Discrete Comput Geom 7:109–123 (1992)

Discrete & Computational
Geometry

An Upper Bound on the Number of Planar *K*-Sets*

János Pach,[1,2] William Steiger,[3] and Endre Szemerédi,[3,1]

[1] Mathematical Institute, Hungarian Academy of Science, H-1364 Budapest, Hungary

[2] Courant Institue, New York University, New York, NY 10012, USA

[3] Computer Science Department, Rutgers University, New Brunswick, NJ 08903, USA

Abstract. Given a set S of n points, a subset X of size k is called a *k-set* if there is a hyperplane Π that separates X from $S - X$. We prove that $O(n\sqrt{k}/\log^* k)$ is an upper bound for the number of k-sets in the plane, thus improving the previous bound of Erdös, Lovász, Simmons, and Strauss by a factor of $\log^* k$.

1. Introduction and Summary

Let $S = \{x_1, \ldots, x_n\}$ denote n points in $\Re^d$. A subset X of size k is called a k-set if there is a hyperplane Π that separates X from $S - X$. The interest centers on $e_d(k; n)$, the maximal number of k-sets over all configurations S of size n in $\Re^d$.

Upper and lower bounds for $e_2(k; n)$ were obtained by Lovász [12] for halving sets (n even, $k = n/2$), and later, for general $k \leq n/2$, by Erdös *et al.* [10]. A simple construction gives a set S with $n \log(k + 1)$ k-sets, while a counting argument shows that $e_2(k; n) = O(n\sqrt{k})$. These bounds were rediscovered by Edelsbrunner and Welzl [9] but have not been improved. The papers [1], [11], and [14] contain results related to the study of $e_2(k; n)$.

Raimund Seidel (see [8]) extended the lower bound construction to $d = 3$ and showed that $e_3(k; n) = \Omega(nk \log(k + 1))$. Clearly, $e_3(k; n) = O(n^3)$. The upper bound

* The research of J. Pach was supported in part by NSF Grant CCR-8901484 and by Grant OTKA-1418 from the Hungarian Foundation for Scientific Research. The research of W. Steiger and E. Szemerédi was supported in part by NSF Grant CCR-8902522. All authors express gratitude to the NSF DIMACS Center at Rutgers.

has been improved when $k = o(n)$. Cole *et al.* [7] proved that $e_3(k; n) = O(n^2k)$ and Chazelle and Preparata [3] proved that $e_3(k; n) = O(nk^5)$. Clarkson [5], [6] has also obtained better bounds in restrictive cases using random sampling methods. For example, he showed that $e_3(k; n) = O(nk^2)$. An interesting recent result of Bárány *et al.* [2] improved the upper bound on $e_3(n/2; n)$ to $n^{3-\varepsilon}$, where $\varepsilon > 0$ is some small constant. This, in turn, was improved by Chazelle *et al.* [4] to $O(n^{8/3+\delta})$, for any $\delta > 0$.

Finally, it is worth pointing out that k-sets have been applied in computer science, e.g., in the papers [3], [7], and [9] cited above. It is likely that further applications will be found. However, the real interest remains in understanding $e_d(k; n)$. Since the original Lovász result in 1971, several attempts have been made to reduce the upper bound for $e_2(k; n)$; these attempts have either failed or merely succeeded in devising new proofs. The difficulty may partly explain why the question has been so tantalizing.

This paper presents the first, modest improvement in the planar upper bound (see also [13]). Specifically we show that

Theorem. $e_2(k; n) = O(n\sqrt{k}/\log^* k)$.

Without loss of generality, we assume that the points are in general position. As in [12], we study the *dissection graph* $\vec{G}_k$ of S. The vertices of $\vec{G}_k$ are the points of S and the edges are the ordered pairs $\overrightarrow{xy}$ corresponding to directed segments $\overrightarrow{xy}$, $x, y \in S$, such that the directed line through x and y has k points in the open half-space to its right. These are the *k-segments*. Note that $\vec{G}_{n-k-2}$ has the same edges as $\vec{G}_k$ but with opposite orientation. Obviously it is enough to bound $|E(\vec{G}_k)|$.

Given a configuration S with at least $\varepsilon n\sqrt{k}$ edges in its dissection graph, we derive a contradiction to the *intersection lemma* that was proved in [12]:

Proposition 1 (Lovász). *Given a directed line l which is not incident with any point of S, the number of k-segments that cross l from left to right is at most $k + 1$.*

The contradiction is achieved by first extracting a large subgraph $\vec{G}_k^*$ (at least $\varepsilon_1 n\sqrt{k}$ edges) which is the union of disjoint, bipartite blocks of size $C_1\sqrt{k}$. (Throughout, ε_i and C_i denote small and large constants, respectively, which are polynomials in ε and $1/\varepsilon$, respectively.) The key result is Lemma 2 in Section 3. Its proof is based on a geometric fact and a careful combinatorial analysis of a single block in $\vec{G}_k^*$. This enables us to find a lot of edges in the graph $\vec{G}_{k\pm j^2} = \bigcup_{i=k-j^2}^{k+j^2} \vec{G}_i$ for any $j \ll k$. In Section 4 we apply this fact for a long sequence of integers $j_1 \ll j_2 \ll j_3 \ll \cdots \ll k$ to obtain a natural number k' and a straight line l intersecting more than $3k'$ edges of $\vec{G}_{k'}$. This contradicts the intersection lemma and completes the proof. A straightforward evaluation of the constants ε_i, C_i reveals that the contradiction will occur whenever $\varepsilon > C/\log^* k$. In fact our method can be extended to the case $\varepsilon > C/(\log k)^\varepsilon$ but the counting is more complicated. An interesting feature of our argument is the slightly stronger blend of geometry and combinatorics that we have been able to use.

2. Preprocessing

Let $\vec{G}_k$ be the dissection graph of S. In order to obtain a contradiction we suppose that, for some $\varepsilon > 0$, $|E(\vec{G}_k)| > \varepsilon n\sqrt{k}$. In this section we simplify the picture of $\vec{G}_k$ by extracting a large subgraph $\vec{G}_k^*$ with a very regular structure.

We place a system $L = \{l_1, \ldots, l_{n-1}\}$ of $n-1$ parallel (say *vertical*) lines over the configuration in such a way that between each adjacent pair of lines there is exactly one point of S. Lines l_i and l_{i+j} are said to be at distance j. The *length* $\|\overrightarrow{xy}\|$ of the edge $xy \in E(\vec{G}_k)$ is the number of lines in the system that intersect xy. By Proposition 1, $2(n-1)(k+1)$ is an upper bound on the number of intersections of edges in $\vec{G}_k$ and lines in L. This fact helps establish the following result.

Lemma 1. *There exist positive constants ε_1 and C_1 depending on ε only (i.e., independent of n, k, and S) such that there is a subgraph $\vec{G}_k^* \subseteq \vec{G}_k$ which splits into disjoint blocks $\vec{B}_1 \cup \vec{B}_2 \cup \cdots \cup \vec{B}_m$ satisfying the following conditions*:

(1) $|E(\vec{G}_k^*)| = \sum_{i=1}^{m} |E(\vec{B}_i)| \geq \varepsilon_1 n\sqrt{k}$.

(2) $|V(\vec{B}_i)| = \lfloor C_1\sqrt{k}\rfloor$ *for every* $1 < i < m$; $|V(\vec{B}_1)|$ *and* $|V(\vec{B}_m)| \leq \lfloor C_1\sqrt{k}\rfloor$, *and the vertices of each $\vec{B}_i$ are consective elements in the ordering of S determined by the x-coordinates of the points.*

(3) *All edges of $\vec{B}_i$ are directed from the left to the right, and all of them intersect a specific vertical line $l_{j_i} \in L$ $(1 \leq i \leq m)$.*

(4) *Every vertex of $\vec{B}_i$ is either isolated or incident to at least $\varepsilon_1\sqrt{k}$ edges of $\vec{B}_i$ $(1 \leq i \leq m)$.*

Proof. We can assume without loss of generality that the subgraph $\vec{G}'_k \subseteq \vec{G}_k$, formed by all edges of $\vec{G}_k$ going from left to right, has at least $(\varepsilon/2)n\sqrt{k}$ edges. Proposition 1 implies that most edges in $\vec{G}'_k$ are short, i.e., of length less that $5\sqrt{k}/\varepsilon$. This follows from

$$\left|\left\{\overrightarrow{xy} \in \vec{G}'_k \colon \|\overrightarrow{xy}\| \geq \frac{5}{\varepsilon}\sqrt{k}\right\}\right| \left(\frac{5}{\varepsilon}\sqrt{k}\right) \leq \sum_{\overrightarrow{xy} \in E(\vec{G}'_k)} \|\overrightarrow{xy}\|$$
$$\leq (n-1)(k+1).$$

Deleting all edges of $\vec{G}'_k$ whose length is at least $(5/\varepsilon)\sqrt{k}$, we obtain a subgraph $\vec{G}''_k \subseteq \vec{G}'_k$, and

$$|E(\vec{G}''_k)| \geq \frac{\varepsilon}{4} n\sqrt{k}.$$

For any $p_i \in S$ let $d^-(p_i)$ denote the outdegree of p_i in $\vec{G}''_k$. The total number of intersections between the edges of $\vec{G}_k''$ and the lines of L is

$$\sum_{\vec{xy} \in E(\vec{G}_k'')} \|\vec{xy}\| \geq \sum_{i=1}^{n} \binom{d^-(p_i)+1}{2} \geq \frac{1}{2n}\left(\sum_{i=1}^{n} d^-(p_i)\right)^2$$
$$\geq \frac{1}{2n}|E(\vec{G}_k'')|^2 \geq \frac{\varepsilon^2}{32} nk. \tag{1}$$

This fact implies that we may find a subsystem L' of L consisting of lines at a distance $\Delta = \lfloor 10\sqrt{k}/\varepsilon \rfloor$ from each other, which intersect at least $\varepsilon^3 n\sqrt{k}/320$ edges of $\vec{G}_k''$, as follows. Define the subsystems

$$L_t = \{l_t, l_{t+\Delta}, l_{t+2\Delta}, \ldots\}, \qquad t = 1, \ldots, \Delta. \tag{2}$$

No edge of $\vec{G}_k''$ can intersect more than one line of L_t. On the other hand, we can fix a t so that the number of edges of $\vec{G}_k''$ intersected by some element of L_t is at least

$$\frac{\varepsilon^2 nk}{32\Delta} \geq \frac{\varepsilon^3}{320} n\sqrt{k}.$$

Let $\vec{G}_k''' \subseteq \vec{G}_k''$ denote the graph consisting of these edges. Finally, let $\vec{G}_k^*$ be the graph obtained from $\vec{G}_k'''$ by sequentially removing all edges incident to at least one vertex whose indegree or outdegree (in $\vec{G}_k'''$) is smaller than $\varepsilon^2\sqrt{k}/640$. Clearly,

$$|E(\vec{G}_k^*)| \geq |E(\vec{G}_k''')| - n\frac{\varepsilon^3}{640}\sqrt{k} \geq \frac{\varepsilon^3}{640} n\sqrt{k}.$$

$\vec{G}_k^*$ satisfies all the requirements of the lemma with $\varepsilon_1 = \varepsilon^3/640$, $C_1 = 10/\varepsilon$. $\vec{B}_i$ denotes the subgraph of $\vec{G}_k^*$ induced by those points which are between $l_{t+(i-1)\Delta-\lfloor\Delta/2\rfloor}$ and $l_{t+(i-1)\Delta+\lfloor\Delta/2\rfloor}$, and all edges of $\vec{B}_i$ cross $l_{t+(i-1)\Delta}$. □

The graph $\vec{G}_k^*$ is represented in Fig. 1.

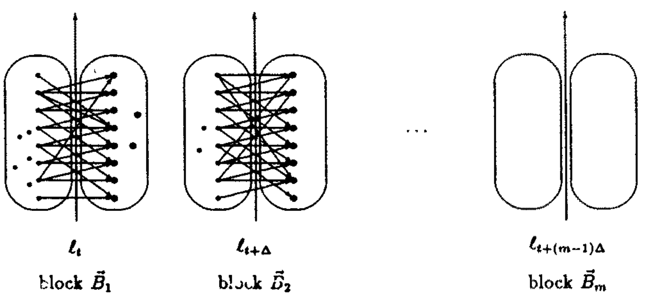

Fig. 1. The graph $\vec{G}_k^*$, a union of independent, nearly complete bipartite blocks.

3. Analysis of the Blocks

Let $\vec{B}$ be a *nonempty* block in $\vec{G}_k^*$, i.e., $\vec{B}$ has at least one nonisolated vertex and satisfies Lemma 1. Therefore its vertex set splits into $Q = \{q_1, q_2, \ldots\}$ and $R = \{r_1, r_2, \ldots\}$, at most $C_1\sqrt{k}$-element subsets of S which are separated from each other by a vertical line l. Any nonisolated vertex q_i is connected to at least $\varepsilon_1\sqrt{k}$ elements of R. Let $r_0^{(i)}, r_1^{(i)}, \ldots$ denote these elements in clockwise order. Similarly, $q_0^{(i)}, q_1^{(i)}, \ldots \in Q$ denotes the neighbors of r_i in clockwise order.

Fix an integer j. The convex cone $r_{(m-1)j}^{(i)} q_i r_{mj}^{(i)}$ with apex at q_i is called a *wedge of size* j and is denoted by $W_m^{(i)}$, $m = 1, 2, \ldots$. Similarly, we define wedges meeting at a point $r_i \in R$ by numbering the neighbors of r_i in Q, and dividing them into groups of size j. Figure 2 shows a typical block and helps understand the following statement.

Lemma 2. *There exists a constant $\varepsilon_2 > 0$ depending only on ε_1 and C_1 (and thus only on ε) such that, for any $j > 0$,*

(1) *either at least $\varepsilon_2\sqrt{k}$ points $q_i \in Q$ have the property that at least $\varepsilon_2\sqrt{k}/j$ wedges (of size j) meeting at q_i contain at least j^2 elements of S in their interiors;*
(2) *or the analogous statement holds for at least $\varepsilon_2\sqrt{k}$ points $r_i \in R$.*

The proof is based on the following geometric observation. Consider all (the at least $\varepsilon_1^2 k$) edges in $\vec{B}$, and write them in order of decreasing slopes. Thus we have

$$E(\vec{B}) = \{\vec{e}_1, \vec{e}_2, \ldots\}.$$

Claim 1. *If two edges $\overrightarrow{q_i r_g^{(i)}}$ and $\overrightarrow{q_i r_h^{(i)}}$ (or $\overrightarrow{q_g^{(i)} r_i}$ and $\overrightarrow{q_h^{(i)} r_i}$) are more than d apart in the ordering of $E(\vec{B})$, then there are at least d points of S (not necessarily in $\vec{B}$) in the interior of the cone $r_g^{(i)} q_i r_h^{(i)}$ (or $q_g^{(i)} r_i q_h^{(i)}$).*

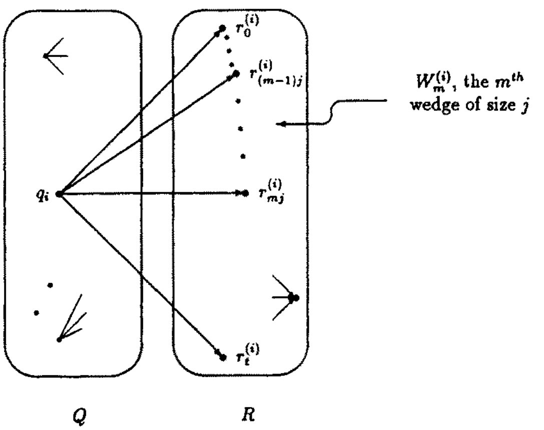

Fig. 2. A typical block.

Proof. Let $S^+(l)$ and $S(\overrightarrow{q_i r}_g^{(i)})$ denote the set of those elements of S that are in the open half-plane to the right of l and $\overrightarrow{q_i r}_g^{(i)}$, respectively ($l$ is the vertical line separating Q and R). The proof of Proposition 1 (see [12]) implies that if $\vec{e}$ and $\vec{e}'$ are *exactly* d' apart in the clockwise ordering of *all* edges of $\vec{G}_k$ that cross l from left to right, then

$$\|S^+(l) \cap S(\vec{e})| - |S^+(l) \cap S(\vec{e}')\| = d'.$$

Thus, for every pair $\vec{e}_a, \vec{e}_b \in E(\vec{B})$, $b > a$,

$$|S^+(l) \cap S(\vec{e}_a)| - |S^+(l) \cap S(\vec{e}_b)| \geq b - a.$$

Hence, if $h > g + d$, then

$$|S^+(l) \cap S(\overrightarrow{q_i r_g^{(i)}})| - |S^+(l) \cap S(\overrightarrow{q_i r_h^{(i)}})| \geq d + 1,$$

which implies that

$$|S^+(l) \cap (S(\overrightarrow{q_i r_g^{(i)}}) - S(\overrightarrow{q_i r_h^{(i)}}))| \geq d + 1. \qquad \square$$

Note that the number of points of S in the angular region $r_g^{(i)} q_i r_h^{(i)}$ is exactly the same as in its vertical angle (its reflection about q_i), because $\overrightarrow{q_i r_g^{(i)}}$ and $\overrightarrow{q_i r_h^{(i)}}$ are k-segments.

Now partition the edges in $E(\vec{B})$ into consecutive groups with $5j^2$ edges in each group:

$$E_i = \{\vec{e}_{5(i-1)j^2+1}, \ldots, \vec{e}_{5ij^2}\}, \qquad i = 1, 2, \ldots. \tag{3}$$

For any edge $\vec{e}_t$ of $\vec{B}$, let $q(\vec{e}_t)$ denote its initial point in Q, and let $r(\vec{e}_t)$ denote its endpoint in R. An edge $\vec{e} \in E_i$ is called *q-regular* if

$$|\{\vec{e}_t \in E_i: q(\vec{e}_t) = q(\vec{e})\}| < 3j,$$

i.e., there are fewer than $3j$ edges in E_i whose starting point is $q(\vec{e})$. Similarly, $\vec{e}$ is called *r-regular* if

$$|\{\vec{e}_t \in E_i: r(\vec{e}_t) = r(\vec{e})\}| < 3j.$$

An edge is called *irregular* if it is neither q-regular nor r-regular.

Claim 2. *Every group E_i contains at least j^2 q-regular edges or at least j^2 r-regular edges.*

Proof. Let I denote the set of irregular edges in E_i and

$$A = \sum_{\vec{e} \in E_i} |\{\vec{e}_s \in E_i: q(\vec{e}_s) = q(\vec{e})\}| \times |\{\vec{e}_t \in E_i: r(\vec{e}_t) = r(\vec{e})\}|.$$

Clearly,

$$A = \sum_{q \in Q} |\{\vec{e}_s \in E_i\colon q(\vec{e}_s) = q\}| \sum_{\vec{qr} \in E_i} |\{\vec{e}_t \in E_i\colon r(\vec{e}_t) = r\}|$$
$$\leq \left(\sum_{q \in Q} |\{\vec{e}_s \in E_i\colon q(\vec{e}_s) = q\}| \right)\left(\sum_{r \in R} |\{\vec{e}_t \in E_i\colon r(\vec{e}_t) = r\}| \right),$$

which gives

$$A \leq |E_i|^2 = 25j^4.$$

On the other hand,

$$A \geq |I| 9j^2.$$

These inequalities combine to show that the number of irregular edges in E_i is at most $25j^2/9 < 3j^2$. The result now follows because $|E_i| = 5j^2$. □

If there is a q-regular edge contained within a wedge of size j, then we can infer that many elements of S are in this, or nearby wedges.

Claim 3. *Let $W_m^{(i)} = r_{(m-1)j}^{(i)} q_i r_{mj}^{(i)}$ be a wedge of size j with apex $q_i \in Q$, and suppose that there is a q-regular edge $\overrightarrow{q_i r_u^{(i)}}$ in $W_m^{(i)}$. Then at least one of the seven consecutive wedges $W_{m-3}^{(i)}, W_{m-2}^{(i)}, \ldots, W_{m+3}^{(i)}$ has at least j^2 elements of S in its interior. (A similar statement is valid for wedges with apex at points of R, and containing an r-regular edge.)*

Proof. Suppose that $\overrightarrow{q_i r_u^{(i)}}$ belongs to the group $E_s = \{\vec{e}_{5(s-1)j^2+1}, \ldots, \vec{e}_{5sj^2}\}$. If neither $\overrightarrow{q_i r_{(m-1)j}^{(i)}}$ nor $\overrightarrow{q_i r_{mj}^{(i)}}$ is contained in E_s, then they are more than $5j^2$ apart in the clockwise order of all edges of $\vec{B}$, and Claim 1 implies that $W_m^{(i)}$ has at least $5j^2$ points of S in its interior. In the opposite case, let $[\mu, \nu]$ be the *maximal* interval of integers such that $\overrightarrow{q_i r_{xj}^{(i)}} \in E_s$ for every $x \in [\mu, \nu]$. Then $\nu - \mu < 3$ by the q-regularity of $\overrightarrow{q_i r_u^{(i)}}$. On the other hand, by the maximality of $[\mu, \nu]$, $\overrightarrow{q_i r_{(\mu-1)j}^{(i)}}$ and $\overrightarrow{q_i r_{(\nu+1)j}^{(i)}}$ are at least $5j^2$ apart in $E(\vec{B})$. Hence, $\overrightarrow{q_i r_{(x-1)j}^{(i)}}$ and $\overrightarrow{q_i r_{xj}^{(i)}}$ are at least $5j^2/4 > j^2$ apart for some $x \in [\mu, \nu + 1]$ and, again, Claim 1 implies that the wedge $W_x^{(i)}$ has at least j^2 elements of S in its interior. □

Proof of Lemma 2. Claim 2 implies that at least half of the groups $E_i \subseteq E(\vec{B})$ contain at least j^2 (say) q-regular edges. Thus, the number of q-regular edges in $\vec{B}$ is at least

$$\frac{|E(\vec{B})|}{10} \geq \frac{|Q| \varepsilon_1 \sqrt{k}}{10} \geq \frac{\varepsilon_1^2}{10} k. \tag{4}$$

If condition (1) of Lemma 2 were false for some $\varepsilon_2 > 0$, fewer than $\varepsilon_2\sqrt{k}$ elements $q_i \in Q$ would have at least $\varepsilon_2\sqrt{k}/j$ wedges $W_m^{(i)}$ with at least j^2 elements of S in their interiors. Claim 3 then implies that there are fewer than $\varepsilon_2\sqrt{k}$ elements $q_i \in Q$ with more than $7\varepsilon_2\sqrt{k}/j$ wedges $W_m^{(i)}$ that contain at least one q-regular edge. Hence, the number of q-regular edges in $\vec{B}$ is at most

$$\varepsilon_2\sqrt{k}|R| + |Q|\,\frac{7\varepsilon_2\sqrt{k}}{j}\,j \le 8\varepsilon_2 C_1 k,$$

contradicting (4) if $\varepsilon_2 < \varepsilon_1^2/(80C_1)$. □

Now we point out an interesting consequence of Lemma 2. Suppose it is condition (1) of Lemma 2 that holds. Consider a wedge $W_m^{(i)} = r_{(m-1)j}^{(i)} q_i r_{mj}^{(i)}$ with at least j^2 elements of S in its interior. Rotate l^*, the line from q_i to $r_{(m-1)j}^{(i)}$, clockwise about q_i to the position $q_i r_{mj}^{(i)}$. Let $p_1, p_2, \ldots, p_{j^2}$ denote the first j^2 elements of S hit by l^*, and note that every point p_t is either in $W_m^{(i)}$ or in its vertical angle $\bar{W}_m^{(i)}$. Also, the number of points of S to the right of the line $q_i p_t$ differs from k by at most t. More precisely, for every $1 \le t \le j^2$,

$$\begin{aligned} |S(\overrightarrow{q_i p_t}) - k| &\le t \qquad \text{if} \quad p_t \in W_m^{(i)}, \\ |S(\overrightarrow{p_t q_i}) - k| &\le t \qquad \text{if} \quad p_t \in \bar{W}_m^{(i)}. \end{aligned} \tag{5}$$

This means that we have identified j^2 edges incident to q_i in the graph

$$\vec{G}_{k \pm j^2} = \bigcup_{s=k-j^2}^{k+j^2} \vec{G}_s.$$

We call $W_m^{(i)} \cup \bar{W}_m^{(i)}$ a *double-wedge* sitting at q_i. For the case when condition (2) holds in Lemma 2, we define double-wedges sitting at r_i in the analogous way.

We apply the same procedure to each point p_i for which the double-wedge sitting at p_i contains at least j^2 elements of S, each time obtaining j^2 new edges of $\vec{G}_{k \pm j^2}$. The graph formed by all these edges is called $\vec{G}'_{k \pm j^2}$. By Lemma 2, each nonempty block of $\vec{G}_k^*$ has at least $(\varepsilon_2\sqrt{k})(\varepsilon_2\sqrt{k}/j) = \varepsilon_2^2 k/j$ double-wedges containing at least j^2 points of S, and the number of nonempty blocks is clearly at least

$$|E(\vec{G}_k^*)| \Big/ \binom{C_1\sqrt{k}}{2} \ge \frac{2\varepsilon_1 n}{C_1^2\sqrt{k}} \qquad \text{(see Lemma 1)}.$$

Therefore $\vec{G}'_{k \pm j^2}$ has at least

$$\frac{1}{2}\,\frac{2\varepsilon_1}{C_1^2}\,\frac{n}{\sqrt{k}}\,\frac{\varepsilon_2^2 k}{j}\,j^2 = \frac{\varepsilon_1\varepsilon_2^2}{C_1^2}\,nj\sqrt{k} \tag{6}$$

edges.

We claim that $\vec{G}'_{k\pm j^2}$ contains a fixed *positive proportion of all edges of* $\vec{G}_{k\pm j^2}$.

Claim 4. $|E(\vec{G}_{k\pm j^2})| \le 5nj\sqrt{k}$.

The argument is similar to that used in (1); another proof is given in [14].

Proof. Recall that L is a system of $n-1$ vertical lines that separate the points of S; we estimate the total number of intersections between the edges of $\vec{G}_{k\pm j^2}$ and lines in L. Applying Proposition 1 to $\vec{G}_{k-j^2}, \dots, \vec{G}_{k+j^2}$,

$$\sum_{\vec{xy}\in E(\vec{G}_{k\pm j^2})} \|\vec{xy}\| = \sum_{s=k-j^2}^{k+j^2} \sum_{\vec{xy}\in E(\vec{G}_s)} \|\vec{xy}\| \le 2 \sum_{s=k-j^2}^{k+j^2} (n-1)(s+1) < 5nkj^2.$$

Note that there is a factor of 2 here because we must count both left–right and right–left edges in $\vec{G}_s$. On the other hand, denoting the outdegree of x in $\vec{G}_{k\pm j^2}$ by $d^-(x)$,

$$\sum_{\vec{xy}\in E(\vec{G}_{k\pm j^2})} \|\vec{xy}\| \ge \sum_{x\in S} \left(\frac{d^-(x)}{2}\right)^2 \ge \frac{(\sum_{x\in S} d^-(x))^2}{4n} = \frac{|E(\vec{G}_{k\pm j^2})|^2}{4n}. \qquad \square$$

Remark. If we could improve j^2 to $j^{2+\delta}$ (with some $\delta > 0$) in Lemma 2, then by the above procedure we could find at least $\varepsilon_1\varepsilon_2^2 nj^{1+\delta}\sqrt{k}/C_1^2$ edges of $\vec{G}_{k+j^2}$. This would contradict Claim 4 and complete the proof of our theorem.

4. The Structure of $\vec{G}_{k\pm j^2}$

In this section we modify the construction described at the end of the last section and extract a large subgraph $\vec{G}^*_{k\pm j^2} \subseteq \vec{G}'_{k\pm j^2} \subseteq \vec{G}_{k\pm j^2}$, which is much easier to handle. The arguments are very similar to those used in the preprocessing of $\vec{G}_k$. As before, $L = \{l_1, l_2, \dots, l_{n-1}\}$ is a system of $n-1$ vertical lines separating the points of S.

In (6) it was shown that $|E(\vec{G}'_{k\pm j^2})| \ge (\varepsilon_1\varepsilon_2^2/C_1^2)nj\sqrt{k}$. Proposition 1 implies that the total length of edges in $\vec{G}_{k\pm j^2}$ is at most

$$\sum_{s=k-j^2}^{k+j^2} \sum_{\vec{xy}\in E(\vec{G}_s)} \|\vec{xy}\| \le 2 \sum_{s=k-j^2}^{k+j^2} (n-1)(s+1) < 5nkj^2.$$

Therefore

$$\left|\left\{\vec{xy} \in E(\vec{G}'_{k\pm j^2}) \colon \|\vec{xy}\| \ge \frac{20C_1^2}{\varepsilon_1\varepsilon_2^2} j\sqrt{k}\right\}\right| \le \frac{5nkj^2}{(20C_1^2/(\varepsilon_1\varepsilon_2^2))j\sqrt{k}} = \frac{\varepsilon_1\varepsilon_2^2}{4C_1^2} nj\sqrt{k}. \tag{7}$$

Also,

$$\left|\left\{\overrightarrow{xy} \in E(\vec{G}'_{k\pm j^2})\colon \|\overrightarrow{xy}\| \le \frac{\varepsilon_1\varepsilon_2^2}{4C_1^2} j\sqrt{k}\right\}\right| \le \frac{\varepsilon_1\varepsilon_2^2}{4C_1^2} nj\sqrt{k}. \tag{8}$$

Letting $\vec{G}''_{k\pm j^2}$ be the graph consisting of those edges $\overrightarrow{xy} \in E(\vec{G}'_{k\pm j^2})$ with

$$\frac{\varepsilon_1\varepsilon_2^2}{4C_1^2} j\sqrt{k} < \|\overrightarrow{xy}\| < \frac{20C_1^2}{\varepsilon_1\varepsilon_2^2} j\sqrt{k},$$

(6), (7), and (8) imply

$$|E(\vec{G}''_{k\pm j^2})| \ge \frac{\varepsilon_1\varepsilon_2^2}{2C_1^2} nj\sqrt{k}.$$

As in (2), let $L' = \{l_\Delta, l_{2\Delta}, l_{3\Delta}, \ldots\}$ with $\Delta = \lfloor (12C_1^2/(\varepsilon_1\varepsilon_2^2))^3 j\sqrt{k} \rfloor$. No line of L' can intersect more than $((20C_1^2/(\varepsilon_1\varepsilon_2^2))j\sqrt{k})^2$ edges of $\vec{G}''_{k\pm j^2}$, so if $\vec{G}'''_{k\pm j^2}$ denotes the graph formed by those edges of $\vec{G}''_{k\pm j^2}$ which intersect no line in L',

$$|E(\vec{G}'''_{k\pm j^2})| \ge |E(\vec{G}''_{k\pm j^2})| - |L'|\left(\frac{20C_1^2}{\varepsilon_1\varepsilon_2^2} j\sqrt{k}\right)^2 \ge \frac{\varepsilon_1\varepsilon_2^2}{4C_1^2} nj\sqrt{k}.$$

Finally, deleting (in sequence) all edges of $\vec{G}'''_{k\pm j^2}$ that are incident to points of degree less than $(\varepsilon_1\varepsilon_2^2/8C_1^2)j\sqrt{k}$, we obtain the graph $\vec{G}^*_{k\pm j^2}$. Its properties are summarized in the following statement, analogous to Lemma 1.

Lemma 3. *There are constants $\varepsilon_3, C_3 > 0$ depending only on $\varepsilon_1, \varepsilon_2, C_1$ (and hence only on ε) such that there is a subgraph $\vec{G}^*_{k\pm j^2} \subseteq \vec{G}'_{k\pm j^2}$ which splits into disjoint blocks $\vec{B}_1^* \cup \vec{B}_2^* \cup \cdots \cup \vec{B}_m^*$ which satisfy the following conditions*:

(1) $|E(\vec{G}^*_{k\pm j^2})| = \sum_{i=1}^m |E(\vec{B}_i^*)| \ge \varepsilon_3 nj\sqrt{k}$.
(2) $|V(\vec{B}_i^*)| = \lfloor C_3 j\sqrt{k} \rfloor$ *for every* $1 \le i < m$, $|V(\vec{B}_m^*)| \le \lfloor C_3 j\sqrt{k} \rfloor$, *and the vertices of every $\vec{B}_i^*$ are consecutive elements in the ordering of S determined by the x-coordinates of the points.*
(3) $\varepsilon_3 j\sqrt{k} < \|\overrightarrow{xy}\| < C_3 j\sqrt{k}$ *for every* $\overrightarrow{xy} \in E(\vec{G}^*_{k\pm j^2})$.
(4) *Every vertex of $\vec{B}_i^*$ is either isolated or incident to at least $\varepsilon_3 j\sqrt{k}$ edges of $\vec{B}_i^*$, $1 \le i \le m$.*

Given a graph $G = (V, E)$ and a subset $X \subseteq V$, let $G - X$ denote the graph obtained by removing all edges incident to at least one element of X. The role of the q's and the r's is clearly symmetric (see Lemma 2). Therefore, to avoid repetition, *from this point on* we use q_i to denote any nonisolated vertex of $\vec{G}_k^*$, whether a q_i or an r_i.

Lemma 4. *There exists a constant $\varepsilon_4 > 0$ depending only on $\varepsilon_1, \varepsilon_2, \varepsilon_3, C_1, C_3$ (and hence only on ε) such that, for any subset $X \subseteq S$ of size at most $\varepsilon_3 n/(2C_3)$, we can find at least $\varepsilon_4 n$ nonisolated points q_i in $\vec{G}_k^*$ satisfying the following property:*

(P_j) *The number of double-wedges $W_m^{(i)} \cup \bar{W}_m^{(i)}$ sitting at q_i that contain at least $\varepsilon_4 j^2$ edges of $\vec{G}_{k\pm j^2}^* - X$ incident to q_i is at least $\varepsilon_4\sqrt{k}/j$.*

Proof. By Lemma 3(2),

$$|E(\vec{G}_{k\pm j^2}^* - X)| \geq |E(\vec{G}_{k\pm j^2}^*)| - |X| C_3 j\sqrt{k} \geq \frac{\varepsilon_3}{2} nj\sqrt{k}.$$

According to the definition of $\vec{G}_{k\pm j^2}'$ following (6), each edge is assigned to a point q_i and to a double-wedge $W_m^{(i)} \cup \bar{W}_m^{(i)}$ whose apex is at q_i. If the lemma were false, then if we added up the number of edges of $E(\vec{G}_{k\pm j^2}^* - X)$ assigned to each of the points q_i that satisfy property (P_j), we would get a sum which is at most $\varepsilon_4 n(C_1\sqrt{k}/j)j^2 = \varepsilon_4 C_1 nj\sqrt{k}$. On the other hand, the number of edges of $E(\vec{G}_{k\pm j^2}^* - X)$ assigned to points not having property (P_j) does not exceed

$$n\left(\frac{\varepsilon_4\sqrt{k}}{j} j^2 + \frac{C_1\sqrt{k}}{j} \varepsilon_4 j^2\right).$$

Setting $\varepsilon_4 = \varepsilon_3/(6C_1)$, we obtain

$$|E(\vec{G}_{k\pm j^2}^* - X)| < \varepsilon_4 C_1 nj\sqrt{k} + n\left(\frac{\varepsilon_4\sqrt{k}}{j} j^2 + \frac{C_1\sqrt{k}}{j} \varepsilon_4 j^2\right) < \frac{\varepsilon_3}{2} nj\sqrt{k},$$

the desired contradiction. □

If $G = (V, E)$ is a graph whose vertices are a subset of the integers, a point $x \in V$ is said to *cut the edge* $yz \in E$, if x is contained in the open interval (y, z). It is easy to prove the following simple fact, needed in the next section.

Lemma 5. *For every $\gamma \in (0, 1)$, there exists a $\delta = \delta(\gamma) > 0$ with the following property. Given any graph G on the vertex set $V = \{1, 2, \ldots, K\}$ with minimum degree at least γK, and given any subset $V' \subseteq V$, $|V'| \geq \gamma K$, we can always find an interval $I \subseteq V$ such that $|I \cap V'| \geq \delta K$ and every element of I cuts at least δK^2 edges of G ($\delta \geq \gamma^3/35$).*

Proof. For simplicity, assume that $|V'| = \gamma K$. For a fixed δ, let $i_1 < i_2 < \cdots < i_{t-1}$ denote those points of V that cut fewer than δK^2 edges of G. Set $i_0 = 1$, $i_t = K$, and let v_j (and v_j') denote the number of elements of V (resp. V') in the closed interval $I_j = [i_{j-1}, i_j]$, $1 \leq j \leq t$. Clearly, $\sum v_j \leq K + t$, $\sum v_j' \leq |V'| + t$. An edge of G is called short if both its endpoints belong to the same I_j, and long otherwise.

Double counting L, the total length of edges of G incident to at least one element of V', we obtain

$$\tfrac{1}{2}|V'|\frac{\gamma^2K^2}{4}=\frac{\gamma^3K^3}{8}\le L=L_{\text{short}}+L_{\text{long}},$$

where

$$L_{\text{short}}\le\sum_{j=1}^{t}v'_jv_j^2$$

and

$$\begin{aligned}L_{\text{long}}&\le\sum_{j=1}^{t-1}(\text{number of edges cut by } i_j)(v_j+v_{j+1}-1)\\&\le\delta K^2(2|V'|+t)<3\delta K^3.\end{aligned}$$

Assume now that $|(i_{j-1},i_j)\cap V'|<\delta K$ for all j. Then

$$L_{\text{short}}\le(\delta K+2)\sum v_j^2\le(\delta K+2)K^2,$$

and choosing $\delta=\gamma^3/35$ we get a contradiction. Thus $|(i_{j-1},i_j)\cap V'|\ge\delta K$ for some j, which completes the proof. □

5. Stabbing Many Edges by a Line

By Proposition 1, every line intersects at most $2\sum_{s=k-j^2}^{k+j^2}(s+1)<5j^2k$ edges of $\vec{G}_{k\pm j^2}$. Now we show how to find an integer j and a straight line that violate this condition, a contradiction that will complete the proof of our theorem. Let $S_j\subseteq S$ denote the set of all *nonisolated* vertices of $\vec{G}^*_{k\pm j^2}$, and write $S_{\le j}=\bigcup_{t=1}^{j}S_t$.

Lemma 6. *Suppose that a point $q_i\in S_0$ has property (P_j) with $X=S_{\le j}-S_{\le h}$ for some $j>h$. Let $r_0^{(i)}, r_1^{(i)},\ldots$ denote the neighbors of q_i in $\vec{G}^*_k$. Let $A=A(j,h,q_i)$ denote the total number of intersections between the lines $q_ir_t^{(i)}$ and the edges $\overrightarrow{pp}'\in E(\vec{G}_{k\pm h^2})$ which satisfy*

$$\varepsilon_3j\sqrt{k}-C_3h\sqrt{k}\le\|\overrightarrow{q_ip}\|,\ \|\overrightarrow{q_ip}'\|\le C_3j\sqrt{k}+C_3h\sqrt{k}.\tag{9}$$

Then

$$A>\left(\frac{\varepsilon_3\varepsilon_4}{10C_3h}\right)^{14}k^{3/2}.$$

Proof. Consider all double-wedges $W_m^{(i)} \cup \bar{W}_m^{(i)}$ of size j sitting at q_i which contain at least $\varepsilon_4 j^2$ edges of $\vec{G}^*_{k\pm j^2} - X$ incident with q_i. The other endpoints of these edges are called *red* points. Since q_i has property (P_j), there are at least $\varepsilon_4^2 j\sqrt{k}$ red points, each being at distance at least $\varepsilon_3 j\sqrt{k}$ and at most $C_3 j\sqrt{k}$ from q_i.

Because every red point belongs to some S_g ($g \le h$), there is a $g \le h$ such that $\vec{G}^*_{k\pm g^2}$ has at least $\varepsilon_4^2 j\sqrt{k}/(h+1) \ge \varepsilon_4^2 j\sqrt{k}/(2h)$ red points among its nonisolated vertices. Let $\vec{B}_1^*, \vec{B}_2^*, \dots$ denote the blocks of $\vec{G}^*_{k\pm g^2}$, each with $\lfloor C_3 g\sqrt{k}\rfloor$ vertices (see Lemma 3). At most $2C_3 j\sqrt{k}/|V(\vec{B}_t^*)| < C_3 j$ blocks $\vec{B}_t^*$ may contain red points. Writing $R(\vec{B}_t^*)$ for the number of red points in $\vec{B}_t^*$ we have

$$\left(\#\vec{B}_t^*\colon R(\vec{B}_t^*) \ge \frac{\varepsilon_4^2}{4C_3}\frac{\sqrt{k}}{h}\right)|V(\vec{B}_t^*)|$$

$$\ge R(\vec{G}^*_{k\pm g^2}) - \left(\#\vec{B}_t^*\colon R(\vec{B}_t^*) < \frac{\varepsilon_4^2}{4C_3}\frac{\sqrt{k}}{h}\right)\frac{\varepsilon_4^2}{4C_3}\frac{\sqrt{k}}{h}$$

$$\ge \frac{\varepsilon_4^2 j\sqrt{k}}{2h} - C_3 j\frac{\varepsilon_4^2}{4C_3}\frac{\sqrt{k}}{h} = \frac{\varepsilon_4^2 j\sqrt{k}}{4h},$$

so that

$$\left(\#\vec{B}_t^*\colon R(\vec{B}_t^*) \ge \frac{\varepsilon_4^2}{4C_3}\frac{\sqrt{k}}{h}\right) \ge \frac{\varepsilon_4^2}{4C_3}\frac{j}{h^2}.$$

Now fix a block $\vec{B}_t^*$ with at least $(\varepsilon_4^2/4C_3)(\sqrt{k}/h)$ red points $b_1, b_2, b_3, \dots, b_s$, $s \ge (\varepsilon_4^2/4C_3)(\sqrt{k}/h)$, listed in clockwise ordering of the rays from q_i. We apply Lemma 5 to the graph G obtained from $\vec{B}_t^*$ by omitting its isolated vertices and regarding its edges as undirected. We view G from q_i and, in applying Lemma 5, we let $V' = \{b_1, b_2, \dots, b_s\}$, $\varepsilon_3 g\sqrt{k} < |V| = K \le C_3 g\sqrt{k}$, $\gamma = \min(\varepsilon_3/C_3,\ \varepsilon_4^2/(4C_3^2h^2)) = \varepsilon_4^2/(4C_3^2h^2)$ (see Lemma 3). We conclude that there are integers u and v such that $v - u \ge \gamma^3K/35$, and every half-line $\overrightarrow{q_ip}$ in the angular region $b_uq_ib_v$ intersects at least $\gamma^3K^2/35$ edges of $\vec{B}_t^* \subseteq \vec{G}^*_{k\pm g^2} \subseteq \vec{G}_{k\pm h^2}$.

It follows from the definitions that every double-wedge $W_m^{(i)} \cup \bar{W}_m^{(i)}$ of size j contains at most j^2 red elements b_w of $\vec{B}_t^*$. This means that for at least $\gamma^3K/(40j^2)$ consecutive integers m, the rays $\overrightarrow{q_ir}{}_{mj}^{(i)}$ are in the angular region $b_uq_ib_v$, and so the number of lines $\overrightarrow{q_ir}{}_p^{(i)}$ that intersect at least $\gamma^3K^2/35$ edges of $\vec{B}_t^*$ exceeds $\gamma^3K/(45j)$. Applying the same argument to all the blocks $\vec{B}_t^*$ having at least $(\varepsilon_4^2/4C_3)(\sqrt{k}/h)$ red points (there are at least $(\varepsilon_4^2/4C_3)(j/h^2)$ of them), we obtain

$$A \ge \left(\frac{\varepsilon_4^2}{4C_3}\frac{j}{h^2}\right)\left(\frac{\gamma^3}{35}K^2\right)\left(\frac{\gamma^3}{45}\frac{K}{j}\right)$$

$$> \left(\frac{\varepsilon_3\varepsilon_4}{10C_3h}\right)^{14}k^{3/2}. \qquad \square$$

Proof of the Theorem. Put $M = \lceil 2C_3/\varepsilon_3 \rceil$, $N = \lceil (10C_3/(\varepsilon_3\varepsilon_4))^{15} \rceil$, and define M sets of integers as follows:

$$J_u = \{h_u M, h_u M^2, \dots, h_u M^{h_u^{16}N}\}, \qquad 1 \le u \le M,$$

where $h_1 = 1$ and, for every $u > 1$, $h_u = h_{u-1} M^{h_{u-1}^{16}N}$.

The function

$$f(j) \equiv |S_{\le j}| = \left| \bigcup_{t=1}^{j} S_t \right|$$

is increasing in j, so there is an integer $u \le M$ such that

$$f(h_{u+1}) - f(h_u) \le \frac{n}{M} \le \frac{\varepsilon_3}{2C_3} n.$$

This implies that

$$|S_{\le j} - S_{\le h_u}| \le \frac{\varepsilon_3}{2C_3} n, \qquad \forall j \in J_u.$$

By Lemma 4, for every $j \in J_u$, at least $\varepsilon_4 n$ points $q_i \in S_0 \subseteq V(\vec{G}_k^*)$ have property (P_j) with $X = S_{\le j} - S_{\le h_u}$. This, in turn, implies that there is a point, say $q_i \in S_0$, satisfying property (P_j) for every j belonging to some subset $J'_u \subseteq J_u$ of size at least $\varepsilon_4 |J_u| = \varepsilon_4 h_u^{16} N$.

We apply Lemma 6 to the point q_i (with $h = h_u$, $j \in J'_u$) and observe that the total number of intersections between the lines $q_i r_t^{(i)}$ and the edges of $\vec{G}_{k \pm h_u^2}$ is at least

$$\sum_{j \in J'_u} A(j, h, q_i) \ge |J'_u| \left(\frac{\varepsilon_3 \varepsilon_4}{10C_3 h_u} \right)^{14} k^{3/2} \ge 10C_1 h_u^2 k^{3/2}. \tag{10}$$

(Note that the condition in (9) guarantees that no intersection is counted more than once.)

By Lemma 3, q_i has at most $C_1\sqrt{k}$ neighbors $r_t^{(i)}$ in $\vec{G}_k^*$. One of them gives a line, $q_i r_t^{(i)}$, with more than the average number of intersections. According to (10) this line intersects at least $10C_1 h_u^2 k^{3/2}/(C_1\sqrt{k}) \ge 10h_u^2 k$ edges of $\vec{G}_{k \pm h_u^2}$. This contradicts the first sentence of this section, completing the proof. A straightforward evaluation of the constants ε_i, C_i, M, N, h_u shows that our arguments always lead to contradiction, provied that $\varepsilon > C/\log^* k$. □

Acknowledgment

We are grateful to Micha Sharir for many valuable comments.

References

1. N. Alon and E. Györi. The number of small semispaces of a finite set of points. *J. Combin. Theory Ser. A*, 41:154–157, 1986.
2. I. Bárány, Z. Füredi, and L. Lovász. On the number of halving planes in R^3. *Proc. Fifth ACM Symposium on Computational Geometry*, pages 140–144, 1989.
3. B. Chazelle and F. Preparata. Halfspace range search: an algorithmic application of k-sets. *Discrete Comput. Geom.*, 1:83–93, 1986.
4. B. Chazelle, H. Edelsbrunner, L. Guibas, and M. Sharir. Points and triangles in the plane and halving planes in space. Preprint, 1989.
5. K. Clarkson. New applications of random sampling in computational geometry. *Discrete Comput. Geom.*, 2:195–222, 1987.
6. K. Clarkson and P. Shor. Applications of random sampling in computational geometry, II. *Discrete Comput. Geom.*, 4:387–421, 1989.
7. R. Cole, M. Sharir, and C. Yap. On k-hulls and related problems. *SIAM J. Comput.*, 16:61–77, 1987.
8. H. Edelsbrunner. *Algorithms in Combinatorial Geometry.* Springer-Verlag, Berlin, 1987.
9. H. Edelsbrunner and E. Welzl. On the number of line separations of a finite set in the plane. *J. Combin. Theory Ser. A*, 38:15–29, 1985.
10. P. Erdös, L. Lovász, A. Simmons, and E. Strauss. Dissection graphs of planar point sets. In J. Srivastava *et al.*, editors, *A Survey of Combinatorial Theory*, pages 139–149. North-Holland, Amsterdam, 1973.
11. J. E. Goodman and R. Pollack. On the number of k-sets of a set of n points in the plane. *J. Combin. Theory Ser. A*, 36:101–104, 1984.
12. L. Lovász. On the number of halving lines. *Ann. Univ. Sci. Budapest Eötvös Sect. Math.*, 14:107–108, 1971.
13. J. Pach, W. Steiger, and E. Szemerédi. An upper bound on the number of planar k-sets. *Proc. 30th IEEE Symposium on Foundations of Computer Science*, pages 72–79, 1989.
14. E. Welzl. More on k-sets of finite sets in the plane. *Discrete Comput. Geom.*, 1:95–100, 1986.

Received December 28, 1989, and in revised form February 28, 1990.