18.409 The Behavior of Algorithms in Practice	4/18/02 and 4/23/02

Lecture 15/16

Lecturer: Dan Spielman *Scribe: Mikhail Alekhnovitch*

Let $a_1, ..., a_n$ be independent random gaussian points in the plane with variance 1. Today lecture is devoted to the following question: what is the expected size of their convex hull?

Theorem 1 (Renyi-Salanke).

$$\mathop{\mathrm{E}}_{a_1,\ldots,a_n}[\text{size of C.H.}] = \Theta(\sqrt{\lg n}).$$

In this lecture we will prove a weaker bound $O(\lg^2 n)$. First we notice, that

$$\Pr[\bar{o} \notin \text{C.H.}] \leq \left(\frac{3}{4}\right)^{\frac{n}{3}}.$$

This is because the probability that $\bar{o} \notin$ C.H. of three points is exactly 3/4, so we can divide all points into $n/3$ groups of 3, and each group covers o with probability 1/4. Thus, with exponentially high probability $\bar{o} \in C.H.$ so we can assume for the rest of the lecture that this is always the case.

For a vector z consider the edge (a_j, a_k) of the convex hull that crosses z clockwise. Denote

$$P_z(\epsilon) = \Pr[ang(z\bar{o}a_k)) < \epsilon].$$

Then clearly

$$\text{E}[\text{size of C.H.}] \leq \lim_{\epsilon \to 0} \frac{2\pi}{\epsilon} P_z(\epsilon) + n\left(\frac{3}{4}\right)^{\frac{n}{3}}.$$

Denote by $CH_{j,k}$ the event that (a_j, a_k) is an edge of $CH(a_1, ..., a_n)$ and other points lie on the origin side of the line $a_j a_k$. For a fixed vector z, $Cross_{jk}$ is the event that the edge (a_j, a_k) crosses z clockwise.

$$P_z(\epsilon) = \sum_{j,k} \Pr\left[CH_{j,k} \wedge Cross_{j,k}\right] \cdot \Pr\left[ang(z\bar{o}a_k) < \epsilon | CH_{j,k} \wedge Cross_{j,k}\right] =$$

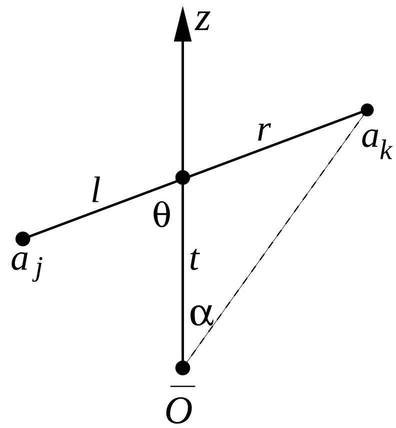

Figure 1:

$$= \Pr\left[ang(z\bar{o}a_k) < \epsilon | CH_{j,k} \wedge Cross_{j,k} \right]$$

for any choice of j and k (the latter equation follows because of the symmetry). Assume that (a_j, a_k) crosses z. It will be convenient to choose the coordinates θ, t, l, r instead of a_j, a_k (see figure 1). Let $\alpha = ang(zoa_k)$. Then the probability $P_z(\epsilon)$ can be expressed as:

$$\frac{\int\limits_{t,\theta} \left(\int\limits_{i \neq j,k} [CH_{j,k}] \right) \cdot \int\limits_{l,r \geq 0} \left([\alpha < \epsilon] \right) (l+r) sin(\theta) \mu(a_j) \mu(a_k) \, d\theta \, dt \, dl \, dr}{\int\limits_{t,\theta} \left(\int\limits_{i \neq j,k} [CH_{j,k}] \right) \cdot \int\limits_{l,r \geq 0} (l+r) sin(\theta) \mu(a_j) \mu(a_k) \, d\theta \, dt \, dl \, dr}$$

We need the following claim that estimates the maximal norm of n gaussian points in the plane.

Claim 2.

$$\Pr\left[\max_i ||a_i|| > \sqrt{8 \lg n} \right] < \frac{1}{n}.$$

In the assumption of the claim, we can bound $t \leq \sqrt{8 \lg n}$; $r, l \leq 2\sqrt{8 \lg n}$. Once again we can assume that this is always the case (it can change the expectation at most by 1). When α is sufficiently small,

$$\alpha > \frac{1}{2} \tan(\alpha) = \frac{1}{2} \cdot \frac{r \sin(\theta)}{t + r \cos(\theta)} \geq \frac{r \sin(\theta)}{6\sqrt{8 \lg n}}.$$

Thus

$$\mathrm{E}[\text{size of C.H.}] \le \lim_{\epsilon\to 0} \frac{2\pi}{\epsilon} P_z(\epsilon) + n\left(\frac{3}{4}\right)^{\frac{n}{3}} \le \lim_{\epsilon\to 0} \frac{2\pi}{\epsilon} \Pr\left[\frac{r\sin(\theta)}{6\sqrt{8\lg n}} < \epsilon\right] + n\left(\frac{3}{4}\right)^{\frac{n}{3}} + 1.$$

We estimate the latter probability using the Combination Lemma from Lecture 19. Namely, we show that

$\Pr[r < \epsilon] < O(\sqrt{\lg n} \cdot \epsilon)$

$\Pr[\sin(\theta) < \epsilon] = O(\lg n \cdot \epsilon^2)$.

Thus, the following two lemmas imply the theorem:

Lemma 1. $\forall t \le \sqrt{8\lg n}$

$$\frac{\int_{r\ge 0} [r<\epsilon](l+r)\mu(a_k)\,dr}{\int_{r\ge 0} (l+r)\mu(a_k)\,dr} \le O(\sqrt{\lg n}\cdot\epsilon) \tag{1}$$

Lemma 2. $\forall t \le \sqrt{8\lg n}$, $l, r \le 2\sqrt{8\lg n}$

$$\frac{\int_{\theta} [\sin(\theta)\le\epsilon]\left(\int_{i\ne j,k} [CH_{j,k}]\right)\sin(\theta)\mu(a_j)\mu(a_k)}{\int_{\theta}\left(\int_{i\ne j,k} [CH_{j,k}]\right)\sin(\theta)\mu(a_j)\mu(a_k)} \le O((\lg n\cdot\epsilon)^2) \tag{2}$$