DISTINCT DISTANCES DETERMINED BY SUBSETS OF A POINT SET IN SPACE

*David Avis**

School of Computer Science
McGill University
3480 University
Montreal, Canada, H3A 2A7

Paul Erdős

Mathematical Institute of the Hungarian Academy of Sciences
Budapest 1364, pf. 127
Hungary

János Pach †

Mathematical Institute of the Hungarian Academy of Sciences
and
Courant Institute of Mathematical Sciences
New York University
251 Mercer Street
New York, NY 10012, U.S.A.

March 18, 1997

ABSTRACT

We answer the following question posed by Paul Erdős and George Purdy: determine the largest number $f_d(k) = f$ with the property that almost all k-element subsets of any n-element set in $\mathbf{R}^d$ determine at least f distinct distances, for all sufficiently large n. For $d = 2$ we investigate the asymptotic behaviour of the maximum number of k-element subsets of a set of n points, each subset determining at most i distinct distances, for some prespecified number i. We also show that if $k = o(n^{1/7})$, almost all k-element subsets of a planar point set determine distinct distances.

* Research supported by the Natural Science and Engineering Research Council Grant number A3013 and the F.C.A.R. Grant number EQ1678.
† Research supported by NSF Grant CCR-89-01484, Hungarian Science Foundation Grant OTKA-1814, and by the DIMACS Center at Rutgers University under NSF Grant STC-88-09648

1. Introduction

Let k, n be natural numbers, k fixed and n very large. Given a set of k points $S_k = \{p_1, \ldots, p_k\}$ in Euclidean d-space, let $|p_i - p_j|$ denote the distance between p_i and p_j, and let $D(S_k)$ denote the number of distinct values of $|p_i - p_j|$, $1 \le i < j \le k$. The following problem was raised by Paul Erdős and George Purdy[5]. Determine the largest number $f_d(k) = f$ with the property that almost all k-element subsets of any n-element set $S \subseteq \mathbf{R}^d$ determine at least f distinct distances, that is

$$\min_{\substack{S \subseteq \mathbf{R}^d \\ |S| = n}} \frac{|\{S_k \subseteq S : |S_k| = k \text{ and } D(S_k) \ge f\}|}{\binom{n}{k}} \to 1$$

as n tends to infinity. In sections 2 and 3 we obtain matching upper and lower bounds for $f_d(k)$ when $d \ge 3$.

It is not difficult to see that $f_1(k) = f_2(k) = \binom{k}{2}$ for every integer k. In fact, these equations follow from the general theory developed in Section 3. As Erdős and Purdy suggested almost 20 years ago, it might be interesting to investigate the following more delicate question. Given $0 < i \le \binom{k}{2}$, determine or estimate the asymptotic behaviour of the function $h_d^{i,k}(n)$, the maximum number of k-element subsets S_k of an n element set $S \subseteq \mathbf{R}^d$ with the property that $D(S_k) \le i$, that is, S_k determines at most i distinct distances. Obviously,

$$h_d^{i,k}(n) = o(n^k) \quad \text{for all } i < f_d(k),$$

$$h_d^{i,k}(n) = \Omega(n^k) \quad \text{for all } i \ge f_d(k).$$

(For non-negative functions f and g defined on the natural numbers, we write $f(n) = \Omega(g(n))$ if $f(n) \ge c\ g(n)$ for some positive constant c and all sufficiently large n.)

However, even in the plane it seems to be a hopelessly difficult task to determine the exact order of magnitude of these functions. For simplicity, we shall write $h^{i,k}(n)$ instead of $h_2^{i,k}(n)$. In particular, we are very far from knowing the answer to the following old problem of Erdős: What is the smallest value $k(i)$

for which

$$h^{i,k(i)}(n) = 0$$

holds for all n? It is conjectured that $k(i) = O(i \log^{1/2} i)$, and the best currently known bound $k(i) < i^{5/4} \log^c i$ is due to Chung, Szemerédi and Trotter[1]. In section 4 we obtain some results on $h^{\binom{k}{2} - i,\, k}$ for $1 \le i \le k - 2$.

2. Upper bound for $f_d(k)$

For $d \ge 3$ upper bounds on $f_d(k)$ come from the so-called Lenz examples (see[3]). For even d, set $t = d/2$, and fix k. For any give large n, distribute n points as evenly as possible amongst t mutually orthogonal unit circles with common centre, so that each circle contains either $\lfloor \frac{n}{t} \rfloor$ or $\lceil \frac{n}{t} \rceil$ points. Every pair of points chosen from distinct circles determines the same distance. Now consider a k-tuple S_k where the points are again chosen as evenly as possible from the t circles. Then it can easily be verified that $D(S_k) \le g_d(k)$ where

$$g_d(k) = \begin{cases} 0 & k = 1 \\ 1 & k = 2 \\ g_d(k-1) + \lfloor \frac{k-1}{t} \rfloor & k \ge 3 \end{cases} \qquad d \text{ even}, \ t = d/2. \tag{1}$$

It is easy to see that we can choose an $\varepsilon > 0$ for which at least εn^k k-subsets of the n points have the above property. This implies that $f_d(k) \le g_d(k)$.

For odd d, set $t = \lceil \frac{d}{2} \rceil$, and fix k. For any given large n, distribute n points as evenly as possible amongst t sets consisting of $t - 1$ mutually orthogonal unit circles with common centre and the line passing through the common centre orthogonal to all of the circles. Now again every pair of points chosen from distinct circles determines the same distance. However the points on the line each determine a distinct distance to the circles. Now consider a k-tuple S_k where the points are again chosen as evenly as possible from the t sets. In case of an uneven division, we choose $\lfloor \frac{k}{t} \rfloor$ points from the line. Then it can easily be verified that $D(S_k) \le g_d(k)$ where

$$g_d(k) = \begin{cases} 0 & k = 1 \\ 1 & k = 2 \\ g_d(k-1) + \lfloor \frac{k}{t} \rfloor & k \ge 3 \end{cases} \qquad d \text{ odd}, \quad t = \lceil \frac{d}{2} \rceil. \tag{2}$$

Again, this implies that $f_d(k) \le g_d(k)$. Our main result is to show that these bounds are in fact tight, that is $f_d(k) = g_d(k)$.

It is convenient to get a closed form expression for $g_d(k)$, which we now do in terms of the Turán number $T(k, t+1)$, which is the maximum number of edges in a graph on k vertices which does not contain a K_{t+1}. Turán showed that

$$T(k, t+1) = \frac{(t-1)(k^2 - r^2)}{2t} + \binom{r}{2}$$

where r is the remainder of k upon division by t. This bound is obtained for the so-called Turán graph, that is, a complete t-partite graph formed by distributing the k points into t parts as evenly as possible. Consider the Lenz example with d even and $t = d/2$. The number of distinct distances in an evenly distributed k-tuple is just the number of edges in the complement of the Turn graph plus one for the distance between the circles, so that

$$g_d(k) = \binom{k}{2} - T(k, t+1) + 1, \qquad d \text{ even.} \tag{1'}$$

When d is odd, let $t = \lceil \frac{d}{2} \rceil$. Then the number of distinct distances in an evenly distributed k-tuple in our construction is the number of edges in the complement of the Turán graph, plus one for each point on the line, and when $d \ge 5$, plus one for the distance between the circles. Therefore

$$g_d(k) = \binom{k}{2} - T(k, t+1) + \lfloor \frac{k}{t} \rfloor + 1_{\{ d > 3 \}}, \qquad d \text{ odd,} \tag{2'}$$

where $1_{\{ d > 3 \}}$ is the indicator function that is zero unless $d > 3$ and one otherwise.

3. Lower Bound on $f_d(k)$

For large n and fixed $d \geq 3$, let S be a set of n points in $\mathbf{R}^d$. Two ordered k-tuples of points $S_k = (p_1, \ldots, p_k)$ and $S_k' = (p_1', \ldots, p_k')$ are said to have the same *type* if

$$|p_i - p_j| = |p_g - p_h| \quad \text{if and only if} \quad |p_i' - p_j'| = |p_g' - p_h'| \qquad \text{for all } i, j, g, h.$$

Recalling that $D(S_k)$ denotes the number of distinct distances in S_k, we clearly have $D(S_k) = D(S_k')$ for any two k-tuples S_k, S_k' having the same type. Let $\mathbf{T} = \{T_1, \ldots, T_k\}$ be a set of k triples, where $T_i \subseteq S$. We say that $\mathbf{T}$ is *type invariant* if all k-tuples $S_k' = (p_1', \ldots, p_k')$ with $p_i \in T_i$ $(1 \leq i \leq k)$ have the same type. If $\mathbf{T}$ is type invariant, we define $D(\mathbf{T}) = D(S_k)$. We call S_k a *representative* k-tuple of $\mathbf{T}$. The role of type invariance is shown in the following lemma.

Lemma 3.1 Let d,k,w be fixed integers, and fix $\varepsilon > 0$. Then there exists an integer $n_0 = n_0(d, k, w, \varepsilon)$ with the following property. Any set S of $n \geq n_0$ points in $\mathbf{R}^d$ for which there are at least εn^k k-tuples that determine w distinct distances contains a type invariant set of k triples $\mathbf{T} = \{T_1, \ldots, T_k\}$, with $D(\mathbf{T}) = w$.

Proof: Let S satisfy the condition of the lemma, and let H denote the k-uniform hypergraph consisting of the k-tuples described. It follows from a well known result of Erdős that we can divide S into disjoint parts $V_1, \ldots, V_k$ and find an $\varepsilon' > 0$ so that there are at least $\varepsilon' n^k$ hyperedges in H having precisely one point from each V_i. In fact $\varepsilon' = \dfrac{k!}{k^k}\varepsilon$ will do. Let H' denote the subhypergraph consisting of these k-tuples. Since the total number of possible types of k-tuples is at most $\binom{k}{2}^{\binom{k}{2}}$, we can find an $\varepsilon^* > 0$ so that at least $\varepsilon^* n^k$ k-tuples in H' have the same type. Let H^* denote the subhypergraph of H' consisting of these k-tuples. Now ε^* is independent of n. It follows from another well known theorem of Erdős[4] that if n is sufficiently large, there is a system of triples $T_i \subseteq V_i$, such that every k-tuple with a point from each T_i is included in H^*. Therefore $\mathbf{T} = \{T_1, \ldots, T_k\}$ is a type invariant triple system with $D(\mathbf{T}) = w$. □

Given a type invariant triple system $\mathbf{T}$, we define an edge coloured graph $G = G(\mathbf{T})$ as follows. The vertices of G are the k triples in $\mathbf{T}$. Let $S_k = \{p_1, \ldots, p_k\}$ be a representative k-tuple for $\mathbf{T}$. Then (T_i, T_j)

is an edge of G whenever $|p_i - p_j| = |p_g - p_h|$ for some $\{g, h\} \neq \{i, j\}$. Furthermore, edges are given the same colour whenever their lengths are the same. Let $\bar{G}$ denote the complement of G. It is immediate from the definitions that $D(\mathbf{T})$ is equal to the number of edges in $\bar{G}$ plus the number of edge colours. Each edge in G implies certain geometric constraints on the triples defining its endpoints. These are summarized in the following lemma.

Lemma 3.2 Let (T_i, T_j) and (T_g, T_h) be monochromatic edges of the graph $G = G(\mathbf{T})$ defined above.

(i) If i, j, g, h are all distinct, then T_i and T_j determine two circles C_i and C_j with common centre and lying in two orthogonal 2-flats.

(ii) If i, j, g are distinct and $h = j$, then either case (i) occurs with T_g lying on C_i, or T_j lies on a line l orthogonal to, and passing through the common centre of, circles C_i and C_g determined by T_i and T_g. C_i and C_g need not be orthogonal. □

We call a vertex T_i that is not isolated in G *linear* if its points lie on a line, and *circular* otherwise. Based on the above lemma, we have that $G(\mathbf{T})$ satisfies the following combinatorial properties. describing $G(\mathbf{T})$.

Lemma 3.3 For $d \geq 3$, let $t = \lceil \frac{d}{2} \rceil$.

(i) $G(\mathbf{T})$ cannot contain K_{t+1} as a subgraph.

(ii) If d is odd then $G(\mathbf{T})$ cannot contain a K_t in which all of the vertices are circular.

(iii) If T_i is linear, then either all edges of a given colour are incident on T_i or none are.

Proof: (i) Suppose $G = G(\mathbf{T})$ contained a K_{t+1}. Each edge colour in G occurs at least twice. Therefore by Lemma 3.2, each vertex of the K_{t+1} corresponds to some triple T_i that lies in a subspace orthogonal to the triples corresponding to each of the other t vertices. Further at most one of these vertices can be linear. Therefore the K_{t+1} spans $2t + 1$ dimensions, a contradiction.

(ii) Arguing similarly to (i), the existence of such a K_t would require $2t$ dimensions, a contradiction.

(iii) Suppose T_i is linear and edge (T_i, T_j) is red. If there is a red edge (T_g, T_h) where $\{i, j\} \neq \{g, h\}$, we can apply Lemma 3.2 (i). But this contradicts the fact that T_i is linear. □

Finally, before proving the main lemma we need a couple of technical results that are contained in the following lemma.

Lemma 3.4 For natural numbers $k \geq t \geq 2$:

(i) $\sum_{i=0}^{t-1} \lfloor \frac{k-i}{t} \rfloor = k - t + 1.$

(ii) $T(k, t+1) \geq T(k, t) + \lfloor \frac{k}{t} \rfloor.$

Proof: (i) This is easily verified for $k = t$. Otherwise

$$\sum_{i=0}^{t-1} \lfloor \frac{k+1-i}{t} \rfloor - \sum_{i=0}^{t-1} \lfloor \frac{k-i}{t} \rfloor = \lfloor \frac{k+1}{t} \rfloor - \lfloor \frac{k+1-t}{t} \rfloor = 1.$$

(ii) Recall that $T(k, t+1)$ counts the number of edges in a complete t-partite graph on k vertices with the vertices as evenly distributed as possible. We construct a complete $t-1$-partite graph with vertices as evenly distributed as possible and $T(k, t)$ edges from this graph as follows: Choose a part with $\lfloor \frac{k}{t} \rfloor$ vertices, and sequentially move each vertex to one of the other $t-1$ parts that has minimum cardinality. The vertex moved is made adjacent to all vertices that are not in its new part. Since the cardinality of the new part is at least $\lfloor \frac{k}{t} \rfloor$, the number of edges in the graph goes down by at least one at each step. The result follows. □

We are now ready to prove the main lemma about type invariant triple systems.

Lemma 3.5 For a type invariant triple system T containing k triples of points in R^d, $d \geq 3$,

$$D(\mathbf{T}) \geq g_d(k).$$

Proof: Let $t = \lceil \frac{d}{2} \rceil$ and let $G = G(\mathbf{T})$ be the edge coloured graph described above defined on $\mathbf{T}$. By Lemma 3.3 (i), G cannot contain a K_{t+1}. Therefore $\bar{G}$, the complement of G, contains at least $\binom{k}{2} - T(k, t+1)$ edges. Recall that the number of distinct distances in $\mathbf{T}$ is the sum of the number of colours in G and the number of edges in its complement. If d is even we have from (1') that $D(\mathbf{T}) \geq \binom{k}{2} - T(k, t+1) + 1 = g_d(k)$, as required.

The case where d is odd is more difficult and we proceed by induction on k. We have trivially that

$D(\mathbf{T}) \geq g_d(k)$ for $k \leq t-1$. Fix some $k \geq t$ and assume inductively that $D(T) \geq g_d(j)$ for $j \leq k-1$. We first suppose that G has no K_t as a subgraph. If $d = 3$ then G has no edges and all distances are distinct. If $d \geq 5$, then the complement $\bar{G}$ has at least $\binom{k}{2} - T(k,t)$ edges, so

$$D(\mathbf{T}) \geq \binom{k}{2} - T(k,t) + 1 \geq \binom{k}{2} - T(k,t+1) + \lfloor \frac{k}{t} \rfloor + 1 = g_d(k),$$

where the second inequality follows from Lemma 3.4(ii).

We now assume that G contains a K_t, which by Lemma 3.3 (ii) has at least one linear vertex, say T_i. If $k = t$ we are done: this case is impossible for $d = 3$, and for $d \geq 5$ G must have at least two colours, so $D(\mathbf{T}) \geq 2 = g_d(t)$. If $k = t+1$ then $\bar{G}$ contains at least one edge (since there can be no K_{t+1}), the edges of G require two colours if $d > 3$ and one colour if $d = 3$, so $D(\mathbf{T}) \geq g_d(t+1)$ as required. So we may assume $k > t+1$. By Lemma 3.3(iii), there must be some colour, say red, for which all the red edges are incident on T_i. Since there can be no K_{t+1}, each of the $k-t$ vertices of G not in the K_t must be non-adjacent to at least one vertex in the K_t. Consider deleting the t triples corresponding to the vertices of the K_t from the k triple system $\mathbf{T}$ obtaining the $k-t$ triple system $\mathbf{T}'$. Since $\mathbf{T}'$ is also type invariant, we may apply the inductive hypothesis to it. Let $G' = G(\mathbf{T}')$. The number of colours in G' is at least one less than the number of colours in G. The number of edges in $\bar{G}'$ is at least $k-t$ less than the number of edges in $\bar{G}$. Therefore

$$\begin{aligned} D(\mathbf{T}) &\geq D(\mathbf{T}') + 1 + k - t \\ &\geq g_d(k-t) + 1 + k - t \\ &= g_d(k) + 1 + k - t - \sum_{i=0}^{t-1} \lfloor \frac{k-i}{t} \rfloor \\ &= g_d(k) \end{aligned}$$

as required. The first equation above comes from iterating equation (2), the second follows from Lemma 3.4(i). □

We can now prove the main theorem.

Theorem 3.1 For all $d \geq 3$, $f_d(k) = g_d(k)$.

Proof: In Section 2 we showed that $f_d(k) \leq g_d(k)$, so it remains to show that $g_d(k)$ is also a lower bound. Let $t = \lceil \frac{d}{2} \rceil$. Assume that there is an $\varepsilon > 0$ for which one can find arbitrarily large sets of n points where at least εn^k of the k-subsets determine w distinct distances. By Lemma 3.1, we can find a type invariant triple system $\mathbf{T} = \{T_1, \ldots, T_k\}$ which also determines w distinct distances. From Lemma 3.5 we have $w \geq g_d(k)$, completing the proof. □

4. Results for $d = 2$

In this section we obtain some results of $h^{i,k}(n)$, the maximum number of k-element subsets S_k of a planar set of n points with the property that $D(S_k) \leq i$.

Proposition 4.1 Let $k \geq 3$, $2 \leq i \leq k-2$ be fixed and let n tend to infinity. Then

(i) $\Omega(n^{k-1} \log n) \leq h^{\binom{k}{2}-1,\, k}(n) \leq O(n^{k-2/3})$,

(ii) $\Omega(n^{k-1}) \leq h^{\binom{k}{2}-i,\, k}(n) \leq O(n^{k-1})$.

Proof: (i) The number of isosceles triangles determined by the points of an $n^{1/2}$ *by* $n^{1/2}$ piece of the integer lattice is at least $cn^2 \log n$, for some positive constant c. For each isosceles triangle we can pick $k-3$ other points arbitrarily, to obtain a k-element subset with at most $\binom{k}{2} - 1$ distinct distances. This shows the lower bound. On the other hand, the number of those k-element subsets of an n-element set S which contain two elements equidistant from a third one is at most

$$(\text{ number of isosceles triangles in S }) \binom{n-3}{k-3} = O(n^{k-2/3}),$$

because the first term of this product is known to be $O(n^{7/3})$[6]. The number of those k-element subsets of S which contain two disjoint pairs determining the same distance, is at most

$$\binom{n-4}{k-4} \sum_{\Delta} \binom{m(s, \Delta)}{2},$$

where $m(s, \Delta)$ denotes the number of point pairs in S determining distance Δ. Thus, $\sum_{\Delta} m(s, \Delta) = \binom{n}{2}$ and by [7] and[2], $m(s, \Delta) \le cn^{4/3}$ for every Δ. This implies that the above expression cannot exceed

$$\binom{n-4}{k-4} \frac{\binom{n}{2}}{cn^{4/3}} \binom{cn^{4/3}}{2} = O(n^{k-2/3}),$$

proving the upper bound in (i). For part (ii) of the proposition, we record the fact that the number of 4-tuples consisting of two disjoint pairs of vertices determining the same distance is $O(n^{10/3})$.

(ii) For the lower bound, it suffices to consider $i = k - 2$. Let S be an n-element set consisting of a point p and $n - 1$ points equidistant from p. Then any k-element subset of S which contains p determines at most $\binom{k}{2} - k + 2$ distinct distances, giving the lower bound. For the upper bound it suffices to consider $i = 2$. Let us consider an n-element point set S in the plane, and let S_k be a k-element subset of S that determines at most $\binom{k}{2} - 2$ distinct distances. Let G_k be an edge coloured graph with vertex set $V(G_k) = S_k = \{p_1, \ldots, p_k\}$, two vertices p_a and p_b being joined by an edge if and only if $|p_a - p_b| = |p_c - p_d|$ for some $\{c, d\} \ne \{a, b\}$. Since S_k determines $\binom{k}{2} - 2$ distinct distances, colours red and blue suffice. We consider five cases.

Case (1): G_k contains three independent monochromatic edges.

We proceed as in part (i) of the proposition. The number of such k-element subsets is at most

$$\binom{n-6}{k-6} \sum_{\Delta} \binom{m(s, \Delta)}{3} \le \binom{n-6}{k-6} \frac{\binom{n}{2}}{cn^{4/3}} \binom{cn^{4/3}}{3} = O(n^{k-1}).$$

Case (2): G_k contains disjoint red and blue isosceles triangles.

The number of 6-tuples determining two disjoint isosceles triangles is at most the square of the total number of isosceles triangles determined by the set of points, which is $O(n^{14/3})$. Thus the number of such k-element subsets is at most $O(n^{k-1})$.

Case (3): G_k contains a (say) red isosceles triangle and two blue edges, which are all mutually disjoint. The number of such 7-tuples is at most the number of isosceles triangles times the number of 4-tuples determining two monochromatic edges. From part (i) of the proposition, this is at most $O(n^{7/3})O(n^{10/3}) = O(n^6)$. Again the number of k-element subsets for this case is at most $O(n^{k-1})$.

Case (4): G_k contains two red edges and two blue edges, which are all mutually disjoint. The number of such 8-tuples is at most $O(n^{10/3})O(n^{10/3}) = O(n^7)$, and the number of k-element subsets for this case is at most $O(n^{k-1})$.

Case (5): There is a point p_a in G_k that is (a) incident to 3 edges of the same colour, or (b) incident to at least one red edge and at least one blue edge, or (c) incident to 2 edges of the same colour and G_k is monochromatic.
We want to assign to each such k-element subset $S_k \subseteq S$ a $(k-1)$-element subset $S_k^{'} \subseteq S_k$ so that every $(k-1)$-tuple is assigned to only a bounded number of k-tuples. This can be done as follows. Set $S_k^{'} = S_k - \{p_a\}$. Given $S_k^{'}$ and G_k, the location of p_a is pretty much determined. If p_a has degree at least 3 in G_k then its position is uniquely determined. Otherwise there are at most 2 possible locations for p_a. Now for each $(k-1)$-tuple $S' \subseteq S$ there are at most $c_1\, k^6$ different k-tuples S_k with $S_k^{'} = S'$, for some positive constant c_1. To see this, note that the position of p_a is determined (up to a constant number of locations) by three points in Case (5)(a), at most six points in Case (5)(b), and at most four points in Case (5)(c). This shows that the number of k-tuples satisfying the condition of Case (4) is at most $c_1\, k^6 \binom{n}{k-1} = O(n^{k-1})$.

Cases (1)-(5) are exhaustive, so we have shown that the number of k-tuples determining at most $\binom{k}{2} - 2$ distinct distances is at most $O(n^{k-1})$ completing the proof of the proposition. □

Until now k was always assumed to be fixed as n tended to infinity. Next we investigate how fast k can grow with n so that it still remains true that almost all (in the sense defined in the introduction) k-element subsets of any set of n points in the plane determine $\binom{k}{2}$ distinct distances.

Theorem 4.2 Let n tend to infinity, $k = o(n^{1/7})$. Then almost all k-element subsets of a set S of n points in the plane determine $\binom{k}{2}$ distinct distances.

Proof: We prove the equivalent statement that there are at least $(1 - o(1))n^k$ *ordered* k-tuples that determine distinct distances. Let S be any set of n points in the plane. A point $p \in S$ is called *central* if there is a circle around p passing through at least $n^{3/7}$ elements of S. According to a theorem of Clarkson et al. [2] the maximum number of incidences among n points and m circles in the plane is $O(n^{3/5}m^{4/5} + n + m)$. Hence, letting m denote the number of central points of S, we obtain

$$mn^{3/7} = O(n^{3/5}m^{4/5} + n + m)$$

and so $m = O(n^{6/7})$. A straight line l is called *rich* if l contains at least $n^{1/2}$ elements of S. It is easy to see that the number of rich lines $r = O(n^{1/2})$.

Let us define a large number of ordered k-tuples of S by the following procedure. Pick any non-central point $p_1 \in S$. If $p_1, p_2, \ldots, p_i$ have already been determined for some $i < k$ so that all distances $d_1, d_2, \cdots, d_{\binom{i}{2}}$ induced by them are different, then we choose p_{i+1} to be any point of S satisfying the following properties.

(i) p_{i+1} is not central;

(ii) $|p_{i+1} - p_a| \neq d_b \qquad (a \leq i,\ b \leq \binom{i}{2})$;

(iii) p_{i+1} is not the mirror image of p_a with respect to a rich line l $(a \leq i)$;

(iv) p_{i+1} does not lie on the perpendicular bisector of p_a and $p_{a'}$ $(a, a' \leq i)$;

Let $n^{(i)}$, $n^{(ii)}$, $n^{(iii)}$, and $n^{(iv)}$ denote the number of points in S violating property (i)-(iv), respectively. Obviously, we have $n^{(i)} \leq m = O(n^{6/7})$, and $n^{(iii)} \leq ir = iO(n^{1/2})$. The bound

$$n^{(ii)} \leq i\binom{i}{2}n^{3/7}$$

follows from the fact that no p_a is central, thus the circle of radius d_b around p_a passes through at most $n^{3/7}$ points of S. Finally

$$n^{(iv)} \le \binom{i}{2} n^{1/2}$$

because the perpendicular bisector of p_a and $p_{a'}$ cannot be rich.

Thus the number of different ways we can define an ordered k-tuple $p_1, \ldots, p_k$ by the above procedure is at least

$$\prod_{0 \le i \le k} (n - i - O(n^{6/7}) - i\binom{i}{2} n^{3/7} - iO(n^{1/2}) - \binom{i}{2} n^{1/2}) \ge n^k e^{-ck/n^{1/7}} = (1 - o(1))n^k,$$

provided that $k = o(n^{1/7})$, and the result follows. □

It is easy to see that if we choose n equidistant points along a line and let $k = \Omega(n^{1/4})$, then a positive percentage of all k-tuples determine fewer than $\binom{k}{2}$ distinct distances. A simple calculation along the above lines shows that this bound on k cannot be improved, as the following holds.

Proposition 4.3 Let n tend to infinity and let $k = o(n^{1/4})$. Then almost all k-element subsets of a set of k points on a line determine $\binom{k}{2}$ distinct distances. □.

References

1. F.R.K. Chung, E. Szemerédi, and W. T. Trotter Jr, ‘‘The Number of Different Distances Determined by a Set of Points in the Euclidean Plane,’’ *Manuscript*, 1989.

2. K. Clarkson, H. Edelsbrunner, L. Guibas, M. Sharir, and E. Welzl, ‘‘Combinatorial Complexity Bounds for Arrangements of Curves and Surfaces,’’ *Discrete Comput. Geom.*, vol. 5, 1990.

3. P. Erdős, ‘‘On Sets of Distances on n Points in Euclidean Space,’’ *Magyar Tud. Akad. Mat. Kutató Int. Kozl*, vol. 5, pp. 165-169, 1960.

4. P. Erdős, ‘‘On Extremal Problems of Graphs and Generalized Graphs,’’ *Israel J. of Math.*, vol. 2, pp. 183-190, 1964.

5. P. Erdős and G. Purdy, ‘‘Some Extremal Problems in Geometry,’’ *J. Combin. Th.*, vol. 10, pp. 246-252, 1971.

6. J. Pach and M. Sharir, "Repeated Angles in the Plane and Related Problems," *J. Combin Th. A*, to appear.

7. J. Spencer, E. Szemerédi, and W. T. Trotter Jr, "Unit Distances in the Euclidean Plane," in *Graph Theory and Combinatorics*, ed. B. Bollobás, pp. 293-308, Academic Press, 1984.