The k-set problem

Michiel Smid*

September 23, 2003

Let S be a set of n points in the plane. We assume that S is in *general position*, in the sense that no two points of S are on a vertical line and no three points of S are collinear. Let S' be a subset of S and let k be an integer with $0 \leq k \leq n-2$.

1. We say that S' is a k-*set* if $|S'| = k$ and there are two distinct points q and q' in S such that $S \cap B(q,q') = S'$, where $B(q,q')$ is the set of all points in $\mathbb{R}^2$ that are strictly below the line through q and q'.

2. We say that S' is a $(\leq k)$-*set* if S' is an ℓ-set for some integer ℓ with $0 \leq \ell \leq k$.

One of the most tantalizing open problems in combinatorial geometry is to determine the maximum number of k-sets that the set S can contain. Dey [2] has shown that the number of k-sets in any set of n points in the plane is $O(nk^{1/3})$. This is the currently best known upper bound. Edelsbrunner [3] gives an example of a set of n points in the plane containing $\Omega(n \log k)$ k-sets. Toth [4] presents a contruction of a set of n points in the plane with $n \cdot 2^{\Omega(\sqrt{\log k})}$ k-sets, for any n and $k < n/2$, which is the currently best known lower bound.

A simpler problem is to determine the maximum number of $(\leq k)$-sets.

Exercise 1 Let S be a set of n points on the lower half of a circle and let k be an integer with $1 \leq k \leq n-2$. Prove that S contains at least ckn $(\leq k)$-sets, where c is a constant that does not depend on k and n.

*School of Computer Science, Carleton University, Ottawa, Ontario, Canada K1S 5B6. E-mail: `michiel@scs.carleton.ca`.

In the rest of this note, we will prove that the lower bound in Exercise 1 is tight. The proof is due to Clarkson and Shor [1].

Let $S = \{q_1, q_2, \ldots, q_n\}$. For $1 \leq i < j \leq n$, let $n(i,j)$ denote the number of points of S that are strictly below the line through q_i and q_j. For $0 \leq k \leq n-2$, let f_k denote the number of k-sets, i.e.,

$$f_k = |\{(i,j) : 1 \leq i < j \leq n \text{ and } n(i,j) = k\}|,$$

and let g_k denote the number of $(\leq k)$-sets, i.e.,

$$g_k = f_0 + f_1 + \ldots + f_k.$$

Hence, our goal is to prove an upper bound on g_k.

We fix an integer k with $2 \leq k \leq n-2$, and a real number p with $0 < p < 1$. Let R be a random sample of S obtained by choosing each point of S with probability p and independently of the other points. Let X be the random variable whose value is equal to the number of edges on the lower hull of R.

We will prove upper and lower bounds on the expected value $E(X)$ of X. By combining these bounds, we will get an upper bound on g_k.

The upper bound is easy: Since $X \leq |R|$, we have

$$E(X) \leq E(|R|) = pn.$$

So it remains to prove a lower bound on $E(X)$. For $1 \leq i < j \leq n$, let X_{ij} denote the random variable whose value is one if $q_i q_j$ is an edge of the lower hull of R, and zero otherwise. Then $X = \sum_{i,j} X_{ij}$. Also, $X_{ij} = 1$ if and only if q_i and q_j are elements of R and none of the $n(i,j)$ points of S that are strictly below the line through q_i and q_j is contained in R. Therefore,

$$E(X_{ij}) = \Pr(X_{ij} = 1) = p^2 (1-p)^{n(i,j)}.$$

Using the linearity of expectation, we obtain

$$\begin{aligned} E(X) &= \sum_{1 \leq i < j \leq n} E(X_{ij}) \\ &= \sum_{1 \leq i < j \leq n} p^2 (1-p)^{n(i,j)} \\ &= \sum_{\ell=0}^{n-2} f_\ell \, p^2 (1-p)^\ell \end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{\ell=0}^{k} f_\ell\, p^2 (1-p)^\ell \\
&\geq \sum_{\ell=0}^{k} f_\ell\, p^2 (1-p)^k \\
&= g_k\, p^2 (1-p)^k.
\end{aligned}
$$

By combining the upper and lower bounds on $E(X)$, we obtain

$$g_k \leq \frac{pn}{p^2(1-p)^k} = \frac{n}{p(1-p)^k}.$$

Observe that this upper bound holds for all p with $0 < p < 1$. For $p = 1/k$, we get

$$g_k \leq kn(1-1/k)^{-k} \leq ckn,$$

for some constant c that does not depend on k and n.

Theorem 1 *Let S be a set of n points in the plane that is in general position and let k be an integer with $1 \leq k \leq n-2$. The number of $(\leq k)$-sets in S is at most ckn, for some constant c that does not depend on k and n.*

Exercise 2 We have proved Theorem 1 for $2 \leq k \leq n-2$. Show that the theorem also holds for $k = 1$.

References

[1] K. L. Clarkson and P. W. Shor. Applications of random sampling in computational geometry, II. *Discrete Comput. Geom.*, 4:387–421, 1989.

[2] T. K. Dey. Improved bounds on planar k-sets and related problems. *Discrete Comput. Geom.*, 19:373–382, 1998.

[3] H. Edelsbrunner. *Algorithms in Combinatorial Geometry.* Springer-Verlag, Heidelberg, Germany, 1987.

[4] G. Toth. Point sets with many k-sets. *Discrete Comput. Geom.*, 26:187–194, 2001.