Reprinted from JOURNAL OF COMBINATORIAL THEORY, Series A
All Rights Reserved by Academic Press, New York and London

Vol. 25, No. 2, September 1978
Printed in Belgium

Note

Some Combinatorial Problems in the Plane

P. ERDÖS

Hungarian Academy of Science

AND

G. PURDY

Texas A & M University, College Station, Texas 77843

Communicated by the Editors

Received March 26, 1977

1

Let there be given n points in the plane. Denote by t_i the number of lines which contain exactly i of the points $(2 \leqslant i \leqslant n)$. The properties of the set $\{t_i\}$ have been studied a great deal. For example, there is the classical result of Gallai and Sylvester: Assume $t_n = 0$ (i.e., the points are not all on one line); then $t_2 > 0$. For the history of this problem see, e.g., Motzkin [6] and Erdös [3, 4]. In this note we prove that some new and perhaps unexpected properties of the family $\{t_i\}$ hold.

Let there be given n distinct points in the plane, not all on a line. We conjectured that for $n > n_0$ there always is an i such that $t_i \geqslant n - 1$. Krier and Straus pointed out that for $n = 6$ and 9 there are counterexamples. For $n = 9$, take the vertices of a square and its center and the four points of infinity determined by the sides and diagonals of the square. For $n = 13$ we also get a counterexample from the vertices of a regular hexagon and its center and the six points of infinity determined by the sides and diagonals of the hexagon. Nevertheless, the conjecture is true for $n \geqslant 25$. In fact we show in Theorem 1 that we can always choose $i = 2$ or 3, and that

$$\max_i t_i = \max(t_2, t_3).$$

Assume now that no line has more than $(1 - \epsilon)\eta$ points on it. Then we are convinced that there is an $\eta = \eta(\epsilon)$ such that

$$t_2 + 3t_3 > \eta \binom{n}{2}. \tag{1}$$

0097-3165/78/0252-0205$02.00/0

Since $\sum_{i=2}^{n} \binom{i}{2} t_i = \binom{n}{2}$ we can rewrite (1) as

$$t_2 + 3t_3 > \frac{\eta}{1-\eta} \sum_{i=4}^{n} \binom{i}{2} t_i . \tag{2}$$

A result of Sylvester and Burr *et al.* [2] shows that there are examples with $t_2 \leqslant n - 3$. Hence (1) is definitely false if $t_2 + 3t_3$ is replaced by t_2. What we are able to show (see Theorem 1) is that $t_3 \geqslant cn^2$ whenever $t_2 < n - 1$. We determine c explicitly, but we probably do not have the best value of c. If $t_2 = n - 1$ we can of course have $t_i = 0$, $3 \leqslant i < n - 1$, $t_{n-1} = 1$, i.e., $n - 1$ points on a line.

We also conjecture that $t \geqslant cn$ whenever $t_3 \geqslant t_2$, where t is the total number of lines, i.e., $t = \sum_{i=2}^{n} t_i$. Perhaps this conjecture is too optimistic. It has been conjectured that $t_2 \geqslant n/2$ always holds, and an example of Motzkin shows that if true it is best possible. It gives a set of $2n$ points with $t_2 = n$, $t_3 = \binom{n}{2}$ and $t_n = 1$. We think that if $t_m > 1$ for some m, $(n/2)(1 + \epsilon) < m < (1 - \epsilon)n$, then $t_2 > c_\epsilon n^2$.

We start with a lemma.

LEMMA 1. *If r points lie on a line l and s points do not lie on l then $t_2 \geqslant rs - s(s-1)$.*

Proof. The result is true if $s = 0$ or $s = 1$. We shall suppose $s > 1$ and use induction on s. Let P be a point not on l. If we remove P, then $t_2 \geqslant r(s-1) - (s-1)(s-2)$. The addition of P will spoil at most $s - 1$ of these lines and will introduce r new lines at most $s - 1$ of which will contain three or more points. Hence $t_2 \geqslant r(s-1) - (s-1)(s-2) - (s-1) + rs - (s-1) = rs - s(s-1)$, and the result follows by induction.

COROLLARY. *Given n points in the plane, not all on a line, if a line contains exactly $(n/2)(1+\epsilon)$ points, then $t_2 \geqslant (1-\epsilon)\,\epsilon n^2/2$.*

Proof. Put $r = (n/2)(1+\epsilon)$ and $s = (n/2)(1-\epsilon)$ in the lemma, and we have $t_2 \geqslant rs - s(s-1) = [(1-\epsilon)\epsilon/2]\, n^2 + (n/2)(1-\epsilon)$.

THEOREM 1. *Given n points in the plane, not all on a line, if $n \geqslant 25$ then* $\max(t_2, t_3) \geqslant n - 1$. *For all n, if $t_2 < n - 1$, then $t_3 \geqslant cn^2$ where c is a positive constant. Also, we always have* $\max_i t_i = \max(t_2, t_3)$.

Proof. We shall suppose that $t_2 < n - 1$. We start by showing that $t_i = 0$ for $i \geqslant \frac{3}{4}n$. To see this, let l be a line with at least $\frac{3}{4}n$ of the points on it. If l has $n - 1$ points, then $t_2 = n - 1$. Hence we may suppose that there are two points P and Q not on l. If we restrict our attention to the points on l, to P and Q, then we get, by Lemma 1, $t_2 \geqslant \frac{3}{2}n - 2$. Each of the $n/4 - 2$ points not in $l \cup \{P, Q\}$ lies on at most two of these lines. Hence $n - 2 \geqslant$

$t_2 \geqslant \frac{3}{2}n - 2 - 2(n/4 - 2) = n + 1$, which is absurd. Hence $t_i = 0$ for $i \geqslant \frac{3}{4}n$.

From [5] we have

$$(n - 2) \geqslant t_2 \geqslant 3 + \sum_{i \geqslant 4} (i - 3)\, t_i\,, \tag{3}$$

and clearly

$$t_2 + 3t_3 + \sum_{i \geqslant 4} \binom{i}{2} t_i = \binom{n}{2}. \tag{4}$$

We consider now how large $S = \sum_{i \geqslant 4} \binom{i}{2} t_i$ can be if (3) holds. If we choose $t_i = 0$, $4 \leqslant i < k$, $k = \frac{3}{4}n$ and choose t_k as large as possible, we will get an upper bound on S which is valid even if k is not an integer. We then have $n - 2 = 3 + (k - 3)\, t_k$, $t_k = (n - 5)/(k - 3) < 4/3$ and $S = \binom{k}{2} t_k \leqslant \frac{1}{2}(\frac{3}{4}n - 1)(\frac{3}{4}n - 2)\,\frac{4}{3} \leqslant \frac{3}{8}(n - 1)(n - 2)$.

To show that $\max(t_2, t_3) \geqslant n - 1$ it is enough to obtain a contradiction from the assumption that $t_3 \leqslant n - 2$. From (4) we obtain

$$n - 2 + 3(n - 2) + S \geqslant \binom{n}{2},$$

$$4(n - 2) + \tfrac{3}{8}(n - 1)(n - 2) \geqslant \tfrac{1}{2}n(n - 2),$$

which reduces to $n^2 - 27n + 58 \leqslant 0$, which is false for $n \geqslant 25$.

To show that $t_3 \geqslant cn^2$, we procede as follows:

We have $S \leqslant \frac{3}{8}n^2$, and from (4)

$$n - 2 + 3t_3 + \frac{3}{8}\, n^2 \geqslant \binom{n}{2},$$

$$3t_3 \geqslant \frac{1}{2}\, n^2 - \frac{1}{2}\, n - \frac{3}{8}\, n^2 - n + 2 = \frac{1}{8}\, n^2 - \frac{3}{2}\, n + 2,$$

$$t_3 \geqslant \frac{1}{24}\, n^2 - \frac{n}{2} - \frac{2}{3} \geqslant cn^2 \qquad \text{for} \quad n > n_0\,,$$

where c is any constant less than $\frac{1}{24}$ and n_0 depends only on c.

Finally, we remark that the assertion $\max_i t_i = \max(t_2, t_3)$ follows from the second inequality in (3).

We conclude this section be stating an old conjecture of one of the authors: Assume $t_l = 0$ for all $l > k \geqslant 5$. Then $t_k = o(n^2)$. This is certainly false for $k = 3$ (Sylvester and Burr *et al.*, see [2]). Croft and Erdös observed that it is false if the assumption $t_l = 0$ is not made. (This is in fact shown by the lattice points in the plane.)

Kártezi proved that $t_k > cn \log n$ is possible and Grünbaum showed recently that $t_k > cn^{1+1/k}$ is possible.

2

Let there be given n distinct $X_1, \ldots, X_n$ in the plane not all on a line, let $L_1, \ldots, L_m$ be the lines determined by the points, and let $|L_i|$ denotc the number of points on the line L_i. Assume $|L_1| \geqslant |L_2| \geqslant \cdots \geqslant |L_m|$. The result of Sylvestor and Gallai states that $|L_m| = 2$, and that $m \geqslant n$.

It would be interesting to determine or estimate the number $A(n)$ of families of sets $\{|L_1|, \ldots, |L_m|\}$ determined by n points in the plane. The difficulty is that it is not at all clear to us (and as far as we know to anybody else) what conditions a sequence $\alpha_1, \alpha_2, \ldots, \alpha_m$ must satisfy in order that there should be a set of points $X_1, \ldots, X_n$ with $|L_i| = \alpha_i$, $1 \leqslant i \leqslant m$.

These problems can be restated in a more combinatorial form. Let $\mathscr{S}$ be a set, $|\mathscr{S}| = n$ and let $A_i \subset \mathscr{S}$, $2 \leqslant |A_i| < n$ for $1 \leqslant i \leqslant m$ be a family of subsets of $\mathscr{S}$ so that every pair $x, y \in \mathscr{S}$ is contained in one and only one of the A_i. A theorem of de Bruijn and Erdös [1] states that $m \geqslant n$. Now let $F(n)$ be the smallest integer for which there is a system $\{A_i\}$ with the above properties so that for every r $(2 \leqslant r \leqslant n - 1)$ the number of indices i with $|A_i| = r$ is at most $F(n)$. It would be interesting to determine or estimate $F(n)$. We conjecture

$$F(n) = cn^{1/2} + o(n^{1/2}). \tag{5}$$

It is a simple consequence of the theorem of de Bruijn and Erdös that $F(n) > c_1 n^{1/2}$, but we could not prove $F(n) < c_2 n^{1/2}$. Perhaps this will not be so very difficult, but we were only able to prove $F(n) < cn^{3/4}$, which is our Theorem 2.

It would be interesting to determine or estimate the number $B(n)$ of sequences

$$\{\alpha_1, \ldots, \alpha_m\}, \qquad |\alpha_1| \geqslant \cdots \geqslant |\alpha_m| \geqslant 2 \tag{6}$$

for which there is a system $A_i \subset \mathscr{S}$, $|A_i| = \alpha_i$, so that every pair (x, y) of $\mathscr{S}$ is contained in one and only one A_i. As in the geometric case we do not have any necessary and sufficient conditions for a sequence (6) which would ensure the existence of a corresponding system $\{A_i\}$. We are convinced that $B(n)$ is very much larger than $A(n)$.

THEOREM 2. $F(n) < cn^{3/4}$.

Proof. We use the probability method and only outline the proof. Let p be the largest prime for which $p^2 + p + 1 < 2n$, and put $m = p^2 + p + 1$. Consider a finite geometry of m points and m $(p + 1)$-tuples. Choose at random n of the points. One can do this in $\binom{m}{n}$ ways. It is not hard to show that all but $o(\binom{m}{n})$ choices have the property that there are fewer than $cn^{3/4}$ lines with the same number of points. The computations are somewhat laborius and we suppress them, since this method cannot give any better result than $F(n) < cn^{3/4}$ and we are certain that $F(n) < cn^{1/2}$ is true.

3

Let there be given n points in the plane, not all on a line, and form all the connecting lines. We ask how many points are needed to represent all the lines, if a point represents a line by being on it, and if none of the original n points can be used for representation. Let $f(n)$ be the minimum of the representation numbers taken over all configurations of n points not all on a line. The example of $n - 1$ points on a line and one other point shows that $f(n) \leqslant n$. We conjecture that $f(n) \geqslant cn$ for some $c > 0$, but we can only prove

THEOREM 3. $f(n) \geqslant n^{1/2}$.

Proof. If there are $n^{1/2}$ points on a line, then take a point P not on the line and at least $n^{1/2}$ points are needed to represent the lines through P. If there are not $n^{1/2}$ points on a line, then pick any point Q. Then Q has at least $n^{1/2}$ lines going through it, and at least $n^{1/2}$ points are needed to represent all of these lines. This concludes the proof.

Let $g(n)$ denote the minimum number of different directions determined by n points in the affine plane not all on a line. Scott has shown [7] that

$$\frac{1}{2}\{1 + (8n - 7)^{1/2}\} \leqslant g(n) \leqslant 2\left[\frac{n}{2}\right].$$

We show here

THEOREM 4. $g(n) \geqslant f(n)$.

Proof. Let n points, not all collinear, be given in the affine plane. Each collection of parallel lines intersects the line at infinity in a single point. If there are m directions, then m points at infinity suffice to represent all the points, and these points are distinct from the original points. Hence $g(n) \geqslant f(n)$.

We may alter the definition of $f(n)$ by insisting that no three points of the original configuration lie on line, obtaining $h(n)$. Clearly $f(n) \leqslant h(n)$. Grünbaum pointed out that the regular n-gon and the regular n-gon with center shows that $h(n) \leqslant 2[n/2]$. We also conjecture that $h(n) \geqslant cn$, and this might be easier to prove than $f(n) \geqslant cn$.

Note added in proof. G. R. Burton and G. Purdy have recently proved that $g(n) \geqslant [n/2]$.

REFERENCES

1. N. G. DE BRUIJN AND P. ERDÖS, On a combinatorial problem, *Nederl. Akad. Wetensch.* **51** (1948), 1277–1279.

2. S. Burr, B. Grünbaum, and N. J. A. Sloane, The orchard problem, *Geometriae Dedicata* **2** (1974), 397–424.
3. P. Erdös, On a problem of Grünbaum, *Canad. Math. Bull.* **15** (1972), 23–25.
4. P. Erdös, On some problems of elementary and combinatorial geometry, *Ann. di Mat. Ser. IV* **103** (1975), 99–108.
5. L. M. Kelly and W. O. J. Moser, On the number of ordinary lines determined by *n* points, *Canad. J. Math.* **10** (1958), 210–219.
6. Th. Motzkin, The lines and planes connecting the points of a finite set, *Trans. Amer. Math. Soc.* **70** (1951), 451–464.
7. P. R. Scott, On the sets of directions determined by *n* points, *Amer. Math. Monthly* **77** (1970), 502–505.