Europ. J. Combinatorics (1983) **4**, 385–394

A Combinatorial Distinction Between the Euclidean and Projective Planes

E. SZEMERÉDI AND W. T. TROTTER, JR*

Let n and m be integers with $n = m^2+m+1$. Then the projective plane of order m has n points and n lines with each line containing $m+1 \approx n^{1/2}$ points. In this paper, we consider the analogous problem for the Euclidean plane and show that there cannot be a comparably large collection of lines each of which contains approximately $n^{1/2}$ points from a given set of n points. More precisely, we show that for every $\delta > 0$, there exist constants c, n_0 so that if $n \geqslant n_0$, it is not possible to find n points in the Euclidean plane and a collection of at least $cn^{1/2}$ lines each containing at least $\delta n^{1/2}$ of the points. This theorem answers a question posed by P. Erdös. The proof involves a covering lemma, which may be of independent interest, and an application of the first author's regularity lemma.

1. INTRODUCTION

Throughout this paper we are concerned with the maximum size of a family $\mathscr{L}$ of lines in the Euclidean plane for which there exist a set $\mathscr{P}$ of n points so that each line in $\mathscr{L}$ contains at least $\delta n^{1/2}$ of the points in $\mathscr{P}$ (δ is fixed positive constant). The points with integer coordinates $\{(i,j)\colon 1 \leqslant i \leqslant n^{1/2}, 1 \leqslant j \leqslant n^{1/2}$ show that when $\delta = 1$, $|\mathscr{L}|$ can be at least $2n^{1/2}$. The principal result of this paper will be to show that this simple construction is essentially best possible.

THEOREM. *For every $\delta > 0$, there exist constants c, n_0 so that if $n \geqslant n_0$, $\mathscr{P} = \{p_1, p_2, \ldots, p_n\}$ is any set of n points in the Euclidean plane, and $\mathscr{L}$ is a collection of lines each containing at least $\delta n^{1/2}$ of these points, then $\mathscr{L}$ contains less than $cn^{1/2}$ lines.*

This theorem settles in the affirmative a conjecture of P. Erdös. The corresponding problem for a projective plane has a dramatically different answer. When $n = m^2+m+1$, a projective plane has n points, n lines and each line contains $m+1 \approx n^{1/2}$ points. The proof of our theorem requires the development of a covering lemma for points in the plane and a subsequent application of the first author's regularity lemma [5]. We refer the reader to Erdös' survey papers [1], [2], [3], and [4] for extensive listings of problems in combinatorial geometry.

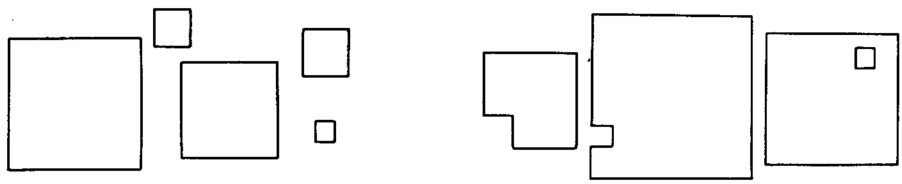
2. THE COVERING LEMMAS

In this section, we consider the problem of covering a set $\mathscr{P}$ of m points in the plane by a family $\mathscr{F}$ of regions with certain restrictions being placed on the shapes of the regions in $\mathscr{F}$ and the number of points from $\mathscr{P}$ which they contain. We shall only be concerned with regions of two different shapes (see Figure 1): Shape 1—a square with sides parallel to the coordinate axes; and Shape 2—a square from which a smaller square has been removed. The sides of the squares are parallel to the coordinate axes.

When we count the number of points in $\mathscr{P}$ contained in a region of one of these two shapes, we include those on the boundary of the region. Keeping in mind the nature of

* Research Supported in Part by NSF Grants ISP-8011451 and MCS-8202172.

0195-6698/83/040385+10 $02.00/0



Shape 1 region   Shape 2 regions

FIGURE 1.

the general problem we are investigating, it is clear that we have the freedom to apply an arbitrary linear transformation to the plane so that we can rotate, translate, compress, or expand the points in $\mathscr{P}$ as desired. Therefore it is natural to assume that all points in $\mathscr{P}$ are contained in the unit square $Q_0 = \{(x, y): 0 \leq x \leq 1, 0 \leq y \leq 1\}$.

For each $i \geq 0$, we will let $\mathscr{G}_i$ denote the family of subsquares of Q_0 obtained by dividing Q_0 into 4^i subsquares of equal size. We will also let $\mathscr{G} = \bigcup_{i=0}^{\infty} \mathscr{G}_i$. Hereafter, all shape 2 regions considered will be of the form $Q - Q'$ where Q, $Q' \in \mathscr{G}$ and Q' is a subsquare of Q. However, unless we specify that a square belongs to $\mathscr{G}$, it will only be assumed that it is contained in the unit square Q_0 and that the sides are parallel to the coordinate axes. A family $\mathscr{F}$ of regions is said to be *almost disjoint* when no point in the plane belongs to the interior of more than one region in $\mathscr{F}$. Our first lemma is concerned with covering the points in $\mathscr{P}$ with an almost disjoint family of regions of these two shapes.

LEMMA 1. *Let r_1 and r_2 be positive integers with $4r_1 \leq r_2$. Then let $m \geq r_2$ and let $\mathscr{P}$ be any set of m points in the unit square. Then there exists an ordered pair (c_0, d_0) of nonnegative integers and an almost disjoint family $\mathscr{F}$ of regions so that:*

1. *Each region is either a square from $\mathscr{G}$ or a shape 2 region of the form $Q - Q'$ where Q, $Q' \in \mathscr{G}$.*
2. *Each region in $\mathscr{F}$ contains at least r_1 but no more than r_2 points from $\mathscr{P}$.*
3. *$\mathscr{F}$ covers at least $c_0 r_2/4$ of the points in $\mathscr{P}$.*
4. *$\mathscr{F}$ fails to cover at most $(d_0 + 1)r_1$ of the points in $\mathscr{P}$.*
5. *$d_0 \leq 4c_0 - 4$.*

PROOF. The argument will involve an inductive construction beginning with small squares in $\mathscr{G}$ and proceeding up through larger squares. We will find it convenient to use the following terminology. If $Q \in \mathscr{G}_i$, the four squares in $\mathscr{G}_{i+1}$ contained in Q are called *immediate subsquares* of Q. A square Q is said to be *dense* if it contains at least $r_2/4$ points from $\mathscr{P}$. Otherwise, the square is said to be *sparse*. For each $i \geq 0$, a square Q in $\mathscr{G}_i$ is said to be *concentrated* when it contains more than r_2 points from $\mathscr{P}$ and has exactly one dense immediate subsquare. Now let k be the least nonnegative integer for which every square in $\mathscr{G}_{k+1}$ is sparse. We now proceed to construct the desired family $\mathscr{F}$ by an algorithm which will involve examining in turn each of the families $\mathscr{G}_k, \mathscr{G}_{k-1}, \ldots, \mathscr{G}_0$. Based on a prescribed set of rules, certain regions will be added to the family $\mathscr{F}$ and others will be deleted. In addition certain dense squares will be assigned labels which will be an ordered pair of nonnegative integers. When a dense square Q is assigned the label (c_Q, d_Q), c_Q will be called the *cover number* of Q and d_Q will be called the *deficiency number* of Q. Once a square is labelled, it retains that label permanently.

We begin by letting $\mathscr{F}$ consist of the collection of dense squares in $\mathscr{G}_k$. Each dense square Q in $\mathscr{G}_k$ will be labelled with the pair $(1, 0)$. We then observe that the following properties concerning labelled dense squares are satisfied when $i = k$.

If Q is a labelled dense square in some family $\mathscr{G}_j$ where $k \geq j \geq i$, then
P_1: $d_Q \leq 4c_Q - 4$,
P_2: $\mathscr{F}$ covers at least $c_Q r_2/4$ of the points in $\mathscr{P}$ which belong to Q, and
P_3: $\mathscr{F}$ fails to cover at most $d_Q r_1$ of the points in $\mathscr{P}$ which belong to Q.

Furthermore, we observe that the following condition concerning unlabelled dense squares is satisfied vacuously when $i = k$.
P_4: If Q is an unlabelled dense square in $\mathscr{G}_j$ where $k \geq j \geq i$, then Q has a labelled subsquare Q' so that $Q - Q'$ contains less than r_1 points from $\mathscr{P}$.

Finally, we observe that $\mathscr{F}$ satisfies the following two conditions.
P_5: Each region in $\mathscr{F}$ is either a square from $\mathscr{G}$ or is a shape 2 region of the form $Q - Q'$ where $Q, Q' \in \mathscr{G}$.
P_6: Each region in $\mathscr{F}$ contains at least r_1 but no more than r_2 points from $\mathscr{P}$.

Next, we suppose that for some integer i_0 with $k > i_0 \geq 0$, we have a family $\mathscr{F}$ and a labelling of certain dense squares in $\mathscr{G}_k \cup \mathscr{G}_{k-1} \cup \cdots \cup \mathscr{G}_{i_0+1}$ so that these six properties are satisfied when $i = i_0 + 1$. We then provide the algorithm for updating the family $\mathscr{F}$ and for labelling certain dense squares in $\mathscr{G}_i$ so that each of the six properties is satisfied when $i = i_0$.

Let Q be a dense square in $\mathscr{G}_{i_0}$.

Case 1. Q contains at most r_2 points from $\mathscr{P}$. In this case, we update $\mathscr{F}$ by adding Q and deleting any region previously in $\mathscr{F}$ contained in Q. The dense square Q is labelled with the pair $(1, 0)$.

Case 2. Q is concentrated. In this case, we let Q' denote the unique immediate subsquare of Q which is dense.
Subcase 2(a). Q' is labelled. If $Q - Q'$ has less than r_1 points then Q is unlabelled. If $Q - Q'$ has at least r_1 points, add $Q - Q'$ to $\mathscr{F}$ and label Q as follows. Set $d_Q = d_{Q'}$. If $Q - Q'$ has at least $r_2/4$ points, set $c_Q = 1 + c_{Q'}$. Otherwise, set $c_Q = c_{Q'}$.
Subcase b. Q' is unlabelled. Let Q'' be a labelled subsquare so that $Q' - Q''$ contains less than r_1 points. If $Q - Q''$ has less than r_1 points, then Q is unlabelled. If $Q - Q''$ has at least r_1 points, add $Q - Q''$ to and label Q as follows. Set $d_Q = d_{Q''}$. If $Q - Q''$ has at least $r_2/4$ points, set $c_Q = 1 + c_{Q''}$. Otherwise, set $c_Q = c_{Q''}$.

In each of the first two cases, it is trivial to verify that the desired conditions are satisfied for the dense square Q. The remaining case requires only a little more effort.

Case 3. Q has more than r_2 points from $\mathscr{P}$ but is not concentrated. In this case we let $Q_1, Q_2, \ldots, Q_t$ be the immediate subsquares of Q which are dense, where $2 \leq t \leq 4$. For each j, we let (c_j, d_j) be the label assigned to Q_j if it has been labelled. When Q_j is unlabelled, we let (c_j, d_j) be the label assigned to a labelled subsquare Q_j' of Q_j for which $Q_j - Q_j'$ has less than r_1 points. The square Q is then labelled with the pair (c_Q, d_Q) where $c_Q = \sum_{j=1}^t c_j$ and $d_Q = 4 + \sum_{j=1}^t d_j$. The family $\mathscr{F}$ is then updated by adding any sparse immediate subsquare of Q which contains at least r_1 points from $\mathscr{P}$.

It is obvious that $\mathscr{F}$ satisfies P_5 and P_6 so it remains only to show that the first three conditions are satisfied for this labelled square. First, we observe that $d_Q = 4 + \sum_{j=1}^t d_j \leq 4 + \sum_{j=1}^t 4c_j - 4 = (8 - 4t) + 4c_Q - 4$. Since $t \geq 2$, we conclude that $d_Q \leq 4c_Q - 4$ and thus P_1 is satisfied.

Property P_2 is clearly satisfied since $\mathscr{F}$ covers at least $c_j r_2/4$ of the points in each dense immediate subsquare Q_j, and $c_Q = \sum_{j=1}^t c_j$. To see that property P_3 is satisfied, we see that for each j, $\mathscr{F}$ fails to cover at most $d_j r_1$ of the points in Q_j when Q_j is labelled. When Q_j is not labelled, $\mathscr{F}$ fails to cover at most $d_j r_1$ of the points in Q_j' but also covers none of the points in $Q_j - Q_j'$. Thus for each j, there are at most an additional r_1 points not covered. $\mathscr{F}$ covers all the points in each sparse immediate subsquare of Q unless it contains less than r_1 points. We conclude that each immediate subsquare of Q can contribute at most r_1 additional points of $\mathscr{P}$ which are not covered by $\mathscr{F}$. Thus $\mathscr{F}$ fails to cover at most $(\sum_{j=1}^{t} d_j r_1) + 4r_1 = d_Q r_1$ points in Q.

With this observation, we have completed the proof that this algorithm may be applied inductively, so we now consider the resulting family when the algorithm terminates at $i = 0$. Let (c_0, d_0) be the label applied to the unit square Q_0 if it is labelled by the algorithm. Otherwise, let (c_0, d_0) be the label assigned to a labelled subsquare Q_0' of Q_0 for which $Q_0 - Q_0'$ has less than r_1 points. It is then immediate that the pair (c_0, d_0) and the current family $\mathscr{F}$ satisfy the conclusion of the lemma.

Our next lemma is concerned with a covering of points from $\mathscr{P}$ by a family $\mathscr{F}$ of regions each of which is a square, although not necessarily a square in $\mathscr{G}$. When we use the term 'rectangle' in the proof of Lemma 2, it will be assumed that the sides are parallel to the coordinate axes.

LEMMA 2. *Let r_1 and r_2 be positive integers with $256r_1 \leq r_2$. Then let $m \geq r_2$ and let $\mathscr{P}$ be any set of m points in the unit square. Then there exists an almost disjoint family $\mathscr{F}$ of squares to that:*
1. *Each square in $\mathscr{F}$ contains at least r_1 but no more than r_2 points from $\mathscr{P}$.*
2. *$\mathscr{F}$ covers at least $m/16$ of the m points in $\mathscr{P}$.*

PROOF. We first apply the algorithm of the preceding lemma for the values $r_1' = 8r_1$ and $r_2' = r_2$ to produce a family $\mathscr{F}'$ of regions some of which are squares in $\mathscr{G}$ and some of which are shape 2 regions of the form $Q - Q'$ where $Q, Q' \in \mathscr{G}$. Let (c_0, d_0) be the pair associated with the family $\mathscr{F}'$. We conclude that $\mathscr{F}'$ covers at least $c_0 r_2'/4$ points in $\mathscr{P}$ and fails to cover at most $(d_0 + 1)\, r_1'$ points in $\mathscr{P}$. Since $c_0 \geq \frac{1}{4} d_0$ we have

$$\frac{c_0 r_2'/4}{(d_0+1) r_1'} \geq \frac{d_0 r_2/(4 \times 4)}{2d_0(8r_1)} \geq 1.$$

Thus $\mathscr{F}$ covers at least $m/2$ of the points in $\mathscr{P}$. It remains only to show that if $Q - Q'$ is a Shape 2 region in $\mathscr{F}'$, then there exists a square Q'' contained in $Q - Q'$ so that Q'' contains at least 1/8-th of the points in $\mathscr{P}$ which belong to $Q - Q'$. However, there is a natural way to subdivide the region $Q - Q'$ into rectangles by extending the sides of the small square.

Depending on whether Q' is located in the corner, on one side, or in the interior of Q, the number of rectangles determined is 3, 5, or 8 (see Figure 2) In any case, one of these rectangles is certain to contain at least 1/8th of the points in $\mathscr{P}$ belonging to $Q - Q'$. Regardless of which rectangle enjoys this property, it is easy to see that there is a square Q'' contained in $Q - Q'$ and containing the specified rectangle.

Note that in the proof of Lemma 2, the existence of Q'' depends heavily on the fact that $Q, Q \in \mathscr{G}$. If Q and Q' are arbitrary squares and Q' is in the interior of Q but is

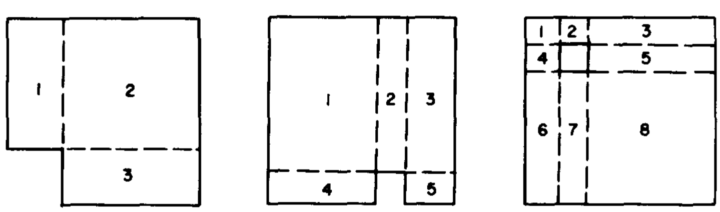


FIGURE 2.

very close to the boundary of Q in comparison to the size of Q', then the existence of Q'' cannot be asserted.

3. The Proof of the Principal Theorem

Throughout this section, we assume that $\mathcal{P} = \{p_1, p_2, \dots, p_n\}$ is a collection of n points in the unit square and that $\mathcal{L} = \{l_1, l_2, \dots, l_t\}$ is a family of $t = cn^{1/2}$ lines each of which contains at least $\delta n^{1/2}$ points from $\mathcal{P}$. Without loss of generality, we may assume δ is small, say $\delta < 1/10$. We will then derive a series of assumptions that can be made concerning the properties of this configuration culminating in a contradiction when c and n are sufficiently large in comparison to $1/\delta$. At a crucial stage of the argument, we will use the second covering lemma developed in the preceding section. We will also require the regularity lemma developed by the first author in [5]. In this paper, we present only the material necessary to understand and apply this lemma and refer the reader to [5] for the argument.

Let G be a graph on a set S, and let A_1 and A_2 be disjoint nonempty subsets of S. Then we denote by $G(A_1, A_2)$ the subgraph of G whose vertex set is $A_1 \cup A_2$ and whose edge set contains those edges in G with one endpoint in each of A_1 and A_2. The *density* of $G(A_1, A_2)$ is the ratio of the number of edges in $G(A_1, A_2)$ divided by $|A_1||A_2|$. The density of $G(A_1, A_2)$ measures the probability that an ordered pair $(a_1, a_2) \in A_1 \times A_2$ chosen at random is an edge in G.

Now let ε be a small positive number. Then we say that $G(A_1, A_2)$ is *ε-regular* provided that the density of $G(A_1', A_2')$ differs from the density of $G(A_1, A_2)$ by at most ε whenever $A_i' \subseteq A_i$ and $|A_i'| \geq \varepsilon |A_i|$ for $i = 1, 2$.

Lemma 3 (Szemerédi [5]). *For every $\varepsilon \in 0$ and every positive integer M, there exist constants N_1 and N_2 so that if $n \geq N_2$ and G is any graph whose vertex set is an n-element set S, then there exists a partition $S = A_1 \cup A_2 \cup \cdots \cup A_k$ so that:*

1. *The partition is equipartite, i.e.* $\| A_{i_1}| - |A_{i_2}\| \leq 1$, for $i_1, i_2 = 1, 2, \dots, k$.
2. $M \leq k \leq N_1$.
3. *For all but at most $\varepsilon \binom{k}{2}$ of the pairs $\{A_{i_1}, A_{i_2}\}$, the subgraph $G(A_{i_1}, A_{i_2})$ is ε-regular.*

This regularity lemma will be applied to a graph whose vertex set is a subset of the set $\mathcal{L}$ of lines in our configuration. But we must first take care of some preliminaries. In particular we must bound the number of lines which are incident with any point in $\mathcal{P}$. We begin by applying an appropriate linear transformation so that each line in $\mathcal{L}$ has a slope from the interval $(-1/2, 1/2)$.

For each line $l_j \in \mathcal{L}$, we choose $\delta n^{1/2}$ points from $\mathcal{P}$ which lie on the line l_j, and hereafter when we refer to an *incidence* between a point in $\mathcal{P}$ and the line l_j, it will be assumed that the point in question is one of the $\delta n^{1/2}$ chosen points. For each i, we let d_i count the number of lines from $\mathcal{L}$ with which p_i is incident (in this restricted sense). We call d_i the *degree* of the point p_i. Then $\sum_{i=1}^{n} d_i = \delta cn$.

Similarly, if l_j and l_k are distinct lines in $\mathcal{L}$, we say that l_j and l_k *intersect* when there is a point $p_i \in \mathcal{P}$ which is incident with both of them. When l_j and l_k are distinct nonparallel lines in $\mathcal{L}$ and there is no point in $\mathcal{P}$ incident with both, we say l_j and l_k *cross*. We then let E denote the set of pairs of intersecting lines. Later, we will consider a subset of E as the edge set of a graph G whose vertices are lines in $\mathcal{L}$. For now, we note the $|E| = \sum_{i=1}^{n} \binom{d_i}{2} \leq \binom{cn^{1/2}}{2}$. On the otherhand

$$\sum_{i=1}^{n} d_i^2 \geq \frac{1}{n}\left(\sum_{i=1}^{n} d_i\right)^2 = \delta^2 c^2 n,$$

so that

$$\sum_{i=1}^{n}\binom{d_i}{2}=\frac{1}{2}\sum_{i=1}^{n}d_i^2-\frac{1}{2}\sum_{i=1}^{n}d_i\geq\tfrac{1}{2}\delta^2c^2n-\tfrac{1}{2}\delta cn>\delta^3c^2n,$$

and

$$\sum_{i=1}^{n}d_i^2=2\sum_{i=1}^{n}\binom{d_i}{2}+\sum_{i=1}^{n}d_i\leq 2c^2n.$$

Since $\sum_{i=1}^{n}d_i^2\leq 2c^2n$, we conclude that less that $2\delta^4n$ points from $\mathscr{P}$ have degree exceeding c/δ^2. Now suppose there are s points in $\mathscr{P}$ which have degree less than δ^2c. Then consider the equation $\sum_{i=1}^{n}d_i=\delta cn$ and the inequality $\sum_{i=1}^{n}d_i^2\leq 2c^2n$. We now proceed to systematically change the numbers $d_1, d_2, \ldots, d_n$, but at all times we will insure that these two constraints are satisfied. Once we begin to change the numbers $d_1, d_2, \ldots, d_n$, their physical significance vanishes, but we are at this moment only concerned with an inequality for s. Should any one of the degrees, say d_i, be less than δ^2c. We then let $\bar{d}$ denote the average degree of the remaining $n-s$ points, and observe We then perform the natural 'local exchange' of increasing d_i by 1 and decreasing d_j by 1. Since $(d_i+1)^2+(d_j-1)^2\leq d_i^2+d_j^2$, we conclude that the two constraints $\sum_{i=1}^{n}d_i=\delta cn$ and $\sum_{i=1}^{n}d_i^2\leq 2c^2n$ are satisfied after this exchange has been performed. After a series of such exchanges and a relabelling of the points, we may assume that $d_1=d_2=\cdots=d_s=\delta^2c$. We then let $\bar{d}$ denote the average degree of the remaining $n-s$ points, and observe that:

$$s(\delta^2c)+(n-s)\bar{d}=\delta cn,\quad\text{and}\quad s(\delta^2c)^2+(n-s)\bar{d}^2\leq 2c^2n.$$

A simple calculation shows that $s(2+\delta^2-2\delta^3)<n(2-\delta^2)$ and thus $s<n(1-\delta^3)$. Since $\delta^3n-\delta^4n>\delta^4n$, it follows that there are δ^4n points in $\mathscr{P}$ whose degree is at least δ^2c but no more than c/δ^2. Let $m=\delta^4n$ and relabel the points in $\mathscr{P}$ so that each of the points in the set $\mathscr{B}=\{p_1, p_2, \ldots, p_m\}$ satisfy the property that $\delta^2c\leq d_i\leq c/\delta^2$. We refer to these points as *bounded* points.

There are at least $\sum_{i=1}^{m}d_i\geq\delta^4n\delta^2c=\delta^6cn$ incidences between bounded points and lines in $\mathscr{L}$. Suppose there are s lines in $\mathscr{L}$ which have less than $\delta^7n^{1/2}$ incidences with bounded points. Then it follows that $s\delta^7n^{1/2}+(cn^{1/2}-s)\delta n^{1/2}>\delta^6cn$. Thus $(1-\delta^5)cn^{1/2}>s(1-\delta^6)$, and $s<(1-\delta^6)cn^{1/2}$. We conclude that there are at least $\delta^6cn^{1/2}$ lines in $\mathscr{L}$ each having at least $\delta^7n^{1/2}$ incidences with points in the set $\mathscr{B}$ of δ^4n bounded points.

At this point, we have arrived at a new configuration with the additional property that the degree of any point does not exceed the average by an appreciable amount. So we can reformulate the properties of this configuration as follows. We have a set $\mathscr{B}=\{p_1, p_2, \ldots, p_m\}$ of m points where $m=\delta^4n$. Setting $c_1=\delta^4c$ and $\delta_1=\delta^5$ so that $\delta^6cn^{1/2}=c_1m^{1/2}$ and $\delta^7n=\delta^1m^{1/2}$, then we have a family $\mathscr{L}_{\mathscr{B}}=\{l_1, l_2, \ldots, l_{t_1}\}$ of $t_1=c_1m^{1/2}$ lines each of which has exactly $\delta^1m^{1/2}$ incidences with points from $\mathscr{B}$.

For each $i=1, 2, \ldots, m$, there are e_i lines from $\mathscr{L}_{\mathscr{B}}$ which are incident with p_i. There are $\sum_{i=1}^{m}e_i=\delta_1c_1m$ incidences and $\sum_{i=1}^{m}\binom{e_i}{2}>\delta_1^3c_1^2m$ pairs of lines from $\mathscr{L}_{\mathscr{B}}$ which intersect in a point from $\mathscr{B}$ incident with both. Most importantly, we know that $e_i\leq d_i\leq c/\delta^2=c_1/\delta_1^{6/5}<c_1/\delta_1^2$. The reader should note that we no longer have a lower bound on the degrees $e_1, e_2, \ldots, e_m$, since we have discarded some of the lines in $\mathscr{L}$.

Next, we let G be the graph whose vertex set is the set $\mathscr{L}_{\mathscr{B}}$ of lines and whose edge set is the set E_{β} of pairs of lines from $\mathscr{L}_{\mathscr{B}}$ which intersect. We know $|E_{\mathscr{B}}|\geq\delta_1^3c_1^2m$. Then let $M=\max\{10^{20}, \delta^{-100}\}$, and set $\varepsilon=1/M$. For these values of ε and M, we let N_1 and N_2 be the constants provided by the regularity lemma. We note that N_2 is very large compared to N_1. Hereafter, we will assume that $n\geq N_1^{21}$ and $c\geq N_2$. With these assumptions, it is certainly true that $c_1m^{1/2}\geq N_2$ so that the graph G has at least N_2 vertices.

We let $\mathcal{L}_{\mathcal{B}}=\mathcal{L}_1\cup\mathcal{L}_2\cup\cdots\cup\mathcal{L}_k$ be the partition guaranteed by the regularity lemma. Note that each $\mathcal{L}_i$ contains $c_1m^{1/2}/k$ lines. Then we let E_1 denote those edges with both end points in the same subset $\mathcal{L}_i$ for some i. It follows that $|E_1|\leq k(c_1m^{1/2}/k)^2\leq c_1^2m/M<\delta_1^4c_1^2m$. Then let E_2 denote the set of edges between distinct subsets $(\mathcal{L}_{i_1},\mathcal{L}_{i_2})$ for which $G(\mathcal{L}_{i_1},\mathcal{L}_{i_2})$ is not ε-regular. Then $|E_2|\leq\varepsilon\binom{k}{2}(c_1m^{1/2}/k)^2\leq\delta_1^4c_1^2m$. So there are at least $\delta_1^3c_1^2m-2\delta_1^4c_1^2m>\delta_1^4c_1^2m$ edges whose end points lie in distinct subsets $\mathcal{L}_{i_1},\mathcal{L}_{i_2}$ for which $G(\mathcal{L}_{i_1},\mathcal{L}_{i_2})$ is ε-regular. We may then choose an ε-regular pair $(\mathcal{L}_{i_1},\mathcal{L}_{i_2})$ so that there are at least $\delta_1^4c_1^2m/\binom{k}{2}\geq\delta_1^4c_1^2m/k^2$ edges between $\mathcal{L}_{i_1}$ and $\mathcal{L}_{i_2}$. We can relabel the subsets so that $i_1=1$ and $i_2=2$. Note that the density of $G(\mathcal{L}_1,\mathcal{L}_2)$ is at least δ_1^4.

Now suppose that $\mathcal{L}'$ is any set of parallel lines in $\mathcal{L}_{\mathcal{B}}$. Then $|\mathcal{L}'|\delta_1m^{1/2}\leq m$ so that $|\mathcal{L}'|\leq(1/\delta_1)m^{1/2}<|\mathcal{L}_1|/3$. Now all lines have slope in the interval $(-\frac{1}{2},\frac{1}{2})$, so we can choose a line l_0 with slope m_0 so that at least one-third of the lines in $\mathcal{L}_1$ have slope in the interval $(m_0,\frac{1}{2})$, and at least one-third have slope in the interval $(-\frac{1}{2},\mathrm{m}_0)$. As for $\mathcal{L}_2$, we know that either one-third of the lines in $\mathcal{L}_2$ have slope in $(m_0,\frac{1}{2})$, or at least one third have slope in $(-\frac{1}{2},\mathrm{m}_0)$. So without loss of generality we may assume that $\mathcal{L}_1'$ and $\mathcal{L}_2'$ each contain $(c_1m^{1/2})/3k$ lines, each line from $\mathcal{L}_1'$ has slope in the interval $(m_0,\frac{1}{2})$ and each line from $\mathcal{L}_2'$ has in the interval $(-\frac{1}{2},m_0)$.

Next, we rotate the points in $\mathcal{B}$ so that the line l_0 has slope zero. Since the rotation is at most 30°, we know that each line in $\mathcal{L}$ has slope in the interval $(-3/2^{1/2},3/2^{1/2})$. Now each line in $\mathcal{L}_1'$ has positive slope and each line in $\mathcal{L}_2'$ has negative slope. For each pair of vertices (l_1,l_2) from $\mathcal{L}_1'\times\mathcal{L}_2'$ regardless of whether this pair intersects or crosses, we let $\alpha(l_1,l_2)$ measure the angle formed by this pair of lines (Figure 3). By convention, we measure the angle which includes the x-axis.

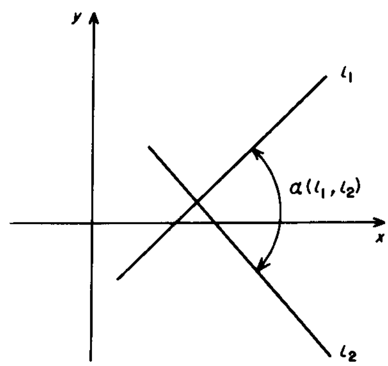


FIGURE 3.

We know that $0°\leq\alpha(l_1,l_2)\leq120°$. So we may apply a linear transformation of the form $(x,y)\rightarrow(ax,by)$ so that the average value of $\alpha(l_1,l_2)$ is 90°. Note that it is not necessary for any one of these angles to be exactly 90° in order for the average to be 90°. However it is easy to see that there must be a relatively large fraction of the angles which are somewhat near 90°, say in the interval [30°, 150°]. More precisely, the pigeon-hole principle shows that there exist subsets $\mathcal{L}_i''\subseteq\mathcal{L}_i'$ with $|\mathcal{L}_i''|=\frac{1}{10}|\mathcal{L}_i'|$ for $i=1,2$ so that $\alpha(l_1,l_2)$ is in the interval [30°, 150°] for every $l_1\in\mathcal{L}_1''$ and every $l_2\in\mathcal{L}_2''$. A second plication of the pigeon-hole principle followed by a suitable linear transformation allows us to conclude that there exist subsets $\mathcal{L}_i'''\subseteq\mathcal{L}_i''$ with $|\mathcal{L}_i'''|=\frac{1}{1000}|\mathcal{L}_i''|$ so that each line l_1 in $\mathcal{L}_1'''$ has slope in the interval [0·99, 1·01] and each line l_2 in $\mathcal{L}_2'''$ has slope in the interval [−1·01, −0·99]. As a consequence, $\alpha(l_1,l_2)$ is very nearly 90°. The essential fact is that the regularity lemma guarantees that many of the pairs $(l_1, l_2) \in \mathscr{L}_1''' \times \mathscr{L}_2'''$ will be edges in the graph G. This follows since $1/3 \cdot 10^4 > \varepsilon$. So the density of $G(\mathscr{L}_1''', \mathscr{L}_2''') \geq$ density of $G(\mathscr{L}_1, \mathscr{L}_2) - \varepsilon \geq \delta_1^4 - \varepsilon \geq \delta_1^5$. Hence there are at least $\delta_1^5 c_1^2 m/9 \cdot 10^8 k^2 \geq c_1^2 m/k^3$ pairs (l_1, l_2) from $\mathscr{L}_1''' \times \mathscr{L}_2'''$ which intersect in a point $p_i \in \mathscr{B}$ which is incident with both of these lines.

For each $p_i \in \mathscr{B}$, we let a_i count the number of lines from $\mathscr{L}_1'''$ with which p_i is incident and b_i count the number of lines from $\mathscr{L}_2'''$ with which p_i is incident. Since $a_i + b_i \leq e_i$, we know that neither a_i nor b_i can exceed c_1/δ_1^2. Now suppose there are s points in $\mathscr{B}$ for which either $a_i < c_1/k^4$ or $b_i < c_1/k^4$. Then it follows that

$$s\frac{c_1}{k^4}\cdot\frac{c_1}{\delta_1^2}+(m-s)\frac{c_1}{\delta_1^2}\frac{c_1}{\delta_1^2}\geq\frac{c_1^2 m}{k^3},$$

and thus $s < m[1-(1/k)]$.

So there m/k^4 points which satisfy the inequalities: $c_1/k^4 \leq a_i$, $b_i \leq c_1/\delta^2$. $c_1/k^4 \leq a_i$, $b_i \leq c_1/\delta_1^2$. We refer to these points as *doubly bounded* and let $\mathscr{D}$ be the set of m/k^4 doubly bounded points. Then there are at least $(m/k^4)(c_1/k^4)(c_1/k^4)$ pairs $(l_1, l_2) \in \mathscr{L}_1''' \times \mathscr{L}_2'''$ which intersect in a doubly bounded point incident with both l_1 and l_2.

Next, we apply the covering lemma of section 2 to the doubly bounded points with $r_1 = N_1^{16}$ and $r_2 = 256 r_1$. Of course this requires $m/k^4 \geq N_1^{16}$ which is certainly true when $n \geq N_1^{21}$. So we may assume that $\mathscr{F}$ is a family of almost disjoint squares covering at least 1/16-th of the m/k^4 doubly bounded points so that each square contains at least r_1 but no more than r_2 doubly bounded points. Now let Q be any square in $\mathscr{F}$. Consider the four triangles into which Q is subdivided by its diagonals (Figure 4)

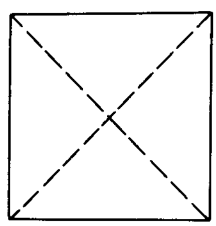


FIGURE 4.

At least one of these triangles contains at least 1/4th of the doubly bounded points in Q. Let T be such a triangle. Then for each doubly bounded point $p_i \in T$ we may choose $a_i - r_2$ lines from $\mathscr{L}_1'''$ which are incident with p_i but are not incident with any other doubly bounded point in Q. Similarly, we may choose $b_i - r_2$ lines from $\mathscr{L}_2'''$ which are incident to p_i but are not incident with any other doubly bounded point in Q. We note that if l_1 is a line from $\mathscr{L}_1'''$ chosen for p_i, l_2 is a line from $\mathscr{L}_2'''$ chosen for a distinct p_j, and these lines intersect in a doubly bounded point p_k incident with both lines, then p_k does not belong to Q (Figure 5).

It is then easy to see that if p_i and p_j are distinct points in the triangle T there are many crossings which occur inside Q (Figure 6). Since $a_i - r_2 \geq c_1/k^5$ and $b_i - r_2 \geq c_1/k^5$, it is easy to see that regardless of the location of p_i and p_j in T, T, there are at least c_1^2/k^{10} such crossings.

Since there are at least $\binom{r_1}{2}/4$ pairs of doubly bounded points to be considered in each square and there are at least $m/16k^4 r_2$ squares, we conclude that there are at least

$$\frac{m}{16k^4 r_2}\frac{\binom{r_1}{2}}{4}\frac{c_1^2}{k^{10}}$$

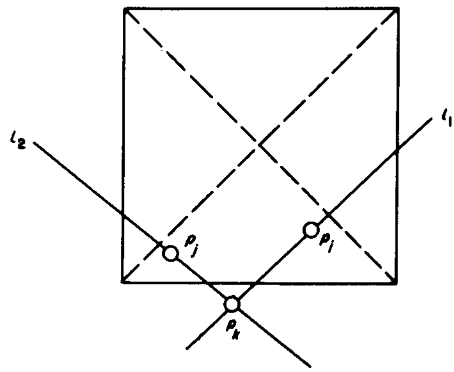


FIGURE 5.

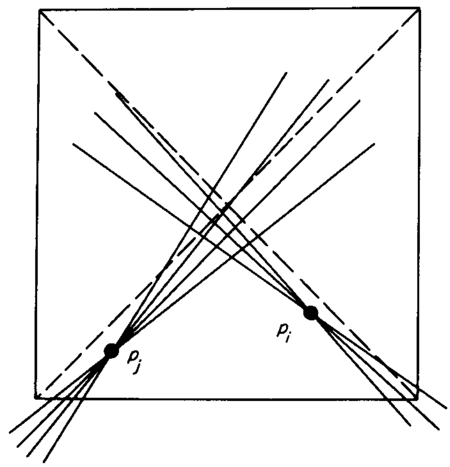


FIGURE 6.

such crossings altogether. However,

$$\frac{m}{16k^4r_2}\frac{\binom{r_1}{2}}{4}\frac{c_1^2}{k^{10}} \geqslant \frac{r_1}{k^{15}}c_1^2 m \geqslant \frac{r_1}{k^{16}}c^2 n \geqslant \frac{r_1}{N_1^{16}}c^2 n = c^2 n > \binom{cn^{1/2}}{2}$$

This implies that the number of crossings exceeds the total number of pairs of lines. Clearly this is impossible, and with this contradiction, the proof of the principal theorem is complete.

REFERENCES

1. P. Erdös, On some problems of elementary and combinatorial geometry, *Annali di Mat., Ser.* **4**, 103 (1974), 99–108.
2. P. Erdös, Some combinatorial problems in geometry, Conference held in Haifa, Israel, 1979. Lecture Notes in Mathematics.
3. P. Erdös, 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.
4. P. Erdös, Problems and results in combinatorial geometry, *Proc. N.Y. Acad. Sci.*, to appear.

5. E. Szemerédi, Regular Partitions of Graphs, in *Proc. Colloque Inter. CNRS* (J.-C. Bermond, J.-C. Fournier, M. Las Vergnas and D. Sotteau, eds), 1978, pp. 399–402.

Received 24 July 1982 and in revised form 21 November 1982

E. Szemerédi
Mathematical Institute of the Hungarian Academy of Science,
Budapest, Hungary
and
W. T. Trotter, Jr
Department of Mathematics and Statistics,
University of South Carolina
Columbia, SC 29208, U.S.A.