Discrete Mathematics 60 (1986) 147–153
North-Holland

ON SOME METRIC AND COMBINATORIAL GEOMETRIC PROBLEMS

P. ERDÖS
Mathematical Institute of the Hungarian Academy of Science, Budapest, Hungary

Received 19 January 1985

Let $x_1, \ldots, x_n$ be n distinct points in the plane. Denote by $D(x_1, \ldots, x_n)$ the minimum number of distinct distances determined by $x_1, \ldots, x_n$. Put

$$f(n) = \min D(x_1, \ldots, x_n).$$

An old and probably very difficult conjecture of mine states that

$$f(n) > cn/(\log n)^{\frac{1}{2}}.$$

$f(5) = 2$ and the only way we can get $f(5) = 2$ is if the points form a regular pentagon. Are there other values of n for which there is a unique configuration of points for which the minimal value of $f(n)$ is assumed? Is it true that the set of points which implements $f(n)$ has lattice structure? Many related questions are discussed.

I have published many papers on this and related topics [1]. Important progress has been made over the last few years on many of these problems and I will give a short review of some of these at the end of this paper and also state there some of the remaining problems, but first of all I will state some new problems. Usually we will restrict ourselves to the plane though many interesting questions can be posed in higher dimensions and even on the line (though the problems on the line are almost entirely of number theoretic and combinatorial character); also I almost entirely ignore our numerous problems and results with George Purdy since we plan to write both a survey paper and a book on these questions, but enough of idle talk and let us see some action.

Let $x_1, x_2, \ldots, x_n$ be n distinct points in the plane, denote by $D(x_1, \ldots, x_n)$ the number of distinct distances determined by $x_1, \ldots, x_n$. Put

$$f(n) = \min D(x_1, \ldots, x_n), \tag{1}$$

where in (1) the minimum is to be taken for all possible choices of $x_1, \ldots, x_n$. Denote by $\mathrm{d}(x_i, x_j)$ the distance from x_i to x_j, and denote by $g(n)$ the largest number of pairs x_i, x_j for which $\mathrm{d}(x_i, x_j) = 1$. The determination of $f(n)$ or $g(n)$ are probably hopeless, and to get good upper and lower bounds for these functions is also very difficult. As far as I know these problems were first stated by me in 1946 [2] and as I stated recently, important progress has been made on them. The strongest conjectures are [2]

$$f(n) > c_1 n/(\log n)^{1/2} \quad (\text{perhaps } f(n) = (1 + \mathrm{o}(1))cn/(\log n)^{1/2} \tag{2}$$

0012-365X/86/$3.50 © 1986, Elsevier Science Publishers B.V. (North-Holland)

and

$$g(n) < n(\exp c_2 \log n/\log\log n). \tag{3}$$

It was already shown in [2] that if (2) and (3) are true they are best possible. I offered (and offer) 500 dollars for a proof or disproof of (2) and (3) and 250 dollars for $g(m) < n^{1+\varepsilon}$. (Szemerédi considers this perhaps attackable.) I have made no progress on these problems since [2]. Quoting from an old paper of E. Landau, so far my remarks do not justify writing a new paper, but now it is time to state some new problems.

Let $x_1, \ldots, x_n$ be a set which implements $f(n)$ (i.e., $D(x_1, \ldots, x_n) = f(n)$). Is it true that $x_1, \ldots, x_n$ has lattice structure?

I really have no idea and the problem is perhaps too vaguely stated. The first step would be to decide if there always is a line which contains $cn^{\frac{1}{2}}$ of the x_i (and in fact n^{ε} instead of $n^{\frac{1}{2}}$ would already be interesting). A stronger result would be that there are $cn^{\frac{1}{2}}$ (or $n^{1-\varepsilon}$) lines which contain all the x_i. The only result in this direction, due to Szemerédi [1], states that if $D(x_1, x_2, \ldots, x_n)$ in $o(n)$ and $n > n_0(k)$ then there always is a line which contains at least k of our points. In fact Szemerédi's result gives that such a line can be chosen as the perpendicular bisector of two of our points, and also that there are $o(n)$ lines which contain all our points. The first new problem I want to state is the following. Assume

$$D(x_1, \ldots, x_n) = o(n). \tag{4}$$

Is it then true that (4) implies that there are always four x_i's which determine less than four distinct distances? I would expect that the answer is negative, i.e., I think one can find, for every $\varepsilon > 0$ and $n > n_0(\varepsilon)$, n points for which $D(x_1, \ldots, x_n) < \varepsilon n$ but any four of our x's determine at least four distinct distances. I got nowhere with this simple and I hope attractive problem, but perhaps I overlook a trivial point.

Denote by $D_k(x_1, \ldots, x_n)$ the smallest value of $D(x_{i_1}, \ldots, x_{i_k})$ for any choice of k of our points $x_1, \ldots, x_n$. Trivially, if (4) holds then

$$D_k(x_1, \ldots, x_n) \leq \binom{k}{2} - k + 2. \tag{5}$$

This is really obvious since (4) implies that every x_i is the center of a circle which contains many of our x_i's. It is conceivable that for every k and $n > n_l(k)$ there are points for which (4) holds and for every $l \leq k$, (5) is best possible, i.e., $D(x_1, \ldots, x_n) < \varepsilon n$ for every $n > n_0(\varepsilon, k)$ and for every $l \leq k$, $D_l(x_1, \ldots, x_n) = \binom{l}{2} - l + 2$. If the answer to this question is affirmative then try to determine or estimate the smallest $k = k(n)$ for which (4) implies $D_k(x_1, \ldots, x_n) < \binom{k}{2} - k + 2$. Clearly many related questions can be asked, but since I have no nontrivial results at the moment we leave the formulation of these to the reader.

Perhaps it is more interesting to ask the inverse problem. The simplest interesting cases probably are the following. Assume

$$D_3(x_1, \ldots, x_n) = 3, \tag{6}$$

i.e., our set contains no isosceles triangles. How small can $D(x_1, \ldots, x_n)$ be? Clearly (6) implies $D(x_1, \ldots, x_n) \geq n-1$ and I believe that (6) implies

$$D(x_1, \ldots, x_n)/n \to \infty \tag{7}$$

If (7) is true, then it is fairly close to being best possible. To see this let $a_1 < a_2 < \cdots < a_n$ be a sequence of integers which contains no three terms in an arithmetic progression and for which a_n is minimal. By a well known result of Behrend [3] $\min a_n < n(\exp(c(\log n)^{\frac{1}{2}}))$. Thus there are n points on the line for which (6) holds and

$$D_n(x_1, \ldots, x_n) < n \exp(c(\log n)^{\frac{1}{2}}).$$

Assume next

$$D_4(x_1, \ldots, x_n) = 5, \tag{8}$$

i.e., every set of four points determines at least five distinct distances. How small can $D(x_1, \ldots, x_n)$ be? I cannot decide whether (8) implies $D(x_1, \ldots, x_n) > cn^2$. Also, if (8) holds what is the largest $h(n)$ so that we must have a subset of $h(n)$ points all of whose distances are distinct? If the x_i are on a line then it is easy to see that $h(n) \geq [\frac{1}{2}(n+1)]$. In fact it is easy to see that the hypergraph formed by the four-tuples $(x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4})$ with $\mathrm{d}(x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}) = 5$ is two chromatic if the x_i are on a line. If the x_i are not all on a line then (8) very likely permits this hypergraph to have arbitrarily large chromatic number. Clearly here too many further problems can be formulated; many of them will have a Ramsey-like flavor but I expect that geometric methods will give better results. Here I state only: It is immediate that if $D_6(x_1, \ldots, x_n) = 14$, then $D(x_1, \ldots, x_n) \geq \frac{1}{2}\binom{n}{2}$ since the same distance can occur only twice. I have not been able to show that $D_5(x_1, \ldots, x_n) = 9$ implies $D(x_1, \ldots, x_n) > cn^2$; perhaps here too I overlook a trivial point. Assume next that $x_1, x_2, \ldots, x_n$ is such that every set of 5 points contains four points for which all the distances are different? What is the largest m so that our set contains m points for which all distances are different. The fact that m tends to infinity with n is of course a consequence of Ramsey's theorem but geometric considerations no doubt will give very much better estimates. Let me remind the reader that the following problem is still far from solved:

Let $h_k(n)$ be the largest integer so that if $x_1, \ldots, x_n$ is any set of n distinct points in k dimensional space one can always find a subset of $h_k(n)$ points for which all distances are different. It is easy to see that $h_k(n) > n^{\varepsilon_k}$ but the best possible value of ε_k is not known. $\varepsilon_1 = \frac{1}{2}$ follows from a result of Ajtai, Komlós, Sulyok and Szemerédi [12]. $h_1(7) = 3$ is the only exact result. Here I just want to remark that infinite problems are often (of course not always) easier than finite ones; e.g. I proved [4] that if S is any set of power m ($m \geq \aleph$) in k-dimensional euclidean space then one can always find a subset $S_1 \subseteq S$, $|S_1| = m$ for which all distances are distinct. The proof uses the axiom of choice but not the continuum hypothesis.

The fact that infinite problems are often simpler than finite ones led Ulam and me to make the following somewhat impudent and conceited joke: "The infinite we do right away, the finite takes somewhat longer." This was stolen from the U.S. Navy where they said during World War II: "The difficult we do right away, the impossible takes somewhat longer."

Let us now return to our original problem. Assume that $x_1, \ldots, x_n$ implements $f(n)$, i.e.,

$$D(x_1, \ldots, x_n) = f(n). \tag{9}$$

Is it then true that for every k if $n > n_0(k)$ there is a subset $x_{i_1}, \ldots, x_{i_t}$ satisfying

$$D_k(x_{i_1}, \ldots, x_{i_t}) = f(k)? \tag{10}$$

The answer may very well be different for different values of k. For example, it is not at all clear if (9) implies (10) for $k = 3$, i.e., is it true that if $x_1, \ldots, x_n$ implements $f(n)$ then our set must contain an equilateral triangle. I expect that the answer is yes. Assuming that I am right, then the following problem must be faced. Assume that $x_1, \ldots, x_n$ contains no equilateral triangle; how small can $D(x_1, \ldots, x_n)$ be? It can certainly be less than $cn/(\log n)^{\frac{1}{2}}$ since the square lattice does not contain an equilateral triangle, but I think that $D(x_1, \ldots, x_n) > (1 + c)f(n)$ in this case. On the other hand I believe that if $x_1, \ldots, x_n$ contains neither squares nor equilateral triangles then

$$D(x_1, \ldots, x_n) \Big/ \frac{n}{\sqrt{\log n}} \to \infty \tag{11}$$

I do not see how to attack (11). If we assume $D_4(x_1, \ldots, x_n) = 4$, then I strongly believe that (11) holds, but I can certainly not justify my belief.

Observe that $f(5) = 2$ is only possible for the regular pentagon. I believe that (9) may very well hold without our set containing a regular pentagon and in fact perhaps if $x_1, \ldots, x_n$ contains a regular pentagon then (9) cannot hold. This is the reason for my belief that (9) may imply (10) for some, but not all, values of k.

How many choices of $x_1, \ldots, x_n$ are there which satisfy (9)? Denote this number by $r(n)$. Two implementations are considered different if there is no similarity transformation which passes one onto the other. Of course $r(3) = r(5) = 1$, while $r(4) = 3$ (the square, the equilateral triangle with its center and two equilateral triangles having a common side). Perhaps for every large n there are two implementations of (9) which have only two points in common. This certainly is the case for $n = 4$. it would be of interest to find all ways of satisfying (9) for some $n \geqslant 6$ for large n; this probably will not be easy. Also; can one implement $f(n)$ and $g(n)$ at the same time for all n? For small values of n this is certainly possible.

The general problem which faces us can perhaps be stated as follows. Let us assume that $D(x_1, \ldots, x_n)$ is given, and is of course $\geqslant f(n)$. What is the range of $D_k(x_1, \ldots, x_n)$? Or, if $D_k(x_1, \ldots, x_n)$ is given what can be said about the range of $D(x_1, \ldots, x_n)$? These problems are perhaps too general to lead to any interesting results.

To end this somewhat confused chapter, I would like to mention a special problem which intrigued me for some time. Let $x_1, \ldots, x_n$ be n points in general position, i.e., no three on a line and no four on a circle. Is it prossible $D(x_1, \ldots, x_n) = n-1$, so that the ith distance occurs i times $1 \leq i \leq n$? (The distances are not ordered by size or in any other way.) This is trivially possible for $n=3$ and $n=4$, and Pomerance showed an example for $n=5$ while two Hungarian students found an example for $n=6$. (Unfortunately I have forgotten both the examples and their names.) Palásthy and Liu independently found such a construction for $n=7$. Palásthy's proof will appear in: Stud. Sci. Math. Hungar.

Perhaps no such examples exist for $n>6$ or at least for sufficiently large n. Denote by $h(n)$ the minimum value of $D(x_1, \ldots, x_n)$ if the x_i are in general position. Perhaps for $n > n_0$, $h(n) \geq n$, but I cannot even prove $h(n) > \frac{1}{2}(1+\varepsilon)n$. I hope in fact that

$$\frac{h(n)}{n} \to \infty \quad \text{but} \quad \frac{h(n)}{n^2} \to 0. \tag{12}$$

Another old conjecture of mine states that if $x_1, \ldots, x_n$ are n points in the plane then there is always one of them which has at most $\exp(c \log n / \log \log n)$ other points equidistant from it. This is easy with $cn^{\frac{1}{2}}$ and Beck proved it recently with $o(n^{\frac{1}{2}})$. The proof is not published and is surprisingly complicated. The square (or triangular) lattice shows that $\exp(c \log n / \log \log n)$ if true is best possible. If the n points form a convex n-gon I conjectured that there is a vertex which has no three other vertices equidistant from it. This was disproved by Danzer (unpublished), but I hope it holds with 4 instead of 3 even though I could not even prove it with n^{ε} instead of 4. Perhaps here too I may overlook a simple argument.

I conjectured and Altman [5] proved that for a convex n-gon $D(x_1, \ldots, x_n) \geq [\frac{1}{2}n]$, equality e.g., for the regular polygon. I further conjectured that in a convex n-gon there is always a vertex x_1 so that there are at least $[\frac{1}{2}n]$ distinct numbers among the number $d(x_1, x_i)$, $2 \leq i \leq n$. This conjecture is still open.

Moser and I conjectured nearly 30 years ago that in a convex n-gon the same distance can occur at most cn times, the best example we had is $3n+1$ points for which $\mathrm{d}(x_i, x_j) = 1$ has $5n$ solutions. I hoped that convexity can be weakened by assuming only that no three of the points are on a line. Purdy and I have an example [11] of n points no 3 on a line with $n2^{n-1}$ pairs of points at unit distance. Is it true that if $x_1, \ldots, x_n$ are n points no three (or more generally: no k) on a line then the number of pairs x_i, x_j with $\mathrm{d}(x_i, x_j) = 1$ is $< c_k\, n \log n$?

Let there be given n points no four on a line. Denote by $l(n)$ the largest integer so that one can always select $l(n)$ points, no three on a line. It is easy to see that

$$l(n) + \binom{l(n)}{2} \geq n \quad \text{or} \quad l(n) \geq \sqrt{2n-1}. \tag{13}$$

I could not improve (13) but could not disprove $l(n) > cn$.

Now I give a short outline of some of the recent progress of my old problems. The first progress on $g(n)$ was due to Beck and Spencer [6] who proved $g(n)<n^{\frac{3}{2}-\varepsilon}$ for some $\varepsilon>0$, this result has been improved and the proof simplified by various authors but at the moment $g(n)<n^{1+\varepsilon}$ still seems out of reach.

In 1952 Moser proved $f(n)>cn^{\frac{2}{3}}$ and recently Fan Chung [7] improved this to $f(n)>n^{\frac{5}{7}}$. This has since been improved by her and various other authors, but $f(n)>n^{1-\varepsilon}$ still seems out of reach. Beck, Szemerédi and Trotter [6] solved several of my problems on lines but the following old problem is still open: Let there be given n points no five on a line, denote by $L(n)$ the largest number of lines which contain four of our points. Is it true that $L(n)=o(n^2)$.

Kárteszi proved more than 20 years ago that $L(n)>cn\log n$ and Grunbaum [8] improved this to $L(n)>n^{\frac{3}{2}}$ which could very well be best possible.

More than 50 years ago Esther Klein (Mrs. Szekeres) asked: Is it true that for every n there is an $f(n)$ so that if $f(n)$ points are given in the plane no three on a line, then one can always find n of them which form the vertices of a convex n-gon. She observed that $f(4)=5$ and Szekeres conjectured $f(n)=2^{n-2}+1$. Makai and Turán confirmed this for $n=5$ and Szekeres and I proved [9]:

$$2^{n-2}+1\leqslant f(n)\leqslant\binom{2n-4}{n-2}. \tag{14}$$

As far as I know this is all that is known about $f(n)$. In 1976 I asked: Is it true that there is an $F(n)$ so that if $F(n)$ points are given in the plane then there are always n of them which form a convex n-gon and no other point is in the interior of this convex polygon. Trivially $F(4)=5$ and Harborth proved $F(5)=10$ and thought that $F(n)$ does not exist for $n\geqslant 7$. He was not sure about $n=6$, $n=6$ is still open but Horton proved Harborth's conjecture [10].

To end this paper let me state a graph theoretial conjecture of Szemerédi: Is it true that for every $\varepsilon>0$ and k there exists n_0 so that for every $n>n_0(\varepsilon,k)$ every $G(n;c)$ ($G(n;c)$ is a graph of n vertices and c edges) $c>n^{1+\varepsilon}$ contains two sets of vertices S_1 and S_2, $|S_1|=|S_2|=k$, $S_1\cap S_2=\emptyset$ so that if $A\subset S_1$, $B\subset S_2$, $|A|=|B|=[\frac{1}{2}k]$ are arbitrary subsets of S_1 and S_2, then there is at least one edge joining a vertex of A and a vertex of B. Szemeredi showed that this remarkable conjecture would imply $f(n)<n^{1+\varepsilon}$ (it seems though that this conjecture has been disproved).

References

[1] P. Erdös, On some problems of elementary and combinatorial geometry, Ann. Math. Ser. 4, 103 (1975) 99–108;
Some applications of graph theory and combinatorial methods to number theory and geometry, Algebraic methods in Graph Theory, Coll. Math. Soc. J. Bolyai 25 (1981) 137–148;
Some combinatorial problems in geometry, Geom. and Diff. Geom., Proc. Haifa, Israel 1979. Lecture Notes in Math. 792 (Springer, Berlin, 1980) 46–53.

[2] P. Erdös, On sets of distances of n points, Amer. Math. Monthly 53 (1946) 248–250.

[3] F.A. Behrend, On sets of integers which contain no three terms in arithmetic progression, Proc. Nat. Acad. Sci. USA 32 (1946) 331–332.

[4] P. Erdös, Some remarks on set theory, Proc. Amer. Math. Soc. 1 (1950).

[5] E. Altman, On a problem of P. Erdös, Amer. Math. Monthly 70 (1963) 148–154. (See also: Some theorems on convex polygons, Canad. Math. Bull. 15 (1972) 329–340.)

[6] T. Beck and J. Spencer, Unit distances, J.C.T. Ser. A 37 (1984), 231–238;
J. Spencer, E. Szemerédi, W. Trotter Jr., Unit distances in the Euclidean Plane, Graph Theory and Combinatorics, Proc. Cambridge Comb. Conf. Acad. Press (1984), 293–303;
E. Szemerédi and W. Trotter, Extremal problems in discrete geometry, Combinatorica 3 (1983) 381–397;
J. Beck, On the lattice property in the plane and some problems of Dirac, Motzkin and Erdös in combinatorial geometry, ibid 281–297.

[7] Fan Chung, The number of different distances determined by n points in the plane, J.C.T. Ser. A. 36 (1984) 392–354.

[8] B. Grübaum, New Views on Some Old Questions of Combinatorial Geometry, Colloquio Internazion ale sulle Teorie Combinatorie, Rome 3–15 Sept. 1973, Tomo 1, pp. 451–468, published in Accademia Nazionale dei Lincei (1976) Vol. 17.

[9] P. Erdös and G. Szekeres, A combinatorial problem in Geometry, Compositio Math. 2 (195), 463–470 and
On some extremum problems in geometry, Ann. Univ. Sci. Univ. Budapest 3–4 (1960–61) 53–62.

[10] H. Harborth, Konvexe Fünfecke in ebenen Punktmengen, Elem. Math. 33 (1978) 116–118;
J.D. Horton, Sets with no empty convex 7-gons, Canad. Math. Bulletin 46 (1983) 482–489. (For a recent collection of geometric problems see W. Moser, Research problems in discrete geometry, Montreal. Mimeograph Notes.)

[11] G. Purdy, P. Erdös, Some Extremal Problems in Geometry IV, Proc. 7th S–E Conf. Combinatories, Graph Theory, and Computing, pp. 307–322.

[12] J. Komlós, M. Sulyok and E. Szemerédi, Linear problems in combinatorial number theory, Acta Math. Acad. Sci. Hungar. 26 (1975) 113–121.