i-th square. There are three situations:

A. $\min_i N_i = 0$: in this case, we can cross at most $3n - 6$ edges.

B. $\min_i N_i > 0$, X is further than $2V_X + 2H_X + 2D\sqrt{\log n/n}$ from the rim: in this case, Y is at least $2D\sqrt{\log n/n}$ from the rim as well (since Y is on the perimeter of the leaf rectangle), so that any Delaunay edge that crosses (X, Y) can be of length at most $r \stackrel{\text{def}}{=} C\sqrt{\log n/n}$ by an argument as in Lemma 10 (see Appendix). Hence, the Delaunay edges are always between points that are within distance $C\sqrt{\log n/n}$ of the leaf rectangle of X. Consider the rectangle of dimensions $(2r+2H_X)\times(2r+2V_X)$ centered at X, and let N_X be the number of data points in that rectangle, which incidentally is entirely contained within the original square. Then (X, Y) cuts at most $3N_X$ Delaunay edges.

C. $\min_i N_i > 0$, X is closer than $2V_X + 2H_X + 2D\sqrt{\log n/n}$ from the rim. Define R_X as the set of all data points that are within distance $3V_X + 3H_X + 3D\sqrt{\log n/n}$ from the rim. Clearly, any Delaunay edge crossing (X, Y) must have both endpoints in R_X by Lemma 10. Let M_X be the number of data points in R_X. Then (X, Y) cuts at most $3M_X$ Delaunay edges.

Recapping the above, we see that the expected number of Delaunay edges cut does not exceed

$$3n\mathbb{P}\{\min_i N_i = 0\} + 3\mathbb{E}\{N_X\} + 3\mathbb{E}\{M_X \mathbb{1}_A\}$$

where A is the event that X is closer than $2V_X + 2H_X + 2D\sqrt{\log n/n}$ from the rim. By Lemma 6, the first term is $o(1)$. For $x \in \mathbb{R}^2$, $a > 0$, let $[x-a, x+a]$ denote the square with opposite vertices $(x - (a, a), x + (a, a))$. Let $N(x, j)$ denote the number of data points in

$$\left[x - 2r - 2\sqrt{\frac{(j+1)\log n}{n}}, x + 2r + 2\sqrt{\frac{(j+1)\log n}{n}}\right] .$$

Set

$$\gamma_{j,n} = \sup_x \sqrt{\mathbb{P}\left\{V_x + H_x > \sqrt{\frac{j \log n}{n}}\right\}} .$$

Observe that

$$
\begin{aligned}
3\sup_x \mathbb{E}\{N_x\} &\le 3\sup_x \left\{ \sum_{j=0}^{\infty} \mathbb{E}\left\{ \mathbb{1}_{\left[V_x+H_x>\sqrt{\frac{j\log n}{n}}\right]} N(x,j) \right\} \right\} \\
&\le 3\sum_{j=0}^{\infty} \gamma_{j,n} \mathbb{E}^{\frac{1}{2}}\left\{ N(x,j)^2 \right\} \\
&\quad \text{(by the Cauchy-Schwarz inequality)} \\
&\le 3\sum_{j=0}^{\infty} \gamma_{j,n} \mathbb{E}\left\{ 1 + N(x,j) \right\} \\
&\quad \text{(because for } B \text{ binomial } (n,p),\ \mathbb{E}\{B^2\} \le (1+\mathbb{E}\{B\})^2 \text{)} \\
&\le 4\sum_{j=0}^{\infty} \gamma_{j,n} \mathbb{E}\left\{ N(x,j) \right\} \\
&\quad \text{(as } \mathbb{E}\{N(x,j)\} \ge r^2 n \ge 3 \text{ for all } n \text{ large enough)} \\
&\le 64n \sum_{j=0}^{\infty} \gamma_{j,n} \left(r + \sqrt{\frac{(j+1)\log n}{n}} \right)^2 \\
&= 64 \log n \sum_{j=0}^{\infty} \gamma_{j,n} \left(C + \sqrt{j+1} \right)^2 \\
&\le 128 \log n \sum_{j=0}^{\infty} \gamma_{j,n} \left(C^2 + j + 1 \right) \ .
\end{aligned}
$$

Before we work this out, we turn to the third term, and argue similarly. Set $r = D\sqrt{\frac{\log n}{n}}$. Now, let $N(j)$ denote the number of data points within distance $3r + 3\sqrt{\frac{(j+1)\log n}{n}}$ from the perimeter of the unit square. Let R be the set of all points in the unit square within distance $2r + 2\sqrt{\frac{(j+1)\log n}{n}}$ from the perimeter. Then,

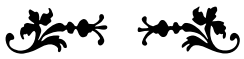
$$
\begin{aligned}
&3\mathbb{E}\{M_X \mathbb{1}_A\} \\
&\le 3\left\{ \sum_{j=0}^{\infty} \mathbb{E}\left\{ N(j) \mathbb{1}_{\left[V_X+H_X>\sqrt{\frac{j\log n}{n}}\right]} \mathbb{1}_{[X\in R]} \right\} \right\} \\
&\le 3\sum_{j=0}^{\infty} \mathbb{P}\left\{ X \in R \right\} \sup_x \mathbb{E}\left\{ N(j) \mathbb{1}_{\left[V_x+H_x>\sqrt{\frac{j\log n}{n}}\right]} \right\} \\
&\le 3\sum_{j=0}^{\infty} 4\left(2r + 2\sqrt{\frac{(j+1)\log n}{n}} \right) \gamma_{j,n} \left(1 + 4n\left(3r + 3\sqrt{\frac{(j+1)\log n}{n}} \right) \right) \\
&\quad \text{(by the Cauchy-Schwarz inequality)}
\end{aligned}
$$

$$
\begin{aligned}
&= 24\sum_{j=0}^{\infty}\left(r+\sqrt{\frac{(j+1)\log n}{n}}+12n\left(r+\sqrt{\frac{(j+1)\log n}{n}}\right)^2\right)\gamma_{j,n}\\
&\le 24\sum_{j=0}^{\infty}\left(r+\sqrt{\frac{(j+1)\log n}{n}}+24nr^2+24(j+1)\log n\right)\gamma_{j,n}\\
&\le 24\sum_{j=0}^{\infty}\left(25D^2\log n+25(j+1)\log n\right)\gamma_{j,n}\quad(\text{as } n\ge 4,\ D\ge 1)\\
&\le 600\log n\sum_{j=0}^{\infty}\left(D^2+j+1\right)\gamma_{j,n}\ .
\end{aligned}
$$

Thus, the three terms taken together are $O(\log n)$ provided that $\sum_{j=0}^{\infty} j\gamma_{j,n} = O(1)$. But by Lemma 12 (see Appendix),

$$
\sum_{j=0}^{\infty}\sup_x j\sqrt{\mathbb{P}\left\{V_x > \sqrt{\frac{j\log n}{4n}}\right\}} = O(1)\ ,
$$

and this implies the result . □



Nearest neighbor data structures

Assume that Y, the place at which the walk is started, is the nearest neighbor of X among $X_1,\ldots,X_n$. We prove the following.

THEOREM 5. *If $X, X_1,\ldots,X_n$ are independent and identically (but not necessarily uniformly) distributed with any density f on $\mathbb{R}^2$, then the expected time needed to walk from Y to X in the Delaunay data structure for $X_1,\ldots,X_n$, where Y is the nearest neighbor of X among $X_1,\ldots,X_n$, is $O(1)$.*

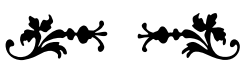
PROOF. If the Delaunay structure would be updated to include X, then (X,Y) would be a Delaunay edge in this new structure. To update that data structure takes time proportional to the degree of X after the update. Now, the number of old Delaunay edges crossed by (Y,X) is not more than this degree, because each crossed edge is necessarily replaced and thus contributes "one" to the degree of X. But by planarity, the number of edges in the new Delaunay triangulation is not more than $3(n+1)-6 < 3n$, and thus

the expected degree of X is not more than $6n/(n+1) < 6$, because the $n+1$ data points are i.i.d. $\square$

In other words, any nearest neighbor data structure would be useful for use in Delaunay point location. To be considered in the context of this paper, it has to be simple and efficient.

A particularly simple data structure is the grid. Assume that the data take values in the unit square, and are hashed to a $\sqrt{n} \times \sqrt{n}$ grid embedded in $[0,1]^2$. Then, assuming that the data are uniformly distributed, the nearest neighbor Y of X can be found in $O(1)$ expected time by spiral search—first search in the cell of X, then in cells at grid distance one, and so forth (Bentley et al, 1980).

We may simplify this method, and stop as soon as a cell point is found (without actually checking that it is the nearest neighbor). Then the distance from that point Y to X is at most $\sqrt{2}\left(\frac{2}{\sqrt{n}} + \Delta\right)$, where Δ is the distance from X to its nearest neighbor. It turns out that this is good enough, and that the expected time needed to go from Y to X is still $O(1)$, and that the expected time to find Y is $O(1)$ as well, assuming hashing in $O(1)$ time. We will not give that simple analysis here.



Conclusion

In a random 2-d tree for points uniformly distributed in the unit square, a nearest neighbor algorithm was proposed by Chanzy, Devroye and Zamora-Cura (2001), whose expected time (under the model studied in this paper) was shown to be $O(n^\rho)$, where $\rho = (\sqrt{17} - 4)/2 = 0.056\ldots$. By Theorem 5, Delaunay point location can be done in $O(n^\rho)$ expected time as well, based on a preliminary nearest neighbor search.

If however we locate a leaf rectangle for X (as we did for balanced 2-d trees), let Y be the parent for the leaf rectangle, and perform a Delaunay triangulation walk from Y to X, then the analysis is slightly trickier. The length of the segment (X, Y) is bounded by $V_X + H_X$. Chanzy, Devroye and Zamora-Cura (2001) showed that $\mathbb{E}\{V_X\} = n^{1/2+\rho}$, so that we obtain an expected complexity for point location of $O(n^\rho)$, once again. The details of the proof are omitted.

Other possible structures of varying complexity include the random quadtree (which behaves as the random 2-d tree and yields similar expected times), the squarish 2-d tree of Devroye, Jabbour and Zamora-Cura (2000) (for which the expected complexity is unknown but probably $O(\log^2 n)$), and the random 2-d trie (or region quadtree).

The grid structure mentioned in the previous section would work fine for the uniform distribution on $[0, 1]^2$, but when the distribution is smeared out over another set, its performance deteriorates. The methods described in this paper adjust themselves automatically and are to some extent less sensible to rescalings of the axes. Hence the interest in them.

All results presented here, possibly with other constants, remain true for uniform distributions in general convex sets in the plane. The boundary effect is minimal for the uniform distribution on the circle, as there are about $n^{1/3}$ convex hull points, and thus, the edges on the convex hull are of length about $n^{-1/3}$, only slightly longer than most edges of the Delaunay triangulation (which are about $1/\sqrt{n}$ in length). On the other hand, the uniform distribution on a convex k-gon with k fixed, has about $\Theta(k \log n)$ convex hull points. With high probability, there are some Delaunay edges of length $\Theta(1)$. In a sense then, the boundary effects are worst for the uniform distribution on these polygons, and the analysis presented here for the square thus shows how to deal with the worst situation.

Finally, nearly all methods presented here walk from a pre-selected point Y to the query point X by following a straight line. Other walks are possible that may deviate from the straight line, but their analysis is more cumbersome. An example is provided by Devillers, Pion and Teillaud (2001).

Appendix

LEMMA 10. *Partition $[0,1]^2$ into a regular $k \times k$ grid of squares. Let N_i denote the number of points from among $X_1, \ldots, X_n$ that fall in the i-th square. Define $C = 16$, $D = 17$. If (x, y) is a Delaunay edge, or a segment of a Delaunay edge, and $\min_i N_i > 0$ with $k = \lfloor \sqrt{n/4 \log n} \rfloor$, then either the length of (x, y) is less than $C\sqrt{\log n/n}$, or both x and y are within distance $D\sqrt{\log n/n}$ from the border of the unit square.*

PROOF. Let x and y be two points in the unit square at distance $2r \geq 2r' = C\sqrt{\log n/n}$ from each other. Consider the open circle S with x and y at opposite ends, and center $z = (x + y)/2$. Its radius is r. If (x, y) is a (piece of a) Delaunay edge, then clearly one of the halfcircles S_1 or S_2 must be empty, where the halfcircles have radius r', are centered at z, and have (all of, or part of) (x, y) as supporting segment. Consider the point w in S_1 at distance $r'/2$ from z such that (w, z) is perpendicular to (x, y). Find the cell in the $k \times k$ grid to which w belongs, if this is possible.

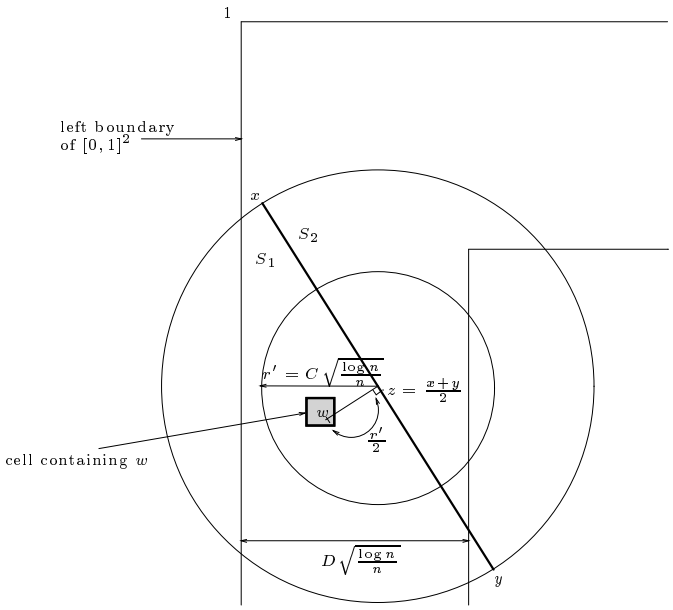


Figure 11.

Assume one of x or y is at distance at least $D\sqrt{\log n/n}$ from the rim. Then z is at distance at least $(D/2)\sqrt{\log n/n}$ from the rim. And w is at distance at least $((D-C)/2)\sqrt{\log n/n}$ from the rim. Thus, w is in the square if $D > C$. We next claim that the entire grid cell to which w belongs falls in the halfcircle, which shows that both x and y must be

within distance $D\sqrt{\log n/n}$ from the rim. Indeed, the maximal distance between z and any point in the cell of w is at most

$$\begin{aligned}\frac{r'}{2}+\frac{\sqrt{2}}{k} &\le \frac{r'}{2}+\frac{\sqrt{2}}{\sqrt{\frac{n}{4\log n}}-1}\\ &\le \frac{r'}{2}+\frac{2}{\sqrt{\frac{n}{4\log n}}} \qquad (\text{if } n \ge 64\log n)\\ &= r'\left(\frac{1}{2}+\frac{8}{C}\right) \le r'\end{aligned}$$

since $C = 16$. □

LEMMA 11. *Let x be arbitrary, and let V_x be the vertical dimension of the leaf rectangle for x. Then*

$$\mathbb{P}\{\sqrt{n}V_x \ge c\} \le \frac{9\log_2(n)e^{-\log^2(c/2)/27}}{\log(c/2)}$$

when $c \ge 2e^6$.

PROOF. Let $n_0 = n$, n_1, $n_2, \ldots$ denote the number of points properly falling in the rectangle to which x belongs as x moves down the 2-d tree, and let k be the first index for which $n_k = 0$. It is clear that $n_i/2 - 1 \le n_{i+1} \le n_i/2$. By induction, we have, $n/2^i - 2 \le n_i \le n/2^i$ for all $k \ge i \ge 0$. As $n_{k-1} \in \{1, 2\}$, we see that $n/4 \le 2^{k-1} \le n$, and thus, $n/2 \le 2^k \le 2n$. The sizes are deterministically fixed of course. If we begin with a vertical cut, then after the first horizontal cut, the vertical size of the rectangle containing x is not more than $\max(B_1, 1-B_1)$, where B_1 is the median of a sample of size n_1. Note that this is $1/2 + |B_1 - 1/2|$. Write $B_1, B_2, \ldots$ for independent random variables distributed as the medians of samples of sizes n_1, n_2, and so forth. As the sample in the rectangle of x remains uniformly distributed conditional on the cut, we see that the vertical size of the rectangle containing x is not more than

$$\prod_{i\ge 0, 2i+1<k}(1/2 + |B_{2i+1} - 1/2|) \le \frac{e^{2\sum_{i\ge 0,2i+1<k}|B_{2i+1}-1/2|}}{2^{\frac{k-1}{2}}} \le \frac{2e^{2\sum_{i\ge 0,2i+1<k}|B_{2i+1}-1/2|}}{\sqrt{n}}\ .$$

Thus, for any x, $\sqrt{n}V_x$ is bounded by a random variable distributed as

$$2\exp\left\{2\sum_{i\ge 0,2i+1<k}|B_{2i+1} - 1/2|\right\}\ .$$

Let $t \geq 2$ be fixed. Then, defining the event

$$A = \bigcup_{i \geq 0, 2i+1<k} \left[|B_{2i+1} - 1/2| \geq \frac{t}{\sqrt{n/4^i}} \right] ,$$

we have

$$\begin{aligned} \mathbb{P}\{A\} &\leq \sum_{i \geq 0, 2i+1<k} \mathbb{P}\left\{ |B_{2i+1} - 1/2| \geq \frac{t}{\sqrt{n/4^i}} \right\} \\ &\leq \sum_{i \geq 0, 2i+1<k, 2t \leq \sqrt{n/4^i}} \frac{3e^{-2t^2(4^i/n)(n_{2i+1}/2)}}{4(t2^i/\sqrt{n})\sqrt{n_{2i+1}/4}} \end{aligned}$$

by Lemma 13 (see below), provided $n_{2i+1}/2 \geq 2$. That we may indeed apply Lemma 13 follows from this simple argument: let i be so large that $2i+1 < k$, yet $n_{2i+1} < 4$. Then clearly, we must have $k > 2i+1 \geq k-2$. Thus, $n/4^i \leq n/4^{(k-3)/2} = 8n/2^k \leq 16 \leq 4t^2$. In the bound above, use $n_{2i+1} \geq n/2^{2i+1} - 2 \geq n/2^{2i+1} - (1/2)n_{2i+1}$, and thus $n_{2i+1} \geq (1/3)n/4^i$ for the i in the range of summation. This replacement shows

$$\begin{aligned} &\mathbb{P}\left\{ \bigcup_{i \geq 0, 2i+1<k} \left[|B_{2i+1} - 1/2| \geq \frac{t}{\sqrt{n/4^i}} \right] \right\} \\ &\leq \sum_{i \geq 0, 2i+1<k} \frac{3\sqrt{12}e^{-t^2/3}}{4t} \\ &\leq \frac{3\log_2(n)e^{-t^2/3}}{t} \end{aligned}$$

because $k/2 \leq \log_2 n$. Bound $\sqrt{n}V_x$ by $\sqrt{n}$ if A holds. On A^c, the complement of A, we have

$$\sqrt{n}V_x \leq 2\exp\left\{ 2 \sum_{i \geq 0, 2i+1<k} \frac{t2^i}{\sqrt{n}} \right\} \leq 2\exp\left\{ \frac{2t2^{k/2}}{\sqrt{n}} \right\} \leq 2e^{t\sqrt{8}} < 2e^{3t} .$$

Replace $c = 2e^{3t}$ and add the condition $c \geq 2e^6$. Thus,

$$\mathbb{P}\{\sqrt{n}V_x \geq c\} \leq \frac{9\log_2(n)e^{-\log^2(c/2)/27}}{\log(c/2)}$$

when $c \geq 2e^6$. □

LEMMA 12. *Let x be arbitrary, and let V_x be the vertical dimension of the leaf rectangle for x. Then*

$$\sum_{j=0}^{\infty} \sup_x j \sqrt{\mathbb{P}\left\{V_x > \sqrt{\frac{j \log n}{4n}}\right\}} = o(1) \ .$$

PROOF. Apply the bound of Lemma 11. For the j-th term in the summation, replace c in Lemma 11 by $\sqrt{j \log n/4}$. Then note that the series

$$\begin{aligned}
&\sum_{j=1}^{\infty} j \sqrt{\frac{9 \log_2(n) e^{-\log^2(\sqrt{j \log n}/4)/27}}{\log(\sqrt{j \log n}/4)}} \\
&\quad \le \sum_{j=1}^{\infty} j \frac{3\sqrt{\log_2(n)} e^{-\log^2(j \log n/16)/216}}{\sqrt{\log(\sqrt{\log n}/4)}} \\
&\quad \le O(\sqrt{\log n}) \sum_{j=1}^{\infty} \frac{j}{(j \log n/16)^{\frac{\log(j \log n/16)}{216}}} \\
&\quad = o(1)
\end{aligned}$$

by noting that for $j \le e^{648}$, the sum is $O((\log n)^{-c \log n})$ for some $c > 0$, and that for the tail, each term is bounded for n large enough by $c\sqrt{\log n}/(j \log n)^2$. □

The following Lemma gives a tail inequality for medians of uniform samples of both odd and even sizes. The beta (a, b) density, $a > 0$, $b > 0$, is the density on $[0, 1]$ given by

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, 0 < x < 1,$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$ is the beta function. The distribution of the k-th smallest of n i.i.d. uniform $[0, 1]$ random variables is beta $(k, n - k + 1)$.

LEMMA 13. *Let B be a beta (n, n) or beta $(n, n + 1)$ random variable. Then, for $u > 0$, $n \ge 2$,*

$$\mathbb{P}\{|B - 1/2| \ge u\} \le \frac{3e^{-2nu^2}}{4u\sqrt{n/2}} \ .$$

PROOF. The simple tail inequality is derived directly from the form of the density of B. Let $B(.,.)$ denote the beta function. Observe that $B(n,n)/B(n,n+1) = 2$, so that when B is beta $(n, n+1)$, its density is

$$B^{-1}(n,n+1)x^{n-1}(1-x)^n \leq 2B^{-1}(n,n)x^{n-1}(1-x)^{n-1} \ , \ 0 < x < 1 \ .$$

That is, at every x, it is at most twice the density of a beta (n,n) random variable. Thus, assume that B is beta (n,n). Note that

$$\frac{B(n,n)}{B(n+1,n+1)} = 4 + \frac{2}{n} \leq 4e^{\frac{1}{2n}} \ .$$

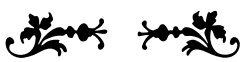
By iterating this, we obtain for $n \geq 3$,

$$B^{-1}(n,n) \leq 4^{n-2}e^{\sum_{j=2}^{n-1}\frac{1}{2j}}B^{-1}(2,2) \leq 4^{n-2}e^{\frac{\log(n-1)}{2}}\,6 = 6 \times 4^{n-2}\sqrt{n-1} \ .$$

This inequality is valid for $n \geq 2$. Thus, for $n \geq 2$,

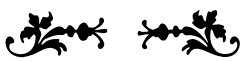
$$\begin{aligned}
\mathbb{P}\left\{|B-1/2| \geq u\right\} &= 2\int_{1/2+u}^{1} B^{-1}(n,n)(x(1-x))^{n-1}\,dx \\
&= 2\int_{u}^{1/2} \frac{B^{-1}(n,n)}{4^{n-1}}\left(1-4y^2\right)^{n-1}\,dy \\
&\leq 3\sqrt{n-1}\int_{u}^{1/2} e^{-4y^2(n-1)}\,dy \\
&\leq 3\sqrt{n-1}\int_{u}^{\infty} \frac{y}{u}e^{-4y^2(n-1)}\,dy \\
&= 3\sqrt{n-1}\frac{e^{-4u^2(n-1)}}{8(n-1)u} \\
&= \frac{3e^{-4u^2(n-1)}}{8u\sqrt{n-1}} \ .
\end{aligned}$$

For $n \geq 2$, we have $n - 1 \geq n/2$. This concludes the proof of Lemma 13. □



Acknowledgment

The authors thank both eagle-eyed referees.



References

D. Angluin and L. G. Valiant, "Fast probabilistic algorithms for Hamiltonian circuits and matchings," *Journal of Computer and System Sciences*, vol. 18, pp. 155–193, 1979.

J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, pp. 509–517, 1975.

J. L. Bentley, B. W. Weide, and A. C. Yao, "Optimal expected-time algorithms for closest point problems," *ACM Transactions on Mathematical Software*, vol. 6, pp. 563–580, 1980.

M. Bern, D. Eppstein, and F. F. Yao, "The expected extremes in a Delaunay triangulation," *International Journal of Computational Geometry and its Applications*, vol. 1, pp. 79–91, 1991.

J.-D. Boissonnat and M. Teillaud, "On the randomized construction of the Delaunay tree," *Theoretical Computer Science*, vol. 112, pp. 339–354, 1993.

J.-D. Boissonnat and M. Yvinec, *Géométrie Algorithmique*, Ediscience, Paris, 1995.

J.-D. Boissonnat and M. Yvinec, *Algorithmic Geometry*, Cambridge University Press, Cambridge, 1998.

H. Borouchaki and P. L. George, *Triangulation de Delaunay et maillages*, Hermès, 1997.

P. Bose and L. Devroye, "Intersections with random geometric objects," *Computational Geometry: Theory and Applications*, vol. 10, pp. 139–154, 1998.

P. Chanzy, L. Devroye, and C. Zamora-Cura, "Analysis of range search for random k-d trees," *Acta Informatica*, vol. 37, pp. 355–383, 2001.

B. Chazelle, O. Devillers, F. Hurtado, M. Mora, V. Sacristan, and M. Teillaud, "Splitting a Delaunay triangulation in linear time," *Proceedings of the Ninth European Symposium on Algorithms*, vol. 2161, pp. 312–320, Lecture Notes in Computer Science, Springer-Verlag, 2001.

T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, MIT Press, Boston, 1990.

O. Devillers, S. Meiser, and M. Teillaud, "Fully dynamic Delaunay triangulation in logarithmic expected time per operation," *Computational Geometry: Theory and Applications*, vol. 2, pp. 55–80, 1992.

O. Devillers, S. Pion, and M. Teillaud, "Walking in a triangulation," *Proceedings of the 17th Annual Symposium on Computational Geometry*, pp. 106–114, 2001.

O. Devillers, "The Delaunay hierarchy," *International Journal of the Foundations of Computer Science*, vol. 13, pp. 163–180, 2002.

L. Devroye, E. Mücke, and B. Zhu, "A note on point location in Delaunay triangulations of random points," *Algorithmica*, vol. 22, pp. 477–482, 1998.

L. Devroye, C. Lemaire, and J.-M. Moreau, *Fast Delaunay point location with search structures*, p. Vancouver, Canada, Eleventh Canadian Conference on Computational Geometry, 15-18 August 1999. 1999.

L. Devroye, J. Jabbour, and C. Zamora-Cura, "Squarish k-d trees," *SIAM Journal on Computing*, vol. 30, pp. 1678–1700, 2000.

G. H. Gonnet and R. Baeza-Yates, *Handbook of Algorithms and Data Structures*, Addison-Wesley, Workingham, England, 1991.

P. Green and R. Sibson, "Computing Dirichlet Tesselations in the plane," *The Computer Journal*, vol. 21, pp. 168–173, 1978.

L. Guibas and J. Stolfi, "Primitives for the manipulation of general subdivisions and the computation of Voronoi diagrams," *ACM Transactions on Graphics*, vol. 4, pp. 74–123, 1985.

L. J. Guibas, D. E. Knuth, and M. Sharir, "Randomized incremental construction of Delaunay and Voronoi diagrams," *Algorithmica*, pp. 381–413, 1992.

K. Hinrichs, J. Nievergelt, and P. Schorn, "An all-round sweep algorithm for 2-dimensional nearest-neighbor problems," *Acta Informatica*, vol. 29, pp. 383–394, 1992.

D. E. Knuth, *The Art of Computer Programming, Vol. 2*, Addison-Wesley, Reading, Mass., 1981. 2nd Ed..

C. L. Lawson, "Software for C1 surface interpolation," in: *Mathematical Software III*, (edited by J. Rice), pp. 161–194, Academic Press, New York, 1977.

C. Lemaire, "Triangulation de Delaunay et Arbres Multidimensionnels," Ph.D. Dissertation, Université Jean Monnet, Saint-Étienne, France, 1997.

E. P. Mücke, I. Saias, and B. Zhu, "Fast randomized point location without preprocessing in two- and three-dimensional Delaunay triangulations," *Computational Geometry: Theory and Applications*, vol. 12, pp. 63–83, 1999.

J. Nievergelt and E. M. Reingold, "Binary search trees of bounded balance," *SIAM Journal on Computing*, vol. 2, pp. 33–43, 1973.

A. Okabe, B. Boots, and K. Sugihara, *Spatial Tessellations*, John Wiley, New York, 1992.

M. Okamoto, "Some inequalities relating to the partial sum of binomial probabilities," *Annals of Mathematical Statistics*, vol. 10, pp. 29–35, 1958.

D. Panchenko, "Some extensions of an inequality of Vapnik and Chervonenkis," *Electronic Communications in Probability*, vol. 7, pp. 55–65, 2002.

E. M. Reingold, J. Nievergelt, and N. Deo, *Combinatorial Algorithms: Theory and Practice*, Prentice Hall, Englewood Cliffs, N.J., 1977.

J. M. Robson, "The height of binary search trees," *The Australian Computer Journal*, vol. 11, pp. 151–153, 1979.

R. F. Sproull, "Refinements to nearest-neighbor searching in k-dimensional trees," *Algorithmica*, vol. 6, pp. 579–589, 1991.

V. N. Vapnik and A. Ya. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability and its Applications*, vol. 16, pp. 264–280, 1971.