Allowable Sequences and $k$-sets

Edward Early

11/4/99

**Abstract.** We prove an improved constant on the $n \log n$ lower bound for the maximum number of $k$-sets of a set of $n$ points in the plane by looking at allowable sequences. In particular, we focus on the case where $k$ is the nearest integer to $n/3$. The fact that there is a lower bound of the form $n \log_b n + Cn$ has been known for almost 30 years [**6**], but the smallest known value of $b$ was 3 [**4**]. In this paper, we show that $b$ can be any number greater than $2\sqrt{2}$. The best known upper bound was recently improved by Dey to $O(n\sqrt[3]{n})$ [**3**], and the best known constant on the upper bound is due to a result of Pach and Tóth [**8**]. The best known lower bound has just been improved by Tóth to $ne^{c\sqrt{\log n}}$ [**9**].

**1. Introduction.** The $k$-set problem, described below, has been called one of the most tantalizing open problems in combinatorial geometry. It is important not only as a purely combinatorial problem, but also as a computational geometry problem due to its applications in analyzing geometric algorithms [**1**], [**2**], [**5**].

Given a collection of $n$ points in the plane, and a positive integer $k$ less than $n$, there will be some lines in the plane that separate the $n$ points into sets of size $k$ and $n-k$. Subsets of $k$ points found this way are called *$k$-sets.* The question we are concerned with is to find the maximum possible number of $k$-sets for a fixed $n$. Ideally, we would like an algorithm for placing $n$ points in the plane that achieve the maximum.

The way we will approach this problem is to look at a closely related problem about allowable sequences, discussed in Section 2. Given $n$ points in the plane, we can label them 1 through $n$ by the order of their appearance from left to right (increasing order of the $x$-coordinates). If we rotate the plane $180^\circ$ about any point, then the order of the points is reversed. Suppose we record the order of the points as they are rotated. Then we start with $1, \ldots, n$, end with $n, \ldots, 1$, and along the way we only switch two adjacent numbers at a time for a total of $\binom{n}{2} = \frac{n\cdot(n-1)}{2}$ switches. Any sequence of permutations satisfying the properties in the last sentence is a *simple allowable sequence* of permutations of the numbers 1 through $n$. This terminology was introduced by Goodman and Pollack [**7**]. A simple allowable sequence that can be obtained from a set of points in the manner described above is called *geometrically realizable.* Unfortunately, some allowable sequences are not geometrically realizable. Therefore this approach to the $k$-set problem can at best produce an upper bound unless we give a proof of geometric realizability, as we will do in Section 3. A good exercise is to find a simple allowable sequence of permutations of 12345 that is not geometrically realizable.

**2. Allowable Sequences.** We start by establishing a recurrence for a lower bound for the maximum number of switches at the $k$'th spot in a simple allowable sequence of permutations of the numbers 1 through $n$. In other words, if we want to reverse the order of $n$ numbers, switching only two adjacent numbers at a time and using as few total switches as possible, then what is the maximum number of times we can switch the $k$'th and $(k+1)$'st numbers for some $k$? Let $f(n)$ be the maximum when $k$ is the nearest integer to $n/3$. We make use of the fact that maximizing the number of switches at the $k$'th spot is equivalent to maximizing the number of switches at the $(n-k)$'th spot by symmetry.

**Proposition 2-1.** $f(2n+1) \geq 2f(n) + \frac{4n}{3}$.

*Proof:* In sorting $2n+1$ numbers, look at the first $n$ as one block and the last $n$ as another block, leaving one in the "middle." We start by sorting the last block of $n$ numbers for $f(n)$ switches at the $(n-k)$'th spot (now $k$ is the nearest integer to $(2n+1)/3$). Then we move the remaining $n+1$ numbers to the right, but the middle number switches back and forth at the $(n-k)$'th spot for an additional $2(n-k)$ switches. The middle number then moves back to the middle, so we end with the first block of $n$ numbers on the right side. Finally, we sort the first block of $n$ numbers for $f(n)$ more switches, as desired. □

For example, with $n = 4$ we have two blocks of four numbers and the number 5 in the middle, so we get 12 switches of the sixth and seventh numbers (or equivalently the 3rd and 4rd) as follows. 987654321, 98765**34**21, 987653**24**1, 98765**23**41, 9876523**14**, 9876521**3**4, 98765**12**34, 98761**5**234, 98761**25**34, 987**16**2534, 98**17**62534, 9**18**762534, **19**8762534, 1987**26**534, 19872**56**34, 198**27**5634, 19**28**75634, 1**29**875634, 129875**36**4, 129873**5**64, 1298**37**564, 12983**57**64, 129**38**5764, 12**39**85764, 1239857**46**, 1239854**76**, 12398**45**76, 1239**48**576, 12394**58**76, 123**49**5876, 1234**59**876, 12345**89**76, 123458**79**6, 12345**78**96, 1234578**69**, 123457**68**9, 12345**67**89

**Lemma 2-2.** *If $f$ is a function defined on the positive integers, $f(2n+1) \geq 2f(n) + kn$, and $b > 4^{1/k}$, then there is a constant $C$ such that $f(n) \geq n \log_b n + Cn$ for all $n$ of the form $2^m - 1$ for some integer $m$.*

*Proof:* It suffices to prove the statement for $n$ sufficiently large, since we can always take a large enough negative value of $C$ to force the lower bound to work up to any finite $n$. Let $g(n) = n \log_b n + Cn$, and assume without loss of generality that $g(n) \leq f(n)$. If we can show that $g(2n+1) \leq 2g(n) + kn$, then it will follow that $g(2n+1) \leq 2g(n) + kn \leq 2f(n) + kn \leq f(2n+1)$, so the bound works. To this end, we see that $g(2n+1) \leq 2g(n) + kn$ iff $(2n+1)\log_b(2n+1) + C(2n+1) \leq 2n \log_b n + 2Cn + kn$ iff $\log_b \frac{(2n+1)^{2n+1}}{n^{2n}} \leq kn - C$. The left side of this inequality equals $\log_b \left( n \left( \frac{2n+1}{n} \right)^{2n+1} \right) = \log_b n + (2n+1)\log_b(2 + \frac{1}{n}) = \log_b n + \frac{(2n+1)}{\log b}(1 + \frac{1}{n} - \frac{(1+\frac{1}{n})^3}{3} + - \cdots) = \log_b n + \frac{(2n+1)}{\log b}(1 - \frac{1}{3} + \frac{1}{5} - + \cdots + O(\frac{1}{n})) = \log_b n + \frac{(2n+1)}{\log b}(\log 2 + O(\frac{1}{n})) = \log_b n + (2n+1)\log_b 2 + O(1)$. Thus $g(2n+1) \leq 2g(n) + kn$ iff $\log_b n + 2n \log_b 2 + O(1) \leq kn - C$. We absorb the $O(1)$ term into $C$ to get the equivalent statement $\log_b n + n \log_b 4 \leq kn - C$, or $\log_b n + C \leq (k - \log_b 4)n$. Since $\log n$ is eventually less than $cn$ for any $c > 0$, it follows that the inequality is true iff $k - \log_b 4 > 0$, or $b > 4^{1/k}$, as desired. □

One convenient way to look at a simple allowable sequence is to use *words*. Instead of recording the entire permutations, we just keep track of what switches we performed (and the order in which we performed them). For example, the sequence written out above is just

**6**7**6**87**6**5**6**43215**6**4327**6**5**6**4387**6**5**6**45**6**7**6**87**6**.

Notice that there are twelve 6's, since we switched the sixth and seventh positions twelve times. Standard notation is to write $w_i$ instead of $i$ above to indicate that we are performing the elementary permutation $w_i$ that switches the $i$'th and $(i+1)$'st numbers, but when there is no risk of confusion it is easier just to omit the $w$'s. In fancier terms, the above word is a *reduced word for the longest element*, where the "longest element" is the permutation that reverses the order of 123456789, and "reduced" means that we can't do it with any fewer elementary permutations.

We can use the weak upper bound

$$f(n) \leq \left\lceil \frac{1}{2} \left( \frac{n(n-1)}{2} - \left( \left\lceil \frac{n}{2} \right\rceil \left\lceil \frac{n}{2} + 1 \right\rceil /2 + \left\lceil \frac{n+1}{3} - 2 \right\rceil \left\lceil \frac{n+1}{3} - 1 \right\rceil \right) \right) \right\rceil$$

to prove that we have the best possible result in some small cases. The upper bound comes from the easily verified facts that if we let $h(n,l)$ be the number of switches at the $l$'th spot when sorting $n$ numbers, then $h(n,l) \geq l$, $h(n, n-l) \geq l$, and $h(n,l) + h(n, n-l) \geq 3l$ for $l \leq n/2$, and

$h(n,l) \le h(n,l-1) + h(n,l+1) + 1$. The verification of these facts is left as an exercise for the reader, with the hint to think about what they are saying in terms of reduced words. The lower bounds come from the proof of Proposition 2-1, and the fact that $f(2) = 1$ and $f(3) = 2$.

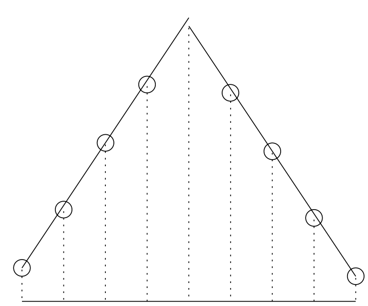
| $n$ | lower | upper |
|---|---|---|
| 2 | 1 | 1 |
| 3 | 2 | 2 |
| 4 | 3 | 3 |
| 5 | 5 | 5 |
| 6 | 6 | 6 |
| 7 | 8 | 8 |
| 8 | 10 | 10 |
| 9 | 12 | 12 |
| 10 | 14 | 14 |
| 11 | 16 | 17 |
| 12 | 19 | 20 |

It is interesting to note that numerical data suggests that the maximum number of $k$-sets of a set of $n$ points may occur when $k$ is the nearest integer to $n/3$, rather than $n/2$, though the order of growth must be the same in both cases.

**3.** ***k*-sets.**

**Lemma 3-1.** *The allowable sequence described in the proof of Proposition 2-1 is geometrically realizable.*

*Proof:* (Due to Geza Tóth) We show that we can always arrange the points so that they almost lie equally spaced on a line, at least when $n$ is of the form $2^m - 1$. For clarity, this means that we have $\epsilon$ neighborhoods of $n$ points spaced one unit apart along a line, and each of those neighborhood contains exactly one of the $n$ points of the set that we are interested in. This is clearly true for $n = 3$, since we can start with the vertices of an equilateral triangle and apply an affine transformation to move the third point as close to the midpoint of the first two as we want. Suppose it works for $n = 2^m - 1$; then we take two lines with $n$ points almost equally spaced and room for one more point at one end, and place them together as shown in Figure 3.



Let $k$ be the nearest integer to $2n/3$. Now we need to show that we can find one more point, which will correspond to the middle number in the allowable sequence point of view, such that the lines connecting it to the top point on the left side, then the $(k-1)$'st point on the right side, the second point the left side, the $(k-2)$'nd point on the right side, etc., have slopes in that order. This produces $k$-sets corresponding to the middle number switching back and forth. Clearly this would be possible if the two lines were parallel and any distance apart, with the right line slightly lower than the left one. Thus by continuity we can do it for some sufficiently small angle, since there are no points in a neighborhood of the intersection of the two lines.

More precisely, suppose the lines actually meet at the top point. Call the angle between the two lines $\theta$; then the balls project down to balls of radius $\epsilon$ around points that are $\sin(\theta/2)$ units apart. Since $\epsilon$ is arbitrary, we can make it small enough relative to $\sin(\theta/2)$, and put a similarly small gap between the lines at the top. We can now flatten the picture with an affine transformation to get a new line with $2n+1$ points approximately equally spaced along it and yielding the desired number of $k$-sets. □

Combining these results yields the following theorem.

**Theorem 3-2.** *Given any $b > 2\sqrt{2}$, there exists a constant $C$ such that for any $n$ of the form $2^m - 1$, there exists a set of $n$ points with at least $n \log_b n + Cn$ $k$-sets, where $k$ is the nearest integer to $n/3$.*

The theorem can be extended to arbitrary $n$, but the proof is not particularly enlightening. We only needed to restrict to $n = 2^m - 1$ to make iterating the algorithm easy. The best known upper and lower bounds are still far apart, so there is plenty of work still to be done. In higher dimensions, we can consider the analogous problem with hyperplanes instead of lines, but very little is known about higher dimensional $k$-sets.

**4. Acknowledgements.** I wish to thank Gil Kalai, Sara Billey, Geza Tóth, and Henry Cohn for their helpful comments and suggestions.

## References

[**1**] B. Chazelle and F. P. Preparata. Half-space range search: An algorithmic application of $k$-sets. *Discrete Comput. Geom.*, 1 (1986), 83–93.

[**2**] R. Cole, M. Sharir, and C. K. Yap. on $k$-hulls and related problems. *SIAM J. Comput.*, 16 (1987), 61–77.

[**3**] T. K. Dey. Improved Bounds for Planar $k$-Sets and Related Problems. *Discrete Comput. Geom.*, 19 (1998), 373–382.

[**4**] H. Edelsbrunner. *Algorithms in Combinatorial Geometry.* Springer-Verlag, Berlin, 1987.

[**5**] H. Edelsbrunner and E. Welzl. On the number of line separations of a finite set in the plane. *J. Combin. Theory, Ser. A*, 38 (1985), 15–29.

[**6**] P. Erdös, L. Lovász, A. Simmons, and E.G. Straus. Dissection graphs of planar point sets. In *A Survey of Combinatorial Theory* (J. N. Srivastava et al., eds.), North-Holland, Amsterdam, 1973, pp. 139–149.

[**7**] J. Goodman and R. Pollack. On the combinatorial classification of nondegenerate configurations in the plane, *J. Combin. Theory, Ser. A*, 29 (1980), 220–235.

[**8**] J. Pach and G. Tóth. Graphs drawn with few crossings per edge. In: *Proc. Graph Drawing 1996*, Berkeley (S. North ed.), Lecture Notes in Computer Science, vol. 1190, Springer-Verlag, Berlin, 1997, pp. 345–354.

[**9**] G. Tóth. Personal communication.