ON THE STABBING NUMBER OF A RANDOM DELAUNAY TRIANGULATION

Prosenjit Bose
School of Computer Science
Carleton University
Herzberg Room 5372, 1125 Colonel By Drive
Ottawa, Canada K1S 5B6
jit@cs.carleton.ca

Luc Devroye
School of Computer Science
McGill University
3480 University Street
Montreal, Canada H3A 2K6
luc@cs.mcgill.ca

ABSTRACT. We consider a Delaunay triangulation defined on n points distributed independently and uniformly on a planar compact convex set of positive volume. Let the stabbing number be the maximal number of intersections between a line and edges of the triangulation. We show that the stabbing number S_n is $\Theta(\sqrt{n})$ in the mean, and provide tail bounds for $\mathbb{P}\{S_n \geq t\sqrt{n}\}$. Applications to planar point location, nearest neighbor searching, range queries, planar separator determination, approximate shortest paths, and the diameter of the Delaunay triangulation are discussed.

KEYWORDS AND PHRASES. Delaunay triangulation, proximity graphs, stabbing number, Voronoi diagram, probabilistic analysis, computational geometry, graph diameter.

CR CATEGORIES: 3.74, 5.25, 5.5.

Authors' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The authors' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.

§1. Introduction.

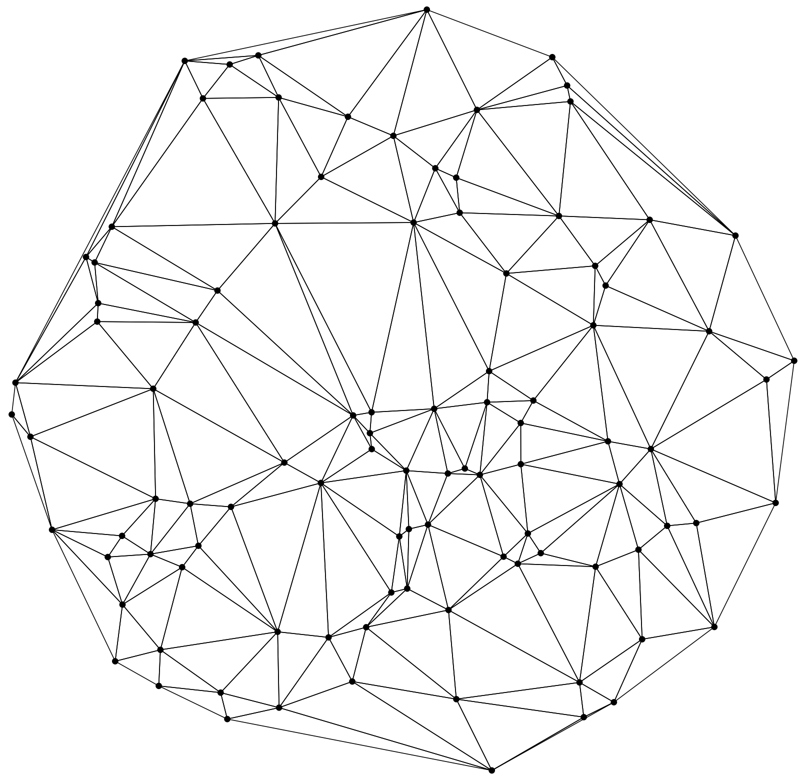
The purpose of this note is to prove a simple result for a random Delaunay triangulation $\mathcal{D}_n$ on n points, $X_1, \ldots, X_n$, that are independently and uniformly distributed on a convex set C of $\mathbb{R}^2$. Throughout the paper, all convex sets are assumed to be compact and of strictly positive volume. The stabbing number S_n of $\mathcal{D}_n$ is the maximal number of Delaunay edges cut by any line. We show the following:

THEOREM 1. *The stabbing number S_n of $\mathcal{D}_n$ is $\Theta(\sqrt{n})$ in the following senses:*

(i) $\mathbb{E} S_n = \Theta(\sqrt{n})$ *;*

(ii) There exist constants $c' > c > 0$ such that $\mathbb{P}\{S_n > c'\sqrt{n}\} \to 0$ and $\mathbb{P}\{S_n < c\sqrt{n}\} \to 0$.

In Theorem 1* below, an explicit tail inequality is derived which shows that it is rather unlikely that S_n is much larger than $t\sqrt{n}$ for a constant t depending upon the shape of C only. Okabe, Boots and Sugihara (1992) describe many of the properties of Delaunay triangulations needed in the proofs below. (see also Bourachaki and George, 1997). We consider a host of problems that directly or indirectly involve a Delaunay triangulation, and whose analysis requires the asymptotic behavior of S_n. Consider for example point location for a query point X. 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. The expected time is $O(\sqrt{n})$ (Devroye, Lemaire and Moreau, 2004). In this paper, we show that the expected time is $O(\sqrt{n})$ even if the query point X is chosen in the worst possible manner after having looked at the data. The bound on the stabbing number allows one to develop simple yet efficient algorithms to solve several other problems such as range queries, shortest-path queries, and nearest neighbor queries. We outline a number of these implications of Theorem 1 in Section 5.



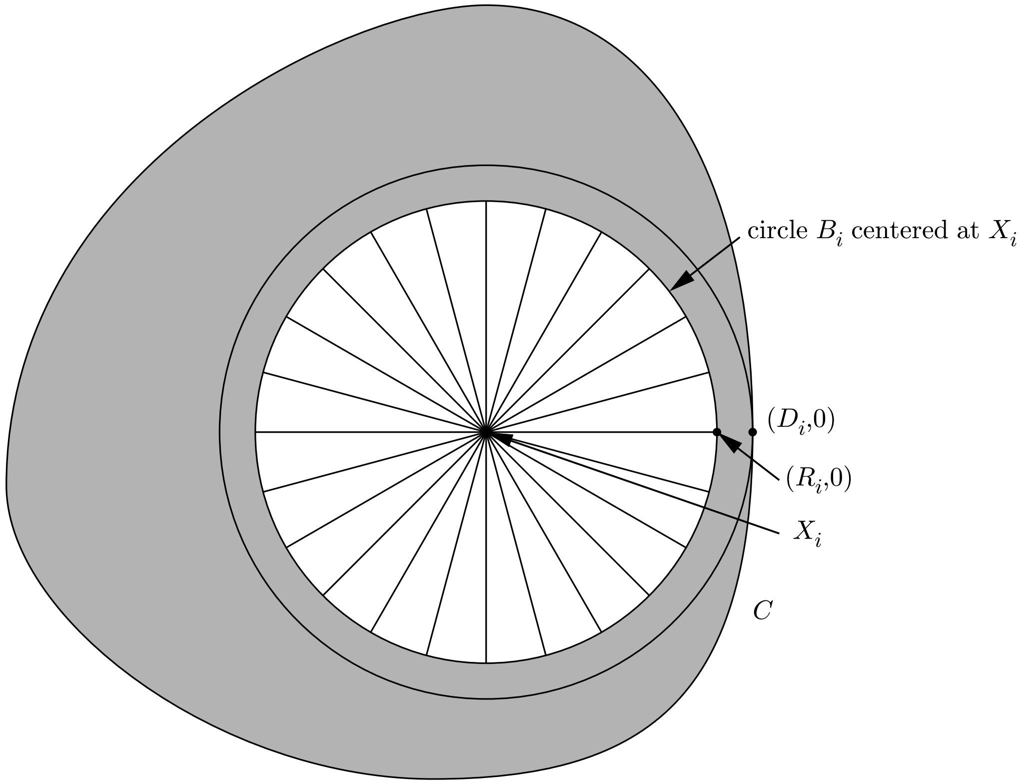
A point set with its Delaunay triangulation. No circle circumscribing any triangle has any data point in its interior.

§2. Border points

We will obtain all our results based on the notion of a *border point.* Let C be a convex set and let $X_1, \ldots, X_n$ be n points in C. Define the distance D_i from X_i to the complement C^c of C:

$$D_i = \inf_{x \notin C} \|X_i - x\| \ .$$

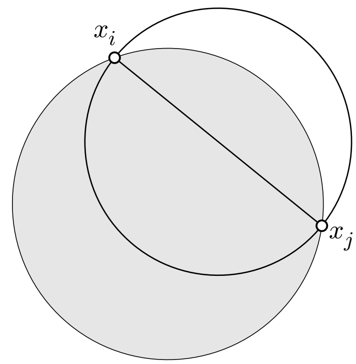
Define the circle B_i centered at X_i of radius $R_i = D_i\sqrt{3}/2$, and partition B_i into 24 cones of equal angle $\pi/12$, with the j-th cone covering all angles in $[((j-1)/24)2\pi, (j/24)2\pi)$. Let $N_{i,j}$ be the cardinality of the j-th cone of B_i, i.e., the number of X_k's with $k \neq i$ that belong to that cone. We call X_i a border point if $\min_j N_{i,j} = 0$.



The convex region C is shaded. A point X_i is a border point if one of the 24 cones centered at it, of radius $R_i = D_i\sqrt{3}/2$, contains no other data point. Here D_i is the distance from X_i to the complement of C.

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) must be empty.*

PROOF. There exists x_k such that the circle through x_i, x_j, x_k is empty. This circle necessarily contains one of the two halfcircles. □



If (x_i, x_j) is a Delaunay edge, then one of two halfcircles supported by (x_i, x_j) must be empty.

Several properties of border points are useful here. The first one shows why we are interested.

LEMMA 2. *Consider the Delaunay triangulation $\mathcal{D}$ for $X_1, \ldots, X_n, Y_1, \ldots, Y_m$, where $Y_1, \ldots, Y_m$ are arbitrary points in C^c. If X_i is not a border point for C, then there is no Delaunay edge from X_i to some Y_j. Thus, all Delaunay edges from X_i to some Y_j must emanate from border points X_i.*

PROOF. For brevity, set $X_i = 0$, $D_i = r$, $B_i = B$. Partition B into 24 equal cones of angle $\pi/12$ each. Assume that $N_{i,j} > 0$ for all j. Let (X_i, Y_k) be a Delaunay edge. Let Z be the point on (X_i, Y_k) at distance r from X_i (so that $Z \in C$). Since (X_i, Y_k) is a Delaunay edge, one of the two halfcircles supported by (X_i, Y_k) must be empty. Thus, one of the halfcircles supported by (X_i, Z) must be empty as well. Fix such a halfcircle H. We claim that H must necessarily contain one of the 24 cones, and thus one of the 24 cones must be empty. Therefore, we obtain a contradiction, and (X_i, Y_k) cannot possibly have been a Delaunay edge. Assume without loss of generality that H is supported by $((0,0),(r,0))$, and faces towards the positive y-axis. Let C be the cone containing $(r\sqrt{3}/2, 0)$. Let C' be the next cone in counterclockwise order. To show that $C' \subseteq H$, it suffices to show that its topmost vertex is in H. This vertex has coordinates $r(\sqrt{3}/2)(\cos\alpha, \sin\alpha)$, where $\pi/12 \le \alpha \le \pi/6$. This square of the distance from this vertex to the center of H, $(r/2, 0)$, is

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

This concludes the proof of Lemma 2. □

The way Lemma 2 will be used is as follows. Consider data $X_1, \ldots, X_n$ on a convex set C. Let L be an infinite line, and let N be the number of Delaunay edge intersections with L. Clearly, L partitions C into two convex sets A and B. Let N_A and N_B be the border points for the data, restricted to A and B respectively. It is clear from Lemma 1 that any intersections with L can only be between border points. But the part of the Delaunay triangulation restricted to these points is a planar graph, and thus, the number of edges in this graph is at most three times the number of vertices. Thus,

$$N \leq 3(N_A + N_B) .$$

By virtue of this, we need only study N_A, the number of border points in a given convex set. To study N_A, we will use specialized versions of the Azuma-Hoeffding method of bounded differences (Azuma, 1967; Hoeffding, 1963; McDiarmid, 1989).

Lemma 3. *Let M_n be the maximum distance from a border point to C^c. Then, if v is the volume of C, and $c > 0$,*

$$\mathbf{P}\left\{ M_n \geq \sqrt{32vc\log n/\pi n} \right\} \leq 24en^{1-c} .$$

Proof. Introduce $u = \sqrt{32vc\log n/\pi n}$. If $M_n \geq u$ then for some $i \leq n$ with $D_i \geq u$ and some cone of radius $u\sqrt{3}/2$ centered at X_i, no other data point falls in this cone. As the probability of this cone is $\pi u^2/32v$, where v is the volume of C, we see that

$$\mathbf{P}\{M_n \geq u\} \leq 24n\left(1 - \pi u^2/32v\right)^{n-1} \leq 24ne\, e^{-n\pi u^2/32v} \leq 24en^{1-c} . \square$$

Lemma 4. *Let $X_1, \ldots, X_n$ be i.i.d. and uniformly distributed in a convex set C, and let $Y_n(X_1, \ldots, X_n)$ be the number of border points. Define $\gamma = 4p/\sqrt{v}$, where v is the volume of C, and p is the length of the perimeter of C. Then*

$$\mathbf{E}\{Y_n\} \leq \gamma\sqrt{n} .$$

Proof. Clearly, $\mathbf{E}\{Y_n\}$ is n times the probability that X_1 is a border point. The latter probability is the probability that one of the 24 cones of the circle of radius $D_1\sqrt{3}/2$ is empty, where D_1 is the distance from X_1 to C^c. For $n = 0$, the inequality is clearly true. For $n = 1$, it is true because $Y_n \leq 1$ and $\gamma \geq 8\sqrt{\pi}$. For $n \geq 2$,

$$\begin{aligned}
\mathbf{E}\{Y_n\} &= n\mathbf{E}\left\{(1 - \pi D_1^2/32v)^{n-1}\right\} \\
&\leq n\mathbf{E}\left\{\exp\left(-(n-1)\pi D_1^2/32v\right)\right\} \\
&= n\int_0^1 \mathbf{P}\left\{\exp\left(-(n-1)\pi D_1^2/32v\right) > t\right\}\, dt \\
&= n\int_0^\infty \mathbf{P}\left\{(n-1)\pi D_1^2/32v < u\right\}\, e^{-u}\, du \\
&= n\int_0^\infty \mathbf{P}\left\{D_1 < \sqrt{\frac{32vu}{(n-1)\pi}}\right\}\, e^{-u}\, du
\end{aligned}$$

$$
\begin{aligned}
&\leq n \int_0^\infty \frac{p}{v} \sqrt{\frac{32vu}{(n-1)\pi}}\, e^{-u}\, du \\
&= \frac{pn\sqrt{\pi}}{2v} \sqrt{\frac{32v}{(n-1)\pi}} \\
&= \sqrt{\frac{8p^2n^2}{(n-1)v}} \\
&\leq \sqrt{\frac{16p^2n}{v}} \\
&= (4p/\sqrt{v})\sqrt{n} \\
&= \gamma\sqrt{n}\ . \ \square
\end{aligned}
$$

LEMMA 5. *Let C_0 be a convex set contained in the convex set C, and let $X_1, \ldots, X_n$ be uniformly distributed in C. Let $Z_n = Z_n(X_1, \ldots, X_n)$ denote the number of border points in C_0. Then*

$$
\mathbf{E}\{Z_n\} \leq \gamma\sqrt{n}\ ,
$$

where the constant $\gamma > 0$ is as in Lemma 4.

PROOF. Let N be the number of X_i's falling in C_0. Let v_0, v be the volumes of C_0 and C, and let p_0 and p be the perimeters of C_0 and C. Then by Lemma 4 and Jensen's inequality,

$$
\begin{aligned}
\mathbf{E}\{Z_n\} &= \mathbf{E}\{\mathbf{E}\{Z_n|N\}\} \\
&\leq \mathbf{E}\{(4p_0/\sqrt{v_0})\sqrt{N}\} \\
&\leq (4p_0/\sqrt{v_0})\sqrt{\mathbf{E}\{N\}} \\
&= (4p_0/\sqrt{v_0})\sqrt{nv_0/v} \\
&= (4p_0/\sqrt{v})\sqrt{n} \\
&\leq (4p/\sqrt{v})\sqrt{n}\ . \ \square
\end{aligned}
$$

LEMMA 6. *Let $X_1, \ldots, X_n$ be uniformly distributed on a convex set C of perimeter p and volume $v > 0$. Define $\gamma = 4p/\sqrt{v}$ and*

$$
W = Y_n(X_1, \ldots, X_n) - Y_{n-m}(X_{m+1}, \ldots, X_n)
$$

for $1 \leq m \leq n/2$. If $c > 1$, $c' > 0$, $n > e$, and

$$
\xi = 512c\log n \left(\gamma m\sqrt{16c\log n/\pi n} + c'\log n\right)\ ,
$$

then we have

$$
\mathbf{P}\{|W| \geq \xi\} \leq \frac{3 + 24e\,2^c}{n^\alpha}
$$

where $\alpha = \min(c-1, c'/3, 128cc'/3)$.

Proof. Define $\delta = \sqrt{32vc\log n/\pi(n-m)}$, where $c > 0$. Define $t = 2pm\delta/v + c'\log n$ for $c' > 0$. Set $\rho = 4t\pi\delta^2/v$. Let C_δ be the collection of all points that are within distance δ of the exterior C^c. Then define the events

$$A = [\text{all border points for } X_1, \ldots, X_n \text{ and } X_{m+1}, \ldots, X_n \text{ are in } C_\delta\ ]\ ;$$

$$B = \left[\sum_{i=1}^{m} 1_{[X_i \in C_\delta]} \le t\right]\ ;$$

$$D = \left[\sum_{i=1}^{m} 1_{[X_i \in C_{2\delta}]} \le 2t\right]\ ;$$

$$E = \left[\sum_{j=m+1}^{n} 1_{\left[X_j \in \cup_{1\le i\le m: X_i \in C_{2\delta}} B(X_i,\delta)\right]} \le 2\rho n\right]\ .$$

We claim that if A, B, and E hold, then

$$|W| \le \max(2\rho n, t) \le \xi\ .$$

To see this, observe that by removing $X_1, \ldots, X_m$ from the data, the number of border points may decrease. This can happen only for X_i's with $i \le m$ that are border points for $X_1, \ldots, X_n$. But under event A, the decrease is not more than the number of points of $X_1, \ldots, X_m$ that are in C_δ, which under event B cannot be more than t. The number of border points may increase. This can only happen if X_j, $j > m$, is not a border point for the full data set but becomes one for $X_{m+1}, \ldots, X_n$. Under event A, each such X_j must be in C_δ. So, $D_j \le \delta$. But then some X_i, $i \le m$, must be within $D_j\sqrt{3}/2 < \delta$ of X_j (otherwise its removal would have no effect on the status of X_j), and thus X_i has to be within distance 2δ from C^c: X_i in $C_{2\delta}$. Let $B(X_i,\delta)$ be the ball of radius δ about X_i. The increase in the number of border points is thus bounded by the number of X_j's, $j > m$, that fall in $\cup_{i\le m} B(X_i,\delta)$, with the union restricted to those X_i's in $C_{2\delta}$. By event E, this number does not exceed $2\rho n$. Thus, $A \cap B \cap E \subseteq [|W| \le \xi]$. Therefore,

$$\begin{aligned}\mathbf{P}\{|W| \ge \xi\} &\le \mathbf{P}\{(A\cap B\cap E)^c\}\\ &\le \mathbf{P}\{A^c\} + \mathbf{P}\{B^c\} + \mathbf{P}\{D^c\} + \mathbf{P}\{D \cap E^c\}\ .\end{aligned}$$

Clearly, with B_n as in Lemma 3,

$$\mathbf{P}\{A^c\} \le \mathbf{P}\{B_n > \delta\} + \mathbf{P}\{B_{n-m} > \delta\} \le 48e(n-m)^{1-c}$$

by a double application of Lemma 3. Next, note that the number of X_i's, $i \le m$ in C_δ is not more in distribution than a binomial $(m, p\delta/v)$ random variable V. Thus, by Lemma A,

$$\begin{aligned}\mathbf{P}\{B^c\} &\le \mathbf{P}\{V \ge 2pm\delta/v + c'\log n\}\\ &\le \min\left(e^{-pm\delta/3v}\ ,\ e^{-(c'\log n/(pm\delta/v))^2 pm\delta/3v}\right)\\ &= \min\left(e^{-pm\delta/3v}\ ,\ e^{-c'^2 v\log^2 n/3pm\delta}\right)\ .\end{aligned}$$

Similarly, replacing δ by 2δ and c' by $2c'$ throughout,

$$\mathbf{P}\{D^c\} \le \min\left(e^{-2pm\delta/3v}\ ,\ e^{-2c'^2 v\log^2 n/3pm\delta}\right)\ .$$

Finally, conditioning on $X_1, \ldots, X_m$ such that D holds, $\sum_{j=m+1}^n 1_{\left[X_j \in \cup_{1 \le i \le m : X_i \in C_{2\delta}} B(X_i,\delta)\right]}$ is stochastically smaller than a binomial (n, ρ) random variable V'. Therefore,

$$\mathbb{P}\{DE^c\} \le \mathbb{P}\{V' \ge 2\rho n\} \le e^{-\rho n/3} \ .$$

Plugging this back into our inequalities, we see that

$$\mathbb{P}\{|W| \ge \xi\} \le 48e(n-m)^{1-c} + 2\min\left(e^{-pm\delta/3v} \ , \ e^{-c'^2 v \log^2 n/3pm\delta}\right) + e^{-\rho n/3} \ .$$

The first term is not more than $24e\,2^c/n^{c-1}$. The middle term has two exponents. Regardless of the value of $pm\delta/v$, one exponent must be smaller than $-c'\log n/3$, so that the middle term in the bound is not more than $2/n^{c'/3}$. Finally, bound the last term by observing that $\rho \ge 4c'\log n \times \pi\delta^2/v = 128cc'\log^2 n/n \ge 128cc'\log n/n$. Thus, the third term does not exceed $1/n^{128cc'/3}$. □

Finally, we turn to the main tail bound for Y_n, derived by means of Lemmas 4, 6 and D.

Lemma 7. *Let $\gamma = 4p/\sqrt{v}$ be as in Lemma 6. Then, there exists a universal integer n_0 such that for $n \ge n_0$,*

$$\mathbb{P}\{Y_n \ge 4(\gamma\sqrt{n} + 129024\log^2 n)\} \le \frac{8358 + 33416\log^5 n}{n^6} \ .$$

Proof. Assume throughout $n \ge 2^{31}$. If $16p/\sqrt{v} > \sqrt{n}$, the probability is clearly zero, as $Y_n \le n$. So, we assume $p/\sqrt{v} \le \sqrt{n}/16$. Note that in any case, $p/\sqrt{v} \ge 2\sqrt{\pi} > 3$ because for fixed volume v, p is minimized for the circle. Define $k = \lfloor \log^5 n \rfloor$, and note that $0.9999\log^5 n \le k \le \log^5 n < n/2$. Define $m = \lfloor n/k \rfloor$ and $n' = mk$. Note that $n/2 \le n-k \le n' \le n$, and that $n'/(n'-m) = k/(k-1) < 1.0001$. Define

$$W = Y_n(X_1, \ldots, X_n) - Y_{n'}(X_1, \ldots, X_{n'})$$

and

$$Z = Y_{n'}(X_1, \ldots, X_{n'}).$$

We have

$$Y_n(X_1, \ldots, X_n) \le (W + Z.$$

Partition the data $X_1, \ldots, X_{n'}$ into k vectors $Z_1, \ldots, Z_k$, where $Z_1 = (X_1, \ldots, X_m)$, $Z_2 = (X_{m+1}, \ldots, X_{2m})$ and so forth. With a slight abuse of notation, we use $Y_{n'}(X_1, \ldots, X_{n'})$ and $Y_k(Z_1, \ldots, Z_k)$ according to whichever is more convenient. With this notation, we have $Z \equiv Y_k$. Clearly, $Y_k \ge 0$ and Y_k is permutation invariant. So to apply Lemma D we need to bound the tail probabilities for

$$W' = Y_k(Z_1, \ldots, Z_k) - Y_{k-1}(Z_2, \ldots, Z_k) \ .$$

We have, for $\xi, \theta > 0$,

$$\begin{aligned} &\mathbb{P}\left\{Y_n \ge 4(\gamma\sqrt{n} + \theta\log^2 n)\right\} \\ &\le \mathbb{P}\left\{W \ge \gamma\sqrt{n} + \theta\log^2 n\right\} + \mathbb{P}\left\{Z \ge 3\left(\gamma\sqrt{n} + \theta\log^2 n\right)\right\} \\ &\le \mathbb{P}\left\{W \ge \gamma\sqrt{n} + \theta\log^2 n\right\} + 4k\mathbb{P}\left\{|W'| \ge \psi/2 + \theta\log^2(n)/2\right\} + 4\exp\left(-\frac{\gamma^2 n}{2k(\psi + \theta\log^2 n)^2}\right) \end{aligned}$$

where we used Lemmas 4 and D. We choose

$$\psi = 60534\gamma\sqrt{n}\log^{-7/2} n$$

and

$$\theta = 2 \times 64512 = 129024$$

and bound each of the terms in the upper bound individually.

THE TERM INVOLVING W. We apply Lemma 6 to W and show the following:

$$\mathbb{P}\left\{W \geq \gamma\sqrt{n} + \theta \log^2 n\right\} \leq \frac{8354}{n^6} .$$

First, we replace m in the definition of ξ in Lemma 6 by $n - n'$, set $c = 7, c' = 18$ there, and define

$$\zeta = 512c\log n \left(\gamma k\sqrt{16c\log n/\pi n} + c'\log n\right) = 64512\log^2 n + 2048\gamma(\log n)^{3/2} k\sqrt{7/\pi n} .$$

By the bound of Lemma 6, if n is so large that

$$2048(\log n)^{3/2} k\sqrt{7/\pi n} < \sqrt{n} ,$$

then

$$\mathbb{P}\left\{W \geq \gamma\sqrt{n} + 64512\log^2 n\right\} \leq \mathbb{P}\{W \geq \zeta\} \leq \frac{3 + 24e\,2^c}{n^\alpha} \leq \frac{8354}{n^6}$$

since $\alpha = \min(c-1, c'/3, 128cc'/3) = 6$.

THE TERM INVOLVING W'. We apply Lemma 6 to W' and show the following:

$$\mathbb{P}\left\{|W'| \geq \psi/2 + \theta\log^2(n)/2\right\} \leq \frac{8354}{n^6} .$$

First we choose ξ as in Lemma 6 (which should be applied with m as in the present context, but with n replaced by n'). Picking $c = 7$, $c' = 18$, we have

$$\begin{aligned} \xi &= 512c\log n' \left(\gamma m\sqrt{16c\log n'/\pi n'} + c'\log n'\right) \\ &\leq 512cc'\log^2 n + 2048c\log n\,\gamma\sqrt{14n\log n/\pi k^2} \\ &\leq 64512\log^2 n + 30267\gamma\sqrt{n}\log^{-7/2} n . \end{aligned}$$

Recalling $\theta/2 = 64512$ and $\psi = 60534\gamma\sqrt{n}\log^{-7/2} n$, we conclude from Lemma 6 the following:

$$\mathbb{P}\left\{|W'| \geq \psi/2 + \theta\log^2(n)/2\right\} \leq \mathbb{P}\{|W'| \geq \xi\} \leq \frac{8354}{n^6} .$$

THE EXPONENTIAL TERM. The last term in the upper bound is

$$\begin{aligned} &4\exp\left(-\frac{\gamma^2 n}{2k(\psi + \theta\log^2 n)^2}\right) \\ &\leq 4\exp\left(-\frac{\gamma^2 n}{4k(\psi^2 + \theta^2\log^4 n)}\right) \\ &\leq 4\exp\left(-\frac{\gamma^2 n}{4\log^5 n(60534^2\gamma^2 n\log^{-7} n + 129024^2\log^4 n)}\right) \\ &\leq 4\exp\left(-\frac{1}{4(60534^2\log^{-2} n + 129024^2\log^9 n/\gamma^2 n)}\right) \end{aligned}$$

$$\leq 4\exp\left(-\frac{1}{4(60534^2\log^{-2}n+129024^2\log^9 n/64\pi n)}\right)$$
$$\leq 4\exp\left(-\frac{\log^2 n}{5\times 60534^2}\right)$$
$$\leq \frac{4}{n^6}$$

provided that n is so large that $4\times 129024^2\log^9 n/64\pi n < 60534^2\log^{-2} n$, which is the case here, and that $\log n > 30\times 60534^2$.

We let $n_0 \geq 2^{31}$ be so large that for all $n\geq n_0$, $n/(\log n)^{3/2}\geq 2048\sqrt{7/\pi}$ and $\log n > 30\times 60534^2$. Collecting bounds, we thus have for $n\geq n_0$,

$$\mathbf{P}\left\{Y_n \geq 4(\gamma\sqrt{n}+\theta\log^2 n)\right\} \leq \frac{4+(1+4k)8354}{n^6} \leq \frac{8358+33416\log^5 n}{n^6}\ .\ \square$$

Lemma 7 provides a useful tail bound for the number of border points in any convex region that has n uniformly distributed points in it. However, we need more, as we will consider all regions that are obtained by intersecting C with a linear halfspace H. Note in particular that the care we took in the previous lemmas with respect to the dependence of various inequalities on the perimeter and volume of C finally pays off. Without it, we would not have been able to handle the boundary effect correctly. Also note the dependence of the final result, once again, on the shape parameter γ. Indeed, the bound below cannot be made uniform over all convex sets: just consider a rectangle C of length n and height 1.

Lemma 8. *Let $X_1,\ldots,X_n$ be i.i.d. and uniformly distributed in a convex region C with perimeter p and volume $v>0$. Set $\gamma = 4p/\sqrt{v}$. Let $\mathcal{H}$ denote the class of all closed halfspaces, and let Y_H denote the number of border points of the subsample that belongs to $C\cap H$. Let n_0 be as in Lemma 7, and define $\theta = 129024$,*

$$u = \max\left(n_0, \sqrt{32}\,(\gamma\sqrt{n}+\theta\log^2 n)\right)\ ,$$

and

$$n \geq \max\left(n_0, 8\gamma^2, 8e^{16}\theta^2\right)\ .$$

Then:

$$\sup_{H\in\mathcal{H}} \mathbf{P}\{Y_H\geq u\} \leq 2e^{-u/6} + \frac{8358+33416\log^5 n}{u^6}\ .$$

Proof. First we note that $n\geq\sqrt{8}\gamma\sqrt{n}$, $n\geq\sqrt{8}\theta\log^2 n$ (the latter follows from the inequality $\log z\leq e^4 z^{1/4}$ for $z>0$), and $n\geq\sqrt{32}\,(\gamma\sqrt{n}+\theta\log^2 n)$, so that $n\geq u$ for all n as in the statement of Lemma 8. We introduce N_H, the number of data points in $H\cap C$, which is a binomial $(n, v_H/v)$ random variable, where v_H denotes the volume of $H\cap C$, and p_H denotes its perimeter. We set $\gamma_H = 4p_H/\sqrt{v_H}$. Our inequality uses the following inclusion of events, after noting that $Y_H\leq N_H$:

$$[Y_H\geq u] \subseteq [v_H/v\leq u/2n, N_H\geq u] \cup \left[v_H/v\geq u/2n, N_H\geq u, Y_H\geq 16(p_H/\sqrt{v_H})\sqrt{N_H}\right]$$
$$\cup\left[v_H/v\geq u/2n, N_H\geq u, u\leq 16(p_H/\sqrt{v_H})\sqrt{N_H}\right]\ .$$

We consider each event separately. By Lemma A, as $u\leq n$,

$$\mathbf{P}\left\{v_H/v\leq u/2n, N_H\geq u\right\} \leq \mathbf{P}\{\text{binomial }(n,u/2n)\geq u\} \leq e^{-u/6}\ .$$

By Lemma 7, if $u \ge n_0$, and $v_H/v \ge u/2n$,

$$\mathbf{P}\left\{N_H \ge u, Y_H \ge 16(p_H/\sqrt{v_H})\sqrt{N_H}\right\} \le \mathbf{E}\left\{1_{[N_H \ge u]}\frac{8358 + 33416\log^5 N_H}{N_H^6}\right\}$$
$$\le \frac{8358 + 33416\log^5 n}{u^6} \ .$$

Finally, by Lemma A again, if $v_H/v \ge u/2n$,

$$\begin{aligned}\mathbf{P}\left\{N_H \ge u, u \le 16(p_H/\sqrt{v_H})\sqrt{N_H}\right\} &\le \mathbf{P}\left\{\text{binomial}(n, v_H/v) \ge u^2 v_H/256p_H^2\right\} \\ &\le \mathbf{P}\left\{\text{binomial}(n, v_H/v) \ge 2p^2 v_H n/v p_H^2\right\} \\ &\le \mathbf{P}\left\{\text{binomial}(n, v_H/v) \ge 2v_H n/v\right\} \\ &\le e^{-nv_H/3v} \\ &\le e^{-u/6} \ .\end{aligned}$$

This concludes the proof of Lemma 8. □

We note that the last inequality is uniform over all C and all H, so the tail of the random variable Y_H/γ behaves in a universal manner. It is precisely this universality that will allow us to derive a number of nice results.

We used a concentration result for Y_H in the proof of Lemma 8. However, we did not present the best possible bounds as that would have made the paper too long. It suffices to say that the variation of Y_H about its mean (which is $\Theta(\sqrt{n})$) is close to $\Theta(n^{1/4})$.

§3. The stabbing number.

In this section, we prove our main result.

THEOREM 1*. *Let S_n be the stabbing number for the Delaunay triangulation of n points that are independent and uniformly distributed on an arbitrary convex set C with perimeter p and volume v. Define $\gamma = 4p/\sqrt{v}$. Let n_0 be as in Lemma 7, and define $\theta = 129024$,*

$$u = \max\left(n_0, \sqrt{32}\,(\gamma\sqrt{n} + 6\theta\log^2 n)\right) \ ,$$

and

$$n \ge \max\left(n_0, 8\gamma^2, 8e^{16}\theta^2\right) \ .$$

Then:

$$\mathbf{P}\left\{S_n \ge 0.1875\gamma^2 + 3.18\gamma\sqrt{\log n} + 6u\right\} \le \frac{2 + 1.3\,10^{-7}\log^5 n}{n} \ .$$

PROOF. Partition the perimeter of C into n pieces of length p/n each, where length is measured along the perimeter. Call the endpoints of these pieces $x_1, \ldots, x_n$, in counterclockwise order. Let $L_{i,j}$ be the line segment joining x_i and x_j, and let $S_{i,j}$ be the number of Delaunay edges encountered by $L_{i,j}$. Take an infinite line L, and let x, y be the points where L enters C and where it leaves C respectively. Locate the two neighbors x_i, x_{i+1} of x along the perimeter, and similarly, find the two neighbors x_j, x_{j+1} for y. Let H be the halfspace supported by $L_{i+1,j}$ that contains the arc from x_{i+1} to x_j in counterclockwise order, and let H' be the halfspace supported by $L_{j+1,i}$ that contains the arc from x_{j+1} to x_i in counterclockwise order. Assume for now that $i \neq j$ and $i+1 \neq j$ and $j+1 \neq i$. We claim that any Delaunay edge reaching L either emanates from a border point in $C \cap H$ or a border point in $C \cap H'$ or a point in $C - H - H'$. Using the notation $N_{C-H-H'}$ to denote the number of data points in $C - H - H'$ and Y_H for the number of border points of $C \cap H$, we see that the number of Delaunay edges reaching L cannot exceed

$$3(N_{C-H-H'} + Y_H + Y_{H'}) \ .$$

If $x_{i+1} = x_j$, then a similar argument yields a bound

$$3(N_{C-H'} + Y_{H'}) \ .$$

If $i = j$, we obtain the bound

$$3(N_{C-H''} + Y_{H''}) \ ,$$

where H'' is the halfspace supported by (x_i, x_{i+1}) that contains the arc (x_{i+1}, x_i) (in counterclockwise order). Note that all sets $C - H', C - H, C - H'', C - H - H'$ have probability not exceeding $p^2/2nv$ because they can be fit into a rectangle of base $p/2$ and height not exceeding p/n. There are at most n^2 such sets, which we might as well number $A_1, \ldots, A_{n^2}$. Similarly, we may label all possible halfspaces $H_1, \ldots, H_{n^2}$. Let S_n be the stabbing number. Observe that

$$S_n \leq 6 \sup_{1 \leq i \leq n^2} Y_{H_i} + 3 \sup_{1 \leq i \leq n^2} N_{A_i} \ .$$

By Lemmas A and 8, if $n \geq n_0$,

$$\begin{aligned}
&\mathrm{P}\left\{ S_n \geq (3p^2/2v) \left(1 + \sqrt{18 v \log n / p^2}\right) + 6u \right\} \\
&\leq \sum_{i=1}^{n^2} \mathrm{P}\left\{ N_{A_i} \geq (p^2/2v) \left(1 + \sqrt{18 v \log n / p^2}\right) \right\} + \sum_{i=1}^{n^2} \mathrm{P}\{Y_{H_i} \geq u\} \\
&\leq n^2 \mathrm{P}\left\{ \text{binomial}(n, p^2/2nv) \geq (p^2/2v) \left(1 + \sqrt{18 v \log n / p^2}\right) \right\} + n^2 \sup_H \mathrm{P}\{Y_H \geq u\} \\
&\leq n^2 \exp\left(-3 \log n\right) + 2n^2 e^{-u/6} + \frac{(8358 + 33416 \log^5 n) n^2}{u^6} \\
&\leq \frac{1}{n} + 2n^2 e^{-\sqrt{32}\,\theta \log^2 n} + \frac{(8358 + 33416 \log^5 n) n^2}{32^3 \gamma^6 n^3} \\
&\leq \frac{2}{n} + \frac{8358 + 33416 \log^5 n}{32^3\, 8^6 \pi^3 n} \\
&\leq \frac{2 + 1.3\, 10^{-7} \log^5 n}{n} \ .
\end{aligned}$$

Replace $p/\sqrt{v}$ by $\gamma/4$ and conclude. □

In the notation of Theorem 1*, we obtain trivially a bound for $\mathrm{E}\{S_n\}$, as $S_n \leq 3n$, valid for all $n \geq n_0$:

$$\begin{aligned}
\mathrm{E}\{S_n\} &\leq 3n\, \mathrm{P}\left\{ S_n \geq 0.1875\gamma^2 + 3.18\gamma\sqrt{\log n} + 6u \right\} + 0.1875\gamma^2 + 3.18\gamma\sqrt{\log n} + 6u \\
&\leq 6 + 3.9\, 10^{-7} \log^5 n + 0.1875\gamma^2 + 3.18\gamma\sqrt{\log n} + 6\sqrt{32}\,\gamma\sqrt{n} + 36\sqrt{32}\,\theta \log^2 n \\
&= O\left(\gamma^2 + \gamma\sqrt{n}\right) \ .
\end{aligned}$$

§4. Applications

In this section, the different applications and consequences of Theorem 1 and 1* are explored. We assume a Delaunay triangulation of the data $X_1, \ldots, X_n$ that are independent and uniformly distributed in a convex set C. We also assume that the Delaunay triangulation is stored in a standard graph structure such as a doubly-connected edge list, a winged-edge structure or a quad-edge structure (Guibas and Stolfi, 1985). All of these structures support such operations as reporting the d edges, vertices and triangles incident to a given vertex in $O(d)$ time or reporting the two triangles incident on one edge in constant time.

Planar Point Location. PLANAR POINT LOCATION, in our case, refers to the problem of determining which triangle in a Delaunay triangulation contains a given query point. Three criteria are usually measured when addressing the problem of planar point location: pre-processing time, storage space of the data structure and query time. Although there exist many optimal solutions in the literature, most of the solutions are complex and require intricate secondary structures to support fast query times (see Snoeyink (1997) for a survey). We will show that in a Delaunay triangulation $\mathcal{D}_n$ for data $X_1, \ldots, X_n$ stored in the standard manner, a very simple algorithm performs quite well.

Given a query point X, the goal is to determine the triangle to which X belongs (if any). This can be achieved by the following simple method suggested by Lawson (1977) and Green and Sibson (1978): take a random point from the X_i's, and follow the triangles intersecting the segment $[X_i, X]$ in order until the triangle containing the query point X is reached. Finding the first triangle out of X_i costs $O(D_i)$ where D_i is the degree of X_i. Each subsequent step across triangles costs $O(1)$, therefore, the total search cost is bounded by a constant times the degree of X_i plus the number of triangles crossed. The number of triangles intersected by the line segment $[X_i, X]$ is bounded by the stabbing number of the line through the two points. Thus, by Theorem 1, this quantity is $O(\sqrt{n})$.

Let D_i denote the degree of X_i in the Delaunay graph, and set $D_n^* = \max_{1 \le i \le n} D_i$. Assume that the X_i's are i.i.d. and uniformly distributed in a convex set C. Bern, Eppstein and Yao (1991) showed that for all points that are at least $\epsilon > 0$ away from the exterior C^c, the expected maximal degree is $\Theta(\log n / \log\log n)$. The maximal degree is greatly influenced by the border effect. In fact, for sufficiently rotund C, $\mathbb{E}\{D_n^*\} = \Theta(\log n)$, the maximum occurring roughly speaking for a convex hull point. However, this result is not of primary interest in this paper, so a weaker result is sufficient for the sequel, and follows very easily from our results on border points.

LEMMA 9. *For any convex set C, $\mathbb{E}\{D_n^*\} \le 1 + 12\log n + 4\mathbb{E}\{S_n\} = O(\sqrt{n})$.*

PROOF. Consider the two vertical lines and the two horizontal lines at distance $u = \sqrt{32cv\log n/\pi n}$ from X_i, where $c = 3\pi/64$, and v is the volume of C. Let B be the intersection of C with the square of side $2u$ centered at X_i. Then the degree of X_i is clearly bounded by the number of Delaunay edges crossing any one of those four lines, plus the number of points (N_i) in B. Thus, if S_n denotes the stabbing number,

$$D_n^* \le 4S_n + \max_i N_i \ .$$

If $p = 4u^2/v = 128c\log n/\pi n$, Lemma A implies that

$$\mathbb{P}\left\{\max_i N_i \ge 2np\right\} \le n\,\mathbb{P}\left\{\text{binomial}(n,p) \ge 2np\right\} \le ne^{-np/3} = n^{1-128c/3\pi} = 1/n \ .$$

Therefore, $\mathbb{E}\{\max_i N_i\} \le 1 + 2np = 1 + 12\log n$. By Lemma 8, $\mathbb{E}\{D_n^*\} = O(\sqrt{n})$. □

We conclude with the following result.

THEOREM 2. *The expected worst-case complexity of the Green-Sibson method for point location when the data points are independent and identically distributed on a convex set C is $\Theta(\sqrt{n})$. Here, worst-case refers to the placement of the query point and selection of the starting point from $X_1, \dots, X_n$, even after having seen the data $X_1, \dots, X_n$.*

PROOF. We will only show the $O(\sqrt{n})$ upper bound. Assume that the starting point is X_1. Then, the complexity is bounded by $D_1 + 1 + S_n$. Note that $\mathbb{E}D_1 \le 6$, 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 $6n$). However, if the starting point is selected maliciously after the data has been shown, then the complexity is bounded by

$$\sup_i D_i + 1 + S_n \ .$$

By Lemma 9, $\mathbb{E}\{\sup_i D_i\} = O(\sqrt{n})$. By Lemma 8, $\mathbb{E}\{S_n\} = O(\sqrt{n})$. Therefore, the expected complexity is $O(\sqrt{n})$. □

Nearest Neighbor Query. NEAREST NEIGHBOR QUERY refers to the problem of deciding, given a query point X, which of the X_i's is closest to X. In this section, distance is measured in the standard Euclidean metric (see Smid (2000) for a survey of closest-point problems). Given a Delaunay triangulation $\mathcal{D}_n$ for data $X_1, \dots, X_n$ stored in the standard manner, the structure of the triangulation allows one to quickly determine the X_i closest to a given query point X.

Before outlining the steps of the algorithm, we introduce the notion of a *legal flip*. Let $\Delta(a, b, c)$ and $\Delta(a, c, d)$ be two triangles sharing the edge $[a, c]$. If the four points a, b, c, d are in convex position, then a *flip* is the operation of replacing edge $[a, c]$ with the edge $[b, d]$ and forming two new triangles. This flip is *legal* in the Delaunay sense if the circle through points a, b, d does not contain point c. The following shows the relation between legal flips and Delaunay triangulations (Okabe, Boots and Sugihara, 1992).

LEMMA 10. *A triangulation that admits no legal flip is a Delaunay triangulation.*

We now outline the steps of the nearest neighbor search algorithm: first, use point location to determine the triangle T of $\mathcal{D}_n$ containing the query point X. Next, add edges from X to the three vertices of T. This forms a new triangulation. The only edges that can possibly admit legal flips are those with X as apex of the triangle. Perform all legal flips. By Lemma 10, this results in a Delaunay triangulation of $X, X_1, ..., X_n$. Since the Delaunay triangulation has the property that every vertex is adjacent to its nearest neighbor (Okabe, Boots and Sugihara, 1992), report the closest point adjacent to X as its nearest neighbor.

The complexity of the search is bounded by the time to locate the query point X in $\mathcal{D}_n$, the number of legal flips and the degree of X. By Theorem 2, the expected cost of the point location is $O(\sqrt{n})$. The number of legal flips is bounded by the degree of X since every legal flip results in adding an edge adjacent to X. Therefore, by Lemma 9, both the degree of X and the number of legal flips is $O(\sqrt{n})$. We conclude with the following:

THEOREM 3. *The expected complexity of determining the nearest neighbor of a given query point X when the data points are independent and uniformly distributed on a convex set C is $O(\sqrt{n})$. Here, worst-case refers to the placement of the query point, even after having seen the data $X_1, \ldots, X_n$.*

Range Queries. A RANGE QUERY refers to the following problem. Let S be a set of points in R^d and let Γ be a set of subsets of R^d. Each element of Γ is referred to as a *range*. Given a range $r \in \Gamma$, report all points in $r \cap S$. Many variations exist depending on the types of ranges and queries (see Agarwal (1997) or Matoušek (1999) for surveys). In this section, we restrict our attention to points in the plane. Assume that the Delaunay triangulation $\mathcal{D}_n$ for the data $X_1, \ldots, X_n$ is stored in the standard manner. We show that certain types of range queries can be solved simply and efficiently without additional pre-processing.

First, consider the case where the ranges are half-spaces, and the query is to report all of the data points lying in a query half-space. Theorem 1 immediately implies a simple algorithm whose expected running time is $O(\sqrt{n}+k)$ where k is the number of reported points. Let H be the query half-space with boundary h. Removing all of the edges of $\mathcal{D}_n$ that intersect h partitions $\mathcal{D}_n$ into two components. One of the components lies completely in H and the other lies completely outside H. To report all the data points in H, simply traverse the component of $\mathcal{D}_n$ in H in a depth-first or breadth-first manner.

The time required to partition $\mathcal{D}_n$ with respect to H is bounded by the number of edges intersecting h which is $O(\sqrt{n})$ by Theorem 1. The traversal of the component in H can be performed in $O(k)$ time where k is the size of the component.

Next, we consider the case where the ranges are axis-parallel rectangles and the query is to report all of the points in the rectangle. Once again, Theorem 1 implies a simple $O(\sqrt{n}+k)$ time algorithm. Let $R(a,b,c,d)$ be the query rectangle with vertices a,b,c,d in clockwise order. Perform a point location query to find which triangle of $\mathcal{D}_n$ contains a. Next, remove all edges that intersect an edge of the query rectangle. This can be done by walking in the triangulation around the boundary of the rectangle. Again, this partitions $\mathcal{D}_n$ into two components, one of which is completely in the query rectangle. Report all of the points by traversing this component. By Theorem 2, the expected cost of locating a in $\mathcal{D}_n$ is $O(\sqrt{n})$. By Theorem 1, the expected number of edges intersecting the boundary of the query rectangle is $O(\sqrt{n})$. We conclude with the following:

THEOREM 4. *The expected complexity of performing a half-space range query or an orthogonal range query or indeed any range query for an ℓ-gon with ℓ fixed, when the n data points are independent and uniformly distributed on a convex set C is $O(\sqrt{n}+\mathbb{E}K)$ where K is the number of reported points.*

Lazy Halfspace Range Search. In a LAZY HALFSPACE RANGE SEARCH, we are asked to report all points in a given halfspace $\mathcal{H}$, but are allowed to report these as a connected graph with a pointer to just one node. We assume that the Delaunay triangulation of the points is given. The cost of finding that triangulation is a one-time set-up cost. Given the line that defines $\mathcal{H}$, we can find all edges that are stabbed by the line in expected time $O(\sqrt{n})$ for uniform distributions on convex sets. It suffices to perform a point location for any point on that line, and then to walk to infinity from triangle to triangle in both directions. All the stabbed edges are removed from the Delaunay triangulation, and the appropriate remaining component is output.

Planar Separator. A PLANAR SEPARATOR is a set of vertices whose removal separates a graph into two subgraphs of roughly equal size. More specifically, a separator in a graph G, is a set S such each component of $G \setminus S$ has at most $2n/3$ vertices. Lipton and Tarjan (1979) were the first to show that every planar graph has an $O(\sqrt{n})$ separator (see also Pach and Agarwal, 1995). Planar separators have found many applications and are generally useful as they often lead to divide-and-conquer solutions to different problems on planar graphs (Lipton and Tarjan (1980), Leiserson (1983), Leighton (1983), Gilbert (1980), Gilbert and Tarjan (1987)).

We present a simple algorithm to compute an $O(\sqrt{n})$ separator of a Delaunay triangulation, $\mathcal{D}_n$. Let X_m be the X_i with median x-coordinate. Let S be the set of Delaunay vertices that has at least one adjacent edge intersecting the vertical line through X_m. The removal of S partitions $\mathcal{D}_n$ such that each component has size at most $n/2$. The set S can be computed in $O(n)$ time and by Theorem 1, S has expected size $O(\sqrt{n})$.

THEOREM 5. *A planar separator S with expected size $O(\sqrt{n})$ can be computed in $O(n)$ time when the data points are independent and uniformly distributed on a convex set C.*

Approximate Shortest Paths. In this subsection, we address the problem of APPROXIMATE SHORTEST PATH QUERIES in a Delaunay triangulation $\mathcal{D}_n$. Given a pair of vertices X_i and X_j, the goal is to quickly compute a path from X_i to X_j in $\mathcal{D}_n$ whose length is close to the actual shortest path. By using structural properties of the Delaunay triangulation, we show how to compute in expected $O(\sqrt{n})$ time, a path that is at most 5.08 times the Euclidean distance between X_i and X_j, and thus at most 5.08 times the actual shortest path.

Given the two query vertices, the first step is to locate one of the two vertices, say X_i, in $\mathcal{D}_n$ using point location. The next step is to compute a special subgraph of $\mathcal{D}_n$. Let S be the set of vertices having at least one adjacent edge intersecting the segment $[X_i, X_j]$. Let D be the subgraph of $\mathcal{D}_n$ induced by the set $S \cup \{X_i, X_j\}$. Bose and Morin (1999) modified an argument by Dobkin, Friedman and Supowit (1990) to show that the length of the shortest path between X_i and X_j in D is at most 5.08 times $||X_i - X_j||$.

We turn to the complexity of this algorithm. By Theorem 2, the point location step takes $O(\sqrt{n})$ time. By Theorem 1, the expected size of D is $O(\sqrt{n})$. Since D is a planar graph, computing the shortest path between two points can be performed simply using Dijkstra's algorithm (Cormen, Leiserson and Rivest, 1990) in $O(\sqrt{n}\log n)$ time or in $O(\sqrt{n})$ time using the slightly more complex algorithm of Klein, Rao, Rauch and Subramanian (1997). We conclude with the following:

THEOREM 6. *Let $\mathcal{D}_n$ be the Delaunay triangulation of n independent and uniformly distributed data points in a convex set C. In $O(\sqrt{n})$ expected time, given two of the data points X_i and X_j, a path between the two points of length at most 5.08 times $||X_i - X_j||$ can be computed.*

The diameter of a random Delaunay triangulation. The distance between two nodes in a graph is the minimal path distance between the two nodes. The diameter of a graph is the the maximum distance between any two nodes in a graph.

THEOREM 7. *Let $X_1, \ldots, X_n$ be i.i.d. and uniformly distributed in a convex region C with perimeter p and volume $v > 0$. Let Δ_n denote the diameter of the random Delaunay triangulation for $X_1, \ldots, X_n$. Then the bound of Theorem 1* applies to Δ_n as well. In particular, $\mathbb{E}\{\Delta_n\} = O(\sqrt{n})$.*

PROOF. Draw a line between points X_i and X_j, and note that the minimal path distance between X_i and X_j is less than the path distance between X_i and X_j if we are forced to only follow edges that are cut by the line segment $[X_i, X_j]$. There are at most S_n such edges, where S_n is the stabbing number, uniformly over all i, j. Thus, $\Delta_n \le S_n$, and the bound of Theorem 1* applies. □

Divide-and-conquer construction of the Delaunay triangulation. Using a hashing model of computation, we can construct the Delaunay triangulation of n points with a uniform distribution on a convex set C in expected time $O(n)$. Just consider the smallest rectangle R enclosing C, and, assuming that $n = 2^{2k}$ for some integer k, consider a $2^k \times 2^k$ regular grid partition of R. This partition can be regarded as a quadtrie, with R corresponding to the root. Place the n data points in the grid cells in $O(n)$ time. As each grid cell receives a binomial number of points with mean bounded by a constant, we can construct the Delaunay triangulations for all the grid cells individually by a simplistic quadratic algorithm in $O(1)$ expected time per cell. From the bottom of the trie upwards, we merge adjacent Delaunay triangulations in time bounded by the sum of the number of border points of the two triangulations (or, put differently, in time bounded by the stabbing number of the resulting triangulation). At every step, the expected time is bounded by the square root of the number of points involved in the merge operation. Thus, a recurrence for the total expected time T_n is roughly of the form $T_n \le 2T_{n/2} + O(\sqrt{n})$, which yields $T_n = O(n)$. The procedure is easy to implement. We recall here that the spiral method of Bentley, Weide and Yao (1980) also has $O(n)$ expected time, under the same distributional and computational models, but it appears a bit more complicated.

Lower bound for the stabbing number.

THEOREM 8. *Let S_n be the stabbing number for a cloud of n i.i.d. points distributed uniformly in a convex set C. Then there exists a positive constant c such that*

$$\mathbf{E}\{S_n\} \geq (c + o(1))\sqrt{n} \ .$$

PROOF. Omitted. □

§5. Appendix: Auxiliary results from probability theory

We need two tail inequalities. First, a rather standard tail bound for binomials will be used in the following format due to Angluin and Valiant (1979) (see also McDiarmid, 1998):

LEMMA A. *Let X be binomial (n, p). Then*

$$\mathbf{P}\{X \geq (1+u)np\} \leq e^{-u^2 np/3}$$

for all $u > 0$.

The next couple of symmetrization inequalities will be needed.

LEMMA B. *Let X, X' be i.i.d. random variables, and let m be a median of X. Then, for $u > 0$,*

$$\mathbf{P}\{|X - m| \geq u\} \leq 2\mathbf{P}\{|X - X'| \geq u\} \ .$$

PROOF. We have

$$\begin{aligned} \mathbf{P}\{|X - X'| \geq u\} &\geq \mathbf{P}\{X - m \geq u, X' - m \leq 0\} + \mathbf{P}\{X - m \leq -u, X' - m \geq 0\} \\ &\geq (1/2)\left(\mathbf{P}\{X - m \geq u\} + \mathbf{P}\{X - m \leq -u\}\right) \\ &= (1/2)\mathbf{P}\{|X - m| \geq u\} \ . \ \square \end{aligned}$$

LEMMA C. *Let X be an arbitrary nonnegative random variable, and let X' be an independent copy of it. Then, for $u.0$,*

$$\mathbf{P}\{X > 2\mathbf{E}\{X\} + u\} \leq 2\mathbf{P}\{|X - X'| \geq u\} \ .$$

PROOF. Assume without loss of generality that X has a unique median m. Then by Markov's inequality, $1/2 = \mathbf{P}\{X \geq m\} \leq \mathbf{E}\{X\}/m$. Thus, by Lemma B,

$$\mathbf{P}\{X > 2\mathbf{E}\{X\} + u\} \leq \mathbf{P}\{X > m + u\} \leq \mathbf{P}\{|X - m| \geq u\} \leq 2\mathbf{P}\{|X - X'| \geq u\} \ . \ \square$$

Finally, we obtain the first tail bound that relates general random functions $Y = Y_n = Y(X_1, \ldots, X_n)$ of i.i.d. random variables $X_1, \ldots, X_n$ to their mean.

LEMMA D. *Let $Y = Y_n = Y(X_1, \dots, X_n)$ be a nonnegative function of i.i.d. random variables $X_1, \dots, X_n$ and let the function be permutation invariant. Let X_1' be independent of the X_i's and distributed as X_1. Then, with*

$$V = Y_n(X_1, X_2, \dots, X_n) - Y_n(X_1', X_2, \dots, X_n) ,$$

we have, for $u, c > 0$,

$$\mathbf{P}\{Y \geq 2\mathbf{E}\{Y\} + u\} \leq 2n\mathbf{P}\{|V| \geq c\} + 4\exp\left(-\frac{u^2}{2nc^2}\right) .$$

Also, if

$$W = Y_n(X_1, X_2, \dots, X_n) - Y_{n-1}(X_2, \dots, X_n) ,$$

then, for $u > 0$,

$$\mathbf{P}\{Y \geq 2\mathbf{E}\{Y\} + u\} \leq 4n\mathbf{P}\{|W| \geq c/2\} + 4\exp\left(-\frac{u^2}{2nc^2}\right) .$$

Finally,

$$\mathbf{P}\{Y \geq 3\mathbf{E}\{Y\}\} \leq 4n\mathbf{P}\{|W| \geq c/2\} + 4\exp\left(-\frac{(\mathbf{E}\{Y\})^2}{2nc^2}\right) .$$

PROOF. By Lemma C, if Y_n' is an independent copy of Y_n,

$$\mathbf{P}\{Y \geq 2\mathbf{E}\{Y\} + u\} \leq 2\mathbf{P}\{|Y_n - Y_n'| \geq u\} .$$

Let $X_1, \dots, X_n$ and $Z_1, \dots, Z_n$ be i.i.d. sequences, and set

$$V_i = Y_n(Z_1, \dots, Z_{i-1}, X_i, X_{i+1}, \dots, X_n) - Y_n(Z_1, \dots, Z_{i-1}, Z_i, X_{i+1}, \dots, X_n) ,$$

so that

$$\sum_{i=1}^n V_i = Y_n(X_1, \dots, X_n) - Y_n(Z_1, \dots, Z_n) .$$

Clearly, the V_i's form a martingale difference sequence with respect to the filtration $(\mathcal{F}_n)$, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$ ($(\Omega, \mathcal{F}, \mathbf{P})$ is our probability space) and $\mathcal{F}_k = \sigma(X_1, \dots, X_k, Z_1, \dots, Z_k)$. If $\mathbf{E}_k V$ denotes the conditional expectation of a random variable V with respect to $\mathcal{F}_k$, then $\mathbf{E}_k V_{k+1} = 0$. Furthermore, given $\mathcal{F}_k$, the conditional distributions of V_{k+1} and $-V_{k+1}$ are identical. Then, by an extension of the Azuma-Hoeffding bounded difference inequality as reported in Godbole and Hitczenko (1998),

$$\mathbf{P}\left\{\left|\sum_{i=1}^n V_i\right| \geq u\right\} \leq \sum_{i=1}^n \mathbf{P}\left\{|V_i| > c\right\} + 2\exp\left(-\frac{u^2}{2nc^2}\right) .$$

As the V_i's are all distributed as V, the first part of the proof is complete. The last part follows from the triangle inequality $|V| \leq |W| + |W'|$, where $W' = Y_n(X_1', X_2, \dots, X_n) - Y_{n-1}(X_2, \dots, X_n)$ is distributed as W. □

Lemma D provides tail bounds if we know the mean of Y_n and have a tail bound for $\mathbf{P}\{|W| \geq u\}$.

§6. References

P. K. Agarwal, "Range searching," in: *Handbook of Discrete and Computational Geometry*, (edited by J. E. Goodman and J. O'Rourke), pp. 575–598, CRC Press, Boca Raton, FL, 1997.

N. Alon, J. Spencer, and P. Erdös, *The Probabilistic Method*, John Wiley, New York, 1992.

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

K. Azuma, "Weighted sums of certain dependent random variables," *Tohoku Mathematical Journal*, vol. 37, pp. 357–367, 1967.

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.

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. Bose and P. Morin, "Online routing in triangulations," *International Symposium on Algorithms and Computation*, pp. 113–122, 1999.

T. K. Chalker, A. P. Godbole, P. Hitczenko, J. Radcliff, and O. G. Ruehr, "On the size of a random sphere of influence graph," *Advances in Applied Probability*, vol. 31, 1999.

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

L. Devroye, C. Lemaire, and J.-M. Moreau, "Expected time analysis for Delaunay point location," *Computational Geometry Theoiry and Applications*, vol. 29, pp. 61–89, 2004.

D. Dobkin, S. Friedman, and K. Supowit, "Delaunay graphs are almost as good as complete graphs," *Discrete Computational Geometry*, vol. 5, pp. 399–407, 1990.

J. Gilbert, "Graph Separator Theorems and Sparse Gaussian Elimination," Ph.D. Dissertation, Stanford University, 1980.

J. Gilbert and R. Tarjan, "The analysis of a nested dissection algorithm," *Numerische Mathematik*, vol. 50, pp. 377–404, 1987.

A. P. Godbole and P. Hitczenko, "Beyond the method of bounded differences," in: *Microsurveys in Discrete Probability*, vol. 41, pp. 43–58, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, American Mathematical Society, Providence, RI, 1998.

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.

W. Hoeffding, "Probability inequalities for sums of bounded random variables," *Journal of the American Statistical Association*, vol. 58, pp. 13–30, 1963.

R. M. Karp, "The transitive closure of a random digraph," *Random Structures and Algorithms*, vol. 1, pp. 73–93, 1990.

P. Klein, S. Rao, M. Rauch, and S. Subramanian, "Faster shortest-path algorithms for planar graphs," *Journal of Computer and System Sciences*, vol. 55, pp. 3–23, 1997.

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

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

T. Leighton, *Complexity Issues in VLSI*, MIT Press, Boston, 1983.

C. Leiserson, *Area Efficient VLSI Computation*, MIT Press, Boston, 1983.

R. J. Lipton and R. E. Tarjan, "A separator theorem for planar graphs," *SIAM Journal of Applied Mathematics*, vol. 36, pp. 177–189, 1979.

R. J. Lipton and R. E. Tarjan, "Applications of a planar separator theorem," *SIAM Journal on Computing*, vol. 9, pp. 615–627, 1980.

J. Matoušek, "Geometric range searching," *ACM Computing Surveys*, vol. 26, pp. 421–461, 1999.

C. McDiarmid, "On the method of bounded differences," in: *Surveys in Combinatorics 1989*, vol. 141, pp. 148–188, London Mathematical Society Lecture Notes Series, Cambridge University Press, Cambridge, 1989.

C. McDiarmid, "Concentration," in: *Probabilistic Methods for Algorithmic Discrete Mathematics*, (edited by M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed), vol. 16, pp. 195–248, Springer Seriues on Algorithms and Combinatorics, Springer-Verlag, Berlin, 1998.

B. McKay and N. Wormald, "The degree sequence of a random graph, I. The models," *Random Structures and Algorithms*, vol. 11, pp. 97–117, 1997.

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

J. O'Rourke, *Computational Geometry in C (2nd edition)*, Cambridge University Press, Cambridge, 1998.

J. Pach and P. K. Agarwal, *Combinatorial Geometry*, John Wiley, New York, 1995.

E. Shamir and J. Spencer, "Sharp concentration of the chromatic number on random graphs $G_{n,p}$," *Combinatorica*, vol. 7, pp. 121–129, 1987.

M. Smid, "Closest-point problems in computational geometry," in: *Handbook of Computational Geometry*, (edited by J. E. Goodman and J. O'Rourke), pp. 877–937, Elsevier, 2000.

J. Snoeyink, "Point location," in: *Handbook of Discrete and Computational Geometry*, (edited by J. E. Goodman and J. O'Rourke), pp. 559–574, CRC Press, Boca Raton, FL, 1997.