Point-sets with few k-sets

Helmut Alt* Stefan Felsner* Ferran Hurtado† Marc Noy†

Abstract

A k-set of a finite set S of points in the plane is a subset of cardinality k that can be separated from the rest by a straight line. The question of how many k-sets a set of n points can contain is a long-standing open problem where a lower bound of $\Omega(n \log k)$ and an upper bound of $O(nk^{1/3})$ are known today.

Under certain restrictions on the set S, for example, if all points lie on a convex curve, a *linear* upper bound can be shown. Here, we will generalize this observation by showing that if the points of S lie on a constant number of convex curves, the number of k-sets remains linear in n.

Keywords: convex curve, k-set, Lovász' procedure.
AMS Subject Classification: 52C10, 52A10, 68R99, 05C99.

1 Introduction and definitions

Let S be a set of n points in the plane. A *k-set* $(1 \leq k \leq n-1)$ is a subset of S of cardinality k that can be separated from the rest by a straight line. The simple and natural question of how many k-sets a set of n points can contain has been considered for more than 25 years and has inspired considerable research. Nevertheless, it is still not solved completely and remains one of the most prominent open problems in combinatorial geometry.

The first upper bound is due to Lovász [4] who showed that the number of k-sets can be at most $O(n\sqrt{k})$ which is $O(n^{3/2})$. Erdős et al [3] constructed a family of sets having $\Omega(n \log k)$ k-sets which is $\Omega(n \log n)$ for suitable values of k. For a long time the gap between lower and upper bound could not be narrowed until Pach, Steiger and Szemerédi [5] showed an upper bound of $O(n\sqrt{k}/\log^* k)$. More significant progress was made only recently by Dey [2] who proved an upper bound of $O(n\sqrt[3]{k})$ which is $O(n^{4/3})$.

*Institut für Informatik, Freie Universität Berlin, Takustr. 9, D-14195 Berlin, E-mail: *name*@inf.fu-berlin.de

†Departament de Matemàtica Aplicada II, Universitat Politècnica de Catalunya, Pau Gargallo 5, 08028-Barcelona, España. Partially supported by Proyecto MEC-DGES-SEUID PB96-0005-C02-02.

Various generalizations of the k-set problem or its dual formulation, namely determining the number of cells at the k-*level* of the arrangement of a set of n straight lines, have been considered. Among them are k-sets in higher dimensions or k-levels of curves in two dimensions and surfaces in three dimensions [1, 6].

Under certain restrictions on the set S it is possible to prove better upper bounds on the number of k-sets. For example, if all points of S lie on a straight line, there are only 2 k-sets. If all points of S lie on a convex curve, S contains at most n k-sets. Here, we will generalize the latter observation by showing that if the points of S lie on a constant number of convex curves, the number of k-sets remains *linear* in n.

In the following let S be a fixed set of n points in the plane and let $k \leq n-2$. An oriented straight line l is called a k-*line* of S exactly if the open halfplane right of l contains k points of S. As easily can be seen, to any k-set M of S there exists either a corresponding k-line incident to two points $p, q \in S \setminus M$ or a $(k-1)$-line incident to points $p \in M$ and $q \in S \setminus M$ having $M \setminus \{p\}$ as its corresponding $(k-1)$-set. Let us call any line segment $\overline{pq}$ with the former property a k-*segment* of S (a segment with the latter property, then, is a $(k-1)$-segment). The idea is to obtain an upper bound on the number of k-segments and $(k-1)$-segments and, thus, on the number of k-lines. We will make use of a powerful lemma, due to Lovász, that was shown in the classical articles [3, 4].

Lovász' Lemma: *Let S be a set of n points and l a straight line containing no points of S and dividing S into two subsets of m and $n-m$ points, respectively. Then for any $k \leq n-2$ the number of k-segments intersecting l is at most* $2\min\{m, n-m, k+1\}$.

A simple consequence of Lovász Lemma is that the number of k-segments intersecting a straight line l is $O(n)$ (in fact, it is at most n). The lemma is proven by giving a procedure that enumerates all k- and $(k-1)$-segments. For convenience we briefly review *Lovász procedure*:

Let l be an oriented line and let S_l be the set of points of S on the (open) right side of l. Assume without loss of generality that no two points in S have the same y-coordinate. Then there is a unique horizontal line l_0 containing a point p_0 of S such that S_{l_0} is of cardinality exactly k, i.e., there are k points of S below l_0. Let $l = l_0$ and repeat the following step: rotate l counterclockwise around p_i until it hits a point p_{i+1}, let l_{i+1} be the line through p_i and p_{i+1} and increment i. This is done until the line returns into the initial horizontal position, i.e., $l_i = l_0$. Let $l_1, l_2, \ldots, l_z$ be the different lines generated by this process. Each line l_i contains a segment $\overline{p_{i-1}p_i}$, we call these segments the L-*segments* of the process. Lovász observed the following fact.

Fact 1. The L-segments are the union of the $(k-1)$- and the k-segments.

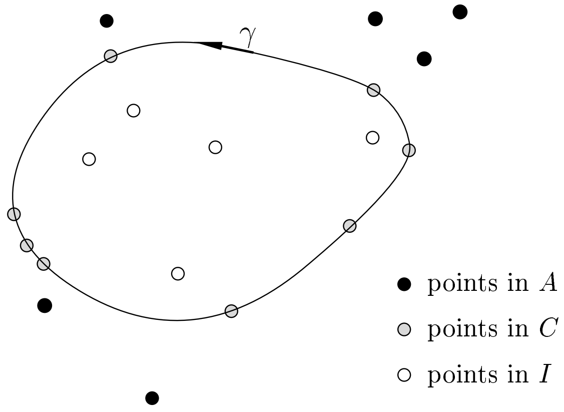


Figure 1: Curve γ and sets A, C and I

So , by the observation above, any upper bound on the number of L-segments will be one on the number of k-sets.

2 The main result

In the following we will consider point sets lying on a fixed collection of curves in the plane. Each curve can be closed or not, bounded or unbounded. We call a curve *convex* exactly if it lies completely on the boundary of its convex hull. Our main result is stated in the following theorem.

Theorem *Consider a fixed, finite collection of p pairwise disjoint convex curves in the plane. Then there is a constant $c_p > 0$ such that any set S of n points, each lying on one of the curves has at most $c_p n$ k-sets.*

We first observe that without loss of generality we may assume that the points in S are in *general position* in the sense that no three of them lie on a straight line. In fact, let $K \subset S$ be a k-set. Then there exists a line l not containing any point of S which separates K and $S \backslash K$. Consequently, under slight perturbations of the points in S any k-set K still remains a k-set. So the points of S can be brought into general position without decreasing the number of k-sets.

The main step towards the proof of the theorem is the following proposition which considers one fixed convex curve and gives an upper bound on the number of k-segments with at least one endpoint on the curve.

Let γ be a convex curve, i.e., the boundary curve of some convex set Γ in the plane. Partition the n points of S into the subset C of points on γ, the subset I of points in the open interior of Γ and the subset A of points outside γ, i.e., $A = S \setminus (C \cup I)$. Figure 1 illustrates the situation.

Proposition 1 *Let r be the number of k-segments of S with at least one endpoint in C and let r_A be the number of k-segments with one endpoint in*

A and one in C. Then

$$r \leq 4|C| + 8r_A.$$

Proof. Let us apply Lovász' procedure to the set S and let $p_0, \dots, p_z$ be the sequence of points and $l_1, \dots, l_z$ the sequence of lines traversed. We will enumerate all L-segments containing a point of C. For any straight line l intersecting γ we call the first intersection point of l with γ when following its orientation its *entry-point*. We will prove an upper bound on the number ρ of L-segments where the entry-point is one of the endpoints of the segment, i.e., on the number of oriented L-segments $\overline{p_i p_{i+1}}$ with p_i in C. For reasons of symmetry this same bound holds for the L-segments with the dually defined exit-point in C.

Let $L_1, L_2 \dots L_\rho$ be the subsequence of $l_1, \dots, l_z$ of those lines L_i with entry-point $q_i \in C$. Let γ be oriented such that the convex region Γ is to the left of γ. The orientation of γ induces a cyclic ordering of the points in C. If during Lovász' procedure every line L between L_i and L_{i+1} intersects γ we call the pair L_i, L_{i+1} a *standard step.*

Claim 1. If L_i, L_{i+1} is a standard step then either q_{i+1} is the point immediately preceding q_i in the cyclic order on C or q_{i+1} is the point immediately following q_i or $q_{i+1} = q_i$.

Proof. During Lovász' procedure the entry-point describes a continuous movement on curve γ. △

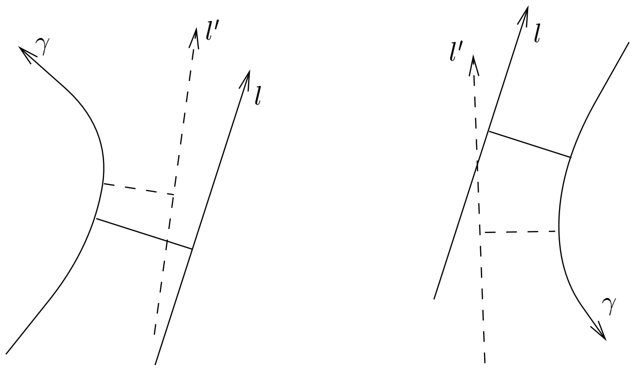
We classify types of standard steps: A standard step L_i, L_{i+1} is a *trivial step* if it corresponds to a rotation around the entry-point $q_i = q_{i+1}$. Note that every second step is a trivial step. A non-trivial standard step is called a *forward step* if q_{i+1} is the point following q_i in the cyclic order on C else, i.e., if $q_{i+1} = q_i$ or if q_{i+1} is the immediate predecessor of q_i in the cyclic order on C, the step is called a *backward step.* A step which is not standard is called a *special step.* If s is the number of special steps, a the number of forward steps and b the number of backward steps, then we have for the total number ρ of steps

$$\rho/2 = a + b + s. \tag{1}$$

Let L_i, L_{i+1} be a backward step and observe the entry-point when leaving q_i. When the point moves backward on γ the rotation of line l is a rotation around some point $p \in A$. In this case we "charge" the step L_i, L_{i+1} to the L-segment $\overline{p\, q_i}$. When the entry-point moves forward on γ after leaving from q_i then it must be $q_{i+1} = q_i$ and the entry point must move backward and return eventually. Again this backward motion is due to rotation around some point p in A and we charge the step to the L-segment $\overline{p\, q_{i+1}}$. This shows

$$b \leq r_A. \tag{2}$$

We now turn to a closer discussion of special steps. To simplify the exposition we assume that γ has nonvanishing curvature everywhere (this



(a) γ left of l (b) γ right of l

Figure 2: Rotating from l to l' the nigh-point moves forward on γ

can be done w.l.o.g. since we assume that S is in general position). Suppose some line L between L_i and L_{i+1} has no entry-point on γ. Then there is a moment when the rotating line l leaves γ, i.e., l is tangent to γ. For a line with less than two points of intersection with γ we define the *nigh-point of l* as the (unique!) closest point to l on γ. There are two important observations about the nigh-point:

Fact 2. For every line l the entry-point or the nigh-point are defined. If they are both defined then they coincide with the tangent point. Hence, observing whichever is defined during Lovász' procedure we obtain a continuously moving *en-point* for l on γ.

Fact 3. While l rotates outside of γ the nigh-point always moves forward in the orientation of γ. This is easily seen to be true in both possible cases: γ left of l and γ right of l (see Figure 2).

As a consequence of Fact 3 we obtain that the en-point can move backwards only when it is the entry-point. Therefore, if L_i, L_{i+1} is a special step then either q_{i+1} is the point immediately preceding q_i in the cyclic order on C or $q_{i+1} = q_i$ or the nigh-point sweeps over a piece of curve γ containing all points of C that are between q_i and q_{i+1} in the cyclic order.

We now refine the classification and add the attribute *backward* to special steps with either q_{i+1} preceding q_i in the cyclic order on C or $q_{i+1} = q_i$. The attribute *forward* is given to all other special steps. Let s_a be the number of forward special steps and s_b be the number of backward special steps, then

$$s = s_a + s_b \tag{3}$$

As in the argument preceding Inequality (2) we find an L-segment involving a point p of A and a point of C for every backward special step.

This improves Inequality 2 to

$$b + s_b \leq r_A. \tag{4}$$

The next lemma will show that the winding number of the en-point is $+1$.

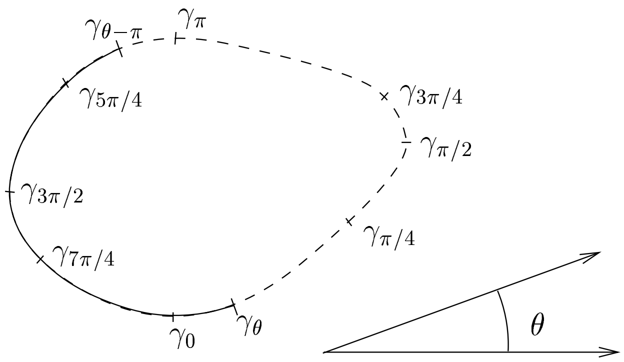


Figure 3: The potential locations for the en-point of a line at angle θ cover the part of γ parametrized by $[\theta - \pi, \theta]$.

Lemma 1 *During Lovász' procedure the effect of the movement of the en-point is one full rotation around* Γ.

Proof. For a fixed angle θ consider the en-points of all θ-oriented lines. These points cover one half of γ if we parametrize the curve by tangential-angle (see Figure 3).

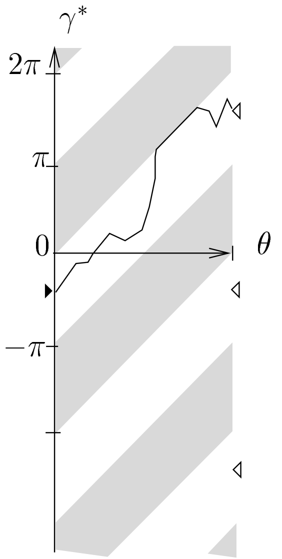


Figure 4: Possible motion of the en-point.

Consider the lifting-space γ^* of curve γ, a doubly infinite spiral where each point $\gamma(t)$ on γ has a copy $\gamma^*(t+k2\pi)$ for every integer k. We identify γ^* with the reals by $t \leftrightarrow \gamma^*(t)$. During Lovász' procedure the rotating line sweeps through all angles θ from 0 to 2π. For every θ the tangential-angle of the en-point of $l(\theta)$ is in the interval $[\theta - \pi, \theta]$. Lines $l(0)$ and $l(2\pi)$ and hence their en-points equal each other. Therefore in the lifting-space γ^* the en-points of $l(0)$ and $l(2\pi)$ differ by multiples of 2π. Together this shows that as a function from the angle into the lifting-space the en-point starts at some value $x \in [-\pi, 0]$, stays in the white area in Figure 4 and ends in one of the copies of x. A possible value for x is indicated by the black triangle in Figure 4; the copies of x on the right side are the white triangles. Since the function stays in the white strip containing x we know at which of the white triangles the function ends. It follows that in the lifting-space of γ the en-point of l moves up exactly one level during Lovász' procedure, this is another way of saying that the en-point makes one full rotation around Γ. $\triangle$

This lemma will be used next to bound the difference between forward and backward steps during the procedure.

Claim 2. If $|C| = t$ then

$$(a + s_a) - (b + s_b) \leq t. \tag{5}$$

Proof. To see this recall three already proven facts: (1) Every forward step moves forward at least one element in the cyclic order on C. (2) Every backward step moves backward at most one element in the cyclic order on C. (3) As a consequence of Lemma 1 the overall surplus of forward steps cannot exceed $|C|$. $\triangle$

We are ready now to come to an end with the proof of Proposition 1.

$$\begin{array}{rcll} \rho/2 & = & (a + s_a) + (b + s_b) & \text{by (1) and (3)} \\ & \leq & t + 2(b + s_b) & \text{by (5)} \\ & \leq & t + 2r_A & \text{by (4)} \end{array}$$

For the number ρ' of L-segments with exit-point in C the dual argument shows that $\rho' \leq 2t + 4r_A$. With $r = \rho + \rho'$ we obtain the bound claimed in the proposition. $\square$

Next we extend the result to sets of nested curves.

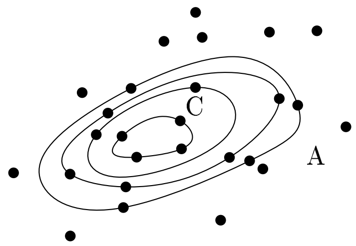


Figure 5: Nested convex curves.

Lemma 2 *Let $\gamma_1 \ldots, \gamma_p$ be closed convex curves where γ_i is contained inside γ_{i-1}, $i = 2, \ldots, p$. The points of S may lie on the curves (subset C) or outside of γ_1 (subset A) see Figure 5. There is a constant c_p such that $r_C \leq c_p(|C| + r_A)$ when r_C denotes the number of k-segments incident to at least one point in C and r_A the number of k-segments with one endpoint in A and one in C.*

Proof. Let r_i be the number of k-segments from γ_i to some $\gamma_j, j \geq i, i = 1, \ldots, p-1$ and let $r_0 = r_A$ be the number of k-segments from A to some

γ_i, $j = 1, \ldots, p$.

Then by Proposition 1

$$r_i \leq c\,(n + s_i)$$

for some constant c, where s_i is the number of segments with one endpoint in $A \cup \gamma_1 \cup \ldots \cup \gamma_{i-1}$ (the outside of γ_i) and one on γ_i.
So $s_i \leq r_0 + \ldots + r_{i-1}$ and we have the recursive inequality

$$\begin{aligned} r_i &\leq c\,(n + \sum_{j=0}^{i-1} r_j) \ \ i = 1, \ldots, p \ \text{ with} \\ r_0 &= r_A. \end{aligned}$$

Certainly $r_i \leq t_i$, $i = 0, 1, \ldots$, where the sequence (t_i) is defined by the same recursive system, except that the $\leq$-inequalities are replaced by equalities.

We have $t_i - t_{i-1} = c\ t_{i-1}$, i.e., $t_i = (c+1)\ t_{i-1}$ and therefore $t_i = (c+1)^{i-1} t_1 = (c+1)^{i-1} c(n + r_A)$ is an upper bound on r_i.

We want an upper bound on the number of segments with at least one endpoint on $\gamma_1 \cup \ldots \cup \gamma_p$, i. e. on

$$\begin{aligned} \sum_{i=0}^{p} r_j &\leq r_A + c(n + r_A) \sum_{i=1}^{p} (c+1)^{i-1} \\ &= r_A + [(c+1)^p - 1](n + r_A) \end{aligned}$$

which proves Lemma 2. $\square$

Next assume that the points of the set S lie on a finite number of closed convex curves which form p nested "clusters" each of the form described in Lemma 2 (see Figure 6). We claim

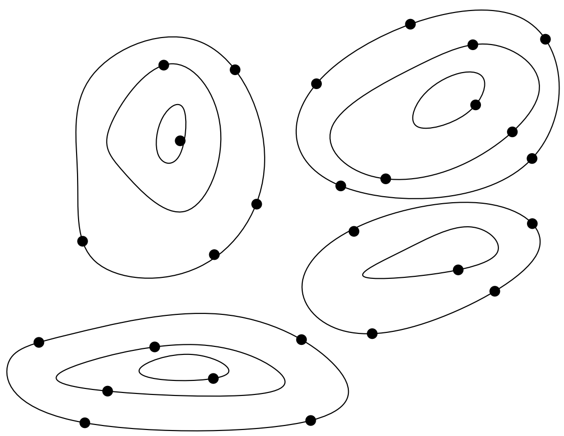


Figure 6: Four clusters of nested convex curves.

Lemma 3 *If the points of S lie on constantly many fixed nested clusters of closed convex curves, then there are $O(n)$ k-sets.*

Proof. Consider one cluster C of the p clusters and let r_C be the number of segments with at least one endpoint on C. C can be separated from any other cluster by a straight line, so by Lovász' Lemma there can be only $O(n)$ segments between the two clusters, so the number r_A of segments with one endpoint in C and one outside of C is $O(n)$. With Lemma 2 we obtain $r_C = O(n + r_A)$ which is $O(n)$. The lemma follows since there are only constantly many clusters. □

Proof **[Theorem].** Let us consider p disjoint convex curves, which may contain the points of the set S. Since S is finite we can assume that all curves are bounded. To the curves we add all vertical lines through their endpoints and all vertical tangents (see Figure 7) The vertical lines decompose the

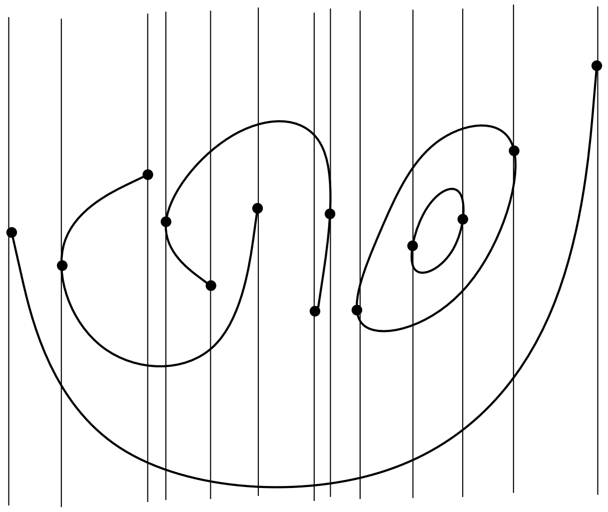


Figure 7: Vertical lines added to a set of convex curves.

plane into constantly many slabs where each slab contains constantly many upward or downward convex segments of curves (see Figure 8). We now add line segments to the curve segments in a slab as shown in Figure 8: If a downward convex segment lies directly above an upward convex segment we connect the endpoints of each of them with a line segment. If the topmost (bottommost) segment is upward (downward) convex we also add the line segment between its endpoints.

Furthermore, from each downward convex segment we construct a closed convex curve by connecting its endpoints by a vertical line segment to the nearest endpoint below (above) of a line segment constructed in the previous phase. Thus, the curve segments obtained by introducing slabs are transformed into a set of clusters of convex curves . The number of clusters and the number of curves within each cluster are bounded by constants.

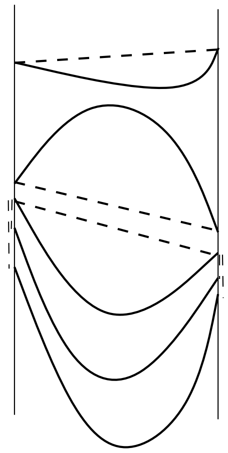


Figure 8: Curve segments and added line segments within a slab.

Observe that, without changing the point set S, the curves can be slightly modified so that they do not touch each other any more. So we can apply Lemma 3 and conclude that S has only $O(n)$ k-segments which proves the theorem. $\square$

Acknowledgment. We would like to thank Michael Godau and Jordi Saludes for helpful discussions concerning this research.

References

[1] P. K. Agarwal, B. Aronov, and M. Sharir, *On levels in arrangements of lines, segments, planes, and triangles*, in Proc. 13th Annu. ACM Sympos. Comput. Geom., 1997.

[2] T. K. Dey, *Improved bounds for k-sets and k-th levels*, in Proc. 38rd Annu. IEEE Sympos. Found. Comput. Sci., 1997.

[3] P. Erdös, L. Lovász, A. Simmons, and E. G. Straus, *Dissection graphs of planar point sets*, in A Survey of Combinatorial Theory, North-Holland, 1973, pp. 139–149.

[4] L. Lovász, *On the number of halving lines*, Ann. Univ. Sci. Budapest, Eötvös, Sect. Math., 14 (1971), pp. 107–108.

[5] J. Pach, W. Steiger, and E. Szemerédi, *An upper bound on the number of planar k-sets*, Discrete Comput. Geom., 7 (1992), pp. 109–123.

[6] M. Sharir, *On k-sets in arrangements of curves and surfaces*, Discrete Comput. Geom., 6 (1991), pp. 593–613.