COMBINATORICA 3 (3—4) (1983) 381—392

EXTREMAL PROBLEMS IN DISCRETE GEOMETRY

ENDRE SZEMERÉDI and WILLIAM T. TROTTER, JR.*

Dedicated to Paul Erdős on his seventieth birthday

Received 19 August 1982

Revised 14 March 1983

In this paper, we establish several theorems involving configurations of points and lines in the Euclidean plane. Our results answer questions and settle conjectures of P. Erdős, G. Purdy, and G. Dirac. The principal result is that there exists an absolute constant c_1 so that when $\sqrt{n} \leq t \leq \binom{n}{2}$, the number of incidences between n points and t lines is less than $c_1 n^{2/3} t^{2/3}$. Using this result, it follows immediately that there exists an absolute constant c_2 so that if $k \leq \sqrt{n}$, then the number of lines containing at least k points is less than $c_2 n^2/k^3$. We then prove that there exists an absolute constant c_3 so that whenever n points are placed in the plane not all on the same line, then there is one point on more than $c_3 n$ of the lines determined by the n points. Finally, we show that there is an absolute constant c_4 so that there are less than $\exp(c_4\sqrt{n})$ sequences $2 \leq y_1 \leq y_2 \leq \ldots \leq y_t$ for which there is a set of n points and a set $l_1, l_2, \ldots, l_t$ of t lines so that l_j contains y_j points.

1. Introduction

Extremal problems in discrete geometry are among the most natural and most tantalizing areas of research in combinatorial mathematics. The reader is encouraged to consult P. Erdős' survey papers [2], [3], [4], [5] and W. Moser's summary [6] of recent progress on a wide range of problems in this area. In this paper, we concentrate on extremal problems in the Euclidean plane. We begin by establishing the following theorem limiting the number of incidences between a set of n points and a family of t lines in the Euclidean plane.

Theorem 1. *There exists a constant c_1 so that if $\mathcal{P}$ is a set of n points and $\mathcal{L}$ is a family of t lines in the Euclidean plane, then the number of incidences between points in $\mathcal{P}$ and lines in $\mathcal{L}$ is at most $c_1 n^{2/3} t^{2/3}$ whenever $\sqrt{n} \leq t \leq \binom{n}{2}$.*

P. Erdős [5] had conjectured the conclusion of Theorem 1 in the special case $t=n$. From Theorem 1, we obtain as an immediate corollary the following theorem settling in the affirmative a conjecture of Erdős and Purdy [5].

* Research supported in part by NSF Grants ISP—8011451 and MCS—8202172.

AMS subject classification (1980): 51 M 05, 05 C 35

Theorem 2. *There exists a constant c_2 so that if $\mathcal{P}$ is a set of n points and $\mathcal{L}$ is a family of t lines each containing at least k points from $\mathcal{P}$ where $k \leqq \sqrt{n}$, then $t < c_2 n^2/k^3$.*

A second corollary is the following result which is a partial solution to Dirac's conjecture [5]. (This result was also proved by J. Beck [1]).

Theorem 3. *There exists a constant $c_3 > 0$ so that if $\mathcal{P}$ is a set of n points, not all on the same line, and $\mathcal{L}$ is the family of all lines determined by $\mathcal{P}$, then there exists at least one point in $\mathcal{P}$ which belongs to more than $c_3 n$ lines from $\mathcal{L}$.*

Let $\mathcal{P}$ be a set of n points and let $\mathcal{L} = \{l_1, l_2, \ldots, l_t\}$ be a family of lines each containing at least two points from $\mathcal{P}$. For each $j = 1, 2, \ldots, t$, let y_j count the number of points from $\mathcal{P}$ which are on line l_j. We may then assume that the lines have been labelled so that $y_1 \leqq y_2 \leqq \ldots \leqq y_t$. Then $\sum_{i=1}^{t} \binom{y_i}{2} \leqq \binom{n}{2}$. However, this condition is not sufficient to insure that an arbitrary nondecreasing sequence arises in this fashion. Let $\mathcal{E}(n)$ count the number of distinct nondecreasing sequences $y_1 \leqq y_2 \leqq \ldots \leqq y_t$ determined by all possible configurations of n points in the plane. We use Theorem 1 to prove the following result which also settles in the affirmative a conjecture of P. Erdős:

Theorem 4. *There exists a constant c_4 so that $\mathcal{E}(n) < 2^{c_4 \sqrt{n}}$ for all $n \geqq 1$.*

2. The Covering Lemma

The principal tool used in establishing Theorem 1 is a covering lemma proved by the authors in [7]. We assume that a pair of coordinate axes has been chosen and when we use the term *square*, it will also be assumed that the sides are parallel to the coordinate axes. The following lemma asserts the existence of a family of squares covering a positive fraction of a set of n points with an additional restriction on the number of points contained in each square. We refer the reader to [7] for the proof of the lemma.

Lemma. *Let r_1, r_2 be integers with $r_2 \geqq 256\, r_1$, and let $\mathcal{P}$ be any set of n points in the plane. Then there exists a family $\mathcal{Q}$ of squares so that:*

1. *No point in the plane is in the interior of more than one square;*
2. *Each square contains at least r_1 but no more than r_2 points of $\mathcal{P}$;*
3. *At least $n/16$ of the points in $\mathcal{P}$ are covered by the squares in $\mathcal{Q}$.* ∎

The covering lemma will be used to localize the intersection patterns of lines. In order to apply the lemma, we take full advantage of the fact that linear transformations can be applied to any configuration of points and lines. Thus whenever we need to, we can rotate the configuration, and we can increase or decrease the slope of the lines in the configuration.

3. The Principal Theorem

In this section we present our principal theorem.

Theorem 1. *There exists a constant c_1 so that if $\mathscr{P}$ is a set of n points and $\mathscr{L}$ is a family of t lines with $\sqrt{n} \leqq t \leqq \binom{n}{2}$, then the number of incidences between the points in $\mathscr{P}$ and the lines in $\mathscr{L}$ is less than $c_1 n^{2/3} t^{2/3}$.*

Proof. We will show that the conclusion holds when $c_1 = 10^{60}$. The proof will be by contradiction. We assume the theorem is false and choose a counterexample with $n+t$ as small as possible. We then show that a large fraction of the points in $\mathscr{P}$ are incident with a large number of lines from two subsets $\mathscr{L}_1$ and $\mathscr{L}_2$. After a suitable linear transformation, the lines in $\mathscr{L}_1$ will have slope very nearly $+1$ and the lines in $\mathscr{L}_2$ will have slope very nearly -1. The covering lemma will then be used to show that more pairs of lines cross where there is no point of $\mathscr{P}$ than there are pairs of lines in $\mathscr{L}$.

Choose a coordinate system consisting of a horizontal line (x-axis) and a vertical line (y-axis). Then we may assume without loss of generality that all lines have slope in the interval $(-1/2, 1/2)$. (If this condition is not satisfied, we apply a linear transformation of the form $(x, y) \to (x, \varepsilon y)$ where ε is a small positive number.)

We label the points $p_1, p_2, \ldots, p_n$ and the lines $l_1, l_2, \ldots, l_t$. Without loss of generality, we may assume that if $j_1 < j_2$, then the slope of l_{j_1} is at least as large as the slope of l_{j_2}. For each $i = 1, 2, \ldots, n$, we define the *degree* of p_i, denoted d_i, as the number of lines in $\mathscr{L}$ which contain p_i. For each $j = 1, 2, \ldots, t$, we define the *density* of l_j, denoted y_j, as the number of points in $\mathscr{P}$ which are on the line l_j. We let I denote the total number of incidences between points and lines. Then

$$c_1 n^{2/3} t^{2/3} \leqq I = \sum_{i=1}^{n} d_i = \sum_{j=1}^{t} y_j .$$

If l_{j_1} and l_{j_2} are lines for which there is some point $p_i \in \mathscr{P}$ common to both lines, we say l_{j_1} and l_{j_2} have a *good* intersection. If l_{j_1} and l_{j_2} intersect where there is no point of $\mathscr{P}$, we say the lines have a *wasted crossing*. We let G denote the total number of good intersections and let W denote the total number of wasted crossings. Clearly, $G = \sum_{i=1}^{n} \binom{d_i}{2}$ and $W \leqq W + G \leqq \binom{t}{2}$. Thus:

$$\frac{t^2}{2} \geqq \binom{t}{2} \geqq \sum_{i=0}^{n} \binom{d_i}{2} = \sum_{i=1}^{n} \frac{1}{2} d_i^2 - \frac{1}{2} \sum_{i=1}^{n} d_i$$

$$\geqq \frac{1}{2n} \left(\sum_{i=1}^{n} d_i \right)^2 - \frac{1}{2} \sum_{i=1}^{n} d_i = \frac{1}{2n} I^2 - \frac{1}{2} I \geqq \frac{1}{3n} I^2 \geqq \frac{c_1^2}{3} n^{1/3} t^{4/3}.$$

In particular, it follows that $t \geqq \frac{c_1^3}{10} \sqrt{n}$. At this stage, it is important to note that t is already known to be a reasonable distance away from the boundary value $t = \sqrt{n}$.

Let y_A denote the *average density* of a line in $\mathscr{L}$, i.e., $y_A = I/t$. We show that every line has density at least as large as $.6y_A$. Suppose to the contrary that there are xt lines which have density less than $.6y_A$ where $0<x<1$. Let $x_1 = \min\{x, 1/2\}$ and choose a set of $x_1 t$ lines each of which has density less than $.6y_A$. Since $\binom{n}{2} > t - x_1 t \geq t/2 \geq \sqrt{n}$, we know that the number of incidences involving the remaining $t - x_1 t$ lines is less than $c_1 n^{2/3}(t-x_1 t)^{2/3}$. Meanwhile, the number of incidences involving the $x_1 t$ lines with small density is less than $(x_1 t)(.6\, y_A)$. This requires:

$$(x_1 t)(.6y) + c_1 n^{2/3}(t - x_1 t)^{2/3} > I.$$

Thus there is a number x_1 with $0 < x_1 \leq 1/2$ for which:

$$.6x_1 + (1-x_1)^{2/3} > 1. \tag{1}$$

However, it is an easy exercise to show that there is no solution to this inequality in the desired range $0<x_1 \leq 1/2$. The contradiction allows us to conclude that all lines have density at least as large as $.6\, y_A$. Since $\sum_{j=1}^{t} \binom{y_j}{2} \leq \binom{n}{2}$, we also know that $t\binom{.6\, y_A}{2} < n^2$. A simple computation shows that $t < 10^3 n^2/c_1^6$, and thus t is significantly less than the upper bound $t = \binom{n}{2}$. Also note that if there are s lines in $\mathscr{L}$ having the same slope, then $s(.6\, y_A) \leq n$. A simple computation shows $s < 10^2\, t/c_1^3$, i.e., a very small fraction of the lines in $\mathscr{L}$ belong to any class of parallel lines.

Let d_A denote the *average degree* of the points in $\mathscr{P}$, i.e., $d_A = I/n$. We will show by a similar argument that all points in $\mathscr{P}$ have degree at least as large as $.6d_A$. To the contrary, suppose there are xn points with degree less than $.6d_A$ where $0<x<1$. As before, let $x_1 = \min\{x, 1/2\}$ and choose a set of $x_1 n$ points having degree less than $.6\, d_A$. Since $\sqrt{n - x_1 n} \leq t \leq \binom{n - x_1 n}{2}$, we know that the number of incidences involving the remaining $n - x_1 n$ points is less than $c_1(n - x_1 n)^{2/3} t^{2/3} = (1-x_1)^{2/3} I$. Since the number of incidences involving the $x_1 n$ points with small degree is less than $(x_1 n)(.6\, d_A)$, this requires that x_1 satisfy inequality (1) which we have already noted is impossible. The contradiction shows that every point in $\mathscr{P}$ has degree at least as large as $.6\, d_A$.

The next step in the proof is to obtain a large subset $\mathscr{P}^*$ of the points so that each point in $\mathscr{P}^*$ is incident with a large number of lines from each of two sets $\mathscr{L}_1$, $\mathscr{L}_2$ of lines. After a suitable linear transformation, all lines in $\mathscr{L}_1$ will have slope in the interval $(.99, 1.01)$ and all lines in $\mathscr{L}_2$ will have slope in the interval $(-1.01, -.99)$. The tight control in the angles is necessary in order to apply the covering lemma. We will proceed in several stages.

First, let $S_1 = \{l_j\colon 1 \leq j \leq t/2\}$ and let $S_2 = \{l_j\colon t/2 < j \leq t\}$. Then let $\mathscr{P}_0$ denote the set of points in $\mathscr{P}$ which are incident with at least $d_A/100$ lines from S_1 and at least $d_A/100$ lines from S_2. Partition $\mathscr{P} - \mathscr{P}_0$ into two subsets $\mathscr{P}_1$, $\mathscr{P}_2$ where $\mathscr{P}_1 = \{p \in \mathscr{P} - \mathscr{P}_0\colon p$ is incident with more lines from S_1 than with $S_2\}$ and $\mathscr{P}_2 = \mathscr{P} - \mathscr{P}_0 - \mathscr{P}_1$. For each $i = 0, 1, 2$, let $|\mathscr{P}_i| = x_i n$ where $x_0 + x_1 + x_2 = 1$. Also for each i, let I_i denote the number of incidences involving the points in $\mathscr{P}_i$. We claim that $x_0 \geq .1$. To the contrary, suppose that $x_0 < .1$. We proceed to a contradiction.

Consider the following inequalities:

(2) $$\sqrt{x_0 n} \leqq t \leqq \binom{x_0 n}{2}.$$

(3) $$\sqrt{x_1 n} \leqq t/2 \leqq \binom{x_1 n}{2}.$$

(4) $$\sqrt{x_2 n} \leqq t/2 \leqq \binom{x_2 n}{2}.$$

If inequalities (2), (3), and (4) are satisfied, then we know:

$$I_0 < c_1(x_0 n)^{2/3} t^{2/3} = x_0^{2/3} I,$$

$$I_1 < (x_1 n)(d_A/100) + c_1(x_1 n)^{2/3}(t/2)^{2/3}, \quad \text{and}$$

$$I_2 < (x_2 n)(d_A/100) + c_1(x_2 n)^{2/3}(t/2)^{2/3}.$$

Since $x_1^{2/3} + x_2^{2/3} \leqq 2[(1-x_0)/2]^{2/3}$, we conclude that:

$$I = I_0 + I_1 + I_2 < [x_0^{2/3} + (1-x_0)/100 + 2^{-1/3}(1-x_0)^{2/3}]\, I.$$

Thus x_0 must be a solution to the inequality

(5) $$1 < x_0^{2/3} + (1-x_0)/100 + 2^{-1/3}(1-x_0)^{2/3}.$$

Again, it is an easy exercise to show that inequality (2) has no solutions in the interval (0, .1]. (However, it does have a solution in the interval (.1, .2).) We may conclude that one or more of the inequalities (2), (3), and (4) does not hold. Since it is assumed that $x_0 < .1$, it is clear that at least one of (3) and (4) is valid. In view of the obvious symmetry, we may therefore assume that (4) is valid, and that at least one of (2) and (3) is invalid.

Suppose first that (2) does not hold, i.e., suppose that $\binom{x_0 n}{2} < t$. Then define a number y_0 by $t = y_0^2 n^2/10$. Then $\sqrt{y_0 n} \leqq t \leqq \binom{y_0 n}{2}$ and $x_0 n < y_0 n$. Since $t < 10^3 n^2/c_1^3$. Thus $I_0 < c_1(y_0 n)^{2/3} t^{2/3} < 10^2 I/c_1^2 < 10^{-100} I$.

Now suppose that (3) is valid. Following reasoning similar to our previous argument, we would conclude that $I_1 + I_2 < c_1(x_1 n)^{2/3}(t/2)^{2/3} + c_1(x_2 n)^{2/3}(t/2)^{2/3} < 2^{-1/3} I$. Thus $I = I_0 + I_1 + I_2 < 10^{-100} I + 2^{-1/3} I$ which is clearly false. It must therefore be the case that when (2) is invalid, (3) is also invalid. However, in this case we know that $x_1 n < y_0 n$ and thus $I_1 < 10^{-100} I$. Also, we know that $I_2 < c_1 n^{2/3}(t/2)^{2/3} + n d_A/100 = 2^{-2/3} I + 10^{-2} I$. Together, these inequalities imply that $I < 2 \cdot 10^{-100} I + 2^{-2/3} I + 10^{-2} I$ which is false.

So it remains to consider the case where (2) and (4) are valid but (3) is not. In this case, we know $I_0 < x_0^{2/3} I$, $I_1 < 10^{-100} I$, and $I_2 < 2^{-2/3} I + 10^{-2} I$. These inequalities require that x_0 satisfy the inequality:

$$1 < x_0^{2/3} + 10^{-100} + 2^{-2/3} + 10^{-2}.$$

However, there is no such number. This completes the proof of our claim that $x_0 \geqq .1$.

Next, let m_0 denote the maximum slope of the lines in S_2 and let $U_0=\{l\in S_1:$ slope $(l)>m_0\}$. Also, let $V_0=S_2$. Then, each of the points in $\mathcal{P}_0$ is incident with at least $d_A/200$ lines in U_0 and at least $d_A/200$ lines in V_0. Each line in U_0 has slope greater than every line in V_0. Rotate the configuration so that the lines in U_0 have positive slope and the lines in V_0 have negative slope. To simplify the inductive construction which follows, we set $M=10^{10}$ and let $n_0=|\mathcal{P}_0|$. We then observe that when $i=0$, we have a configuration of points and lines satisfying the following properties.

1. A set $\mathcal{P}_i$ of at least n_i points where $n_i=.1n\left(1-\frac{2}{M}\right)^i 2^{-i}$.
2. A set U_i of lines with positive slope.
3. A set V_i of lines with negative slope.
4. $|U_i|\leqq t_i$ and $|V_i|\leqq t_i$ where $t_i=t/2^{i+1}$.
5. Each point in $\mathcal{P}_i$ is incident with at least e_i lines from U_i and at least e_i lines from V_i where $e_i=\frac{d_A}{200}\left(1-\frac{2}{M}\right)^i$.

We say that a point $p\in\mathcal{P}_i$ is *closed* when it is incident with less than e_i/M lines from U_i with slope greater than 1/2, and it is incident with less than e_i/M lines from V_i with slope less than $-1/2$. Dually, we say that p is *open* if it is incident with less than e_i/M lines from U_i with slope less than $\sqrt{3}/2$ and it is incident with less than e_i/M lines from V_i with slope greater than $-\sqrt{3}/2$. We hope to find some linear transformation of the form $(x, y)\to(x, my)$ so that there are at least n_i/M points in $\mathcal{P}_i$ which are neither open nor closed. If we can find such a transformation, we terminate this inductive construction.

Suppose that there is no transformation which results in at least n_i/M points which are neither open nor closed. Then we choose a transformation for which the number of open points is as nearly equal to the number of closed points as possible. In this case, there are at least $(n_i-2n_i/M)/2$ open points and at least $(n_i-2n_i/M)/2$ closed points.

To continue the construction, we will either choose $\mathcal{P}_{i+1}$ to be the set of open points or the set of closed points. With either choice, we note that $|\mathcal{P}_{i+1}|\geqq n_{i+1}=(.1n)\left(1-\frac{2}{M}\right)^{i+1}2^{-i-1}$. The choice is determined by the distribution of lines. If less than half the lines in $U_i\cup V_i$ have slope greater than 1 in absolute value, we choose then open points and set $U_{i+1}=\{l\in U_i:\text{slope}(l)>1\}$ and $V_{i+1}=\{l\in V_i:\text{slope}(l)<-1\}$. On the other hand, if at least half the lines have slope greater than 1 in absolute value, we choose the closed points and set $U_{i+1}=\{l\in U_i:\text{ slope }(l)<1\}$ and $V_{i+1}=\{l\in V_i:\text{ slope }(l)>-1\}$. In either case, the desired inequalities hold for $t_{i+1}=t_i/2$.

Now we pause to observe that the number of incidences J_i between the points in $\mathcal{P}_i$ and the lines in U_i satisfies $J_i\geqq n_ie_i$. Now suppose that i is relatively small, say $i\leqq 30$. Then $\left(1-\frac{2}{M}\right)^i\geqq\frac{1}{10}$ and $2^i\leqq M$. Furthermore, we must have $t_i\geqq\sqrt{n_i}$. For if $t_i<\sqrt{n_i}$, then $J_i<c_1(n_i)^{2/3}(n_i^{1/2})^{2/3}=c_1n_i$. This in turn implies that $c_1>e_i=\frac{d_A}{200}\left(1-\frac{2}{M}\right)^i\geqq\frac{d_A}{2000}=\frac{c_1}{2000}n^{-1/3}t^{2/3}\geqq\frac{c_1^3}{20\,000}$, i.e, $c_1^2<2\cdot 10^4$ which is false. On the other hand, we must also have $t_i \leqq \binom{n_i}{2}$. For if $\binom{n_i}{2} < t_i$, then 10^{-18}

$$< \frac{2}{3 \cdot 10^2 \cdot 10^2 \cdot 10^3 2^i} \leqq \frac{1}{3} \cdot n^2 (.1)^2 \left(1 - \frac{2}{M}\right)^{2i} 2^{-2i} \frac{2^{i+1}}{n^2} < \frac{1}{3} n_i^2 \frac{2^{i+1}}{t} \frac{10^3}{c_1^6} < \frac{\binom{n_i}{2}}{t_i} \frac{10^3}{c_1^6}.$$

This requires $c_1^6 < 10^{21}$ which is false.

Since $\sqrt{n_i} \leqq t_i \leqq \binom{n_i}{2}$, we know that $J_i < c_1 (n_i)^{2/3} (t_i)^{2/3}$, and thus $n_i e_i < c_1 (n_i)^{2/3} (t_i)^{2/3}$. Substituting, we obtain the inequality:

$$2^{i/3} \left(1 - \frac{2}{M}\right)^{4i/3} < \frac{200}{(.1)^{1/3} 2^{2/3}}.$$

A simple calculation shows that this inequality fails if $i \geqq 30$. So we conclude that our algorithm terminates after i steps where $i < 30$. We obtain a set of at least n_i/M points which are neither open nor closed.

Divide the quadrant $[0°, 90°]$ into M sectors $\alpha_1, \alpha_2, \ldots, \alpha_M$ of equal size. Divide $[-90°, 0°]$ into $\beta_1, \beta_2, \ldots, \beta_M$ similarly. Then for each point p which is neither open nor closed, we can choose a pair (α_j, β_k) so that:

1. p is incident with at least e_i/M^2 lines passing through p at an angle belonging to sector α_j and with at least e_i/M^2 lines passing through p at an angle belonging to sector β_k.
2. The angle between α_j and β_k is at least 30° and at most 150°.

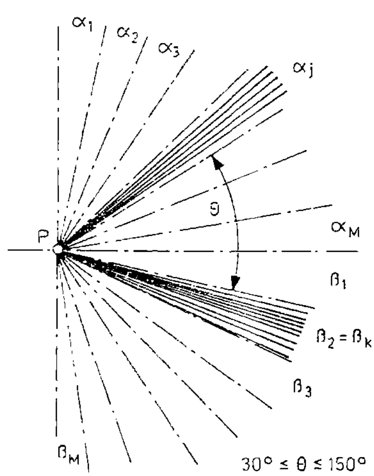

Fig. 1

It follows from the pigeonhole principle, that there is some pair $(\alpha_{j_0}, \beta_{j_0})$ of sectors so that there is a subset $\mathscr{P}^*$ containing at least n_i/M^3 points which satisfy these two conditions for $(\alpha_{j_0}, \beta_{j_0})$. Rotate the configuration so that the x-axis is midway between these two sectors and then apply a linear transformation so that the sectors are centered on the lines with slope $+1$ and -1. Note that $e_i/M^2 \geqq d_A/M^3$.

So we now have a set $\mathscr{P}^*$ of at least n/M^3 points each which is incident with at least d_A/M^3 lines from a set $\mathscr{L}_1$ of lines each of which have slope in the interval (.99, 1.01) and with at least d_A/M^3 lines from a set $\mathscr{L}_2$ of lines each of which has slope in the interval (−1.01, −.99).

Now we are ready to apply the covering lemma. We set $r_2=d_A/2M^3$ and $r_1=d_A/M^4$. We then cover at least $n/16M^3$ of the points in $\mathscr{P}^*$ with squares containing at least r_1 but no more than r_2 points from $\mathscr{P}^*$. Now consider a square Q of this covering. Q is partitioned into four triangles by its diagonals. At least one of these triangles contains at least $r_1/4$ points from $\mathscr{P}^*$. Choose one such triangle and call it T. For each point $p\in T$, there are least d_A/M^3 lines from $\mathscr{L}_1$ which contain p. At most r_2 of these lines pass through other points of $\mathscr{P}^*$ which belong to Q. So there are at least $d_A/M^3-r_2=d_A/2M^3$ lines in $\mathscr{L}_1$ which contain p but do not contain any other point of $\mathscr{P}^*$ inside Q. We can make a similar observation for the lines from $\mathscr{L}_2$.

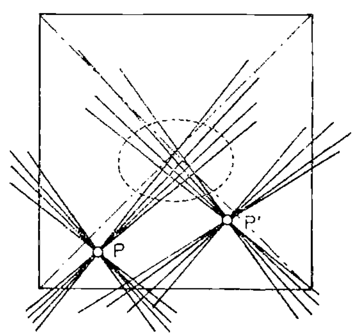

Fig. 2

Now suppose that $p, p'\in T$ and that these lines not containing other points of $\mathscr{P}^*\cap Q$ have been chosen for each of the two points. Then regardless of the location of p and p' inside T, there are at least $d_A^2/4M^6$ wasted crossings produced by these lines. Thus each T produces at least $\binom{r_1/4}{2} d_A^2/4M^6$ wasted crossings. Since there are at least $n/16M^3r_2$ squares, we conclude that the total number of wasted crossings, which we denote by W, satisfies:

$$W \geqq \frac{n}{16M^3r_2}\binom{r_1/4}{2}\frac{d_A^2}{4M^6} > \frac{n}{M^{15}}d_A^3 \geqq \frac{n}{M^{15}}c_1^3n^{-1}t^2 \geqq t^2,$$

since $c_1^3>10^{150}=M^{15}$. The inequality $w>t^2$ is a contradiction since it implies that there are more wasted crossings than there are pairs of lines. With this observation, the proof of Theorem 1 is complete. ∎

Before we close this section we comment that the special case of Theorem 1 when $t=n$, namely that the number of incidences between n points and n lines is at most $c_1n^{4/3}$ was conjectured by Erdős and Purdy [5].

4. Consequences of the Principal Theorem

We begin this section with the following result which follows immediately from Theorem 1.

Theorem 2. *There exists a constant c_2 so that if $2 \leqq k \leqq \sqrt{n}$ and $\mathcal{L}$ is a family of t lines each containing at least k points from a set $\mathcal{P}$ of n points, then $t < c_2 n^2/k^3$.*

Proof. Set $c_2 = c_1^3$ where c_1 is the absolute constant in Theorem 1. Then suppose that for some k with $2 \leqq k \leqq \sqrt{n}$, there is a family of $t = c_1^3 n^2/k^3$ lines each containing at least k points from a set of n points. Then the number of incidences is at least $c_1^3 n^2/k^2$. Furthermore, since $k \geqq 2$, $t \leqq \binom{n}{2}$ and since $k \leqq \sqrt{n}$, $t \geqq \sqrt{n}$. Thus,

$$\frac{c_1^3 n^2}{k^2} < c_1 n^{2/3} t^{2/3} \leqq c_1 n^{2/3} \left(\frac{c_1^3 n^2}{k^3}\right)^{2/3} = \frac{c_1^3 n^2}{k^2}.$$

The contradiction completes the proof. ∎

P. Erdős conjectured the result in Theorem 2 when $k = \sqrt{n}$ and this special case was settled by the authors in [7]. Erdős also conjectured Theorem 2 when $k = \sqrt{n}/2^u$. In addition, Croft and Erdős also conjectured that for every $\varepsilon > 0$ and every $k \geqq 2$, there exists a constant $n(\varepsilon, k)$ so that when $n \geqq n(\varepsilon, k)$, the number of lines each containing at least k points is less than $\varepsilon n^2/k^2$. Theorem 2 shows that this last conjecture is also valid.

Theorem 3. *There exists a constant $c_3 > 0$ so that if $\mathcal{P}$ is a set of n points not all on the same line and $\mathcal{L}$ is the family of all lines determined by $\mathcal{P}$, then there is at least one point in $\mathcal{P}$ which belongs to more than $c_3 n$ of the lines in $\mathcal{L}$.*

Proof. We show that the theorem holds for $c_3 = 10^{-7} c_2^{-6}$ where c_2 is the absolute constant in Theorem 2. Suppose to the contrary that $\mathcal{P} = \{p_1, p_2, \ldots, p_n\}$ is a set of n points not all on the same line, $\mathcal{L} = \{l_1, l_2, \ldots, l_t\}$ is the family of all lines determined by, but that no point belongs to more than $c_3 n$ lines in $\mathcal{L}$. For each $j = 1, 2, \ldots, t$, let y_j count the number of points on line l_j. For each $i = 1, 2, \ldots, n$, let d_i count the number of lines containing p_i. Then $d_i \leqq c_3 n$ for each i. We may then write the following inequalities.

$$\sum_{j=1}^{t} y_j^2 > \sum_{j=1}^{t} \binom{y_j}{2} = \binom{n}{2} > \frac{4n^2}{10}, \quad \text{and}$$

$$\sum_{j=1}^{t} y_j = \sum_{i=1}^{n} d_i \leqq c_3 n^2.$$

Next, let $M = 100 c_2^3$. Split the sum $\sum_{j=1}^{t} y_j^2$ into four terms:

$$S_1 = \sum_{2 \leqq y_j < M} y_j^2; \quad S_2 = \sum_{M \leqq y_j < \sqrt{n}} y_j^2; \quad S_3 = \sum_{\sqrt{n} \leqq y_j < n^{2/3}} y_j^2; \quad \text{and} \quad S_4 = \sum_{n^{2/3} \leqq y_j} y_j^2.$$

First, we show that $S_1 < n^2/10$. To the contrary suppose that $S_1 \geqq n^2/10$. For each $j = 2, 3, \ldots, M-1$ suppose there are s_j lines containing j points. Then $\sum_{j=2}^{M-1} s_j j^2 \geqq \frac{n^2}{10}$. It follows that there is some j with $2 \leqq j \leqq M-1$ so that $s_j \geqq \frac{n^2}{10M^3}$. These s_j lines account for at least $\frac{n^2}{10M^3} j > \frac{n^2}{10M^3} \geqq \frac{n^2}{10^7 c_2^6} \geqq c_3 n^2$ incidences which is a contradiction. Thus $S_1 < n^2/10$.

We next show that $S_2 < n^2/10$. For each $i \geqq 0$ with $M2^i < \sqrt{n}$, let t_i count the number of lines containing at least $M2^i$ but less than $M2^{i+1}$ points. Then $t_i < c_2 n^2/M^3 2^{3i}$. Since these lines contain at most $M \cdot 2^{i+1}$ points, it follows that $S_2 < \sum_{i=0}^{\infty} \frac{c^2 n^{3i}}{M^3 2^{3i}} (M \cdot 2^{i+1})^2 = \frac{8c_2 n^2}{M} < \frac{n^2}{10}$.

Next, we show that $S_3 < n^2/10$. Suppose there are t_0 lines containing at least $\sqrt{n}$ but less than $n^{2/3}$ points, and that $t_0 > 0$. Then $t_0 \leqq c_2 \sqrt{n}$ so $S_3 < c_2 \sqrt{n} \cdot n^{4/3} = c_2 n^{11/6}$. We claim that $c_2 n^{11/6} < n^2/10$, i.e., $10c_2 < n^{1/6}$. To see that this claim is valid, suppose that $10c_2 \geqq n^{1/6}$. Then $10^6 c_2^6 \geqq n$ so $c_3(10^6 c_2^6) = .1 \geqq c_3 n \geqq d_i$ for each $i = 1, 2, \ldots, n$. This is a contradiction since each d_i is a positive integer. This completes the proof of our claim that $c_2 n^{11/6} < n^2/10$. We conclude that $S_3 < n^2/10$.

Since $S_1 + S_2 + S_3 + S_4 > 4n^2/10$, we are left to conclude that $S_4 > n^2/10$. For each $i \geqq 0$, let u_i count the number of lines which contain at least $n^{2/3}2^i$ but less than $n^{2/3}2^{i+1}$ points. We claim that $u_i \leqq 2n^{1/3}2^{-i}$ for each $i \geqq 0$. To see that this is true, suppose that $L_1, L_2, \ldots, L_{u_i}$ are lines each containing at least $n^{2/3}2^i$ points. Then for each $j = 1, 2, \ldots, u_i$ there are $n^{2/3}2^i - j + 1$ points on L_j which do not belong to any L_k with $1 \leqq k < j$. Therefore $\sum_{j=0}^{u_i} n^{2/3}2^i - j + 1 \leqq n$. Summing, we obtain (being generous) $u_i \leqq 2n^{1/3}2^{-i}$.

On the other hand, we observe that $u_i = 0$ whenever $n^{2/3}2^i > c_3 n$. To see that this is true, observe that if there is a line containing more than $c_3 n$ points, then any point not on the line has more than $c_3 n$ lines passing through it.

Now let α be the least integer for which $n^{2/3}2^{\alpha+1} > c_3 n$. Then $2^\alpha \leqq c_3 n^{1/3}$. It follows that we must have $\sum_{i=0}^{\alpha} u_i n^{1/3} 2^{2i+2} \geqq S_4 > n^2/10$, and thus $S_4 \leqq \sum_{i=0}^{\alpha} 8 \cdot n^{5/3} 2^i = 8 \cdot n^{5/3}(2^{\alpha+1} - 1) \leqq 16 \cdot n^{5/3} 2^\alpha \leqq 16 \cdot c_3 n^2$. This is a contradiction since it requires $c_3 > 1/160$. With this observation, the proof is complete. ∎

Theorem 3 is a partial solution to an old problem of Dirac [5] who conjectured that there is always a point on at least $(n/2) - c$ lines. It also shows that n points not all on the same line determine at least $c_3 n$ different angles. Corrádi, Hajnal and Erdős conjecture that n points not on the same line determine at least $n-2$ different angles.

Finally, we turn our attention to the following problem. Let $\mathscr{E}(n)$ denote the number of distinct nondecreasing sequences $y_1 \leqq y_2 \leqq \ldots \leqq y_t$ for which there is a set $\mathscr{P}$ of n points and a family $\mathscr{L} = \{l_1, l_2, \ldots, l_t\}$ of t lines so that l_j contains y_j points from $\mathscr{P}$ for each $j = 1, 2, \ldots, t$.

Theorem 4. *There exists a constant c_4 so that $\mathscr{E}(n) < 2^{c_4 \sqrt{n}}$ for all $n \geqq 1$.*

Proof. We show that the theorem is valid when $c_4 = 3c_2$ where c_2 is the absolute constant in Theorem 2. To count the number of sequences associated with configuration of points and lines, we first choose the value of t, the number of lines, as an integer not exceeding $\binom{n}{2}$. For each $i=1, 2, 3, \ldots, \log n$, we then let s_i denote the number of lines which are to contain at least 2^i but less than 2^{i+1} points. The nonnegative numbers $s_1, s_2, \ldots, s_{\log n}$ sum to t so the number of different choices for the sequence of s_i's is at most $\binom{t+\log n-1}{\log n-1}$. For each i, we choose a sequence $y_j, y_{j+1}, \ldots, y_{j+s_i-1}$ of nondecreasing integers with $2^i \leqq y_j \leqq y_{j+s_i-1} < 2^{i+1}$. Clearly, the number of such sequences is at most $\binom{s_i+2^i}{2^i}$. Therefore, we can write:

$$\mathscr{E}(n) \leqq \binom{n}{2}\binom{t+\log n-1}{\log n-1}\prod_{i=1}^{\log n}\binom{s_i+2^i}{2^i}.$$

Now

$$\binom{n}{2}\binom{t+\log n-1}{\log n-1} < n^2\binom{n^2+\log n}{\log n} < n^2 \cdot n^{3\log n} < 2^{c_2\sqrt{n}}$$

Now let i be an integer for which $2^i < 2\sqrt{n}$. Then $s_i < 10c_2n^2/2^{3i}$ and $10c_2n^2/2^{3i} \geqq 100 \cdot 2^i$. Therefore,

$$\binom{s_i+2^i}{2^i} < \binom{\frac{10c_2n^2}{2^{3i}}+2^i}{2^i} < \left(\frac{100c_2n^2}{2^{4i}}\right)^{2^i}.$$

It follows that

$$\prod_{i:2^i<2\sqrt{n}}\binom{s_i+2^i}{2^i} < \prod_{i:2^i<2\sqrt{n}}\left(\frac{100c_2n^2}{2^{4i}}\right)^{2^i} < 2^{c_2\sqrt{n}}.$$

On the other hand, consider an integer i for which $2\sqrt{n} \leqq 2^i \leqq n$. In this case, we have the trivial inequality $s_i < 2n/2^i < 2^i$. Thus

$$\binom{s_i+2^i}{2^i} < \binom{\frac{2n}{2^i}+2^i}{\frac{2n}{2^i}} < \left(\frac{100 \cdot 2^{2i}}{n}\right)^{\frac{2n}{2^i}}.$$

It follows that

$$\prod_{i:2\sqrt{n}\leqq 2^i\leqq n}\binom{s_i+2^i}{2^i} < \prod_{i:2\sqrt{n}\leqq 2^i\leqq n}\left(\frac{100 \cdot 2^{2i}}{n}\right)^{\frac{2n}{2^i}} < \prod_{j=0}^{\infty}(1000 \cdot 2^{2j})^{\sqrt{n}2^{-j}} < 2^{c_2\sqrt{n}}.$$

Combining these inequalities, we obtain the desired result $\mathscr{E}(n) < 2^{3c_2\sqrt{n}}$. ∎

Acknowledgement. The authors gratefully acknowledge the contributions of P. Erdős, J. Ross, J. Griggs, J. Stahl, C. Tator, and D. Sturtevant through many stimulating conversations on these problems. The authors also wish to thank the referees for their careful readings of a preliminary version of this manuscript.

References

[1] J. Beck, On the lattice property of the plane and some problems of Dirac, Motzkin and Erdős in combinatorial geometry, *Combinatorica* **3** (3—4) (1983), 281—297.
[2] P. Erdős, On some problems of elementary and combinatorial geometry, *Annali di Mat., Ser. 4*, **103** (1974), 99—108.
[3] P. Erdős, Some combinatorial problems in geometry, *Conference held in Haifa, Israel, 1979, Lecture Notes in Mathematics*, **792**, Springer, 1980, 46—53.
[4] P. Erdős, Some applications of graph theory and combinatorial methods to number theory and geometry, *Algebraic methods in graph theory, Coll. Math. Soc. T. Bolyai* **25** (1981), 131—148.
[5] P. Erdős, Problems and results in combinatorial geometry, *Proceedings of New York Academy of Science*, to appear.
[6] William Moser, *Research Problems in Discrete Geometry*, Mimeograph Notes (1981).
[7] E. Szemerédi and William T. Trotter, Jr., A Combinatorial Distinction Between the Euclidean and Projective Planes, *submitted*.

E. Szemerédi

Mathematical Institute of the
Hungarian Academy of Sciences
H—1376 Budapest, Pf. 428, Hungary

W. T. Trotter, Jr.

Dept. of Mathematics and Statistics
University of South Carolina
Columbia, SC 29208, USA