



Contents lists available at ScienceDirect

European Journal of Combinatorics

journal homepage: <https://www.sciencedirect.com/journal/european-journal-of-combinatorics>

Subset selection problems in planar point sets

József Balogh^a, Felix Christian Clemen^b,
Adrian Dumitrescu^{c,d,e}, Dingyuan Liu^{f,*}^a University of Illinois at Urbana-Champaign, USA^b Institute for Basic Science, South Korea^c Algorsearch L.L.C., Milwaukee, WI, USA^d Alfréd Rényi Institute of Mathematics, Budapest, Hungary^e Research Institute of the University of Bucharest, Romania^f Karlsruhe Institute of Technology, Germany

ARTICLE INFO

Article history:

Received 7 December 2025

Accepted 7 September 2026

ABSTRACT

Given a finite point set satisfying condition $\mathcal{A}$, the subset selection problem asks, how large of a subset satisfying condition $\mathcal{B}$ can be extracted? Problems of this nature are notoriously difficult and have attracted extensive study. In this paper, we make progress on three instances of subset selection problems in planar point sets. Let $n, s \in \mathbb{N}$ with $n \geq s$, and let $P \subseteq \mathbb{R}^2$ be a set of n points, where at most s points lie on the same line.

Firstly, we select a general position subset of P , i.e., a subset containing no 3 points on the same line. This problem was proposed by Erdős under the regime when s is a constant. For s being non-constant, we give new lower and upper bounds on the maximum size of such a subset. In particular, we show that in the worst case such a set can have size at most $O(n^{5/6+o(1)}/\sqrt{s})$ when $3 \leq s \leq n^{1/3}$ and $O(n/s)$ when $n^{1/3} \leq s \leq n$.

Secondly, we select a monotone general position subset of P , that is, a subset in general position where the points are ordered from left to right and their y -coordinates are either non-decreasing or non-increasing. We present bounds on the maximum size of such a subset. In particular, when $s = \Omega(\sqrt{n})$, our upper and lower bounds differ at most by a logarithmic factor.

Lastly, we select a subset of P with pairwise distinct slopes. This problem was initially studied by Erdős, Graham, Ruzsa, and Taylor on the grid. We show that for $s = O(\sqrt{n})$ such a subset of

* Corresponding author.

E-mail addresses: jobal@illinois.edu (J. Balogh), felix.clemen@ibs.re.kr (F.C. Clemen), ad.dumitrescu@algorsearch.org (A. Dumitrescu), liu@mathe.berlin (D. Liu).

size $\Omega((n/\log s)^{1/3})$ can always be found in P . When $s = \Theta(\sqrt{n})$, this matches a lower bound given by Zhang on the grid. As for the upper bound, we show that in the worst case such a subset has size at most $O(\sqrt{n})$ for $2 \leq s \leq n^{3/8}$ and $O((n/s)^{4/5})$ for $n^{3/8} \leq s = O(\sqrt{n})$.

The proofs use a wide range of tools such as incidence geometry, probabilistic methods, the hypergraph container method, and additive combinatorics.

© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For an integer $k \geq 3$, a *collinear k -tuple* is a set of k distinct points lying on the same line. A planar point set is said to be in *general position* if it contains no collinear triple. The problem of selecting a large general position subset from an $n \times n$ integer grid, also known as the *no-three-in-line problem*, goes back more than 100 years, see for examples [9] and [28, Ch. 10]. A pigeonhole principle argument shows that one can select at most $2n$ points. Dudeney [9] asked whether one can always select $2n$ points and this question remains open.

The no-three-in-line problem is a typical instance of the *subset selection problem*. In general, the subset selection problem asks: Given a finite point set satisfying condition $\mathcal{A}$, how large of a subset satisfying condition $\mathcal{B}$ can be extracted?

Such problems are known for their intrinsic difficulty and therefore often serve as important benchmark tasks in machine learning. For example, recent AI-assisted methods [5,17] have been used to search for a large subset of an $n \times n$ grid that spans no isosceles triangle, as well as a large subset of an $n \times n \times n$ grid containing no 5 points on a common sphere. In this paper, we focus on three subset selection problems in planar point sets, two of which have received considerable attention in extremal combinatorics and discrete geometry (see Sections 1.1 and 1.3), while the remaining one is a new variation (see Section 1.2).

1.1. Selecting a general position subset

A natural generalization of the no-three-in-line problem is to select a large general position subset from an arbitrary planar point set satisfying a certain collinearity condition. Let $n, s \in \mathbb{N}$ with $n \geq s \geq 2$. The function $f(n, s)$ is defined as the largest integer, such that every set $P \subseteq \mathbb{R}^2$ of n points with at most s points collinear, i.e., without collinear $(s+1)$ -tuples, contains a general position subset of size $f(n, s)$. It follows from the definition that $f(n, 2) = n$. Erdős [11] studied the first non-trivial case $f(n, 3)$ and showed that $f(n, 3) = \Omega(\sqrt{n})$ by a greedy selection. Later Füredi [15] proved that $f(n, 3) = \Omega(\sqrt{n \log n})$ using the lower bound on the independence number of partial Steiner triple systems [29]. Employing a similar idea, Lefmann [25] generalized Füredi's lower bound to all $f(n, s)$ with fixed $s \geq 3$.

On the other hand, Füredi [15] constructed a set $P \subseteq \mathbb{R}^2$ of n points containing no collinear 4-tuple and used the density Hales–Jewett theorem [16] to show that P contains at most $o(n)$ points in general position, namely, $f(n, 3) = o(n)$. Balogh and Solymosi [3] proved that a random n -element subset of a large 3-dimensional grid contains no collinear 4-tuple and at most $n^{5/6+o(1)}$ points in general position with high probability. After projecting this subset into the plane they obtained a better upper bound construction, implying that $f(n, 3) \leq n^{5/6+o(1)}$. Note that $f(n, s)$ is monotone in s , namely, $f(n, s) \leq f(n, s')$ holds for all $s \geq s'$. Thus the aforementioned results¹ established that

$$\Omega(\sqrt{n \log n}) = f(n, s) \leq f(n, 3) \leq n^{5/6+o(1)} \quad \text{for any fixed } s \geq 3. \quad (1)$$

¹ Further discussion on these aspects can be found in [10, Ch. 9].

Payne and Wood [27] were the first to study the function $f(n, s)$ when s is a function of n . They showed that

$$f(n, s) = \begin{cases} \Omega(\sqrt{n \log_s n}) & \text{for } 3 \leq s \leq n^{1/2-c} \text{ with any fixed constant } c > 0, \\ \Omega(\sqrt{n/\log n}) & \text{for } n^{1/2-o(1)} \leq s \ll \sqrt{n \log n}, \\ \Omega(n/s) & \text{for } s = \Omega(\sqrt{n \log n}). \end{cases} \quad (2)$$

We present new upper bounds on $f(n, s)$ in the non-constant regime of s .

Theorem 1. *We have*

$$f(n, s) = \begin{cases} O(n^{5/6+o(1)}/\sqrt{s}) & \text{for } 3 \leq s \leq n^{1/3}, \\ O(n/s) & \text{for } n^{1/3} \leq s \leq n. \end{cases} \quad (3)$$

In particular, $f(n, s) = \Theta(n/s)$ when $s = \Omega(\sqrt{n \log n})$.

Our proof of Theorem 1 in the case $s \leq n^{1/3}$ follows the approach of Balogh and Solymosi [3] on the upper bound on $f(n, 3)$, using the hypergraph container method. This method has been widely employed in geometric settings, see, e.g., [6,31,32,36]. The main novelty in our proof is a balanced supersaturation result, which allows us to iteratively apply the hypergraph lemma with improved parameters and thereby establish the desired upper bound over a wider range of s . Indeed, the supersaturation result of Balogh and Solymosi [3] (see also Keller and Smorodinsky [24]) yields the upper bound $f(n, s) = O(n^{5/6+o(1)}/\sqrt{s})$ only for $s \leq n^{1/6}$, leaving a weaker upper bound on $f(n, s)$ in the range $n^{1/6} < s \leq n^{1/3}$.

A result of Hajnal and Szemerédi [21, Theorem 3] yields an improved lower bound $f(n, s) = \Omega(\sqrt{n \log \log n / \log n})$ for $n^{1/2-o(1)} \leq s = O(\sqrt{n \log \log n / \log n})$. Here we extend the range of s where a similar lower bound holds.

Proposition 2. *Let $n, s \in \mathbb{N}$ with $3 \leq s \ll \sqrt{n \log n}$. Then*

$$f(n, s) = \Omega\left(\sqrt{\frac{n \log \frac{n \log s}{s^2}}{\log s}}\right). \quad (4)$$

In particular, for $n^{1/2-o(1)} \leq s = O(\sqrt{n \log \log n})$ we have $f(n, s) = \Omega\left(\sqrt{\frac{n \log \log n}{\log n}}\right)$.

Note that our lower bound in (4) is consistent with (2) when $s \leq n^{1/2-c}$ for any fixed constant $c > 0$ and better in the critical range $n^{1/2-o(1)} \leq s \ll \sqrt{n \log n}$.

1.2. Selecting a monotone general position subset

According to a well-known theorem due to Erdős and Szekeres [14], for any $\ell, m \geq 1$, every sequence $(x_i)_{i=1}^n$ of real numbers with $n > \ell m$ contains a non-decreasing subsequence of length larger than ℓ , or a non-increasing subsequence of length larger than m . For a point $v \in \mathbb{R}^2$ denote by v_x and v_y the x -coordinate and y -coordinate of v , respectively. We say that a point set $P \subseteq \mathbb{R}^2$ is *monotone* if one of the following holds:

- (i) $\forall u, v \in P : u_x < v_x \implies u_y \leq v_y$;
- (ii) $\forall u, v \in P : u_x < v_x \implies u_y \geq v_y$.

See Fig. 1 for an illustration of a monotone general position subset. A simple corollary of the Erdős–Szekeres theorem states that any set $P \subseteq \mathbb{R}^2$ of n points contains a monotone subset of size at least $\sqrt{n}$. Let $g(n, s)$ be the largest integer, such that any set $P \subseteq \mathbb{R}^2$ of n points with at most s collinear points contains a monotone general position subset of size $g(n, s)$. It is not difficult to show that $g(n, 2) = \Theta(\sqrt{n})$. We are mainly interested in the following question:

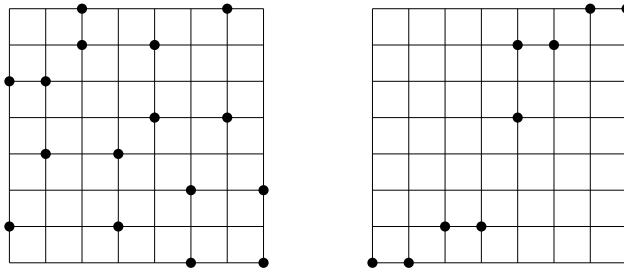


Fig. 1. Left: Dudeney's maximum-sized general position subset in the 8×8 grid. Right: a monotone general position subset in the 8×8 grid.

Question 3. Let $n, s \in \mathbb{N}$ with $n \geq s \geq 3$. What is the growth order of $g(n, s)$?

We give a partial answer to this question as follows.

Theorem 4. Let $n, s \in \mathbb{N}$ with $n \geq s \geq 3$. Then

$$g(n, s) = \begin{cases} \Omega \left(\left(\frac{n \log \frac{n \log s}{s^2}}{\log s} \right)^{1/4} \right) & \text{for } 3 \leq s \ll \sqrt{n \log n}, \\ \Omega(\sqrt{n/s}) & \text{for } s = \Omega(\sqrt{n \log n}). \end{cases} \quad (5)$$

On the other hand, we have $g(n, s) = O(\sqrt{n/s})$.

It is worth noting that the gap between the bounds in Theorem 4 is large for small s but almost vanishes when s is large. In particular, when $s = \Omega(\sqrt{n})$, the upper and lower bounds differ only by a logarithmic factor, and when $s = \Omega(\sqrt{n \log n})$ we have $g(n, s) = \Theta(\sqrt{n/s})$. Additionally, for the n points in a $\sqrt{n} \times \sqrt{n}$ integer grid,² we deduce better bounds.

Theorem 5. Let G be a $\sqrt{n} \times \sqrt{n}$ grid of n points. Then the largest size of a monotone general position subset of G is at least $\Omega((n/\log n)^{2/5})$ and at most $o(n^{1/2})$.

The construction obtaining the lower bound in Theorem 5 is a random subset taken from the intersection of the grid and a sector of an annulus with a carefully chosen radial width.

1.3. Selecting a subset with pairwise distinct slopes

A strengthening of the general position requirement is that of *pairwise distinct slopes*. Indeed every point set with pairwise distinct slopes has no collinear triple.

For the $\sqrt{n} \times \sqrt{n}$ grid G , Erdős, Graham, Ruzsa, and Taylor [12] showed that the largest size of a subset of G with pairwise distinct slopes is at least $\Omega(n^{1/4})$ and at most $O(n^{2/5})$. Soon after, Zhang [39] raised the lower bound to $\Omega((n/\log n)^{1/3})$, and very recently Clemen [7] sharpened it to $\Omega((n \log \log n / \log n)^{1/3})$, and these are the current records.

We consider a generalization of the distinct slope problem to arbitrary planar point sets. Let $h(n, s)$ be the largest integer, such that any set $P \subseteq \mathbb{R}^2$ of n points with at most s points collinear contains a subset of $h(n, s)$ points that determine pairwise distinct slopes. We prove the following bounds on $h(n, s)$ for different ranges of s . In particular, when $s = \Theta(\sqrt{n})$, our lower bound on $h(n, s)$ matches the bound by Zhang [39] for the $\sqrt{n} \times \sqrt{n}$ grid.

² Throughout this paper, whenever we refer to the $\sqrt{n} \times \sqrt{n}$ grid we assume that n is a perfect square. One could work with a $t \times n/t$ grid instead, where $t = \Theta(\sqrt{n})$ and t divides n , but the results would be very similar.

Theorem 6. Let $n, s \in \mathbb{N}$ with $2 \leq s = O(\sqrt{n})$. We have that

$$h(n, s) = \Omega((n/\log s)^{1/3}). \quad (6)$$

On the other hand, we have

$$h(n, s) = \begin{cases} O(\sqrt{n}) & \text{for } 2 \leq s \leq n^{3/8}, \\ O((n/s)^{4/5}) & \text{for } n^{3/8} \leq s = O(\sqrt{n}). \end{cases} \quad (7)$$

The lower bound in (6) is derived by selecting a random subset, with key parameters controlled using the Szemerédi–Trotter theorem from incidence geometry. The construction proving the first upper bound in (7) relies on the theory of Sidon sets from additive combinatorics. Meanwhile, the construction obtaining the second upper bound in (7) is a random subset of the integer grid $[n]^2$.

The organization of the paper is as follows. We first prove [Proposition 2](#) in [Section 2](#). [Theorem 1](#) is established in [Section 3](#). In [Section 4](#) we prove [Theorems 4](#) and [5](#). [Section 5](#) is dedicated to proving [Theorem 6](#). Finally, [Section 6](#) explores additional consequences of our main results.

2. Selecting a general position subset: lower bound

In this section we prove [Proposition 2](#). The main tool behind our proof is the following result of Cooper and Mubayi [8] on the chromatic number of sparse hypergraphs. Given positive integers $j \leq k$ and a k -uniform hypergraph $\mathcal{H}$, we denote the *maximum j -degree* of $\mathcal{H}$ by

$$\Delta_j(\mathcal{H}) := \max_{A \subseteq V(\mathcal{H}), |A|=j} |\{e \in E(\mathcal{H}) : A \subseteq e\}|. \quad (8)$$

The *maximum degree* of $\mathcal{H}$ is denoted by $\Delta(\mathcal{H}) := \Delta_1(\mathcal{H})$.

Lemma 7 (Cooper–Mubayi [8, Theorem 5]). Let $k \geq 3$ be a fixed integer and $\mathcal{H}$ be a k -uniform hypergraph with maximum degree $\Delta = \Delta(\mathcal{H})$. If

$$\Delta_j(\mathcal{H}) \leq \Delta^{\frac{k-j}{k-1}}/\gamma \quad (9)$$

for $j = 2, \dots, k-1$, where $\gamma = \gamma(\mathcal{H}) > 1$, then

$$\chi(\mathcal{H}) = O\left(\left(\frac{\Delta}{\log \gamma}\right)^{\frac{1}{k-1}}\right). \quad (10)$$

To apply [Lemma 7](#), for any given point set $P \subseteq \mathbb{R}^2$ we define a 3-uniform hypergraph on P , whose edges are the collinear triples in P . The following result of Payne and Wood [27] gives an upper bound on the number of edges in this auxiliary hypergraph.

Lemma 8 (Payne–Wood [27, Lemma 2.1]). Let $P \subseteq \mathbb{R}^2$ be a set of n points with at most s points collinear. Then the number of collinear triples in P is at most

$$O(n^2 \log s + ns^2). \quad (11)$$

In our proof, we want to sparsify our auxiliary hypergraph so that its maximum degree is not too large. Meanwhile, we do not want its maximum degree to be too small. To this end, we prove the following lemma, which allows us to boost the maximum degree of our auxiliary hypergraph without increasing the maximum 2-degree too much.

Lemma 9. Let $n, s \in \mathbb{N}$ with $n \geq s \geq 2$. There exists a 3-uniform hypergraph $\mathcal{H}$ on n vertices, such that

$$\Delta(\mathcal{H}) = \Theta(n \log s + s^2) \quad \text{and} \quad \Delta_2(\mathcal{H}) = O(s). \quad (12)$$

Proof of Lemma 9. We may assume that n is sufficiently large and $n = 2m + 1$. Let

$$q = \Theta(\log s + s^2/n) \leq m \quad (13)$$

and let $S_1, \dots, S_q$ be pairwise disjoint perfect matchings in a clique on the vertex set $[2m]$. Let $\mathcal{H}$ be a 3-uniform hypergraph on the vertex set $[n]$, whose edge set is

$$\left\{ \{n, x, y\} : \{x, y\} \in \bigcup_{i=1}^q S_i \right\}. \quad (14)$$

Then $\Delta(\mathcal{H})$ is equal to the degree of vertex n , namely, $\Delta(\mathcal{H}) = mq = \Theta(n \log s + s^2)$. For any two fixed vertices, the number of edges containing them is at most the number of perfect matchings, hence, $\Delta_2(\mathcal{H}) \leq q = O(s)$. $\square$

2.1. Proof of Proposition 2

Suppose that n is sufficiently large and $s \ll \sqrt{n \log n}$. Let $P \subseteq \mathbb{R}^2$ be an arbitrary set of n points with at most s points collinear. By Lemma 8 there are at most $O(n^2 \log s)$ collinear triples in P . Let $\tilde{P} \subseteq P$ be a subset of P obtained by only retaining $|P|/2$ points that are contained in the least number of collinear triples. Now every point in $\tilde{P}$ is contained in at most $O(|\tilde{P}| \log s)$ collinear triples. Let $\mathcal{H} = \mathcal{H}(\tilde{P})$ be a 3-uniform hypergraph on the vertex set $\tilde{P}$, whose edges are the collinear triples in $\tilde{P}$. As there is a one-to-one correspondence between the general position subsets of $\tilde{P}$ and the independent sets in $\mathcal{H}$, it suffices to show that

$$\alpha(\mathcal{H}) = \Omega \left(\sqrt{\frac{n \log \frac{n \log s}{s^2}}{\log s}} \right). \quad (15)$$

Recall that

$$\Delta(\mathcal{H}) = O(|\tilde{P}| \log s). \quad (16)$$

Moreover, since $\tilde{P}$ does not contain a collinear $(s+1)$ -tuple, any pair of points appears in at most $s-2$ collinear triples. This gives

$$\Delta_2(\mathcal{H}) \leq s-2. \quad (17)$$

By Lemma 9 there exists a 3-uniform hypergraph $\mathcal{H}'$ on the vertex set $\tilde{P}$ satisfying

$$\Delta(\mathcal{H}') = \Theta(|\tilde{P}| \log s) \quad \text{and} \quad \Delta_2(\mathcal{H}') = O(s). \quad (18)$$

Let $\mathcal{H}''$ be the union of $\mathcal{H}$ and $\mathcal{H}'$, that is, a 3-uniform hypergraph with the vertex set $\tilde{P}$ and the edge set $E(\mathcal{H}) \cup E(\mathcal{H}')$. Then we have, letting $\Delta = \Delta(\mathcal{H}'')$,

$$\Delta = \Theta(|\tilde{P}| \log s) \quad \text{and} \quad \Delta_2(\mathcal{H}'') = O(s) \leq \sqrt{\Delta}/\gamma, \quad (19)$$

where $\gamma = \Theta \left(\sqrt{\frac{|\tilde{P}| \log s}{s^2}} \right) > 1$. Since $\mathcal{H}$ is a subgraph of $\mathcal{H}''$, Lemma 7 implies that

$$\chi(\mathcal{H}) \leq \chi(\mathcal{H}'') = O \left(\sqrt{\frac{\Delta}{\log \gamma}} \right) = O \left(\sqrt{\frac{|\tilde{P}| \log s}{\log \frac{|\tilde{P}| \log s}{s^2}}} \right). \quad (20)$$

We complete the proof by noting that $\alpha(\mathcal{H}) \geq v(\mathcal{H})/\chi(\mathcal{H})$ and $v(\mathcal{H}) = |\tilde{P}| = n/2$. $\square$

3. Selecting a general position subset: upper bounds

In this section we prove Theorem 1, i.e.,

$$f(n, s) = \begin{cases} O(n^{5/6+o(1)}/\sqrt{s}) & \text{for } 3 \leq s \leq n^{1/3}. \\ O(n/s) & \text{for } n^{1/3} \leq s \leq n. \end{cases} \quad \begin{array}{l} \text{part (i)} \\ \text{part (ii)} \end{array} \quad (21)$$

3.1. Proof of Theorem 1 part (ii)

Consider the 2-dimensional grid $[n]^2$. Let $\mathcal{L}$ be the set of lines that go through at least two points in $[n]^2$. Note that $|\mathcal{L}| \leq n^4$ and every line in $\mathcal{L}$ contains at most n points in $[n]^2$. Moreover, by the pigeonhole principle, every set of $2n + 1$ points in $[n]^2$ contains a collinear triple. Let P be a random subset of $[n]^2$ by retaining each point independently with probability $n^{\alpha-1}/2$, where $\alpha \in [1/2, 1]$. Then by Chernoff's bound, see, e.g., [26, Chap. 4], it holds that $|P| = \Theta(n^{1+\alpha})$ with high probability. Furthermore, using the union bound and Chernoff's bound we have

$$\begin{aligned} \mathbb{P}(P \text{ contains no collinear } n^\alpha\text{-tuple}) &= 1 - \mathbb{P}(\exists \ell \in \mathcal{L} : |P \cap \ell| \geq n^\alpha) \\ &\geq 1 - \sum_{\ell \in \mathcal{L}} \mathbb{P}\left(|P \cap \ell| \geq \left(1 + \frac{n^\alpha}{2\mathbb{E}(|P \cap \ell|)}\right) \mathbb{E}(|P \cap \ell|)\right) \\ &\geq 1 - n^4 \exp(-n^\alpha/10), \end{aligned} \quad (22)$$

which tends to 1 as $n \rightarrow \infty$. Namely, there exists a point set $P \subseteq [n]^2$ of size $\Theta(n^{1+\alpha})$, so that P contains no collinear n^α -tuple and every subset of P of size $2n + 1$ contains a collinear triple. After scaling we obtain

$$f(n, n^{\frac{\alpha}{1+\alpha}}) = O(n^{\frac{1}{1+\alpha}}) \quad \text{for} \quad \frac{1}{2} \leq \alpha \leq 1, \quad (23)$$

or equivalently

$$f(n, s) = O(n/s) \quad \text{for} \quad n^{1/3} \leq s \leq n^{1/2}. \quad (24)$$

For $n^{1/2} \leq s \leq n$, we simply consider the grid $[s] \times [n/s]$. Obviously, this grid is a set of n points with at most s points on the same line. Furthermore, by the pigeonhole principle any $2n/s + 1$ points contain a collinear triple. Hence, we have $f(n, s) = O(n/s)$. $\square$

3.2. Proof of Theorem 1 part (i)

Recall that $3 \leq s \leq n^{1/3}$. For this proof we shall use the hypergraph container method. This method was introduced independently by Balogh, Morris, and Samotij [2], and by Saxton and Thomason [35]. Here we use a version by Saxton and Thomason. For a hypergraph $\mathcal{H}$ and $C \subseteq V(\mathcal{H})$, let $\mathcal{H}[C]$ denote the subhypergraph of $\mathcal{H}$ induced by C .

Lemma 10 (Saxton–Thomason [35, Corollary 3.6]). *Let $\mathcal{H}$ be an N -vertex k -uniform hypergraph, and $0 < \varepsilon, \tau < 1/2$. Define*

$$\Delta(\mathcal{H}, \tau) := 2^{\binom{k}{2}-1} \sum_{j=2}^k \frac{\Delta_j(\mathcal{H})}{d\tau^{j-1}2^{\binom{j-1}{2}}}, \quad (25)$$

where $d = d(\mathcal{H})$ denotes the average degree of $\mathcal{H}$. Suppose that $\tau < 1/(1000 \cdot k \cdot k!^2)$ and $\Delta(\mathcal{H}, \tau) \leq \varepsilon/(100k!)$. Then there exists $c = c(k) \leq 1000 \cdot k \cdot k!^3$ and a collection $\mathcal{C} \subseteq 2^{V(\mathcal{H})}$, such that

- (1) every independent set in $\mathcal{H}$ is a subset of some $C \in \mathcal{C}$;
- (2) for every $C \in \mathcal{C}$, $e(\mathcal{H}[C]) \leq \varepsilon e(\mathcal{H})$;
- (3) $\log|C| \leq cN\tau \log(1/\varepsilon) \log(1/\tau)$.

Given $S \subseteq [n]^3$, we define $\mathcal{H}_S$ as the 3-uniform hypergraph on the vertex set S , whose edges are the collinear triples spanned by S . A supersaturation result by Balogh and Solymosi [3], see also Keller and Smorodinsky [24], shows that $e(\mathcal{H}_S)$ is large whenever $|S| \geq n^{2+c}$ for any fixed $c > 0$. We further prove that in this case not only $\mathcal{H}_S$ contains many edges, but also there is a spanning subhypergraph $\mathcal{H}' \subseteq \mathcal{H}_S$, i.e., a subhypergraph on the same vertex set S , such that $\mathcal{H}'$ has almost as many edges as $\mathcal{H}_S$ and the co-degrees of $\mathcal{H}'$ are well-bounded.

Lemma 11. Fix any $0 < f < 1$. Let $n \in \mathbb{N}$ be sufficiently large and $0 \leq x \leq 1 - f$. For every set $S \subseteq [n]^3$ with $|S| = n^{3-x}$, there exists a spanning subhypergraph $\mathcal{H}' \subseteq \mathcal{H}_S$ such that

$$e(\mathcal{H}') \geq \frac{n^{6-4x}}{10^7 \log n}, \quad \Delta(\mathcal{H}') \leq \frac{n^{3-3x}}{10^3 f \log n}, \quad \Delta_2(\mathcal{H}') \leq 1. \quad (26)$$

Proof of Lemma 11. Set $t = 100n^x$ and let

$$U := \{(a, b, c) \in \mathbb{Z}^3 : 1 \leq a \leq 2n/t, -n \leq b, c \leq n\} \quad (27)$$

and

$$V := \{(a, b, c) \in \mathbb{N}^3 : n/t \leq a \leq 2n/t, a > \max\{b, c\}, a \text{ is a prime}\}. \quad (28)$$

We define a family $\mathcal{L} = \mathcal{L}(t)$ of lines, whose starting points are in U and whose directions are in V . We regard $\mathcal{L}$ as a family of geometric lines, so repeated descriptions of the same line are identified. Note that these lines are defined in $\mathbb{R}^3$ and we are interested in their intersection with S . It is easy to see that every line in $\mathcal{L}$ passes through at most t points of $[n]^3$. Since the number of primes smaller than m is at least $m/\log m$ and at most $1.1m/\log m$ when m is sufficiently large (see [33, Theorem 1]), we have that

$$\frac{n^3}{t^3 \log(n/t)} \leq |V| \leq \frac{100n^3}{t^3 \log(n/t)}. \quad (29)$$

Moreover, the family $\mathcal{L}$ has the following properties.

Claim 12. For each $(a, b, c) \in V$,

- (1) the lines in $\mathcal{L}$ with direction (a, b, c) cover the points in $[n]^3$;
- (2) the number of lines in $\mathcal{L}$ with direction (a, b, c) is at most $10n^3/t$;
- (3) every point in $[n]^3$ is contained in at most $\frac{100n^3}{t^3 \log(n/t)}$ lines from $\mathcal{L}$.

Proof of Claim 12. Fix an arbitrary $(a, b, c) \in V$. For any $(p_1, p_2, p_3) \in [n]^3$, let $q = \lfloor \frac{p_1-1}{a} \rfloor$. Then

$$1 \leq p_1 - qa \leq a \leq \frac{2n}{t}.$$

Moreover, since $1 \leq b, c < a$ and $qa \leq p_1 - 1 \leq n - 1$, we have

$$2 - n \leq p_2 - qb \leq n \quad \text{and} \quad 2 - n \leq p_3 - qc \leq n.$$

Thus, the point (p_1, p_2, p_3) is contained in the line in $\mathcal{L}$ with direction (a, b, c) that goes through $(p_1 - qa, p_2 - qb, p_3 - qc) \in U$. This proves item (1).

For item (2), since each point of U determines exactly one line with direction (a, b, c) , the number of lines in $\mathcal{L}$ with direction (a, b, c) is at most

$$|U| \leq \frac{2n}{t} \cdot (2n + 1)^2 \leq \frac{10n^3}{t}.$$

Finally, for each $v \in V$ there is at most one line with direction v through any fixed point of $[n]^3$. Therefore, by (29) every point of $[n]^3$ is contained in at most

$$|V| \leq \frac{100n^3}{t^3 \log(n/t)}$$

lines from $\mathcal{L}$, proving item (3). $\square$

We now construct the spanning subhypergraph $\mathcal{H}' \subseteq \mathcal{H}_S$ as follows. For each line $\ell \in \mathcal{L}$ we place an arbitrary matching (each matching edge is a collinear triple) on the vertices $\ell \cap S$ covering

all but at most 2 vertices. The edge set of $\mathcal{H}'$ is the set of all the matching edges. Then it holds that $\Delta_2(\mathcal{H}') \leq 1$ and by Claim 12 item (3),

$$\Delta(\mathcal{H}') \leq \frac{100n^3}{t^3 \log(n/t)} \leq \frac{n^{3-3x}}{10^3 f \log n}. \quad (30)$$

Let $\mathcal{L}_v \subseteq \mathcal{L}$ be the family of lines with direction $v \in V$. By Claim 12 item (2), $|\mathcal{L}_v| \leq 10n^3/t$. Together with Claim 12 item (1) we have that the number T_v of edges in $\mathcal{H}'$ coming from triples on lines in $\mathcal{L}_v$ is at least

$$T_v = \sum_{\ell \in \mathcal{L}_v} e(\mathcal{H}'[\ell \cap S]) = \sum_{\ell \in \mathcal{L}_v} \left\lfloor \frac{|\ell \cap S|}{3} \right\rfloor \geq \frac{|S|}{3} - |\mathcal{L}_v| \geq \frac{n^{3-x}}{5}. \quad (31)$$

Since the directions in V are pairwise non-parallel, we have $\mathcal{L}_v \cap \mathcal{L}_{v'} = \emptyset$ for any distinct $v, v' \in V$. Then, since $|V| \geq \frac{n^3}{t^3 \log(n/t)}$ and $t = 100n^x$, we obtain

$$e(\mathcal{H}') = \sum_{v \in V} T_v \geq \frac{n^3}{t^3 \log(n/t)} \cdot \frac{n^{3-x}}{5} \geq \frac{n^{6-4x}}{10^7 \log n}. \quad (32)$$

This completes the proof of Lemma 11. $\square$

Our proof of Theorem 1 part (i) extends the framework of that by Balogh and Solymosi [3], which consists of three steps. We start with the 3-dimensional grid $[n]^3$. In the first step, we iteratively apply the hypergraph container lemma on an auxiliary hypergraph until we obtain a collection $\mathcal{C} \subseteq 2^{[n]^3}$ of small containers, such that every subset of $[n]^3$ without collinear triples is a subset of one of these containers. In the second step, we take a random subset of $[n]^3$ and show that with high probability this set contains no collinear s -tuple and also does not contain any large subset without collinear triples. Note that the collection $\mathcal{C}$ from the first step will be used here to conquer the union bound. In the last step, we project this set from $[n]^3$ into the plane and obtain an upper bound construction. The main novelty in our proof is that in each iteration we do not apply the hypergraph container lemma on $\mathcal{H}_S$ with $S \subseteq [n]^3$ but on a sparsened auxiliary hypergraph $\mathcal{H}' \subseteq \mathcal{H}_S$ given by Lemma 11. The merit of this step is that the co-degrees of $\mathcal{H}'$ are well-bounded, which enables us to get a better collection of containers and ultimately obtain better upper bounds.

Proof of Theorem 1 part (i). Note that for $3 \leq s \leq \log n$, by the monotonicity of the function $f(n, s)$ we have that

$$f(n, s) \leq f(n, 3) \leq n^{5/6+o(1)} = O(n^{5/6+o(1)}/\sqrt{s}). \quad (33)$$

Hence, it suffices to prove the claimed upper bounds for $\log n \leq s \leq n^{1/3}$. Fix any $0 < f < 1/100$ and assume that $n \in \mathbb{N}$ is sufficiently large. Let

$$\alpha \in \left[\frac{2 \log \log n}{\log n}, 1 \right] \quad \text{and} \quad \gamma = \frac{2\alpha + 1}{3} \leq 1. \quad (34)$$

For every subset $S \subseteq [n]^3$ of size n^{3-x} with $0 \leq x \leq \gamma - f$, we can do the following. Let $\mathcal{H}_S$ be the 3-uniform hypergraph on the vertex set S , whose edges are the collinear triples spanned by S . Let $\mathcal{H}' \subseteq \mathcal{H}_S$ be the spanning subhypergraph guaranteed by Lemma 11. Since $e(\mathcal{H}') \geq \frac{n^{6-4x}}{10^7 \log n}$, the average degree $d = d(\mathcal{H}')$ of $\mathcal{H}'$ is at least $\frac{n^{3-3x}}{10^7 \log n}$. Moreover, we have that

$$\Delta_2(\mathcal{H}') \leq 1 \quad \text{and} \quad \Delta_3(\mathcal{H}') \leq 1. \quad (35)$$

Let $\tau = n^{x+\gamma/2-3/2}$. Since $x \leq \gamma - f$ and $\gamma \leq 1$, it holds that $\tau \leq n^{-f} \ll 1$. Furthermore,

$$\begin{aligned} \Delta(\mathcal{H}', \tau) &= \frac{4\Delta_2(\mathcal{H}')}{d\tau} + \frac{2\Delta_3(\mathcal{H}')}{d\tau^2} \leq \frac{4}{\frac{n^{3-3x}}{10^7 \log n} \cdot n^{x+\frac{\gamma}{2}-\frac{3}{2}}} + \frac{2}{\frac{n^{3-3x}}{10^7 \log n} \cdot n^{2x+\gamma-3}} \\ &\leq \frac{10^8 \log n}{n^{2f}} + \frac{10^8 \log n}{n^f} \ll \varepsilon, \end{aligned} \quad (36)$$

where $\varepsilon = n^{-f/2} < 1/2$. Then by Lemma 10 there exists a collection $\mathcal{C} \subseteq 2^S$, such that

- (1) every independent set in $\mathcal{H}'$ is a subset of some $C \in \mathcal{C}$;
- (2) for every $C \in \mathcal{C}$, $e(\mathcal{H}'[C]) \leq n^{-f/2} e(\mathcal{H}')$;
- (3) $|\mathcal{C}| \leq \exp(10^7 n^{(3+\gamma)/2} (\log n)^2)$.

Since $\mathcal{H}' \subseteq \mathcal{H}_S$, every independent set in $\mathcal{H}_S$ is a subset of some $C \in \mathcal{C}$. By item (1) we have for each container $C \in \mathcal{C}$ that

$$e(\mathcal{H}') - |S \setminus C| \cdot \Delta(\mathcal{H}') \leq e(\mathcal{H}'[C]) \leq n^{-f/2} e(\mathcal{H}'). \quad (37)$$

Recall that $|S| = n^{3-x}$ and by Lemma 11, $e(\mathcal{H}') \geq \frac{n^{6-4x}}{10^7 \log n}$ and $\Delta(\mathcal{H}') \leq \frac{n^{3-3x}}{10^3 f \log n}$. After rearranging (37) we get

$$|S \setminus C| \geq \frac{e(\mathcal{H}')}{\Delta(\mathcal{H}')} (1 - n^{-f/2}) \geq \frac{f n^{3-x}}{10^5} = \frac{f}{10^5} \cdot |S|, \quad (38)$$

which implies that $|C| \leq (1 - 10^{-5}f)|S|$.

We apply the above procedure first on $S_0 = [n]^3$ and obtain a collection of containers. Then, we iterate the procedure on the containers until every container has size at most $n^{3-\gamma+f}$. This process takes at most $(\log n)^2$ iterations. Indeed, in each iteration we shrink the container size by a $(1 - 10^{-5}f)$ factor. Since f is fixed and n is sufficiently large, we have $(1 - 10^{-5}f)^{(\log n)^2} < 1/n$. Namely, after at most $(\log n)^2$ iterations the size of every container is at most $n^{3-\gamma+f}$.

Note that any subset of $[n]^3$ with no collinear triple is an independent set in $\mathcal{H}_{S_0}$ and thus is contained in one of the final containers. Overall, we obtain a collection $\mathcal{C}$ of

$$\exp(10^7 n^{(3+\gamma)/2} (\log n)^4) \leq \exp(n^{(3+\gamma)/2+2f}) \quad (39)$$

containers, such that every subset of $[n]^3$ without collinear triples is contained in one of these containers. Moreover, each container has size at most $n^{3-\gamma+f}$.

Let P be a random subset of $[n]^3$ by retaining each grid point independently with probability

$$p = \frac{1}{2} n^{\alpha-1} = \frac{1}{2} n^{(3\gamma-3)/2}. \quad (40)$$

Then by Chernoff's bound we have that $|P| = \Theta(n^{2+\alpha})$ with high probability. Applying the union bound and Chernoff's bound as in (22) we also have that P contains no collinear n^α -tuple with high probability. Furthermore, we say that a subset of P is *bad* if it has size $n^{(3+\gamma)/2+2f}$ and spans no collinear triple. When $\gamma > 1 - 2f/3$, since each container has size at most $n^{3-\gamma+f} < n^{(3+\gamma)/2+2f}$, P contains no bad subset. When $0 \leq \gamma \leq 1 - 2f/3$, the expected number of bad subsets of P is at most

$$\sum_{C \in \mathcal{C}} \binom{|C|}{n^{(3+\gamma)/2+2f}} p^{n^{(3+\gamma)/2+2f}} \leq \exp(n^{(3+\gamma)/2+2f}) \left(\frac{n^{3-\gamma+f}}{n^{(3+\gamma)/2+2f}} \right)^{n^{(3\gamma-3)/2}} \ll 1. \quad (41)$$

By the first moment method, we have that with high probability P does not contain any bad subset. Consequently, there exists a set $P \subseteq [n]^3$ of size $\Theta(n^{2+\alpha})$, so that P contains no collinear n^α -tuple and every subset of P of size $n^{(3+\gamma)/2+2f}$ spans a collinear triple. We can project P into $\mathbb{R}^2$ in such a way that all the collinear relations remain the same. After scaling we obtain

$$f(n, n^{\frac{\alpha}{2+\alpha}}) = O\left(n^{\frac{(3+\gamma)/2+2f}{2+\alpha}}\right) = O\left(n^{\frac{5+\alpha+6f}{6+3\alpha}}\right) \quad \text{for } \alpha \in \left[\frac{2 \log \log n}{\log n}, 1\right], \quad (42)$$

which implies

$$f(n, n^\beta) = O\left(n^{\frac{5-3\beta+6f}{6}}\right) \quad \text{for } \beta \in \left[\frac{\log \log n}{\log n}, \frac{1}{3}\right]. \quad (43)$$

Further, letting $s = n^\beta$,

$$f(n, s) = O\left(n^{\frac{5+6f}{6}} / \sqrt{s}\right) \quad \text{for } \log n \leq s \leq n^{1/3}. \quad (44)$$

Since f can be taken arbitrarily small in the beginning, we obtain the claimed upper bound. $\square$

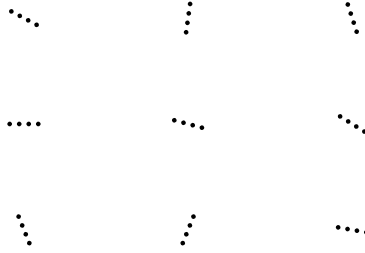


Fig. 2. Upper bound construction for Theorem 4 when $n = 36$ and $s = 4$.

4. Selecting a monotone general position subset

In this section we prove Theorems 4 and 5.

4.1. Proof of Theorem 4

Lower bound. Let $P \subseteq \mathbb{R}^2$ be an arbitrary set of n points without collinear $(s + 1)$ -tuples. First, we have a general position subset of P of size $f(n, s)$. Then, by the Erdős–Szekeres theorem we can extract a monotone position subset from it, of size at least $\sqrt{f(n, s)}$. Recall the lower bounds on $f(n, s)$ from (2) and Proposition 2, we obtain

$$g(n, s) = \begin{cases} \Omega \left(\left(\frac{n \log \frac{n \log s}{s^2}}{\log s} \right)^{1/4} \right) & \text{for } 3 \leq s \ll \sqrt{n \log n}, \\ \Omega(\sqrt{n/s}) & \text{for } s = \Omega(\sqrt{n \log n}). \end{cases} \quad (45)$$

Upper bound. Let $k = \sqrt{n/s}$, and assume without loss of generality that $k \in \mathbb{N}$. Consider a $k \times k$ grid and replace each grid point by a circle of diameter $\epsilon > 0$, where ϵ is chosen to be sufficiently small. Slightly perturb these circles so that no line intersects with three of them. Then, we place a collinear s -tuple spanning a line segment of length ϵ inside every circle and slightly rotate these line segments so that the extension of each line segment does not intersect with other line segments. See Fig. 2 for an illustration.

In this construction there are n points with at most s collinear points. Let P denote the resulting set and let $A \subseteq P$ be a monotone general position subset. Note that

- (i) A contains at most two points from each collinear s -tuple;
- (ii) the largest monotone subset of the $k \times k$ grid that we started with has at most $2k - 1$ points.

Thus, $|A| \leq 4k - 2 = O(\sqrt{n/s})$ as claimed. $\square$

4.2. Proof of Theorem 5

Lower bound. According to a classical result of Jarník [23], a strictly convex curve of length L in the plane can be incident to at most $3(2\pi)^{-1/3} L^{2/3} (1 + o(1))$ lattice points and this estimate is the best possible. In particular, Jarník's polygon in the $\sqrt{n} \times \sqrt{n}$ grid matches this bound; see also [4]. One can then select a monotone arc of this polygon with $\Omega(n^{1/3})$ vertices. Since the polygon is convex, the arc contains no collinear triple. We next derive a better lower bound, namely $\Omega((n/\log n)^{2/5})$.

Let $n \in \mathbb{N}$ be sufficiently large and assume without loss of generality that $\sqrt{n} = 2m + 1$ with $m \in \mathbb{N}$. Let $x = x(n) \geq 1$ with $x \ll \sqrt{n}$ to be determined later. Let U be an axis-aligned square of side length $2m$ containing a $\sqrt{n} \times \sqrt{n}$ grid G . Let o be the center point of the square U and let A be the annulus formed by two concentric circles of radii m and $m - x$, centered at o .

Gauss' circle problem asks to determine the number of points of the integer lattice $\mathbb{Z}^2$ contained in a circle of radius R around the origin. A result by Gauss says that this number is $\pi R^2 + O(R)$, see, e.g., [22]. Thus, $|A \cap G| = \Theta(x\sqrt{n})$. Next we show some important properties of the annulus A .

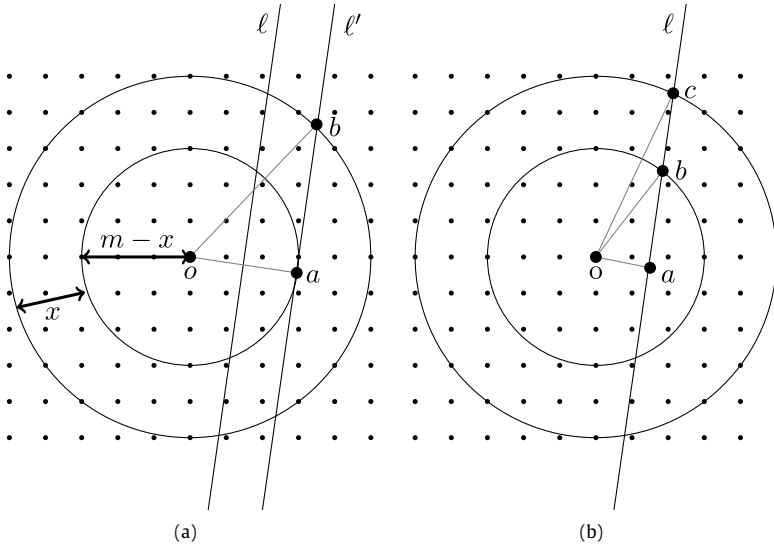


Fig. 3. Part (a) illustrates the two parallel lines ℓ, ℓ' from the proof of Claim 14. Part (b) illustrates the points a, b, c on the line ℓ from the proof of Claim 15.

Lemma 13. *The number of collinear triples in $A \cap G$ is $O(x^{5/2}n^{3/4} \log n)$.*

Before proving Lemma 13, we note that its bound is best possible up to a constant factor. More precisely, for $x \ll \sqrt{n}$, the number of collinear triples in $A \cap G$ is $\Theta(x^{5/2}n^{3/4} \log n)$. A proof of the matching lower bound is included in the appendix.

Proof of Lemma 13. Fix any slope $(q, r) \in \mathbb{Z}^2$ with q and r being coprime. Let $\mathcal{L}(q, r)$ be the family of lines with slope (q, r) that are incident to at least two points of $A \cap G$, and $T(q, r)$ be the number of collinear triples in $A \cap G$ coming from lines $\ell \in \mathcal{L}(q, r)$. Due to the symmetry of A we may assume that $0 \leq q \leq r$. For a set I of pairwise disjoint intervals on a line, let $\mu(I)$ denote their total length. Note that each line $\ell \in \mathcal{L}(q, r)$ contains at most $O(\mu(\ell \cap A)/r)$ points in $A \cap G$. Hence, we first bound $\mu(\ell \cap A)$ from above.

Claim 14. *Let ℓ be an arbitrary line. Then $\mu(\ell \cap A) \leq 2x^{1/2}n^{1/4}$.*

Proof of Claim 14. Let ℓ' be the line which is parallel to ℓ and a tangent line to the inner circle of A . Since $\mu(\ell \cap A) \leq \mu(\ell' \cap A)$, it remains to show that $\mu(\ell' \cap A) \leq 2x^{1/2}n^{1/4}$. Let a denote the tangent point and let b denote one of the intersection points of ℓ' with the outer circle. See Fig. 3 (a) for an illustration.

Recall that the radii of the inner circle and the outer circle are $m-x$ and m , namely, $\mu(oa) = m-x$ and $\mu(ob) = m$. A straightforward trigonometric calculation shows that

$$\mu(\ell' \cap A) = 2\mu(ab) = 2\sqrt{\mu(ob)^2 - \mu(oa)^2} \leq 2\sqrt{2mx} \leq 2x^{1/2}n^{1/4}, \quad (46)$$

as desired. $\square$

Claim 15. *Let $t = t(n) \geq 6$ with $tx \leq m$. If the distance $\text{dist}(\ell, o)$ between a line $\ell \in \mathcal{L}(q, r)$ and the center point o is $m - tx$, then $\mu(\ell \cap A) \leq 4\sqrt{mx/t}$.*

Proof of Claim 15. Let $a \in \ell$ be the point such that the segment oa is orthogonal to ℓ . Since $t > 1$, the line ℓ intersects the annulus A in two segments. Let b and c be the intersection points of ℓ with

the inner and outer circles on one of the sides. See Fig. 3 (b) for an illustration. It follows that

$$\mu(oa) = m - tx, \quad \mu(ob) = m - x, \quad \mu(oc) = m \quad (47)$$

and by symmetry that $\mu(\ell \cap A) = 2\mu(bc)$. According to the Pythagorean formula we have

$$\mu(ab)^2 = \mu(ob)^2 - \mu(oa)^2 = 2mtx - t^2x^2 - 2mx + x^2, \quad \mu(ac)^2 = \mu(oc)^2 - \mu(oa)^2 = 2mtx - t^2x^2. \quad (48)$$

Since $\mu(ab)^2 + 2\mu(ab)\mu(bc) \leq (\mu(ab) + \mu(bc))^2 = \mu(ac)^2$, it holds that

$$\mu(bc) \leq \frac{\mu(ac)^2 - \mu(ab)^2}{2\mu(ab)} = \frac{2mx - x^2}{2\sqrt{2mtx - t^2x^2 - 2mx + x^2}} \leq \frac{mx}{\sqrt{2mtx - t^2x^2 - 2mx}} \leq 2\sqrt{\frac{mx}{t}}. \quad (49)$$

Thus, $|\mu(\ell \cap A)| \leq 2\mu(bc) \leq 4\sqrt{mx/t}$. $\square$

Next we want to bound the number of lines $\ell \in \mathcal{L}(q, r)$ with given $\text{dist}(\ell, o)$.

Claim 16. Let $0 \leq t \leq m/x$ and $\varepsilon > 0$. Then the number of lines $\ell \in \mathcal{L}(q, r)$ with $\text{dist}(\ell, o) \in [m - (t + \varepsilon)x, m - tx]$ is at most $O(\varepsilon xr)$.

Proof of Claim 16. It suffices to prove that the distance between any two distinct lines in $\mathcal{L}(q, r)$ is $\Omega(1/r)$, whereupon a simple calculation shows that the number of lines $\ell \in \mathcal{L}(q, r)$ with $\text{dist}(\ell, o) \in [m - (t + \varepsilon)x, m - tx]$ is $O(\varepsilon xr)$.

Let $\ell_1, \ell_2 \in \mathcal{L}(q, r)$ be two distinct lines and let $\text{dist}(\ell_1, \ell_2)$ denote the distance between ℓ_1 and ℓ_2 . Since ℓ_1 and ℓ_2 both have the same slope (q, r) and each passes through at least two points of $A \cap G$ by definition, we can write them as

$$\ell_1 : rx - qy + k_1 = 0 \quad \text{and} \quad \ell_2 : rx - qy + k_2 = 0, \quad (50)$$

where $k_1, k_2 \in \mathbb{Z}$ and $k_1 \neq k_2$. Then the formula of the distance between two parallel lines in the plane gives that

$$\text{dist}(\ell_1, \ell_2) = \frac{|k_1 - k_2|}{\sqrt{r^2 + q^2}} \geq \frac{1}{\sqrt{r^2 + q^2}} = \Omega\left(\frac{1}{r}\right), \quad (51)$$

where in the last step we used that $0 \leq q \leq r$. $\square$

Now we write $T(q, r)$ as follows and compute the two parts separately.

$$\begin{aligned} T(q, r) &= \sum_{\ell \in \mathcal{L}(q, r)} \binom{|\ell \cap A \cap G|}{3} \\ &\leq \sum_{\substack{\ell \in \mathcal{L}(q, r) \\ \text{dist}(\ell, o) \in [m-6x, m]}} O\left(\frac{\mu(\ell \cap A)^3}{r^3}\right) + \sum_{\substack{\ell \in \mathcal{L}(q, r) \\ \text{dist}(\ell, o) \in [0, m-6x]}} O\left(\frac{\mu(\ell \cap A)^3}{r^3}\right), \end{aligned} \quad (52)$$

where we used in the last step that $|\ell \cap A \cap G| = O(\mu(\ell \cap A)/r)$. By Claim 16 the number of lines $\ell \in \mathcal{L}(q, r)$ with $\text{dist}(\ell, o) \in [m - 6x, m]$ is at most $O(xr)$ and by Claim 14 $\mu(\ell \cap A) \leq 2x^{1/2}n^{1/4}$ holds for all $\ell \in \mathcal{L}(q, r)$. Hence, we have that

$$\sum_{\substack{\ell \in \mathcal{L}(q, r) \\ \text{dist}(\ell, o) \in [m-6x, m]}} O\left(\frac{\mu(\ell \cap A)^3}{r^3}\right) \leq O(xr) \cdot O\left(\frac{x^{3/2}n^{3/4}}{r^3}\right) = O\left(\frac{x^{5/2}n^{3/4}}{r^2}\right). \quad (53)$$

For any $0 \leq t \leq m/x$ and $\varepsilon > 0$, by Claim 16 the number of lines $\ell \in \mathcal{L}(q, r)$ with $\text{dist}(\ell, o) \in [m - (t + \varepsilon)x, m - tx]$ is at most $O(\varepsilon xr)$ and we have that $\mu(\ell \cap A) \leq 4\sqrt{mx/t}$ for every such line ℓ by Claim 15. Therefore,

$$\begin{aligned} \sum_{\substack{\ell \in \mathcal{L}(q, r) \\ \text{dist}(\ell, o) \in [0, m-6x]}} o \left(\frac{\mu(\ell \cap A)^3}{r^3} \right) &\leq \lim_{\varepsilon \rightarrow 0} \sum_{t \in \{6, 6+\varepsilon, 6+2\varepsilon, \dots, m/x\}} \sum_{\substack{\ell \in \mathcal{L}(q, r) \\ \text{dist}(\ell, o) \in [m-(t+\varepsilon)x, m-tx]}} o \left(\frac{\mu(\ell \cap A)^3}{r^3} \right) \\ &\leq \lim_{\varepsilon \rightarrow 0} \sum_{t \in \{6, 6+\varepsilon, 6+2\varepsilon, \dots, m/x\}} O(\varepsilon xr) \cdot o \left(\frac{m^{3/2} x^{3/2}}{t^{3/2} r^3} \right) \\ &= \int_{t=6}^{m/x} O \left(\frac{m^{3/2} x^{5/2}}{t^{3/2} r^2} \right) dt = O \left(\frac{m^{3/2} x^{5/2}}{r^2} \right) \cdot \int_{t=6}^{m/x} t^{-3/2} dt \\ &= O \left(\frac{x^{5/2} n^{3/4}}{r^2} \right). \end{aligned} \quad (54)$$

Note that in the calculation of the integral in (54) we used that $x \ll m \approx \sqrt{n}/2$ and thus the last integral in the penultimate line of (54) is $O(1)$. By combining (52), (53) and (54) we obtain

$$T(q, r) = O \left(\frac{x^{5/2} n^{3/4}}{r^2} \right). \quad (55)$$

Now let

$$S = \{(q, r) \in \{-2m, \dots, 2m\} \times \{-2m, \dots, 2m\}, \text{ where } q \text{ and } r \text{ are coprime}\} \quad (56)$$

be the set of all possible slopes of lines that go through at least two points of $A \cap G$. The number of collinear triples in $A \cap G$ is

$$\sum_{(q, r) \in S} T(q, r) \leq 2 \sum_{\substack{(q, r) \in S, \\ 0 \leq q \leq r}} T(q, r) = O \left(\sum_{r=1}^{\sqrt{n}} r \cdot \frac{x^{5/2} n^{3/4}}{r^2} \right) = O(x^{5/2} n^{3/4} \log n), \quad (57)$$

completing the proof of Lemma 13. $\square$

Recall that our goal is to select a large general position subset P of $A \cap G$, such that for any $u, v \in P$ with $u_x < v_x$ we have $u_y \leq v_y$. To achieve this, one needs to avoid two types of obstacles, collinear triples and descending pairs. A *descending pair* is a pair (u, v) of points with $u_x < v_x$ and $u_y > v_y$. In Lemma 13 we have bounded the number of collinear triples in $A \cap G$. Next we want to bound the number of descending pairs in a certain section of $A \cap G$.

Let v_1 be the lower right vertex of the square U . Moreover, let v_2 and v_3 be two points on the right and bottom sides of the square U , respectively, with $\angle v_2 o v_3 = 30^\circ$ and $o v_1$ being the bisector. Let Q be the quadrilateral formed by $o v_1 v_2 v_3$ and let $B = A \cap Q$ be a sector of the annulus A . See Fig. 4 for an illustration. Note that $|B \cap G| = \Theta(|A \cap G|) = \Theta(x\sqrt{n})$.

Lemma 17. *The number of descending pairs in $B \cap G$ is $O(x^3 \sqrt{n})$.*

Proof of Lemma 17. Fix any $u \in B \cap G$, and let $v \in B \cap G$ be such that (u, v) forms a descending pair. By the choice of the 30° sector, the angle $\angle u v o$ is at most 60° , yielding that

$$\|v - o\| - \|u - o\| \geq \frac{\|u - v\|}{2}.$$

Since u and v lie in an annulus of width x , we also have $\|v - o\| - \|u - o\| \leq x$ and hence $\|u - v\| \leq 2x$. Therefore, for each fixed u , there are $O(x^2)$ possible choices for v . As $|B \cap G| = O(x\sqrt{n})$, the number of descending pairs is $O(x^2) \cdot O(x\sqrt{n}) = O(x^3 \sqrt{n})$. $\square$

We can now finalize the proof of the lower bound. Take a random subset P of $B \cap G$ by retaining each point independently with probability $p \in (0, 1)$. We have that $\mathbb{E}(|P|) = \Omega(x\sqrt{n}p)$. Moreover,

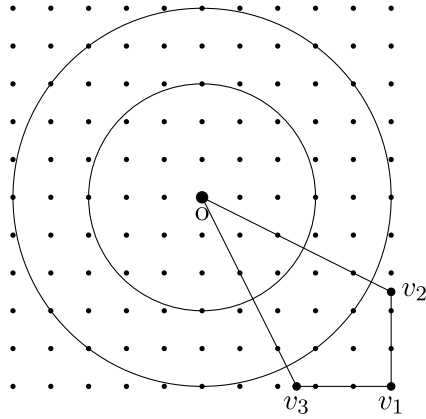


Fig. 4. The sector B of the annulus A .

letting D and T denote the numbers of descending pairs and collinear triples in P , respectively, by Lemmas 13 and 17 it holds that

$$\mathbb{E}(D) = O(p^2 x^3 \sqrt{n}) \quad \text{and} \quad \mathbb{E}(T) = O(p^3 x^{5/2} n^{3/4} \log n). \quad (58)$$

Setting $x = n^{1/10}(\log n)^{2/5}$, we have $\mathbb{E}(|P|) = \Omega(pn^{3/5}(\log n)^{2/5})$ and

$$\mathbb{E}(D) = O(p^2 n^{4/5}(\log n)^{6/5}) \quad \text{and} \quad \mathbb{E}(T) = O(p^3 n \log n). \quad (59)$$

Let $p = \frac{c}{n^{1/5}(\log n)^{4/5}}$ with a sufficiently small constant $c > 0$. Then

$$\mathbb{E}(|P| - D - T) = \mathbb{E}(|P|) - \mathbb{E}(D) - \mathbb{E}(T) \geq \frac{1}{3}\mathbb{E}(|P|) = \Omega\left(\frac{n^{2/5}}{(\log n)^{2/5}}\right). \quad (60)$$

Since each descending pair and collinear triple can be eliminated by deleting one point, there exists a monotone general position subset of $B \cap G$ of size $\Omega((n/\log n)^{2/5})$, proving the lower bound.

Upper bound. Let G be a $\sqrt{n} \times \sqrt{n}$ grid of n points. A result of Pomerance [30] states that for every $k \geq 3$, $b > 0$, there exists a number $m_0(k, b)$ such that if $m \geq m_0(k, b)$ and $\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_m$ are points in $\mathbb{Z}^2$ with $\sum_{i=1}^m |\mathbf{u}_i - \mathbf{u}_{i-1}| \leq bm$, then $\{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_m\}$ contains a collinear k -tuple.

Assume for contradiction that there exists an absolute constant $c > 0$ such that there exists a monotone general position set $P := \{\mathbf{u}_0, \mathbf{u}_1, \dots, \mathbf{u}_m\} \subseteq G$ with $m \geq c\sqrt{n}$. Let $k = 3$, $b = 2/c$ and assume that n is sufficiently large so that $m \geq c\sqrt{n} \geq m_0(k, b)$. Since the set P is monotone, we have

$$\sum_{i=1}^m |\mathbf{u}_i - \mathbf{u}_{i-1}| \leq 2\sqrt{n} \leq \frac{2m}{c} = bm. \quad (61)$$

By the above result (with $k = 3$ and $b = 2/c$), P contains a collinear triple, a contradiction. This concludes the proof of the upper bound and the theorem. $\square$

5. Selecting a subset with pairwise distinct slopes

In this section we prove Theorem 6. First we need an upper bound on the number of trapezoids (each trapezoid is an obstacle in achieving the final goal). A closely related bound was previously obtained by Rudnev and Selig [34] using the Guth–Katz incidence theorem [20]. Our proof below applies under a slightly more general non-collinearity assumption and uses only the Szemerédi–Trotter theorem [37]. Given the elementary nature of the argument, it may also be of independent interest.

Lemma 18. Let $P \subseteq \mathbb{R}^2$ be a set of n points with at most s points collinear. Then the number of trapezoids spanned by P is at most

$$O(n^3 \log s + n^2 s^2). \quad (62)$$

Proof of Lemma 18. Let T denote the number of trapezoids spanned by P . Every trapezoid contains a pair of parallel segments, which determine two parallel lines. Assign every trapezoid to the parallel line that is incident to the largest number of points in P , with ties broken arbitrarily.

For $2 \leq i \leq s$, let k_i be the number of lines containing exactly i points in P . A well-known corollary of the Szemerédi–Trotter theorem [37] states that

$$b_i := \sum_{j \geq i} k_j = O\left(\frac{n^2}{i^3} + \frac{n}{i}\right). \quad (63)$$

Note that the sequence $(b_i)_{i \geq 2}$ is non-increasing. For a given line ℓ with $|\ell \cap P| = i$, let $m = m(\ell)$ be the number of lines determined by P that are parallel to ℓ and incident to at most i points of P . Let $a_1, \dots, a_m \leq i$ be the number of points contained in these lines. Then the number of trapezoids assigned to ℓ is at most

$$\binom{i}{2} \sum_{j=1}^{m(\ell)} \binom{a_j}{2} \leq \binom{i}{2} \frac{n}{i} \binom{i}{2} \leq ni^3. \quad (64)$$

Therefore,

$$T \leq \sum_{i=2}^s \sum_{\ell: |\ell \cap P|=i} ni^3 = n \sum_{i=2}^s i^3 k_i. \quad (65)$$

Note that $i^3 - (i-1)^3 \leq 3i^2$, thus by Abel's partial summation formula we obtain

$$\begin{aligned} \sum_{i=2}^s i^3 k_i &= \sum_{i=2}^s i^3 (b_i - b_{i+1}) = O\left(\sum_{i=2}^s i^2 b_i\right) = O\left(\sum_{i=2}^s i^2 \left(\frac{n^2}{i^3} + \frac{n}{i}\right)\right) \\ &= O\left(n^2 \sum_{i=2}^s \frac{1}{i} + n \sum_{i=2}^s i\right) = O(n^2 \log s + ns^2). \end{aligned} \quad (66)$$

Plugging the above estimate into (65) yields

$$T \leq n \sum_{i=2}^s i^3 k_i = O(n^3 \log s + n^2 s^2). \quad \square \quad (67)$$

5.1. Proof of Theorem 6

Lower bound. Let $P \subseteq \mathbb{R}^2$ be an arbitrary set of n points with at most s points collinear, where $s = O(\sqrt{n})$. We take a random subset $X \subseteq P$ by selecting points independently with probability $p = k/n$. Note that $\mathbb{E}(|X|) = k$.

To select a subset with distinct slopes, one needs to avoid two types of obstacles, collinear triples and trapezoids. Let L and T be the number of collinear triples and trapezoids spanned by P . One obstacle can be eliminated by deleting one point from the subset in the second step. In particular, it suffices to choose k so that

$$\mathbb{E}(Lp^3) \leq k/3 \quad \text{and} \quad \mathbb{E}(Tp^4) \leq k/3, \quad (68)$$

since this implies that the expected number of remaining points after deleting one point from each trapezoid and collinear triple is at least $k - k/3 - k/3 = k/3$. Using the upper bounds in Lemma 8 and Lemma 18, it suffices to choose k so that

$$O(k^2) \leq n/\log s \quad \text{and} \quad O(k^3) \leq n/\log s. \quad (69)$$

Setting $k = c(n/\log s)^{1/3}$ with a sufficiently small constant $c > 0$ satisfies both requirements, proving the lower bound. *Upper bound (i).* Let $P = \{(x, x^2) : x \in [n]\} \subseteq \mathbb{R}^2$. The point set P is of size n and contains no collinear triple. Let $Q \subseteq P$ be a subset of size $|Q| = 2\sqrt{n} + 2$. We show that Q contains a trapezoid.

A *Sidon set* $S \subseteq [n]$ is a subset of the integers such that $a + b = c + d$ for $a, b, c, d \in S$ with $a \neq b, c \neq d$ implies $\{a, b\} = \{c, d\}$; see, e.g., [13, Chap. 6] or [38, Chap. 2]. Note that the largest size of a Sidon set $S \subseteq [n]$ is at most $2\sqrt{n} + 1$, because $\binom{|S|}{2} \leq 2n$. Let $S = \{x : (x, x^2) \in Q\}$. Since $|S| = 2\sqrt{n} + 2$, it is not a Sidon set and thus contains 4 distinct elements $x_1, x_2, x_3, x_4 \in S$ such that $x_1 + x_2 = x_3 + x_4$. Then, the 4 points $(x_1, x_1^2), (x_2, x_2^2), (x_3, x_3^2), (x_4, x_4^2) \in Q$ form a trapezoid. Indeed,

$$\frac{x_2^2 - x_1^2}{x_2 - x_1} = x_1 + x_2 = x_3 + x_4 = \frac{x_4^2 - x_3^2}{x_4 - x_3}. \quad (70)$$

Upper bound (ii). The proof is similar to that of Theorem 1 part (ii). Consider the 2-dimensional grid $[n]^2$. Let $\mathcal{L}$ be the set of lines that go through at least two points in $[n]^2$. Note that $|\mathcal{L}| \leq n^4$ and every line in $\mathcal{L}$ contains at most n points in $[n]^2$. Moreover, it was shown in [12] that every set of $\Omega(n^{4/5})$ points in $[n]^2$ contains a trapezoid. Let P be a random subset of $[n]^2$ by retaining each point independently with probability $n^{\alpha-1}/2$, where $\alpha \in [3/5, 1]$. By Chernoff's bound we have that $|P| = \Theta(n^{1+\alpha})$ with high probability. Furthermore, using the union bound and Chernoff's bound we have

$$\begin{aligned} \mathbb{P}(P \text{ contains no collinear } n^\alpha\text{-tuple}) &= 1 - \mathbb{P}(\exists \ell \in \mathcal{L} : |P \cap \ell| \geq n^\alpha) \\ &\geq 1 - \sum_{\ell \in \mathcal{L}} \mathbb{P}\left(|P \cap \ell| \geq \left(1 + \frac{n^\alpha}{2\mathbb{E}(|P \cap \ell|)}\right) \mathbb{E}(|P \cap \ell|)\right) \geq 1 - n^4 \exp(-n^\alpha/10), \end{aligned} \quad (71)$$

which tends to 1 as $n \rightarrow \infty$. Namely, there exists a point set $P \subseteq [n]^2$ of size $\Theta(n^{1+\alpha})$, so that P contains no collinear n^α -tuple and every subset of P of size $\Omega(n^{4/5})$ contains a trapezoid. After scaling we obtain

$$h(n, n^{\frac{\alpha}{1+\alpha}}) = O(n^{\frac{4/5}{1+\alpha}}) \quad \text{for } 3/5 \leq \alpha \leq 1. \quad (72)$$

Note that $\frac{3}{8} \leq \frac{\alpha}{1+\alpha} \leq \frac{1}{2}$ in this interval. By letting $s = n^{\frac{\alpha}{1+\alpha}}$ we have

$$h(n, s) = O((n/s)^{4/5}) \quad \text{for } n^{3/8} \leq s \leq \sqrt{n}. \quad (73)$$

Aggregating the bounds in (i) and (ii) yields the bounds in the theorem. $\square$

6. Further consequences

Gowers [19] introduced the following natural notion. For $s \in \mathbb{N}$, the (*diagonal*) *Ramsey number of planar points* $r(s)$ is the smallest integer, such that every set $P \subseteq \mathbb{R}^2$ of $r(s)$ points contains either a collinear s -tuple or a general position subset of size s . It follows from the definition that $r(s) = s$ when $s \leq 3$. With a simple drawing one can show that $r(4) = 5$, but the value of $r(5)$ is already unknown. For the asymptotic behavior of $r(s)$, Gowers [19] observed that

$$\Omega(s^2) = r(s) = O(s^3). \quad (74)$$

The lower bound construction comes from the grid $[m]^2$, where $m = \lceil s/2 \rceil - 1$. Obviously, $[m]^2$ contains at most $s/2$ collinear points and by the pigeonhole principle at most $s - 1$ points in general position. The upper bound in (74) is derived by a greedy algorithm. Gowers [19] asked to determine the correct order of $r(s)$ or at least to show which one of the trivial lower and upper bounds is closer to the truth. This question has been partially answered by Payne and Wood [27], who proved that

$$r(s) = O(s^2 \log s), \quad (75)$$

and further conjectured that $r(s)$ is quadratic in s .

Conjecture 19 (Payne–Wood [27, Conjecture 4.1], see also [18]).

$$r(s) = \Theta(s^2). \quad (76)$$

As a consequence of Proposition 2, if a set P of n points in the plane contains no collinear s -tuple with $s = O(\sqrt{n})$, then it contains a general position subset of size $\Omega\left(\sqrt{\frac{n \log \log s}{\log s}}\right)$. This yields the current best upper bound on $r(s)$, first proved by Hajnal and Szemerédi [21].

Corollary 20 (Hajnal–Szemerédi [21, Theorem 3]).

$$r(s) = O\left(\frac{s^2 \log s}{\log \log s}\right). \quad (77)$$

Our proof of Proposition 2 can be regarded as an extension of the arguments presented in [21]. Payne and Wood [27] proposed the following strengthening of Conjecture 19.

Conjecture 21 (Payne–Wood [27, Conjecture 4.2]). Every set $P \subseteq \mathbb{R}^2$ of n points with at most $\sqrt{n}$ points collinear can be colored with $O(\sqrt{n})$ colors such that each color class is in general position.

Conjecture 21, if true, would be sharp since at least $\sqrt{n}/2$ colors are needed to properly color the $\sqrt{n} \times \sqrt{n}$ grid. Payne and Wood showed that $O(\sqrt{n}(\log n)^{3/2})$ colors are sufficient to color P . Here we improve this result by a factor of $\sqrt{\log n}$. Moreover, note that Theorem 22 is a strengthening of Corollary 20.

Theorem 22. Every set $P \subseteq \mathbb{R}^2$ of n points with at most $\sqrt{n}$ collinear points can be colored with $O(\sqrt{n} \log n)$ colors such that each color class is in general position. In addition, one can properly color $n - o(n)$ points of P with $O(\sqrt{n} \log n / \log \log n)$ colors.

Proof of Theorem 22. Assume that n is sufficiently large and let $s = \lfloor \sqrt{n} \rfloor$. Then $P \subseteq \mathbb{R}^2$ is a set of n points with at most s points collinear. Our process of coloring P consists of two phases.

Phase 1:

Let $k = \lfloor \log \log n - 2C \rfloor$ with some sufficiently large constant $C > 0$. For each $i \in [k]$, we let P_i be a subset of $P \setminus (\bigcup_{j < i} P_j)$ obtained by selecting half of the remaining points, namely those that are contained in the least number of collinear triples. Note that $|P_i| \geq 4^C n / \log n$ holds for all $i \in [k]$, namely, $s \leq 2^{-C} \sqrt{|P_i| \log |P_i|}$. Let $\mathcal{H}_i = \mathcal{H}(P_i)$ be a 3-uniform hypergraph with the vertex set P_i , whose edges are the collinear triples in P_i . Observe that the least number of colors needed to color P_i with each color class being in general position is equal to $\chi(\mathcal{H}_i)$. Recall (20), we have that

$$\chi(\mathcal{H}_i) = O\left(\sqrt{\frac{|P_i| \log s}{\log \frac{|P_i| \log s}{s^2}}}\right) = O\left(\sqrt{\frac{n \log s}{2^i \log \frac{n \log s}{2^i s^2}}}\right) = O\left(\sqrt{\frac{n \log n}{2^{i-1} \log \log n}}\right). \quad (78)$$

Hence, we can properly color each P_i using $\chi(\mathcal{H}_i)$ new colors for $1 \leq i \leq k$. The total number of colors to properly color $\bigcup_{i=1}^k P_i$ is at most

$$\sum_{i=1}^k O\left(\sqrt{\frac{n \log n}{2^{i-1} \log \log n}}\right) \leq O\left(\sqrt{\frac{n \log n}{\log \log n}}\right) \sum_{i=1}^{\log \log n} \left(\frac{1}{\sqrt{2}}\right)^{i-1} = O\left(\sqrt{\frac{n \log n}{\log \log n}}\right). \quad (79)$$

Phase 2:

Let P' be the set of remaining uncolored points. It holds that $|P'| = O(n / \log n)$. We will color P' as follows. First, select a largest general position subset I_1 of P' and let $P'_1 = P' \setminus I_1$. For $i > 1$, as long as $P'_{i-1} \neq \emptyset$, we select a largest general position subset I_i of P'_{i-1} and let $P'_i = P'_{i-1} \setminus I_i$. Let m be the stopping time of this process, i.e., $P'_m = \emptyset$, equivalently,

$$P' = \bigcup_{i=1}^m I_i. \quad (80)$$

By assigning a new color to each I_i , we finish coloring P' with m colors. It remains to bound the value of m . Let $P'_0 = P'$. By (2) we have that, for each $i \in [m]$,

$$|I_i| = \Omega\left(\frac{|P'_{i-1}|}{s}\right) \geq \frac{|P'_{i-1}|}{D\sqrt{n}}, \quad (81)$$

for some constant $D > 0$. Accordingly, for each $i \in [m]$,

$$|P'_i| \leq |P'| \left(1 - \frac{1}{D\sqrt{n}}\right)^i. \quad (82)$$

When $m = \lceil D\sqrt{n} \log n \rceil$, it holds that $|P'_{m-1}| \leq 2$. Then, we can choose I_m to be P'_{m-1} and the process stops, namely, $m = O(\sqrt{n} \log n)$. Overall, we have properly colored P using

$$O\left(\sqrt{n \log n / \log \log n}\right) + O(\sqrt{n} \log n) = O(\sqrt{n} \log n) \quad (83)$$

colors. $\square$

Theorem 22 guarantees that most of the points from P can be properly colored with

$$O\left(\sqrt{n \log n / \log \log n}\right) \quad (84)$$

colors. This suggests that extending the result to all points may be achievable with further refinements in the proof.

Acknowledgments

The first author is supported in part by U.S. National Science Foundation grants DMS-1764123 and RTG DMS-1937241, FRG DMS-2152488 and the Arnold O. Beckman Research Award (UIUC Campus Research Board RB 24012), and the Simons Fellowship. The research of the second author is supported by the Institute for Basic Science (IBS-R029-C4). The authors thank the anonymous referees for their useful comments on the manuscript.

Appendix. A matching lower bound for Lemma 13

The estimate in Lemma 13 is tight up to the constant factor. Recall that we assume $\sqrt{n} = 2m + 1$ and let $x = x(n) \geq 1$ with $x \ll \sqrt{n}$. Let G be a $\sqrt{n} \times \sqrt{n}$ grid, and let o be the center point of G . Let A be the annulus formed by two concentric circles of radii m and $m - x$, centered at o . Our goal is to show that the number of collinear triples in $A \cap G$ is $\Omega(x^{5/2} n^{3/4} \log n)$.

After a translation, assume that $G = [-m, m]^2 \cap \mathbb{Z}^2$ and $o = (0, 0)$. Fix coprime integers q and r with $1 \leq q \leq r \leq c\sqrt{mx}$, where $c > 0$ is a sufficiently small constant.

By the first equation in (51), consecutive lattice lines of slope (q, r) are at distance $1/\sqrt{q^2 + r^2} = \Theta(1/r)$. Thus, $\Omega(xr)$ such lines have distance from o in $[m - 2x, m - x]$. For any such line ℓ , write $\text{dist}(\ell, o) = m - tx$, where $1 \leq t \leq 2$. Each component of $\ell \cap A$ has length

$$\sqrt{m^2 - (m - tx)^2} - \sqrt{(m - x)^2 - (m - tx)^2} = \Omega(\sqrt{mx}),$$

where we used $x \ll m$. Since consecutive lattice points on ℓ are at distance $\sqrt{q^2 + r^2} = \Theta(r)$, the line ℓ contains $\Omega(\sqrt{mx}/r)$ points of $A \cap G$. Choosing c sufficiently small, this quantity is at least three, and therefore the number $T(q, r)$ of collinear triples contained in lines of slope (q, r) is

$$T(q, r) = \Omega\left(xr \left(\frac{\sqrt{mx}}{r}\right)^3\right) = \Omega\left(\frac{x^{5/2} m^{3/2}}{r^2}\right).$$

Let $\phi(r)$ denote Euler's totient function, that is, $\phi(r)$ counts the number of positive integers smaller than and coprime to r . Using $\sum_{r \leq R} \phi(r)/r^2 = \Theta(\log R)$ (see, e.g., [1, Theorem 3.2]), we obtain

$$\sum_{\substack{1 \leq r \leq c\sqrt{mx} \\ 1 \leq q \leq r, (q,r)=1}} T(q, r) = \Omega(x^{5/2} m^{3/2} \log(mx)) = \Omega(x^{5/2} n^{3/4} \log n).$$

Therefore, the number of collinear triples in $A \cap G$ is $\Omega(x^{5/2} n^{3/4} \log n)$.

References

- [1] Tom M. Apostol, *Introduction To Analytic Number Theory*, Springer Science & Business Media, 2013.
- [2] József Balogh, Robert Morris, Wojciech Samotij, Independent sets in hypergraphs, *J. Amer. Math. Soc.* 28 (3) (2015) 669–709.
- [3] József Balogh, József Solymosi, On the number of points in general position in the plane, *Discret. Anal.* 16 (2018) 1–20.
- [4] Imre Bárány, Maria Prodromou, On maximal convex lattice polygons inscribed in a plane convex set, *Israel J. Math.* 154 (1) (2006) 337–360.
- [5] François Charton, Jordan S. Ellenberg, Adam Zolt Wagner, Geordie Williamson, Patternboost: constructions in mathematics with a little help from ai, 2024, Preprint [arXiv:2411.00566](https://arxiv.org/abs/2411.00566).
- [6] Yaobin Chen, Xizhi Liu, Jiaxi Nie, Ji Zeng, Random turán and counting results for general position sets over finite fields, *Sci. China Math.* 68 (2025) 3043–3062.
- [7] Felix Christian Clemen, Applications of sparse hypergraph colorings, *Discrete Math.* 349 (2) (2026).
- [8] Jeff Cooper, Dhruv Mubayi, Coloring sparse hypergraphs, *SIAM J. Discrete Math.* 30 (2) (2016) 1165–1180.
- [9] Henry Dudeney, A puzzle with pawns, *Amusements Math.* (1917) 94.
- [10] David Eppstein, *Forbidden Configurations in Discrete Geometry*, Cambridge University Press, Cambridge, 2018.
- [11] Paul Erdős, On some metric and combinatorial geometric problems, *Discrete Math.* 60 (1986) 147–153.
- [12] Paul Erdős, Ron Graham, Imre Ruzsa, Herbert Taylor, Bounds for arrays of dots with distinct slopes or lengths, *Combinatorica* 12 (1) (1992) 39–44.
- [13] Paul Erdős, János Surányi, *Topics in the Theory of Numbers*, Springer, 2003.
- [14] Paul Erdős, George Szekeres, A combinatorial problem in geometry, *Compos. Math.* 2 (1935) 463–470.
- [15] Zoltán Füredi, Maximal independent subsets in Steiner systems and in planar sets, *SIAM J. Discrete Math.* 4 (2) (1991) 196–199.
- [16] Hillel Furstenberg, Yitzhak Katznelson, A density version of the Hales–Jewett theorem for $k = 3$, *Discrete Math.* 75 (1989) 227–241.
- [17] Bogdan Georgiev, Javier Gómez-Serrano, Terence Tao, Adam Zolt Wagner, Mathematical exploration and discovery at scale, 2025, Preprint [arXiv:2511.02864](https://arxiv.org/abs/2511.02864).
- [18] William Timothy Gowers, General position subsets, 2024, available on http://www.openproblemgarden.org/op/general_position_subsets. (Accessed November 2024).
- [19] William Timothy Gowers, A geometric ramsey problem, 2024, available on <https://mathoverflow.net/questions/50928/a-geometric-ramsey-problem>. (Accessed September 2024).
- [20] Larry Guth, Nets Hawk Katz, On the Erdős distinct distances problem in the plane, *Ann. Math.* 181 (1) (2015) 155–190.
- [21] Péter Hajnal, Endre Szemerédi, Two geometrical applications of the semi-random method, *New Trends Intuitive Geom.* (2018) 189–199.
- [22] Godfrey Harold Hardy, On the expression of a number as the sum of two squares, *Q. J. Math.* 46 (1915) 263–283.
- [23] Vojtěch Jarník, Über die Gitterpunkte auf konvexen Kurven, *Math. Z.* 2 (1926) 500–518.
- [24] Chaya Keller, Shakhar Smorodinsky, A new lower bound on Hadwiger–Debrunner numbers in the plane, *Israel J. Math.* 244 (2021) 649–680.
- [25] Hanno Lefmann, Extensions of the no-three-in-line problem, 2024, Preprint available at https://www.tu-chemnitz.de/informatik/ThIS/downloads/publications/lefmann_no_three_submitted.pdf. (Accessed October 2024).
- [26] Michael Mitzenmacher, Eli Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, Cambridge, 2017.
- [27] Michael Payne, David Wood, On the general position subset selection problem, *SIAM J. Discrete Math.* 27 (4) (2013) 1727–1733.
- [28] William Moser Peter Braß, János Pach, *Research Problems in Discrete Geometry*, Springer, New York, 2005.
- [29] Kevin T. Phelps, Vojtěch Rödl, Steiner triple systems with minimum independence number, *Ars Combin.* 21 (1986) 167–172.
- [30] Carl Pomerance, Collinear subsets of lattice point sequences – An analog of Szemerédi's theorem, *J. Combin. Theory Ser. A* 28 (2) (1980) 140–149.
- [31] Oliver Roche-Newton, Krishnendu Bhowmick, Counting arcs in $\mathbb{F}_q^2$, *Discrete Comput. Geom.* 72 (4) (2024) 1630–1646.
- [32] Oliver Roche-Newton, Audie Warren, Arcs in $\mathbb{F}_q^2$, *European J. Combin.* 103 (2022) 1–15.
- [33] Barkley J. Rosser, Lowell Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* 6 (1) (1962) 64–94.

- [34] Misha Rudnev, Jonathan M. Selig, On the use of the klein quadric for geometric incidence problems in two dimensions, *SIAM J. Discrete Math.* 30 (2) (2016) 934–954.
- [35] David Saxton, Andrew Thomason, Hypergraph containers, *Invent. Math.* 201 (2015) 925–992.
- [36] Andrew Suk, Ji Zeng, On higher dimensional point sets in general position, *Combin. Probab. Comput.* 35 (2026) 134–148.
- [37] Endre Szemerédi, William T. Trotter, Extremal problems in discrete geometry, *Combinatorica* 3 (1983) 381–392.
- [38] Terence Tao, Van Vu, *Additive Combinatorics*, Cambridge University Press, Cambridge, 2006.
- [39] Zhen Zhang, A note on arrays of dots with distinct slopes, *Combinatorica* 13 (1) (1993) 127–128.