COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA 15 (3) (1995) 379–408

POINT SETS WITH DISTINCT DISTANCES

HANNO LEFMANN and TORSTEN THIELE

Received April 15, 1993

For positive integers d and n let $f_d(n)$ denote the maximum cardinality of a subset of the n^d-grid $\{1,2,\dots,n\}^d$ with distinct mutual euclidean distances. Improving earlier results of Erdős and Guy, it will be shown that $f_2(n) \geq c \cdot n^{2/3}$ and, for $d \geq 3$, that $f_d(n) \geq c_d \cdot n^{2/3} \cdot (\ln n)^{1/3}$, where $c, c_d > 0$ are constants. Also improvements of lower bounds of Erdős and Alon on the size of Sidon-sets in $\{1^2, 2^2, \dots, n^2\}$ are given.

Furthermore, it will be proven that any set of n points in the plane contains a subset with distinct mutual distances of size $c_1 \cdot n^{1/4}$, and for point sets in general position, i.e. no three points on a line, of size $c_2 \cdot n^{1/3}$ with constants $c_1, c_2 > 0$. To do so, it will be shown that for n points in $\mathbb{R}^2$ with distinct distances $d_1, d_2, \dots, d_t$, where d_i has multiplicity m_i, one has $\sum_{i=1}^t m_i^2 \leq c \cdot n^{3.25}$ for a positive constant c. If the n points are in general position, then we prove $\sum_{i=1}^t m_i^2 \leq c \cdot n^3$ for a positive constant c and this bound is tight.

Moreover, we give an efficient sequential algorithm for finding a subset of a given set with the desired properties, for example with distinct distances, of size as guaranteed by the probabilistic method under a more general setting.

1. Introduction

In [12] Erdős and Guy considered the following problem: Determine the maximum size of a subset X of the $n \times n$-grid, that is the set $\{1,2,\dots,n\} \times \{1,2,\dots,n\}$, such that all mutual euclidean distances between different points of X are distinct, compare also [18]. Denoting the cardinality of such a set X by $f_2(n)$, they proved the following:

Theorem 1.1. [12] *For every integer $n \geq 3$,*

$$n^{\frac{2}{3} - \frac{c_1}{\ln \ln n}} \leq f_2(n) \leq c_2 \cdot \frac{n}{(\ln n)^{1/4}}, \tag{1}$$

where $c_1, c_2 > 0$ are constants.

To obtain the lower bound for $f_2(n)$, Erdős and Guy used Greedy-type arguments. The upper bound for $f_2(n)$ follows from a result of Landau [20], namely, that the number of integers less than x, which are representable as a sum of two squares, is asymptotically $c \cdot \frac{x}{(\ln x)^{1/2}}$, where c is a positive constant.

Recently, in [24] by using more refined counting techniques, the lower bound from (1) has been improved:

Mathematics Subject Classification (1991): 52 C 10

0209–9683/95/$6.00 ©1995 Akadémiai Kiadó, Budapest

Theorem 1.2. [24] *For all integers $n \geq 2$,*

$$f_2(n) \geq c \cdot \frac{n^{2/3}}{(\ln n)^{1/3}} ,$$

where $c > 0$ is a constant.

In this paper, we will further improve the lower bound on $f_2(n)$ by using *uncrowded hypergraphs*, cf. [2] and [4], as well as some results from number theory, namely we will show:

Theorem 1.3. *For integers $n \geq 1$,*

$$f_2(n) \geq c \cdot n^{2/3} ,$$

where $c > 0$ is a constant.

In order to prove this, we will show a so-called *anti-Ramsey* theorem, which will be given in Section 2. This anti-Ramsey result, together with some *number theoretic* results which we deduce in Section 3, also yields lower bounds for the analog of the problem of Erdős and Guy in higher dimensions. Let $f_d(n)$ denote the maximum size of a subset X of the d-dimensional grid $\{1,2,\ldots,n\}^d$ such that all mutual euclidean distances within X are distinct.

We remark that for $d = 1$ one has $f_1(n) = \theta(\sqrt{n})$ by using perfect difference sets, cf. [12]. For $d \geq 3$, Erdős and Guy showed the following:

Theorem 1.4. [12] *Let $d \geq 3$ be a positive integer and $\epsilon > 0$ a fixed real. Then, for positive integers n sufficiently large,*

$$n^{2/3-\epsilon} \leq f_d(n) \leq c \cdot \sqrt{d} \cdot n , \tag{2}$$

where $c > 0$ is a constant.

Indeed, in [12] it has been conjectured that

$$f_d(n) \leq c \cdot d^{2/3} \cdot n^{2/3} \cdot (\ln n)^{1/3}$$

for $d \geq 3$.

In Section 4 we improve the lower bound (2) on $f_d(n)$:

Theorem 1.5. *Let $d \geq 3$ be a positive integer. Then for every integer $n \geq 1$,*

$$f_d(n) \geq c_d \cdot n^{2/3} \cdot (\ln n)^{1/3} ,$$

where $c_d > 0$ is a constant only dependent on d.

In Section 5 we will consider the corresponding *selection problems* for points in the plane in arbitrary position. We will show that every n-point set in the euclidean plane $\mathbb{R}^2$ contains a subset X with mutual distinct distances such that $|X| \geq c \cdot n^{1/4}$ for some constant $c > 0$. This improves a lower bound of $c \cdot n^{1/5}$ which follows from [6, proof of Theorem 4.2], as communicated to us by János Pach, and the lower bound of $c \cdot n^{2/9}$ given in [24]. Moreover, we will show that under the assumption that the n points are in general position (no three on a line) the lower bound on $|X|$ can be improved to $c \cdot n^{1/3}$. To do so we will prove a conjecture of Erdős and Fishburn [14], [11]. Namely, we will show the following: if n points in general position in $\mathbb{R}^2$ are given with distinct distances $d_1, d_2, \ldots, d_t$, where d_i occurs with multiplicity m_i, $i=1,2,\ldots,t$, then $\sum_{i=1}^{t} m_i^2 \leq c \cdot n^3$ for some positive constant c. The regular n-gon shows that this bound is tight up to a constant factor. Moreover, we will show that for the corresponding problem for n arbitrary points in $\mathbb{R}^2$ one has $\sum_{i=1}^{t} m_i^2 \leq c \cdot n^{3.25}$, where c is a positive constant.

In Section 6 we will give the new lower bound $c \cdot n^{2/3}$ (c a positive constant) on the size of a B_2-subset of the set $\{1^2, 2^2, \ldots, n^2\}$. This improves earlier results of Alon and Erdős [3].

Finally, in Section 7 we consider some *algorithmic aspects* of these selection problems under a more general setting. In particular, using derandomization we will give an efficient sequential algorithm that finds in every edge coloring of the complete graph K_n a totally multicolored complete subgraph of size at least as large as guaranteed by the probabilistic method. This algorithm has running time $O\left(n^2 \ln n + \sum_i m_i^2\right)$, where m_i is the number of edges in color i.

2. An Anti-Ramsey Result

In this section we will prove a so-called anti-Ramsey theorem, which we will use for the proofs of Theorems 1.3 and 1.5. Before stating it we will introduce some notation. For further references to anti-Ramsey results we refer to [4].

A graph G with vertex set V and edge set E is denoted by $G=(V,E)$. By K_n we denote the complete graph on n vertices. A mapping $f: E(K_n) \longrightarrow T$ is called an edge coloring of K_n with colors $t \in T$. For $t \in T$, $f^{-1}(t)$ is the set of all edges colored by color t. By $\overline{d}_t = \frac{2 \cdot |f^{-1}(t)|}{n}$ we denote the average degree of color $t \in T$. Let d_t be the maximum degree of color t, i.e. the maximum number of edges in color t, incident at some vertex, and let $\Delta = \max\{d_t \mid t \in T\}$. Finally, a complete subgraph K_k of K_n is called *totally multicolored* if the restriction $f|E(K_k)$ to the edge set of K_k is a one-to-one coloring.

Theorem 2.1. *For every $\gamma > 0$ there exists a constant $C = C(\gamma) > 0$, such that for all integers $n \geq 2$ the following holds.*

Let $f: E(K_n) \longrightarrow T$ be a coloring and suppose τ satisfies the following conditions

(i) $\tau \geq \sum_{t \in T} \overline{d}_t^2$, *and*

(ii) $\tau \geq n^{1/2+\gamma} \cdot \Delta^{3/2}$.

Then there exists a totally multicolored subgraph K_k of K_n with

$$k \geq C \cdot \left(\frac{n^2}{\tau}\right)^{1/3} \cdot (\ln n)^{1/3}. \tag{3}$$

For the proof of Theorem 2.1 we will use the concept of uncrowded hypergraphs. Let $\mathcal{G} = (V, \mathcal{E})$ be a hypergraph with vertex set V and edge set $\mathcal{E}$. For a vertex $v \in V$

let $\deg_{\mathcal{G}}(v)$ denote the degree of v in $\mathcal{G}$, i.e. the number of edges $E \in \mathcal{E}$ containing v. By $\mathrm{Deg}(\mathcal{G}) = \max\{\deg_{\mathcal{G}}(v) \mid v \in V\}$ we denote the maximum degree of $\mathcal{G}$. A hypergraph is k-uniform if each edge $E \in \mathcal{E}$ has cardinality k. A 2-cycle in $\mathcal{G}$ is given by a set of two distinct edges from $\mathcal{E}$, which intersect in at least two elements. A hypergraph is called *uncrowded* if it contains no 2-cycle. Finally, the *independence number* $\alpha(\mathcal{G})$ is the maximum cardinality of a subset of V which contains no edges from $\mathcal{E}$.

We will use the following result from [9], which is an extension of a theorem from [2]:

Theorem 2.2. [9] *Let $\mathcal{G} = (V, \mathcal{E})$ be a k-uniform hypergraph, $k \geq 3$, with $|V| = n$ and maximum degree $\mathrm{Deg}(\mathcal{G}) \leq t^{k-1}$.*
If
(i) *$\mathcal{G}$ contains no 2-cycles, and*
(ii) *t is sufficiently large, i.e. $t \geq t_0(k)$*
then

$$\alpha(\mathcal{G}) \geq c_k \cdot \frac{n}{t} \cdot (\ln t)^{\frac{1}{k-1}} ,$$

where $c_k > 0$ is a constant depending only on k.

We will now prove Theorem 2.1.

Proof. It is sufficient to prove the theorem for sufficiently large n, say $n \geq n_0$. To see this assume the theorem holds for $n \geq n_0$ for some n_0 and some constant $C > 0$. For values of n less than n_0 the lower bound of the theorem is less than $C n_0^{1/2} (\ln n_0)^{1/3}$. By adapting the constant C the inequality (3) holds for all n. Thus we can assume that n is sufficiently large throughout the proof.

Let $V = \{1, 2, \ldots, n\}$ be the vertex set of a complete graph K_n and let $f : E(K_n) \longrightarrow T$ be an edge coloring. Let τ satisfy requirements (i) and (ii) in Theorem 2.1. We can also assume that τ satisfies

$$\tau < n^2 \ln n , \tag{4}$$

since otherwise the assertion (3) is trivial and we are done.

We will construct i-uniform hypergraphs $\mathcal{G}_i = (V, \mathcal{E}_i)$, $i = 3, 4$, with the same vertex set as follows:

$$\{v_1, v_2, v_3\} \in \mathcal{E}_3 \Leftrightarrow f(\{v_1, v_2\}) = f(\{v_1, v_3\})$$
$$\{v_1, v_2, v_3, v_4\} \in \mathcal{E}_4 \Leftrightarrow f(\{v_1, v_2\}) = f(\{v_3, v_4\}).$$

Observe that a subset $X \subseteq V$ yields a totally multicolored complete subgraph if and only if X is an independent set in both $\mathcal{G}_3$ and $\mathcal{G}_4$. Our aim will be to give a lower bound for the maximum size of such an independent set. We cannot apply Theorem 2.2 directly, as the $\mathcal{G}_i$, $i = 3, 4$, are in general not uncrowded. To come to such an uncrowded situation we will pick a random subset of the vertex set V, and show that an induced subhypergraph can be made uncrowded.

First we will give upper bounds for the cardinalities of $\mathcal{E}_3$ and $\mathcal{E}_4$. For $|\mathcal{E}_3|$ note that every pair $\{v, w\}$ of vertices can be extended in at most $2\Delta - 2$ ways to an edge $E \in \mathcal{E}_3$. Thus,

$$|\mathcal{E}_3| < \binom{n}{2} \cdot 2 \cdot \Delta < n^2 \cdot \Delta . \tag{5}$$

Concerning the size of $\mathcal{E}_4$ we obviously have

$$|\mathcal{E}_4| \le \sum_{t\in T} \binom{|f^{-1}(t)|}{2} .$$

As $2\cdot|f^{-1}(t)|=\overline{d}_t\cdot n$, it follows with (i) that

$$|\mathcal{E}_4| \le \sum_{t\in T} \binom{\frac{\overline{d}_t\cdot n}{2}}{2} < \frac{n^2}{8}\cdot\sum_{t\in T}\overline{d}_t^2 \le \frac{1}{8}\cdot n^2\cdot\tau . \tag{6}$$

Next we will count the number of 2-cycles in $\mathcal{G}_4$. Let $c_2(\mathcal{G})$ denote the number of 2-cycles in a hypergraph $\mathcal{G}=(V,\mathcal{E})$. We will count the 2-cycles more carefully: for $j=2,3$ let $c_{2,j}(\mathcal{G})$ be the number of $(2,j)$-cycles, i.e. the number of pairs $\{E,E'\}\in[\mathcal{E}]^2$ with $|E\cap E'|=j$. Clearly, $c_2(\mathcal{G})=c_{2,2}(\mathcal{G})+c_{2,3}(\mathcal{G})$.

Concerning $c_{2,2}(\mathcal{G}_4)$, choose an edge $E\in\mathcal{E}_4$ and then pick a pair $\{v,w\}\subset E$ of vertices. The number of edges $E'\in\mathcal{E}_4$ with $E\cap E'=\{v,w\}$ is less than the number of pairs $\{x,y\}$ with $f(\{v,w\})=f(\{x,y\})$ or $f(\{v,x\})=f(\{w,y\})$. There are at most $2n\Delta$ such pairs, hence with (6) we have

$$c_{2,2}(\mathcal{G}_4) \le |\mathcal{E}_4|\cdot\binom{4}{2}\cdot 2\cdot n\cdot\Delta \le \frac{3}{2}\cdot n^3\cdot\tau\cdot\Delta . \tag{7}$$

To count the number of $(2,3)$-cycles, we fix an edge $E\in\mathcal{E}_4$ and a three-element subset $S\subset E$. Then S can be extended in at most $\binom{3}{2}\cdot\Delta$ ways to an edge $E'\in\mathcal{E}_4$, hence

$$c_{2,3}(\mathcal{G}_4) \le |\mathcal{E}_4|\cdot\binom{4}{3}\cdot\binom{3}{2}\cdot\Delta \le \frac{3}{2}\cdot n^2\cdot\tau\cdot\Delta . \tag{8}$$

Now we choose a random subset of V by picking each vertex with probability

$$p = n^{-1/3+\epsilon}\cdot\tau^{-1/3} ,$$

where $0<\epsilon<\gamma/12$, independently of the other vertices. For the random subset $V'\subseteq V$ consider the induced random subhypergraphs $\mathcal{G}'_i=(V',\mathcal{E}'_i)$ for $i=3,4$, where $\mathcal{E}'_i=\mathcal{E}_i\cap[V']^i$. Moreover, let $c_{2,i}(V')$, $i=2,3$ be random variables counting the number of $(2,i)$-cycles in $\mathcal{G}'_4$.

Assumption (4) makes sure that $pn\to\infty$ as $n\to\infty$, so we have

$$\text{Prob }(|V'|\approx p\cdot n) = 1-o(1) \tag{9}$$

by Chernoff's inequality.

For a random variable X let $E(X)$ denote its expectation. From (5) and (ii) we obtain for $\epsilon<\frac{\gamma}{3}$ that

$$E(|\mathcal{E}'_3|) = p^3\cdot|\mathcal{E}_3| < p^3\cdot n^2\cdot\Delta = pn\cdot\frac{n^{1/3+2\epsilon}\cdot\Delta}{\tau^{2/3}} \le pn\cdot\frac{1}{n^{2/3\cdot\gamma-2\epsilon}} = o(pn) . \tag{10}$$

By (7) and (8) we infer for $\epsilon < \frac{\gamma}{12}$

$$
\begin{aligned}
E(c_2(\mathcal{G}_4')) &= p^6 \cdot c_{2,2}(\mathcal{G}_4) + p^5 \cdot c_{2,3}(\mathcal{G}_4) \leq \frac{3}{2} \cdot p^6 \cdot n^3 \cdot \tau \cdot \Delta + \frac{3}{2} \cdot p^5 \cdot n^2 \cdot \tau \cdot \Delta \\
&= \frac{3pn}{2} \cdot \left(\frac{n^{1/3+5\epsilon} \cdot \Delta}{\tau^{2/3}} + \frac{\Delta}{n^{1/3-4\epsilon} \cdot \tau^{1/3}} \right) \qquad (11) \\
&\leq \frac{3pn}{2} \cdot \left(n^{5\epsilon - 2/3 \cdot \gamma} + \frac{\Delta^{1/2}}{n^{1/2+\gamma/3-4\epsilon}} \right) = o(pn) \, .
\end{aligned}
$$

Moreover, we have

$$
E(|\mathcal{E}_4'|) = p^4 \cdot |\mathcal{E}_4| \, . \qquad (12)
$$

Using Markov's inequality, we infer with (9), (10), (11) and (12) that there exists a subset $V' \subseteq V$ with $|V'| \approx p \cdot n$, such that the induced hypergraphs $\mathcal{G}_i' = (V', \mathcal{E}_i')$, $i = 3,4$, satisfy the following: $|\mathcal{E}_3'| = o(pn)$ and $c_2(\mathcal{G}_4') = o(pn)$ and also $|\mathcal{E}_4'| \leq 2 \cdot p^4 \cdot |\mathcal{E}_4|$. Now, delete one vertex from each triple $E \in \mathcal{E}_3'$ and from each 2-cycle in $\mathcal{G}_4'$. Moreover, delete all vertices of degree bigger than

$$
\frac{24 \cdot p^4 \cdot |\mathcal{E}_4|}{pn} = \frac{24 \cdot p^3 \cdot |\mathcal{E}_4|}{n} \, .
$$

For n sufficiently large, we obtain a subset $V^* \subseteq V'$ of at least $\frac{pn}{2}$ vertices containing no edge from $\mathcal{E}_3$ and such that the induced subhypergraph $\mathcal{G}_4^* = (V^*, [V^*]^4 \cap \mathcal{E}_4)$ has no 2-cycle and has maximum degree

$$
\operatorname{Deg}(\mathcal{G}_4^*) \leq \frac{24 \cdot p^3 \cdot |\mathcal{E}_4|}{n} \leq 3p^3 n\tau = 3n^{3\epsilon} = t^3 \, ,
$$

by (6), with t defined by the equation. We apply Theorem 2.2 to the hypergraph $\mathcal{G}_4^*$ and infer

$$
\alpha(\mathcal{G}_4^*) \geq c_4 \cdot \frac{pn/2}{3^{1/3} \cdot pn^{1/3} \tau^{1/3}} \cdot \left(\ln \left(3^{1/3} \cdot n^{\epsilon} \right) \right)^{1/3} \geq C \cdot \left(\frac{n^2}{\tau} \right)^{1/3} \cdot (\ln n)^{1/3} \, . \quad \blacksquare
$$

Corollary 2.3. *Let $f : E(K_n) \longrightarrow T$ be a coloring of the edges of the complete graph on n vertices, where $\Delta = O(n^{1-\beta})$ for a fixed $\beta > 0$.*

Then, for $n \geq 2$ there exists a totally multicolored subgraph K_k with

$$
k \geq c \cdot \left(\frac{n}{\Delta} \right)^{1/3} \cdot (\ln n)^{1/3} \, ,
$$

where $c = c(\beta) > 0$ is a constant.

Proof. By Theorem 2.1 with $\tau = n \cdot \Delta$ and taking $\gamma < \frac{\beta}{2}$. $\blacksquare$

3. Two Results from Number Theory

In this section we are concerned with two results from analytic number theory that we will need for the proofs of Theorem 1.3 and 1.5. For convenience we will use Vinogradov's notation $f(n) \ll g(n)$ for $f(n)=O(g(n))$.

Definition 3.1. Let $r_d(m)$ be the number of representations of m in the form

$$m = x_1^2 + x_2^2 + \ldots + x_d^2 \ ,$$

where $x_1, x_2, \ldots, x_d$ are integers.

The following result for the 2-dimensional case is due to Ramanujan (see also [26]).

Theorem 3.2. [22]

$$\sum_{m=1}^{n} (r_2(m))^2 = \Theta(n \ln n).$$

In fact Ramanujan determines also the leading constant.

Here we will give an alternative proof for the upper bound. Our approach uses simple geometric considerations and might be of interest by itself. For doing so and for later purposes we will use the following definition and lemma.

Definition 3.3. Let P be a finite set of points in the plane. Consider the bipartite graph $\mathscr{B} = ([P]^2 \cup P, I)$ with

$$(\{p, q\}, z) \in I \iff z \text{ lies on the } \textit{perpendicular bisector} \text{ of } p \text{ and } q.$$

Then define $\Delta(P) := |I|$.

Roughly speaking, $\Delta(P)$ is the number of incidences between perpendicular bisectors determined by P and points of P (each bisector can be generated by several pairs of points!). Note that $\Delta(P)$ is nearly the same as the number of isosceles triangles determined by P apart from the fact that equilateral triangles are counted 3 times.

Lemma 3.4. *Let P be a set of n points in the plane $\mathbb{R}^2$. Let the points of P determine distinct distances $d_1, d_2, \ldots, d_t$, where d_i occurs with multiplicity m_i for $i=1,2,\ldots,t$. Then,*

$$\sum_{i=1}^{t} m_i^2 \leq \frac{n}{2} \cdot \left(\Delta(P) + \binom{n}{2} \right) .$$

The idea of the proof is similar to an argument of Szemerédi (see [10]).

Proof. Let P be the set of n given points in the plane. Notice that a point z lies on the bisector of p and q if and only if p and q have the same distance from z.

For $z \in P$ and $i=1,2,\dots,t$ let $m_i(z)$ denote the number of points in P, which have distance d_i from z. Using Jensen's inequality we infer that

$$\Delta(P) = \sum_{z\in P}\sum_{i=1}^{t}\binom{m_i(z)}{2} = \sum_{i=1}^{t}\sum_{z\in P}\binom{m_i(z)}{2}$$
$$\geq \sum_{i=1}^{t} n\cdot\binom{\frac{2m_i}{n}}{2} = \frac{2}{n}\cdot\sum_{i=1}^{t} m_i^2 - \sum_{i=1}^{t} m_i = \frac{2}{n}\cdot\sum_{i=1}^{t} m_i^2 - \binom{n}{2}.$$

Thus,

$$\sum_{i=1}^{t} m_i^2 \leq \frac{n}{2}\cdot\left(\Delta(P)+\binom{n}{2}\right). \qquad \blacksquare$$

Lemma 3.5. *Let G_n be the set of points of the $n\times n$-grid $\{0,1,\dots,n-1\}\times\{0,1,\dots,n-1\}$. Then,*

$$\Delta(G_n) \leq c\cdot n^4\cdot\ln n,$$

for some positive constant c.

Note that $\Delta(G_n)$ is equal to the number of isosceles triangles since G_n contains no equilateral triangle.

Proof. For distinct points p_1,p_2 in the $n\times n$ grid G_n, where $p_1=(x_1,y_1)$ and $p_2=(x_2,y_2)$, let l be the line through p_1 and p_2 and define

$$s(l) = \max\left\{\frac{1}{g}\cdot|x_2-x_1|, \frac{1}{g}\cdot|y_2-y_1|\right\},$$

where $g=\gcd(x_2-x_1,y_2-y_1)$. Note that $s(l)$ only depends on l and is independent of the choice of p_1 and p_2. We observe that $|l\cap G_n|\leq\frac{n}{s(l)}$. Let l' be the perpendicular bisector of p_1 and p_2, then we also have $|l'\cap G_n|\leq\frac{n}{s(l)}$.

To bound $\Delta(G_n)$ we fix an arbitrary point $p_1\in P$ and an integer $1\leq s\leq n$. Choose a line l through p_1 with $s(l)=s$. The number of such lines is at most $4s$. Then we select a point $p_2\in l\cap G_n$ and a point $z\in l'\cap G_n$, where l' is the perpendicular bisector of p_1 and p_2. For both p_2 and z there are at most $\frac{n}{s}$ possibilities. Thus

$$\Delta(G_n) \leq n^2\cdot\sum_{s=1}^{n} 4s\cdot\frac{n}{s}\cdot\frac{n}{s} = 4n^4\cdot\sum_{s=1}^{n}\frac{1}{s} \leq 4n^4\cdot\left(1+\int_1^n \frac{1}{x}dx\right) \leq c\cdot n^4\cdot\ln n. \qquad \blacksquare$$

Corollary 3.6. *Let G_n be the set of points of the $n\times n$ grid. For $i=1,2,\dots,2(n-1)^2$ let m_i denote the occurrence of distance $\sqrt{i}$ between different points of G_n. Then,*

$$\sum_{i=1}^{2(n-1)^2} m_i^2 \leq c_2\cdot n^6\cdot\ln n$$

for some positive constant c_2.

Proof. By Lemmas 3.4 and 3.5 we obtain

$$\sum_{i=1}^{2(n-1)^2} m_i^2 \leq \frac{n^2}{2} \cdot \left(c \cdot n^4 \cdot \ln n + \binom{n^2}{2}\right) \leq c_2 \cdot n^6 \cdot \ln n\,. \qquad \blacksquare$$

Corollary 3.7. *There exists a positive constant* c_3 *such that for every positive integer* n

$$\sum_{i=1}^{n} (r_2(i))^2 \leq c_3 \cdot n \cdot \ln n\,.$$

Proof. Consider a circle of radius at most $\frac{n-1}{2}$ around an arbitrary point of the $n \times n$-grid. Then at least a quarter of this circle lies inside the $n \times n$-grid. Thus

$$\sum_{i=1}^{(\frac{n-1}{2})^2} \left(n^2 \cdot \frac{r_2(i)}{4}\right)^2 \leq \sum_{i=1}^{2(n-1)^2} m_i^2\,,$$

where m_i denotes the occurrence of distance $\sqrt{i}$ between different points in the $n \times n$ grid. By Corollary 3.6 we deduce $\sum_{i=1}^{n} (r_2(i))^2 \leq c_3 \cdot n \cdot \ln n$ for some positive constant c_3. $\blacksquare$

Remark. By using the same ideas it is also possible to prove the lower bound in Theorem 3.2.

For higher dimensions we will show

Theorem 3.8.

$$\sum_{m=1}^{n} (r_d(m))^2 = O(n^{d-1}), \quad \textit{for } d \geq 3.$$

Remark. It can be shown that the bound given in Theorem 3.8 is asymptotically sharp by using for example Cauchy's inequality together with the fact that $\sum_{m=1}^{n} r_d(m) \approx \frac{\pi^{d/2}}{\Gamma(d/2+1)} n^{d/2}$.

We are going to prove Theorem 3.8 by using the *Hardy-Littlewood circle-method* (see [8], [25]). Throughout the remaining part of this section let $N := \lfloor \sqrt{n} \rfloor$ and suppose $d \geq 3$. Define $R_d(m, N)$ as the number of representations of m in the form

$$m = x_1^2 + x_2^2 + \ldots + x_d^2, \qquad 1 \leq x_i \leq N.$$

Lemma 3.9. *Let* $\epsilon > 0$ *and* $N = \lfloor \sqrt{n} \rfloor$, *then*

$$\sum_{m=1}^{n} (r_d(m))^2 \leq c_d \sum_{m=1}^{n} (R_d(m, N))^2 + O(n^{d-3/2+\epsilon})$$

with a constant c_d depending only on d.

Proof. It is easy to see that we can estimate

$$R_d(m,N) \le r_d(m) \le 2^d \cdot R_d(m,N) + d \cdot r_{d-1}(m)$$

for $m \le n$. As $r_2(m) = O(m^{\epsilon'})$ for every $\epsilon' > 0$, we have

$$r_s(m) \le (\sqrt{m})^{s-2} \cdot O(m^{\epsilon'}) = O(m^{s/2-1+\epsilon'}) \tag{13}$$

for $s \ge 2$. We can then estimate

$$(r_d(m))^2 \le 2^{2d} \cdot (R_d(m,N))^2 + O(m^{d-5/2+\epsilon})$$

and the assertion follows. ∎

Define $e(y) := e^{2\pi i y}$ and $f(\alpha) := \sum_{x=1}^{N} e(\alpha x^2)$ for real numbers α.

Lemma 3.10.

$$\sum_{m=1}^{dn} (R_d(m,N))^2 = \int_0^1 |f(\alpha)|^{2d}\, d\alpha.$$

Proof. First we observe that

$$(f(\alpha))^d = \sum_{x_1=1}^{N} \cdots \sum_{x_d=1}^{N} e\left(\alpha \cdot (x_1^2 + \ldots + x_d^2)\right) = \sum_{m=1}^{dn} R_d(m,N) \cdot e(\alpha m),$$

hence,

$$\left|(f(\alpha))^d\right|^2 = \sum_{m_1=1}^{dn} R_d(m_1,N) \cdot e(\alpha m_1) \sum_{m_2=1}^{dn} R_d(m_2,N) \cdot e(-\alpha m_2).$$

If we integrate this we obtain

$$\int_0^1 |f(\alpha)|^{2d}\, d\alpha = \sum_{1 \le m_1, m_2 \le dn} R_d(m_1,N) \cdot R_d(m_2,N) \cdot \int_0^1 e(\alpha \cdot (m_1 - m_2))\, d\alpha.$$

The integral on the right hand side is 0 for $m_1 \ne m_2$ and 1 for $m_1 = m_2$, thus

$$\int_0^1 |f(\alpha)|^{2d}\, d\alpha = \sum_{m=1}^{dn} (R_d(m,N))^2 \,.$$ ∎

In order to estimate $\int_0^1 |f(\alpha)|^{2d}\, d\alpha$ we break the integration interval into two parts called major-arcs and minor-arcs. The major-arcs will determine the order of magnitude while the contribution of the minor-arcs will be negligible.

In the following calculations let $0<\delta<1/5$ be a fixed real number. For integers a,q with $1\leq a\leq q\leq N^{\delta}$ and $(a,q)=1$ define the *major-arcs* as

$$\mathcal{M}_{a,q} := \{\alpha \in \mathbb{R} : |\alpha - a/q| < N^{-2+\delta}\}.$$

We observe that these major-arcs are disjoint since their lengths are much smaller than the distances between their centers. Let $\mathcal{M}$ denote the union of the $\mathcal{M}_{a,q}$. It is convenient to shift the integration interval $(0,1]$ to the right to $U:=(N^{-2+\delta},1+N^{-2+\delta}]$. As $f(\alpha)=f(\alpha+1)$ we have

$$\int_0^1 |f(\alpha)|^{2d}\, d\alpha = \int_U |f(\alpha)|^{2d}\, d\alpha.$$

Now $\mathcal{M}\subset U$ by definition. The set $\mathrm{M}:=U\setminus\mathcal{M}$ forms the *minor-arcs*.

To see that the minor-arcs can be neglected in order to prove Theorem 3.8 we use the following

Lemma 3.11. *Suppose $s\geq 5$, then*

$$\int_{\mathrm{M}} |f(\alpha)|^{s}\, d\alpha = O(n^{s/2-1-\delta'})\ ,$$

where δ' is a constant depending on δ. ∎

This lemma can be found for example in [8] and [25]. Using this, we immediately get for $d\geq 3$, that

$$\int_{\mathrm{M}} |f(\alpha)|^{2d}\, d\alpha = O(n^{d-1-\delta'})\ , \tag{14}$$

where δ' is a constant depending on δ.

Thus it remains to prove that

$$\int_{\mathcal{M}} |f(\alpha)|^{2d}\, d\alpha = O(n^{d-1})\ .$$

Define

$$S_{a,q} := \sum_{z=1}^{q} e(az^2/q) \qquad \text{and} \qquad I(\beta) := \int_0^N e(\beta\xi^2)\, d\xi.$$

For the treatment of the major-arcs we use the following lemmas.

Lemma 3.12. [8] *If $\alpha\in\mathcal{M}_{a,q}$, then*

$$f(\alpha) = V(\alpha) + O(N^{2\delta})$$

with $V(\alpha):=q^{-1}\cdot S_{a,q}\cdot I(\alpha-a/q)$. ∎

The notation $V(\alpha)$ is a bit sloppy because V also depends on a and q. Nevertheless there should be no confusion since α determines the major-arc $\mathcal{M}_{a,q}$ in which it lies.

Lemma 3.13. [8] *Let a, q be relative prime integers with $q>0$. Then for every $\epsilon>0$,*

$$|S_{a,q}| = O(q^{1/2+\epsilon}).$$ ∎

Now let $\alpha \in \mathcal{M}_{a,q}$. Lemma 3.12 and the obvious fact that

$$|V(\alpha)| = |q^{-1} \cdot S_{a,q} \cdot I(\alpha - a/q)| \leq N$$

yields

$$(f(\alpha))^{2d} = (V(\alpha))^{2d} + O(N^{2d-1+2\delta}).$$

Since the measure of $\mathcal{M}$ is bounded by $O(N^{-2+3\delta})$ and since $\delta < 1/5$ we can again neglect the O-term and consider only $\int_{\mathcal{M}} |V(\alpha)|^{2d}\, d\alpha$.

By Lemma 3.13 we see that

$$|V(\alpha)|^{2d} \leq |q^{-1} S_{a,q}|^{2d} \cdot |I(\alpha - a/q)|^{2d} \ll q^{-d+\epsilon'} \cdot |I(\alpha - a/q)|^{2d}$$

for $\alpha \in \mathcal{M}_{a,q}$. Substituting $\alpha = a/q + \beta$ implies

$$\int_{\mathcal{M}_{a,q}} |V(\alpha)|^{2d}\, d\alpha \ll q^{-d+\epsilon'} \cdot \int_{|\beta|<N^{-2+\delta}} |I(\beta)|^{2d}\, d\beta.$$

Summing up over all major-arcs gives

$$\sum_{q=1}^{N^\delta} \sum_{\substack{a=1 \\ (a,q)=1}}^{q} \int_{\mathcal{M}_{a,q}} |V(\alpha)|^{2d}\, d\alpha \ll \sum_{q=1}^{N^\delta} \sum_{\substack{a=1 \\ (a,q)=1}}^{q} q^{-d+\epsilon'} \cdot \int_{|\beta|<N^{-2+\delta}} |I(\beta)|^{2d}\, d\beta$$

$$\leq \left(\sum_{q=1}^{\infty} q^{1-d+\epsilon'} \right) \cdot \int_{|\beta|<N^{-2+\delta}} |I(\beta)|^{2d}\, d\beta = const \cdot \int_{|\beta|<N^{-2+\delta}} |I(\beta)|^{2d}\, d\beta.$$

The infinite sum converges since $d \geq 3$. By substituting $\xi = N \cdot t$ and $\beta = N^{-2} \cdot \gamma$ in $I(\beta) = \int_0^N e(\beta\xi^2)\, d\xi$ we obtain

$$\int_{|\beta|<N^{-2+\delta}} |I(\beta)|^{2d}\, d\beta = N^{2d-2} \cdot \int_{|\gamma|<N^\delta} \left| \int_0^1 e(\gamma t^2)\, dt \right|^{2d} d\gamma$$

$$\leq n^{d-1} \cdot \int_{|\gamma|<N^\delta} \left| \int_0^1 e(\gamma t^2)\, dt \right|^{2d} d\gamma\,, \tag{15}$$

since $N = \lfloor \sqrt{n} \rfloor$.

We want to show that the outer integral is bounded from above independent of n. By substituting $t=\gamma^{1/2}z$ and by the fact that $\int_0^\infty \cos(x^2)\,dx$ and $\int_0^\infty \sin(x^2)\,dx$ are bounded one can show that

$$\left|\int_0^1 e(\gamma t^2)\,dt\right| = \left|\gamma^{-1/2}\cdot \int_0^{\gamma^{1/2}} e(z^2)\,dz\right| \ll \gamma^{-1/2}\,, \tag{16}$$

for $\gamma \geq 0$. On the other hand, obviously

$$\left|\int_0^1 e(\gamma t^2)\,dt\right| \leq 1\,. \tag{17}$$

This enables us to extend the integration in (15) to infinity:

$$\int_{-\infty}^{+\infty}\left|\int_0^1 e(\gamma t^2)\,dt\right|^{2d} d\gamma =$$

$$\int_{-1}^{+1}\left|\int_0^1 e(\gamma t^2)\,dt\right|^{2d} d\gamma + 2\cdot \int_1^{+\infty}\left|\int_0^1 e(\gamma t^2)\,dt\right|^{2d} d\gamma \ll 2\cdot 1 + 2\cdot \int_1^{+\infty} \gamma^{-d}\,d\gamma = O(1)$$

since $d \geq 3$. There we made use of (16) and (17). Thus we have proved

Lemma 3.14.

$$\int_{\mathcal{M}} |f(\alpha)|^{2d}\,d\alpha = O(n^{d-1}).$$

Because of the fact that $\int_0^1 = \int_{\mathcal{M}} + \int_{\mathrm{M}}$ and by (14) we have

Corollary 3.15.

$$\int_0^1 |f(\alpha)|^{2d}\,d\alpha = O(n^{d-1}).$$

Now Lemma 3.9 and 3.10 and Corollary 3.15 imply Theorem 3.8. ∎

4. Grid Points

In this section we will give the proofs for Theorems 1.3 and 1.5. First we will show Theorem 1.3.

Proof. Given the $n \times n$-grid G_n, we form a complete graph K_{n^2} with vertex set $\{1,2,\ldots,n\} \times \{1,2,\ldots,n\}$. We color the edges $\{x,y\}$ by the square of the euclidean distance of their endpoints x and y. For fixed positive integer t and every grid point v, notice that the number of grid points w with euclidean distance $\sqrt{t}$ from v is bounded from above by the number of representations of t as a sum of two squares. Hence, the average degrees $\overline{d}_t$ of color t satisfy $\overline{d}_t \leq r_2(t)$, and for the maximum degree Δ we have that

$$\Delta \leq n^{c_1/\ln\ln n}\ , \tag{18}$$

where c_1 is a positive constant, by a result of Wigert, cf. [19]. By Theorem 3.2 we have

$$\sum_{t=1}^{2(n-1)^2} \overline{d}_t^2 \leq \sum_{t=1}^{2(n-1)^2} (r_2(t))^2 \leq c_2 \cdot n^2 \cdot \ln n\ ,$$

where c_2 is a positive constant. Setting $\tau = C \cdot n^2 \cdot \ln n$ for a positive constant C, which is large enough, we see with (18) that (i) and (ii) in Theorem 2.1 are satisfied (notice that the number of vertices is n^2). Hence, there exists a totally multicolored complete subgraph on k vertices with

$$k \geq c' \cdot \left(\frac{n^4}{n^2 \cdot \ln n}\right)^{1/3} \cdot (\ln n^2)^{1/3} \geq c \cdot n^{2/3}\ .$$

The vertices of this subgraph determine k points in the $n \times n$-grid with mutual distinct distances. We remark that we could have also used Corollary 3.6 to obtain the same conclusion. ∎

Next we will prove Theorem 1.5:

Proof. Fix a positive integer $d \geq 3$. We proceed as in the proof given above by coloring the edges of the complete graph on n^d vertices by the square of the euclidean distances of the corresponding endpoints. Clearly, $\overline{d}_t \leq r_d(t)$ and

$$\Delta \leq n^{d/2-1+\epsilon'}$$

for any fixed $\epsilon' > 0$ by (13).

Now,

$$\sum_{t=1}^{d(n-1)^2} \overline{d}_t^2 \leq \sum_{t=1}^{d(n-1)^2} (r_d(t))^2 = O\left(d^{d-1} \cdot n^{2d-2}\right) \tag{19}$$

by Theorem 3.8. With $\tau = C \cdot n^{2d-2}$, where $C > 0$ is a large enough constant, requirements (i) and (ii) of Theorem 2.1 are satisfied. Hence there exists a totally multicolored complete subgraph on k vertices with

$$k \geq c(d) \cdot \left(\frac{n^{2d}}{n^{2d-2}}\right)^{1/3} \cdot \left(\ln n^d\right)^{1/3} \geq c_d \cdot n^{2/3} \cdot (\ln n)^{1/3}\ ,$$

which yields the desired result. ∎

5. Points in Arbitrary Position

In this section we will study the maximum cardinality of a subset of n arbitrary points in the euclidean plane $\mathbb{R}^2$, such that the mutual distances among the points of X are distinct. Moreover, we will consider the same question for n points in general position (no three on a line) in $\mathbb{R}^2$.

Theorem 5.1. *Let n arbitrary points in the plane $\mathbb{R}^2$ be given. Let the n points determine distinct distances $d_1, d_2, \ldots, d_t$, where distance d_i occurs with multiplicity m_i for $i = 1, 2, \ldots, t$.*
Then

$$\sum_{i=1}^{t} m_i^2 \le c \cdot n^{13/4} ,$$

where $c > 0$ is a constant.

Remarks. (1) Spencer, Szemerédi and Trotter proved in [23] that under the assumptions of Theorem 5.1 one has $m_i \le c' \cdot n^{4/3}$ for $i = 1, 2, \ldots, t$, where c' is a positive constant. Their result applied in the straightforward way yields $\sum_{i=1}^{t} m_i^2 \le \max\{m_1, m_2, \ldots, m_t\} \cdot \sum_{i=1}^{t} m_i \le c' \cdot n^{4/3} \cdot \binom{n}{2} \le cn^{10/3}$ for a positive constant c. Another way to get this upper bound is by using the result of Pach and Sharir [21] that the number of isosceles triangles is bounded by $O(n^{7/3})$. Then by Lemma 3.4 one obtains that $\sum_{i=1}^{t} m_i^2 \le c_2 n^{10/3}$, where c_2 is a positive constant.

(2) For the points of the $\sqrt{n} \times \sqrt{n}$-grid we have $\sum m_i^2 = \Theta(n^3 \ln n)$. One might conjecture that this upper bound holds for any set of n points in the euclidean plane.

For the proof of Theorem 5.1 we will use

Lemma 5.2. *Let $0 = S_0 \le S_1 \le \ldots \le S_t$ and $m_1 \ge m_2 \ge \ldots \ge m_t \ge 0$ be sequences of real numbers such that*

$$\sum_{i=1}^{j} m_i \le S_j \qquad \text{for } j = 1, 2, \ldots, t. \tag{20}$$

Then

$$\sum_{i=1}^{t} m_i^2 \le \sum_{i=1}^{t} (S_i - S_{i-1})^2 . \tag{21}$$

Proof. We will apply induction on t. For $t = 1$ the assertion is trivial, hence assume $t > 1$ and that the conclusion of the lemma holds for all values $1, 2, \ldots, t-1$. For given $S_0, S_1, \ldots, S_t$ it suffices to show (21) for any sequence $m_1 \ge m_2 \ge \ldots \ge m_t$ satisfying (20) and maximizing the expression $\sum_{i=1}^{t} m_i^2$. Let $m_1, m_2, \ldots, m_t$ be such a sequence. If $m_t = 0$, then (21) is clearly satisfied by the induction hypothesis, hence we can assume that $m_t \ne 0$.

Suppose first that $\sum_{i=1}^{j} m_i < S_j$ for all $j = 1,2,\ldots,t-1$. Then define a new sequence $m_1^*, m_2^*, \ldots, m_t^*$ as follows

$$\begin{aligned} m_1^* &= m_1 + \epsilon \\ m_i^* &= m_i \qquad\qquad \text{for } i = 2,3,\ldots,t-1 \\ m_t^* &= m_t - \epsilon \end{aligned}$$

where $\epsilon = \min\{m_t, S_1 - m_1, S_2 - (m_1 + m_2), \ldots, S_{t-1} - \sum_{i=1}^{t-1} m_i\}$. By assumption $\epsilon > 0$, and hence $\sum_{i=1}^{t} m_i^{*2} > \sum_{i=1}^{t} m_i^2$, which contradicts the maximality of the sequence $m_1, m_2, \ldots, m_t$.

Thus there is a k, $1 \le k \le t-1$, with

$$\sum_{i=1}^{k} m_i = S_k \ . \tag{22}$$

By the induction assumption it follows that

$$\sum_{i=1}^{k} m_i^2 \le \sum_{i=1}^{k} (S_i - S_{i-1})^2 \ . \tag{23}$$

By (20) and (22) we have $\sum_{i=k+1}^{j} m_i \le S_j - S_k$ for $j = k+1, k+2, \ldots, t$. Using again the induction assumption we infer that

$$\sum_{i=k+1}^{t} m_i^2 \le \sum_{i=k+1}^{t} ((S_i - S_k) - (S_{i-1} - S_k))^2 = \sum_{i=k+1}^{t} (S_i - S_{i-1})^2 \ , \tag{24}$$

thus combining (23) and (24) we obtain $\sum_{i=1}^{t} m_i^2 \le \sum_{i=1}^{t} (S_i - S_{i-1})^2$, which finishes the induction step. ∎

Let P be a set of points in the plane $\mathbb{R}^2$ and let C be a set of circles in $\mathbb{R}^2$. Define a bipartite graph G with vertex set $P \cup C$ and edge set E, where $(p,c) \in E$, $p \in P$ and $c \in C$, if and only if p lies on the circle c. Let $I(P,C)$ denote the number of *incidences* between points and circle, that is, the number of edges in this bipartite graph G. In our arguments we will use the following result of Clarkson, Edelsbrunner, Guibas, Sharir and Welzl:

Theorem 5.3. [7] *Let P be a set of points in $\mathbb{R}^2$ and let C be a set of circles in $\mathbb{R}^2$. Then*

$$I(P,C) = O\left(|P|^{3/5} \cdot |C|^{4/5} + |P| + |C|\right) \ . \tag{25}$$

Now we are ready to prove Theorem 5.1:

Proof. Let $P \subset \mathbb{R}^2$ be a set of n points in the plane. Assume that $m_1 \ge m_2 \ge \ldots \ge m_t$. Around each point $p \in P$ draw circles with radius $d_1, d_2, \ldots, d_t$. For $j = 1,2,\ldots,t$

let C_j be the set of all such circles with radius $d_1, d_2, \ldots, d_j$. Then we have $|C_j| = jn$ and

$$I(P, C_j) = 2 \cdot \sum_{i=1}^{j} m_i \ . \tag{26}$$

Thus, by (25) we have

$$I(P, C_j) \leq c_1 \cdot n^{3/5} \cdot (jn)^{4/5} \ , \tag{27}$$

where $c_1 > 0$ is a constant.

Combining (26) and (27), we infer that

$$\sum_{i=1}^{j} m_i \leq \frac{c_1}{2} \cdot n^{7/5} \cdot j^{4/5}$$

for $j = 1, 2, \ldots, t$.

On the other hand, we have

$$\sum_{i=1}^{l} m_i \leq \binom{n}{2} < \frac{n^2}{2} \tag{28}$$

for $l = 1, 2, \ldots, t$.

Put $c_2 = \max\{1/2, c_1/2\}$ and

$$S_j = \begin{cases} c_2 \cdot n^{7/5} \cdot j^{4/5} & \text{if } 0 \leq j \leq n^{3/4} \\ c_2 \cdot n^2 & \text{if } n^{3/4} < j \leq t. \end{cases}$$

Clearly, the sequences $S_0, S_1, \ldots, S_t$ and $m_1, m_2, \ldots, m_t$ satisfy the assumptions of Lemma 5.2. Hence,

$$\sum_{i=1}^{t} m_i^2 \leq \sum_{i=1}^{t} (S_i - S_{i-1})^2 \leq \sum_{i=1}^{n^{3/4}} (S_i - S_{i-1})^2$$

$$\leq c_2^2 \cdot n^{14/5} \cdot \sum_{i=1}^{n^{3/4}} \left(i^{4/5} - (i-1)^{4/5} \right)^2 . \tag{29}$$

Since $i^{4/5} - (i-1)^{4/5} \leq i^{-1/5}$ for $i \geq 1$, (29) becomes

$$\sum_{i=1}^{t} m_i^2 \leq c_2^2 \cdot n^{14/5} \cdot \sum_{i=1}^{n^{3/4}} i^{-2/5} \ . \tag{30}$$

The function $g(x) = x^{-2/5}$, $x > 0$, is decreasing, thus

$$\sum_{i=1}^{n^{3/4}} i^{-2/5} \leq 1^{-2/5} + \int_1^{n^{3/4}} x^{-2/5} dx = 1 + \frac{5}{3}(n^{9/20} - 1) < \frac{5}{3} n^{9/20} \ .$$

We infer with (30) that

$$\sum_{i=1}^{t} m_i^2 \le c \cdot n^{13/4}$$

for some constant $c>0$. ∎

Surprisingly, the situation changes radically if our point set is in *general position*, i.e. no three points lie on a common line. More generally, we have

Theorem 5.4. *Let P be a set of n points in the plane $\mathbb{R}^2$ such that at most s points are on a line. If these points determine distinct distances $d_1, d_2, \ldots, d_t$ with corresponding multiplicities $m_1, m_2, \ldots, m_t$, then*

$$\sum_{i=1}^{t} m_i^2 \le \frac{(s+1)n}{2} \cdot \binom{n}{2} . \tag{31}$$

Remark. If one assumes in contrast to Theorem 5.4 that the n points lie on s lines (instead of 'at most s points on a line'), then one can show by a similar argument that $\sum m_i^2 \le sn\binom{n}{2}$ holds.

Proof. We will give an upper bound for $\Delta(P)$ (see Definition 3.3). By assumption the perpendicular bisector of any two points p and q contains at most s points of P. Thus,

$$\Delta(P) \le s \cdot \binom{n}{2} . \tag{32}$$

By Lemma 3.4 we infer that

$$\sum_{i=1}^{t} m_i^2 \le \frac{n}{2} \cdot \left(\Delta(P) + \binom{n}{2} \right) \le \frac{(s+1)n}{2} \cdot \binom{n}{2} .$$ ∎

Theorem 5.5. *Let P be a set of n points in the plane $\mathbb{R}^2$ in general position. If these points determine the distinct distances $d_1, d_2, \ldots, d_t$ with corresponding multiplicities $m_1, m_2, \ldots, m_t$, then*

$$\sum_{i=1}^{t} m_i^2 \le \frac{3}{4} n^2 (n-1) . \tag{33}$$

If we further assume that the n points are in convex position, then

$$\sum_{i=1}^{t} m_i^2 \le \frac{3}{4} n^2 (n-1) - \frac{n^2}{2} .$$

This proves a conjecture of Erdős and Fishburn [14], [11]. In their paper [11] an even stronger statement is conjectured namely that for convex n-gons one has

$$¿ \quad \sum_{i=1}^{t} m_i^2 \le \frac{n^2 \cdot (n-1)}{2} \quad ?$$

for $n \geq 3$ being an odd integer. For $n \geq 10$ an even integer it is stated in [11] that perhaps

$$¿ \quad \sum_{i=1}^{t} m_i^2 \leq \frac{n^2 \cdot (2n-3)}{4} \quad ?$$

In both cases the regular convex n-gon would be an extremal configuration attaining the upper bounds. For $n=4,6,8$ they proved that this bound does not hold.

Remarks. (1) Füredi proved in [15] that under the assumptions of Theorem 5.5, i.e., P is a convex n-gon, one has $m_i \leq 12n \log n$ for $i = 1,2,\ldots,t$. Applying this in the straightforward way, one gets $\sum_{i=1}^{t} m_i^2 \leq \sum_{i=1}^{t} m_i \cdot \max\{m_1,m_2,\ldots,m_t\} \leq 6n^3 \cdot \log n$.

(2) According to the second remark after Theorem 5.1 we see, that it really makes a difference to assume general position.

Proof. If the points of P are in general position, then (33) follows by Theorem 5.4 with $s=2$. So assume that the points of P determine a convex n-gon. Observe that the bisector of x and y contains at most one point from P if x and y are adjacent along the boundary of the convex hull of P. Thus,

$$\Delta(P) \leq 2\binom{n}{2} - n .$$

By Lemma 3.4 we infer that

$$\sum_{i=1}^{t} m_i^2 \leq \frac{n}{2}\left(2\binom{n}{2} - n + \binom{n}{2}\right) = \frac{3n^2(n-1)}{4} - \frac{n^2}{2} . \qquad \blacksquare$$

It might be worth noting that the following upper bounds for the sum $\sum_{i=1}^{t} m_i^2$ match the conjectured upper bounds of Erdős and Fishburn stated above.

Theorem 5.6. *Let P be a set of n points in $\mathbb{R}^2$, which has the following property:*
(∗) no circle with center $p \in P$ contains three or more other points of P.

Let these n points determine distinct distances $d_1,d_2,\ldots,d_t$ with corresponding multiplicities $m_1,m_2,\ldots,m_t$. Then,

$$\sum_{i=1}^{t} m_i^2 \leq \begin{cases} \frac{n^2(n-1)}{2} & \text{if } n \text{ is odd} \\ \frac{n^2(2n-3)}{4} & \text{if } n \text{ is even.} \end{cases}$$

Proof. By (∗) we have $m_i \leq n$ for $i=1,2,\ldots,t$. Thus

$$\sum_{i=1}^{t} m_i^2 \leq \frac{\binom{n}{2}}{n} n^2 = \frac{n^2(n-1)}{2} ,$$

which shows the assertion for n being an odd integer.

For n even, consider as above the bipartite graph $G = \left([P]^2 \cup P, E\right)$ with $\{\{x,y\},z\} \in E$ if and only if z lies on the perpendicular bisector of x and y. To determine $|E|=\Delta(P)$, fix a point $p\in P$. Let $d(p)$ denote the degree of p in G. Then $d(p)$ is equal to the number of unordered pairs $\{x,y\}\in[P\setminus\{p\}]^2$ such that p lies on the perpendicular bisector of x and y. Property $(*)$ implies that these pairs form a matching. As n is even, we obtain $d(p)\leq(n-2)/2$ and therefore

$$\Delta(P)=\sum_{p\in P} d(p)\leq\frac{n(n-2)}{2}\,.$$

With Lemma 3.4 we infer that

$$\sum_{i=1}^{t} m_i^2\leq\frac{n}{2}\left(\Delta(P)+\binom{n}{2}\right)\leq\frac{n^2(2n-3)}{4}\,. \qquad \blacksquare$$

Next we will consider the corresponding selection problems.

Theorem 5.7. *Let P be a set of n points in general position in the plane. Then there exists a subset $X\subseteq P$ with mutual distinct distances such that*

$$|X|\geq c\cdot n^{1/3}$$

for some positive constant $c>0$.

Remark. An upper bound of $O(n^{1/2})$ is given by the regular n-gon.

Proof. Let $P=\{p_1,p_2,\dots,p_n\}$ be given as above. Let $d_1,d_2,\dots,d_t$ be the occurring distinct distances with corresponding multiplicities $m_1,m_2,\dots,m_t$. We will construct a hypergraph $\mathcal{H}=(P,\mathcal{E}_3\cup\mathcal{E}_4)$ as follows.

Let $\{p_i,p_j,p_k\}\in\mathcal{E}_3\subseteq[P]^3$ if and only if $d(p_i,p_j)=d(p_i,p_k)$, where $d(p,q)$ denotes the euclidean distance between p and q. Moreover, $\{p_i,p_j,p_k,p_l\}\in\mathcal{E}_4\subseteq[P]^4$ if and only if $d(p_i,p_j)=d(p_k,p_l)$.

Consider two points $p,p'\in P$ and let $Q=\{q\in P: d(q,p)=d(q,p')\}$. We observe that all points in Q lie on the perpendicular bisector of p and p'. Since P is in general position we know that $|Q|\leq 2$, thus

$$|\mathcal{E}_3|\leq\binom{n}{2}\cdot 2<n^2\,. \tag{34}$$

Concerning $|\mathcal{E}_4|$, we have by Theorem 5.1 that

$$|\mathcal{E}_4|\leq\sum_{i=1}^{t}\binom{m_i}{2}\leq c_1\cdot n^3\,. \tag{35}$$

We make a random experiment consisting of two steps. First choose a random subset $X\subset P$ by selecting points independently with probability $p=c_2\cdot n^{-2/3}$, where $c_2>0$ is a constant that will be specified later. In the second step we delete from X one point from each edge in $\mathcal{E}_3^*=[X]^3\cap\mathcal{E}_3$ and $\mathcal{E}_4^*=[X]^4\cap\mathcal{E}_4$. By (34) and (35) this results in an independent set $Y\subset X$ with average size

$$\begin{aligned}E(|Y|)&\geq E(|X|)-E(|\mathcal{E}_3^*|)-E(|\mathcal{E}_4^*|)\\&=p\cdot n-p^3\cdot|\mathcal{E}_3|-p^4\cdot|\mathcal{E}_4|>c_2\,n^{1/3}-c_2^3-c_1\,c_2^4\,n^{1/3}.\end{aligned}$$

Choosing $c_2 = \min\{3^{-1/2}, (3c_1)^{-1/3}\}$, we obtain that

$$E(|Y|) > \frac{c_2}{3} \cdot n^{1/3}.$$

Hence there exists a set $Y \subset P$ of the desired size and with distinct mutual distances. ∎

Using Theorem 5.4 one can show using (31) and (32) in a similar fashion the following

Theorem 5.8. *Let P be set of n points in the plane $\mathbb{R}^2$ such that at most s points of P lie on any line. Then there exists a subset $X \subseteq P$ with mutual distinct distances such that*

$$|X| \geq c \cdot \left(\frac{n}{s}\right)^{1/3},$$

where $c>0$ is an absolute constant (independent of s).

For n points in arbitrary position we have the following result.

Theorem 5.9. *Let P be a set of n points in $\mathbb{R}^2$. Then there exists a subset $X \subseteq P$ with mutual distinct distances such that*

$$|X| \geq c \cdot n^{1/4},$$

where $c>0$ is a constant.

This improves a lower bound of $c \cdot n^{1/5}$ which follows from [6, proof of Theorem 4.2], as communicated to us by János Pach, and the lower bound of $c \cdot n^{2/9}$ given in [24]. An upper bound of $O(n^{1/2}/(\log n)^{1/4})$ follows from the $\sqrt{n} \times \sqrt{n}$-grid.

Proof. The arguments are similar to those used in the proof of Theorem 5.7. We form a hypergraph $\mathcal{H} = (P, \mathcal{E}_3 \cup \mathcal{E}_4)$ as before. Pach and Sharir have shown in [21] that n points in the plane determine $O(n^{7/3})$ isosceles triangles, hence

$$|\mathcal{E}_3| \leq c_1 \cdot n^{7/3}$$

and by Theorem 5.1

$$|\mathcal{E}_4| \leq c_2 \cdot n^{13/4}.$$

By choosing vertices at random with probability $p = c_3 \cdot n^{-3/4}$ for some small enough positive constant c_3 we obtain as above a subset $Y \subseteq P$ with mutual distinct distances such that $|Y| \geq c \cdot n^{1/4}$. ∎

With Theorem 5.1 we see that for an n-point set P in $\mathbb{R}^2$ the fraction of those k-element subsets of P which determine less than $\binom{k}{2}$ distinct distances is bounded from above by

$$\frac{\sum_{i=1}^{t} m_i^2 \cdot \binom{n-4}{k-4} + O(n^{7/3}) \cdot \binom{n-3}{k-3}}{\binom{n}{k}} \leq O\left(\frac{k^4}{n^{3/4}} + \frac{k^3}{n^{2/3}}\right).$$

Thus if $k = o(n^{3/16})$ then *almost all* k-element subsets of P determine distinct mutual distances. This improves former results from [6] and [24], where $k=o(n^{1/7})$ respective $k=o(n^{1/6})$ has been shown. In [6] it was stated with respect to an upper bound for this problem that for n equidistant points on a line and supposing $k = \Omega(n^{1/4})$, then a positive percentage of all k-sets determine less than $\binom{k}{2}$ distinct distances. Here we will make this statement more precise in the following form:

Theorem 5.10. *Let $p_1,p_2,\dots,p_n$ be n equidistant points on a line. Then the number of k-element subsets of $\{p_1,p_2\dots,p_n\}$, which determine less than $\binom{k}{2}$ distinct distances, is at least*

$$\left(1-\frac{c}{k}-\frac{cn}{k^4}\right)\cdot\binom{n}{k},$$

where c is a positive constant.

Proof. Let $P=\{p_1,p_2,\dots,p_n\}$ be the set of equidistant points on a line. Form a hypergraph $\mathcal{G}=(P,\mathcal{E}_3\cup\mathcal{E}_4)$ with $\mathcal{E}_3\subseteq[P]^3$ and $\mathcal{E}_4\subseteq[P]^4$ as follows: $\{p_i,p_j,p_k\}\in\mathcal{E}_3$ if and only if $d(p_i,p_j)=d(p_i,p_k)$. Moreover, let $\{p_i,p_j,p_k,p_l\}\in\mathcal{E}_4$ if and only if $d(p_i,p_j)=d(p_k,p_l)$ (where d denotes the euclidean distance). Clearly,

$$c_4n^3 \le |\mathcal{E}_4| \le c_4'n^3 . \tag{36}$$

Let $\mathcal{E}_4 = \{S_1,S_2,\dots,S_t\}$, where $c_4 \le \frac{t}{n^3} \le c_4'$. Now pick a k-element subset K uniformly at random among the set of all k-element subsets of $\mathcal{V}$. Let E denote the event that $[K]^3\cap\mathcal{E}_3=\emptyset$ and $[K]^4\cap\mathcal{E}_4=\emptyset$. In the following we will determine an upper bound for the probability that E occurs. For $i=1,2,\dots,t$ let z_i be indicator random variables for the events $S_i\subseteq K$, i.e.

$$z_i = \begin{cases} 1 & \text{if } S_i \subseteq K \\ 0 & \text{else.}\end{cases}$$

Define another random variable $Z=\sum_{i=1}^t z_i$ and let $E(Z)$ be its expected value. Then by Chebychev's inequality

$$\text{Prob }(E) \le \text{ Prob }(Z=0) \le \frac{\text{Var }(Z)}{E(Z)^2} . \tag{37}$$

By linearity of expectation we have

$$E(Z)=\sum_{i=1}^t E(z_i) = t\cdot\frac{\binom{n-4}{k-4}}{\binom{n}{k}} = t\cdot\frac{[k]_4}{[n]_4} , \tag{38}$$

where $[n]_l=n\cdot(n-1)\cdot\ldots\cdot(n-l+1)$ denotes the falling factorial.

For $s=5,6,7,8$ let a_s denote the number of unordered pairs $\{S_i,S_j\}$, $1\le i< j\le t$, with $|S_i\cup S_j|=s$. Then we have for the variance that

$$\text{Var }(Z) = E\left((Z-E(Z))^2\right)$$

$$= 2 \cdot \sum_{1 \le i < j \le t} E(z_i z_j) + E(Z) - E(Z)^2$$

$$= 2 \cdot \sum_{s=5}^{8} a_s \cdot \frac{[k]_s}{[n]_s} + E(Z) - E(Z)^2 \,. \tag{39}$$

Next we will give upper bounds on a_s, $5 \le s \le 8$. First fix an edge $E \in \mathcal{E}_4$. For the following considerations notice that for a fixed distance $d > 0$ and every integer i, $1 \le i \le n$, the number of points p_j, $1 \le j \le n$, with $d(p_i, p_j) = d$ is at most two.

To bound a_5, choose a three-element subset $R \subset E$. Then R can be extended to an edge in $\mathcal{E}_4 \setminus \{E\}$ in at most 6 ways, hence

$$a_5 \le 24t \,. \tag{40}$$

To bound a_6, choose a two-element subset R of E. Then R can be extended in at most $2(n-4)$ ways to an edge in $\mathcal{E}_4 \setminus \{E\}$, hence

$$a_6 \le t \cdot \binom{4}{2} \cdot 2 \cdot (n-4) \le 12nt \,. \tag{41}$$

To bound a_7, take an element $x \in E$. Then the number of edges $E^* \in \mathcal{E}_4 \setminus \{E\}$ with $x \in E^*$ is at most $(n-4) \cdot (n-5) \cdot 2$, thus

$$a_7 \le 8n^2 t \,. \tag{42}$$

Finally, we have

$$a_8 \le \binom{t}{2} \,. \tag{43}$$

Inserting (40), (41), (42) and (43) in (39) yields for $k, n \ge 7$

$$\mathrm{Var}(Z)$$

$$\le 48t \cdot \frac{[k]_5}{[n]_5} + 24tn \cdot \frac{[k]_6}{[n]_6} + 16tn^2 \cdot \frac{[k]_7}{[n]_7} + 2 \cdot \binom{t}{2} \cdot \frac{[k]_8}{[n]_8} + t \cdot \frac{[k]_4}{[n]_4} - \left(t \cdot \frac{[k]_4}{[n]_4} \right)^2$$

$$\le 88tn^2 \cdot \frac{[k]_7}{[n]_7} + t \cdot \frac{[k]_4}{[n]_4} - t \cdot \frac{[k]_8}{[n]_8} \le 88tn^2 \cdot \frac{[k]_7}{[n]_7} + t \cdot \frac{[k]_4}{[n]_4} \,. \tag{44}$$

With (36), (37), (38) and (44) we infer for $k, n \ge 7$ that

$$\mathrm{Prob}\,(E) \le \mathrm{Prob}\,(Z = 0) \le \frac{t \cdot \left(88n^2 \cdot \frac{[k]_7}{[n]_7} + \frac{[k]_4}{[n]_4} \right)}{t^2 \cdot \left(\frac{[k]_4}{[n]_4} \right)^2}$$

$$\le 88 \cdot \frac{n^3}{t \cdot k} + \frac{[n]_4}{[k]_4 \cdot t} \le \frac{88}{k \cdot c_4} + \frac{81 \cdot n}{k^4 \cdot c_4} \le \frac{c}{k} + \frac{c \cdot n}{k^4} \,,$$

where $c = \frac{88}{c_4}$. ∎

Corollary 5.11. *For $k = \omega(n^{1/4})$ and n equidistant points on a line almost all k-element subsets determine less than $\binom{k}{2}$ distinct distances.* ∎

Considerations, similar to those in the proof of Theorem 5.10 yield the following

Corollary 5.12. *Let $p_1, p_2, \ldots, p_n$ be the points of the regular n-gon. Then the number of k-element subsets of $\{p_1, p_2, \ldots, p_n\}$, which determine less than $\binom{k}{2}$ distinct distances is at least*

$$\left(1 - \frac{c}{k} - \frac{cn}{k^4}\right) \cdot \binom{n}{k},$$

for a positive constant c. ∎

By Theorem 5.1 and (34) it follows that for every n-point set P in $\mathbb{R}^2$ in general position the fraction of those k-element subsets of P which determine less than $\binom{k}{2}$ distinct distances is bounded from above by

$$\frac{\sum_{i=1}^{t} m_i^2 \cdot \binom{n-4}{k-4} + n^2 \cdot \binom{n-3}{k-3}}{\binom{n}{k}} \leq O\left(\frac{k^4}{n} + \frac{k^3}{n}\right) = O\left(\frac{k^4}{n}\right).$$

Thus, for $k = o(n^{1/4})$ almost all k-element subsets of P determine distinct mutual distances. By Corollary 5.12 this bound is tight since almost all k-sets of the points of the regular n-gon determine less than $\binom{k}{2}$ distinct distances for $k = \omega(n^{1/4})$.

Similar conclusions can be obtained for the $n \times n$-grid G_n. Namely, construct as in the proof of Theorem 5.10 a hypergraph $\mathcal{H} = (G_n, \mathcal{E}_3 \cup \mathcal{E}_4)$ with vertex set being the points of the $n \times n$-grid. Then by Theorem 3.2 we infer

$$|\mathcal{E}_3| \leq c_3 \cdot n^4 \cdot \ln n \tag{45}$$

$$|\mathcal{E}_4| \leq c_4 \cdot n^6 \cdot \ln n\,. \tag{46}$$

Thus, by (45) and (46) the fraction of those k-element subsets of the $n \times n$-grid G_n with less than $\binom{k}{2}$ distinct distances is bounded from above by

$$\frac{|\mathcal{E}_3| \cdot \binom{n^2-3}{k-3} + |\mathcal{E}_4| \cdot \binom{n^2-4}{k-4}}{\binom{n^2}{k}} \leq c_3 \cdot \frac{k^3 \cdot \ln n}{n^2} + c_4 \cdot \frac{k^4 \cdot \ln n}{n^2}.$$

Hence, for $k = o\left(\frac{n^{1/2}}{(\ln n)^{1/4}}\right)$ almost all k-element subsets of G_n determine distinct mutual distances.

On the other hand, using the ideas of the proof of Theorem 5.10 with corresponding random variable Z for the $n \times n$-grid and using Theorem 3.2 and (18) one can show by Chebychev's inequality, that

$$\operatorname{Prob}(Z = 0) \leq \frac{c \cdot n^{\frac{c^*}{\ln \ln n}}}{k \cdot \ln n} + \frac{c \cdot n^2}{k^4 \cdot \ln n},$$

where c, c^* are positive constants. Thus for $k = \omega\left(\frac{n^{1/2}}{(\ln n)^{1/4}}\right)$ almost all k-element subsets of the $n \times n$-grid determine less than $\binom{k}{2}$ distinct distances.

We remark that one can show that for the corresponding problem for the n^d-grid, $d \geq 3$, we also have a 0–1 law with threshold function $f(n) = \sqrt{n}$, as can be seen along the lines above using (13), Theorem 3.8 and the remark after Theorem 3.8. In particular, for $k = o(n^{1/2})$ almost all k-element subsets of the n^d-grid determine distinct mutual distances, while for $k = \omega(n^{1/2})$ almost all k-element subsets of the n^d-grid determine less than $\binom{k}{2}$ distinct distances.

6. B_2-Sets

For finite sets $X \subset \mathbb{N}$ a subset $S \subseteq X$ is called a B_2-set (or Sidon set) if all pairwise sums $s+s'$, $s \neq s'$, are distinct. One is interested in the maximum size of S. For the case $X=\{1,2,\dots,n\}$ the maximum size of a B_2-set $S \subseteq X$ is asymptotically well known by results from Erdős and Turán to be $(1+o(1)) \cdot n^{1/2}$. In [3] Alon and Erdős considered the maximum size of B_2-subsets of the set $\{1^2,2^2,\dots,n^2\}$ consisting of the first n squares. Using an idea similar to the one given in the proof of Theorem 5.7 they showed the following:

Theorem 6.1. [3] *For every $\epsilon>0$ there exists $c=c(\epsilon)>0$ such that for every positive integer n there exists a B_2-set $S \subset \{1^2,2^2,\dots,n^2\}$ with*

$$|S| \geq c \cdot n^{2/3-\epsilon} . \tag{47}$$

As already observed in [3], by a theorem of Landau [20] one has the upper bound $|S| \leq c' \cdot \frac{n}{(\ln n)^{1/4}}$. Here we will improve inequality (47), namely we will show:

Theorem 6.2. *For every integer $n \geq 1$ there exists a B_2-set $S \subset \{1^2,2^2,\dots,n^2\}$ with*

$$|S| \geq c \cdot n^{2/3} , \tag{48}$$

where $c>0$ is a constant.

The first idea to prove Theorem 6.2 might be to consider a complete graph with vertex set $V=\{1^2,2^2,\dots,n^2\}$ and a coloring of the edges, where the edge $\{i^2,j^2\}$ receives color i^2+j^2. Then a totally multicolored complete subgraph on k vertices gives rise to a B_2-subset of V of cardinality k. But Theorem 2.1 is not applicable to prove Theorem 6.2, as by condition (ii) we can only guarantee a totally multicolored complete subgraph of size less than $c \cdot n^{1/2}$. But it turns out, that with more refined counting arguments a similar strategy as used for the proof of Theorem 2.1 will show (48):

Proof. As in the proof of Theorem 2.1 we can assume that n is sufficiently large. In the following $c_1,c_2,\dots,c_{10}$ are positive constants. We construct a 4-uniform hypergraph $\mathcal{G}=(V,\mathcal{E})$ with vertex set $V=\{1^2,2^2,\dots,n^2\}$ and edges $\{i^2,j^2,k^2,l^2\} \in \mathcal{E} \subseteq [V]^4$ if and only if $i^2+j^2=k^2+l^2$. As the number of representations of any positive integer x by a sum of two squares is given by $r_2(x)$, we have by Theorem 3.2 that

$$|\mathcal{E}| \leq \sum_{i=1}^{2n^2} \binom{r_2(i)}{2} \leq c_1 \cdot n^2 \cdot \ln n . \tag{49}$$

Next we will count the number of 2-cycles in $\mathcal{G}$. To count $c_{2,2}(\mathcal{G})$ choose an edge $E \in \mathcal{E}$, say $E=\{i^2,j^2,k^2,l^2\}$ where $i^2+j^2=k^2+l^2$. There are six possibilities to choose a two-element subset of E, say we choose $\{i^2,j^2\}$. Then the number of pairs $\{x^2,y^2\}$ with $i^2+j^2=x^2+y^2$ is bounded from above by $r_2(i^2+j^2) \leq n^{\frac{c_2}{\ln \ln n}}$

(cf. [19]). Now consider those pairs $\{x^2, y^2\}$ with $i^2 + x^2 = j^2 + y^2$. Assuming $j > i$, we have $j^2 - i^2 = (x+y)\cdot(x-y)$, i.e. $(x+y)$ divides $j^2 - i^2$. Fixing this divisor fixes both x and y. Hence the number of such pairs $\{x^2, y^2\}$ is bounded from above by the number of divisors of $(j^2 - i^2)$, which is at most $n^{\frac{c_3}{\ln\ln n}}$ (see [19]). Summarizing these considerations, we have

$$c_{2,2}(\mathscr{G}) \le c_4 \cdot n^2 \cdot \ln n \cdot n^{\frac{c_5}{\ln\ln n}} . \tag{50}$$

Concerning $c_{2,3}(\mathscr{G})$, we choose an edge $E \in \mathscr{E}$ and a three-element subset $T \subset E$. Then T can be extended in at most two ways to an edge $E' \in \mathscr{E} \setminus \{E\}$, thus

$$c_{2,3}(\mathscr{G}) \le c_6 \cdot n^2 \cdot \ln n . \tag{51}$$

As in the proof of Theorem 2.1 we choose a random subset of V by picking vertices $v \in V$, independently of the others, with probability

$$p = n^{-1/3+\epsilon} \cdot (\ln n)^{-1/3} ,$$

where $\epsilon < \frac{1}{18}$. Let R be the arising random subset of V. Then,

$$\text{Prob } (|R| \approx pn) = 1 - o(1) \tag{52}$$

and by (49) we have

$$E(|[R]^4 \cap \mathscr{E}|) = p^4 \cdot |\mathscr{E}| \le c_1 \cdot \frac{n^{2/3+4\epsilon}}{(\ln n)^{1/3}} . \tag{53}$$

The expected number $E(c_2(R))$ of 2-cycles in the subhypergraph induced on R can be bounded from above by (50) and (51) as follows:

$$E(c_2(R)) = p^6 \cdot c_{2,2}(\mathscr{G}) + p^5 \cdot c_{2,3}(\mathscr{G}) \le c_4 \cdot \frac{n^{6\epsilon + \frac{c_5}{\ln\ln n}}}{\ln n} + c_6 \cdot \frac{n^{1/3+5\epsilon}}{(\ln n)^{2/3}} = o(pn) \tag{54}$$

for $\epsilon < \frac{1}{18}$.

As in the proof of Theorem 2.1 we infer with (52), (53) and (54) by using Chernoff's and Markov's inequality, deleting one point from each 2-cycle and deleting points of degree bigger than, say, twice the average degree, that there exists a subset $Y \subset V$, $|Y| \ge c_7 \cdot p \cdot n$ such that the subhypergraph $\mathscr{G}'$ of $\mathscr{G}$ induced on Y has no 2-cycles, has at most $c_8 \cdot p^4 \cdot |\mathscr{E}|$ edges and has maximum degree at most $t^3 = c_9 \cdot n^{3\epsilon}$. By Theorem 2.2 applied to $\mathscr{G}'$ we see that

$$\alpha(\mathscr{G}) \ge \alpha(\mathscr{G}') \ge c_{10} \cdot \frac{n^{2/3+\epsilon}}{t \cdot (\ln n)^{1/3}} \cdot (\ln t)^{1/3} \ge c \cdot n^{2/3} ,$$

which finishes the proof. ∎

7. Algorithmic Aspects

In this section we will discuss some algorithmic aspects of the selection problems considered in this article. All these selection problems can be formulated in terms of edge colorings of complete graphs. The general question is: given an edge coloring f of the complete graph K_n, what is the maximum size $r(f)$ of a totally multicolored complete subgraph, i.e. a set of vertices determining mutually distinct edge colors. Clearly, this problem is NP-hard. With an edge coloring we associate a hypergraph $\mathcal{H} = (V(K_n), \mathcal{E}_3 \cup \mathcal{E}_4)$, where $\mathcal{E}_3$ is the family of 3-sets of vertices determining two equal edge colors and $\mathcal{E}_4$ is the family of 4-sets of vertices that determine two equal colors but do not contain a 3-set from $\mathcal{E}_3$. Hence we have $r(f) = \alpha(\mathcal{H})$, where $\alpha(\mathcal{H})$ is the *independence number* of $\mathcal{H}$.

We define the *probabilistic bound*

$$\tilde{\alpha}(\mathcal{H}) := \max_{p \in [0,1]} (pn - p^3 |\mathcal{E}_3| - p^4 |\mathcal{E}_4|),$$

which is a lower bound for $\alpha(\mathcal{H})$, since we can pick each vertex independently with probability p and then delete one vertex from each edge occurring in the resulting subhypergraph. This gives an independent set of size at least $pn - p^3|\mathcal{E}_3| - p^4|\mathcal{E}_4|$ in the average. By using derandomization techniques (see [5]) we can turn this probabilistic argument into a deterministic algorithm that computes an independent set of the hypergraph and thus a totally multicolored complete subgraph of the original graph of size at least $\tilde{\alpha}(\mathcal{H})$. In the following, we will describe the algorithm.

Let the vertex set of K_n be $V = \{1, 2, \ldots, n\}$. Let $f : E(K_n) \longrightarrow T$ be an edge coloring and assume that T is totally ordered. For $t \in T$ let $m_t = |f^{-1}(t)|$ be the number of edges in color t. In a preprocessing we form our hypergraph $\mathcal{H} = (V, \mathcal{E}_3 \cup \mathcal{E}_4)$ by collecting pairs of edges of the same color. By first sorting the set of edges with respect to their colors this can be done in time $O(n^2 \ln n + \sum_{t \in T} m_t^2)$. Moreover, we use the following data structure. There is a list of the vertices $v \in V$ and a list of the edges $e \in \mathcal{E}_3 \cup \mathcal{E}_4$. For each vertex $v \in V$ there are pointers to all edges containing v. For each edge there are pointers to all vertices contained in it.

Knowing $|\mathcal{E}_3|$ and $|\mathcal{E}_4|$, we can easily compute that value $p \in [0,1]$, which maximizes the expression $pn - p^3|\mathcal{E}_3| - p^4|\mathcal{E}_4|$. Fix this value of p.

In the following we will examine the vertices of V one by one and decide whether each vertex belongs to our independent set or not.

Set $\mathcal{E} = \mathcal{E}_3 \cup \mathcal{E}_4$. Suppose that we already made a partial selection of vertices and let $\epsilon_1, \epsilon_2, \ldots, \epsilon_j$ be the 0,1-sequence representing this selection, that is, for some vertices we determined whether they do ($\epsilon_i = 1$) or do not ($\epsilon_i = 0$) belong to our independent set. Define weight functions f_j and F_j depending on $\epsilon_1, \epsilon_2, \ldots, \epsilon_j$ as follows. For vertices $v \in V$ let

$$f_j(v) = \begin{cases} \epsilon_v & \text{if } v \leq j \\ p & \text{if } v > j. \end{cases}$$

For edges $e \in \mathcal{E}$ set

$$f_j(e) = \prod_{v \in e} f_j(v) \,.$$

Finally, set

$$F_j = F_j(\epsilon_1, \epsilon_2, \ldots, \epsilon_{j-1}) = \sum_{v \in V} f_j(v) - \sum_{e \in \mathscr{E}} f_j(e) \,.$$

Observe that F_j is the expected value of the number of vertices minus the number of edges in a random extension of the selection $\epsilon_1, \epsilon_2, \ldots, \epsilon_j$.

At the beginning, for $j=0$, we have $f_0(v)=p$ and $f_0(e)=p^{|e|}$ for $v \in V$ and $e \in \mathscr{E}$, thus $F_0 = \tilde{\alpha}(\mathscr{H})$. We will construct a 0,1-sequence $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ such that the values $F_j = F_j(\epsilon_1, \ldots, \epsilon_j)$ are nondecreasing for $j = 0, 1, \ldots, n$. In particular, we will have $F_n \geq F_0$.

Now assume that $\epsilon_1, \epsilon_2, \ldots, \epsilon_{j-1}$ and $F_{j-1} \geq F_{j-2} \geq \ldots \geq F_0$ are given. To determine ϵ_j and thus F_j, we compute the two values

$$W_j^0 = F_j(\epsilon_1, \epsilon_2, \ldots, \epsilon_{j-1}, 0)$$
$$W_j^1 = F_j(\epsilon_1, \epsilon_2, \ldots, \epsilon_{j-1}, 1).$$

This can be done in time $O(1+\deg_{\mathscr{H}}(j))$. If $W_j^0 \geq W_j^1$, then we set $\epsilon_j = 0$ and $F_j = W_j^0$. Otherwise, if $W_j^1 > W_j^0$, then set $\epsilon_j = 1$ and $F_j = W_j^1$.

By straightforward calculations or by interpreting F_j as an expected value, we derive

$$F_{j-1} = (1-p) \cdot W_j^0 + p \cdot W_j^1 \,. \tag{55}$$

This implies $F_j \geq F_{j-1}$.

Continuing in this way, we obtain a 0,1-sequence $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ with $F_n \geq F_0$. Set $I = \{v \in V \mid \epsilon_v = 1\}$. We claim that I is an independent set in $\mathscr{H}$. Assume this is not the case. Thus there is an edge $e \in \mathscr{E}$ contained in I. Let j be the last vertex of e chosen by our algorithm. Since j was chosen, we know that $W_j^1 > W_j^0$, which with (55) implies that

$$F_j = W_j^1 > F_{j-1} \,.$$

On the other hand, since j is the *last* chosen vertex of e, we infer that

$$F_j - F_{j-1} \leq (f_j(j) - f_{j-1}(j)) - (f_j(e) - f_{j-1}(e)) = (1-p) - (1-p) = 0 \,,$$

a contradiction. Thus I is an independent set with

$$|I| = F_n \geq F_0 = \tilde{\alpha}(\mathscr{H})$$

as desired.

Without the preprocessing this algorithm has a linear running time of $O(n + |\mathscr{E}_3| + |\mathscr{E}_4|)$. Since $|\mathscr{E}_3|, |\mathscr{E}_4| < \sum_{t \in T} m_t^2$, we have an overall running time of $O(n^2 \ln n + \sum_{t \in T} m_t^2)$.

By our former results we obtain the following typical consequences. By Theorem 5.5, given a set P of n points in $\mathbb{R}^2$ in general position, this algorithm finds in sequential time $O(n^3)$ a subset $X \subseteq P$ with mutual distinct distances of size at least $c \cdot n^{1/3}$, where c is a positive constant. Moreover, by Theorem 5.1, if the n points of P are in arbitrary position, then the algorithm finds in time $O(n^{13/4})$ a subset $X \subseteq P$ with mutual distinct distances of size $c \cdot n^{1/4}$ for some positive constant c.

On the other hand, we remark that by the method described by Alon, Babai and Itai [1] there is an NC-algorithm that computes a totally multicolored complete subgraph of size at least $c \cdot \tilde{\alpha}(\mathcal{H})$ for some constant $c > 0$.

Remark. This research was partly motivated by related problems considered in [13] which have applications to the problem of distance measuring by using radar or sonar signals, see [16] and [17].

Acknowledgement. We would like to thank Kent D. Boklan for drawing our attention to the Hardy-Littlewood method. We also thank Emo Welzl for suggesting Lemma 5.2.

References

[1] N. Alon, L. Babai, and A. Itai: A Fast and Simple Randomized Parallel Algorithm for the Maximal Independent Set Problem, *Journal of Algorithms*, **7** (1986), 567–583.

[2] M. Ajtai, J. Komlós, J. Pintz, J. Spencer, and E. Szemerédi: Extremal Uncrowded Hypergraphs, *J. Comb. Theory (Ser. A)*, **32** (1982), 321–335.

[3] N. Alon, and P. Erdős: An Application of Graph Theory to Additive Number Theory, *European Journal of Combinatorics*, **6** (1985), 201–203.

[4] N. Alon, H. Lefmann, and V. Rödl: On an Anti-Ramsey Type Result, *Colloquia Mathematica Societatis János Bolyai, 60. Sets, Graphs and Numbers*, Budapest (1991), 9–22.

[5] N. Alon, and J. Spencer: The Probabilistic Method, *Wiley & Sons, New York*, (1992).

[6] D. Avis, P. Erdős, and J. Pach: Distinct Distances Determined by Subsets of a Point Set in Space, *Computational Geometry* **1**, (1991), 1–11.

[7] K. Clarkson, H. Edelsbrunner, L. Guibas, M. Sharir, and E. Welzl: Combinatorial Complexity Bounds for Arrangements of Curves and Spheres, *Discrete and Computational Geometry*, **5** (1990), 99–160.

[8] H. Davenport: Analytic Methods for Diophantine Equations and Diophantine Inequalities, *Campus Publishers* (1962), 1–30.

[9] R. A. Duke, H. Lefmann, and V. Rödl: On Uncrowded Hypergraphs, *Random Structures and Algorithms*, to appear.

[10] P. Erdős: On Some Problems of Elementary and Combinatorial Geometry, *Annali di Mathematica Pura et Applicata Ser. 4*, **103** (1975), 99–108.

[11] P. Erdős, and P. C. Fishburn: Multiplicities of Interpoint Distances in Finite Planar Sets, *preprint*, (1991), to appear in *Discrete and Applied Math.*

[12] P. Erdős, and R. Guy: Distinct Distances between Lattice Points, *Elemente der Mathematik*, **25** (1970), 121–123.

[13] P. Erdős, R. L. Graham, I. Ruzsa, and H. Taylor: Bounds for Arrays of Dots with Distinct Slopes or Lengths, *Combinatorica*, **12** (1992), 39–44.

[14] P. Fishburn: Convex Polygons with Few Vertices, *Dimacs Technical Report*, **92–17** (1992).

[15] Z. Füredi: The Maximum Number of Unit Distances in a Convex n-gon, *J. Comb. Theory (Ser. A)*, **55** (1990), 316–320.

[16] S. W. GOLOMB: Construction of Signals with Favourable Correlation Properties, *Surveys in Combinatorics*, London Mathematical Society Lecture Note Series, **166** (1991), 1–39.

[17] S. W. GOLOMB, and H. TAYLOR: Two-dimensional Synchronization Patterns for Minimum Ambiguity, *IEEE Transactions Information Theory*, **IT-28** (1982), 600–604.

[18] R. K. GUY: Unsolved Problems in Number Theory, *Springer Verlag, New York*, (1981), 132–133.

[19] G. H. HARDY, and E. M. WRIGHT: An Introduction to the Theory of Numbers, *Oxford University Press*, (1979).

[20] E. LANDAU: Handbuch der Lehre von der Verteilung der Primzahlen, *Teubner Verlag, Leipzig*, (1909).

[21] J. PACH, and M. SHARIR: Repeated Angles in the Plane and Related Problems, *J. Comb. Theory (Ser. A)*, **59** (1992), 12–22.

[22] S. RAMANUJAN: Collected Papers, *Chelsea Publishing Company*, (1962), 133–135.

[23] J. SPENCER, E. SZEMERÉDI, and W. T. TROTTER: Unit Distances in the Euclidean Plane, *Graph Theory and Combinatorics, Academic Press*, London (1984), 253–278.

[24] T. THIELE: Point Sets with Distinct Slopes or Lengths, *preprint*, (1993).

[25] R. C. VAUGHAN: The Hardy-Littlewood Method, *Cambridge University Press*, (1981), 1–25.

[26] B. M. WILSON: Proofs of Some Formulæ Enunciated by Ramanujan, *Proceedings London Mathematical Society*, **21** (1923), 235–255.

Hanno Lefmann

Universität Dortmund,
Fachbereich Informatik,
LS II, D-44221 Dortmund,
Germany
`lefmann@bienayme.informatik.uni-dortmund.de`

Torsten Thiele

Freie Universität Berlin,
Institut für Mathematik II,
Arnimallee 3, D-14195 Berlin,
Germany
`thiele@math.fu-berlin.de`