Bounds on 2-Rich Points

Maximus Lu and Aloysius Ng
Mentor: Jiahui Yu

February 27, 2026

Abstract

We give a new construction with $|P_2(\mathcal{L})| = \omega(LB)$ 2-rich points in $\mathbb{C}^3$ with no $B = O(L^{1/3-\varepsilon})$ lines in the same plane or regulus, better than the previous best construction of $|P_2(\mathcal{L})| = O(LB)$. In particular, when $B \sim \log L$, our construction achieves $|P_2(\mathcal{L})| \sim L^{7/6}$. We also prove bounds on 2-rich points in general $\mathbb{F}^n$ using graph-theoretic methods, showing a bound of $|P_2(\mathcal{L})| \lesssim \min(B^{1/2}L^{3/2}, BL + L^{5/3})$ when there are at most B lines in any 3-flat.

Contents

1 Introduction

Let $\mathcal{L}$ be a set of lines and denote by $P_r(\mathcal{L})$ the set of points that lie on at least r of the lines in $\mathcal{L}$. As $|\mathcal{L}|$ grows, there are a number of incidence bounds on $|P_r(\mathcal{L})|$. For example, in Euclidean space $\mathbb{R}^2$, we have the Szemerédi–Trotter theorem [ST83].

Theorem 1.1. *An incidence is a tuple* (p, ℓ) *of a point* p *and a line* ℓ *where* p *lies in* ℓ*. The number of incidences between* P *points and* L *lines on the Euclidean plane is at most*

$$O(P^{2/3}L^{2/3} + P + L).$$

Incidences are counted with multiplicity at each intersection, and by taking $P = P_r(\mathcal{L})$ the upper bound on incidences gives an upper bound on the number of r-rich points.

Theorem 1.2 ([ST83])**.** *Given* L *lines on the Euclidean plane, we have*

$$|P_r(\mathcal{L})| = O\left(\frac{L^2}{r^3} + \frac{L}{r}\right).$$

In higher dimensions, much work has been done on incidence bounds. For example, [GK15, ST12] use the polynomial ham sandwich theorem to bound the number of incidences in $\mathbb{R}^d$ and [Wal23] uses algebro-geometric methods to give a bound in general $\mathbb{F}^d$, expressed in terms of the concentration of $\mathcal{L}$ on higher-dimensional varieties.

However when r is constant, $P_r(\mathcal{L})$ ignores multiplicities, while incidence bounds do not. As a result these incidence bounds often only give an upper bound for $P_r(\mathcal{L})$ on the order of L^2.

Interest in proving bounds on $P_r(\mathcal{L})$ in the space $\mathbb{C}^3$ originated from work in the polynomial method. In 2008, the finite field Kakeya problem was solved by Dvir [Dvi09] by considering a polynomial that vanishes on the Kakeya set. Later that year, Guth and Katz adapted the polynomial method to solve the joints problem and a conjecture of Bourgain [GK10]. The advances in these tools prompted the paper by Elekes and Sharir in [ES11] that reduces a near-linear bound on the Erdős distinct distances problem to upper bounds on the number of r-rich points in $\mathbb{R}^3$ for $r \geq 2$.

Following this, Guth and Katz proved a bound on r-rich points in [GK15] for all $r \geq 2$, which is sufficient to attain a $\Omega(n/\log n)$ bound on the distinct distances problem.

In this paper we specifically consider the case of $r = 2$. The main results still hold for $r \geq 2$, but we focus on this case, as the restriction on degree 2 surfaces is only necessary for the upper bound when $r = 2$.

In particular, if we consider a set of lines $\mathcal{L} \subset \mathbb{C}^3$ while only restricting the number of lines in a plane, it is still possible to achieve $P_2(\mathcal{L}) \sim L^2$.

Example 1.3. Consider the set of lines

$$\{(x, t, xt) \mid x \in \{1, \ldots, n\}\} \cup \{(t, y, yt) \mid y \in \{1, \ldots, n\}\}.$$

There are $2n$ lines, and these lines have the n^2 intersections (x, y, xy) for $x, y \in \{1, \ldots, n\}$.

In the above example, the lines all lie on the surface $z = xy$. This is an example of a regulus, which is defined as the degree 2 surface containing all lines which simultaneously intersect three skew lines ℓ_1, ℓ_2, ℓ_3. (See [Gut16, Section 8.4] for a more detailed description.)

Importantly, planes and reguli are the only doubly ruled sufaces in $\mathbb{C}^3$, i.e. surfaces which contain two lines through every point. Once we restrict doubly ruled surfaces, we get an improvement on the exponent in the bound for $P_2(\mathcal{L})$.

Theorem 1.4 ([GK15, Theorem 2.10]). *Let $\mathcal{L}$ be any set of L lines in $\mathbb{R}^3$ for which no more than $L^{1/2}$ lie in a common plane and no more than $O(L^{1/2})$ lie in a common regulus. Then,*

$$P_2(\mathcal{L}) = O(L^{3/2}).$$

We can also allow more lines to lie on a plane or regulus than $L^{1/2}$. This gives a linear bound on $P_2(\mathcal{L})$ with respect to L.

Theorem 1.5 ([Gut16, Theorem 13.1]). *If $\mathcal{L}$ is a set of L lines in $\mathbb{C}^3$ with at most B lines in any plane or degree 2 surface, then*

$$P_2(\mathcal{L}) = O(LB + L^{3/2}).$$

1.1 Notation

Throughout this paper, we take $L = |\mathcal{L}|$, and use the notation $f \lesssim g$ to mean $f = O(g)$ (and similarly for $f \gtrsim g$). We also use the shorthand of $f \lesssim_x g$ to mean $f = O_x(g)$, i.e. the constant may depend on x.

We also let $\text{Poly}(\mathbb{F}^n)$ denote the set of polynomials in n variables over $\mathbb{F}$, and $\text{Poly}_D(\mathbb{F}^n)$ be the set of polynomials of degree at most D in $\text{Poly}(\mathbb{F}^n)$.

Finally, a k-flat of a vector space refers to an affine k-dimensional subspace.

1.2 Overview

The current upper bounds on 2-rich points, from Theorem 1.5, give an upper bound of $O(LB + L^{3/2})$. However, the best known constructions only give $O(LB)$, which leaves a gap when $B \leq L^{1/2}$.

In Section 2, we give a new construction for 2-rich points in $\mathbb{C}^3$ (Theorem 2.2), which achieves $|P_2(\mathcal{L})| \sim L^{7/6}$ with at most $B \sim \log L$ lines in any surface of degree at most 2.

The idea of the construction is to first produce a lattice-like construction in $\mathbb{C}^n$, for $n > 3$, then take a "random" projection to $\mathbb{C}^3$, which preserves the property of having few lines on any surface of degree ≤ 2.

In Section 3, we use graph-theoretic methods to bound $|P_2(\mathcal{L})|$ while restricting the number of lines in any affine 3-dimensional subspace, in any vector space $\mathbb{F}^n$ (Theorem 3.1).

In Section 4, we discuss some interesting questions for further investigation, particularly relating to closing the gap between the lower bounds and known constructions.

1.3 Acknowledgments

We would like to thank Larry Guth for suggesting this project, and Roman Bezrukavnikov and Jonathan Bloom for their advice and suggestions throughout the SPUR program.

2 Constructions

In this section we consider certain configurations of lines in $\mathbb{C}^n$ that contain many 2-rich points, while not having many lines on a low-degree algebraic variety. We first give such a construction in $\mathbb{C}^n$ for $n \geq 3$.

Theorem 2.1. *There exists a set of lines $\mathcal{L} \subset \mathbb{C}^n$ such that for all $0 < c < 1$, $|P_2(\mathcal{L})| \gtrsim_n L^{1+\frac{c}{n-1}}$ and no B lines lie on the same k-flat, where*

$$B \sim_{n,k} \left(1 + L^{\frac{(k-1)-(1-c)(n-k)}{n-1}}\right) \cdot \log L.$$

Remark. Throughout this section, we focus on the case of $\mathbb{C}^n$, and $\mathbb{C}^3$, as this is the space considered in previous authors, such as in [Gut16]. However, all our results still hold when $\mathbb{C}$ is replaced by $\mathbb{R}$.

Although this construction is done in high-dimensional $\mathbb{C}^n$, it is also possible to project it down to a good construction in lower dimensions. In particular by taking $n = 4$, $k = 2$, and projecting down to $\mathbb{C}^3$, we get the following.

Theorem 2.2. *There exists a set of lines $\mathcal{L} \subset \mathbb{C}^3$ such that for all $1/2 \leq c < 1$, $|P_2(\mathcal{L})| \sim L^{1+\frac{c}{3}}$ and no B lines lie on the same degree ≤ 2 surface where*

$$B \sim L^{\frac{2c-1}{3}} \cdot \log L.$$

The two endpoints are particularly interesting. When $c = 1/2$, we get the following.

Corollary 2.3. *There exists a set of lines $\mathcal{L} \subset \mathbb{C}^3$ such that no $B \sim \log L$ lines lie in a degree ≤ 2 surface and $|P_2(\mathcal{L})| \sim L^{7/6}$.*

When $c = 1 - 3\varepsilon$, we get the following.

Corollary 2.4. *Take $\varepsilon > 0$. There exists a set of lines $\mathcal{L} \subset \mathbb{C}^3$ such that no $B \sim L^{1/3-2\varepsilon} \cdot \log L$ lines lie in a degree ≤ 2 surface and $|P_2(\mathcal{L})| \sim L^{4/3-\varepsilon} = \omega(LB)$.*

We may compare this construction to the results of [GK15]. The upper bound and constructions in $\mathbb{C}^3$ may be visualized as in Fig. 1.

Remark. The analogues of the above results still hold when we replace P_2 with P_r for constant $r \geq 2$, although the constants may depend on r. The proofs are essentially the same in both cases.

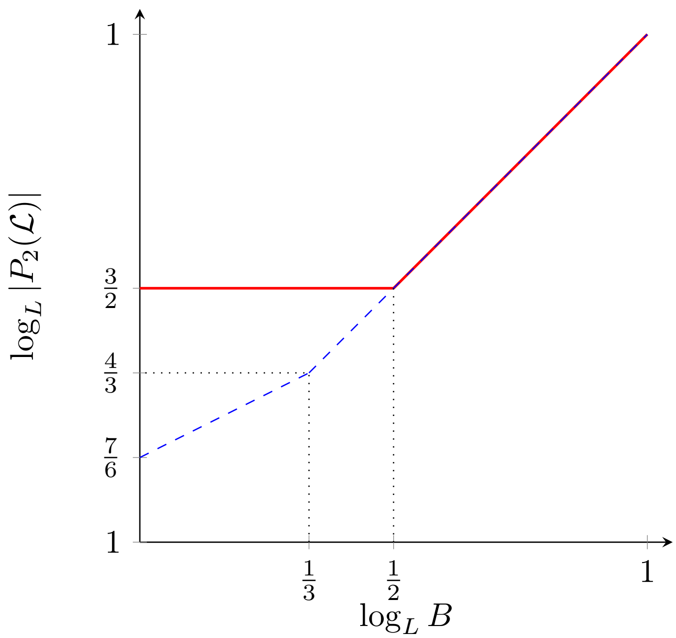

Figure 1: The red line is the upper bound on $P_2(\mathcal{L})$ proven in [GK15], and the dashed blue line shows the construction of Corollary 2.3.

2.1 Probability Lemmas

The constructions below are randomized, and so we now show that properties of these constructions hold with high or constant probability. To do this we use this set of standard concentration bounds.

Lemma 2.5 (Multiplicative Chernoff Bound)**.** *Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables, and let $X = \sum_{i=1}^n X_i$. Suppose $\mu = \mathbb{E}[X]$. Then,*

$$\mathbb{P}\left[X \geq (1+\delta)\mu\right] \leq e^{-\frac{\delta^2}{2+\delta}\mu}, \qquad 0 < \delta,$$
$$\mathbb{P}\left[X \leq (1-\delta)\mu\right] \leq e^{-\frac{\delta^2}{2}\mu}, \qquad 0 < \delta < 1.$$

We also have another lemma that will be used to bound the value of B needed for our construction to work with high probability.

Lemma 2.6. *Suppose there are m subsets $A_1, A_2, \ldots, A_m$ of S, such that $|A_i| < a$ for all i. Let $T \subset S$ be a random subset of S where every element independently selected with probability p, and let $X = \max_i |T \cap A_i|$. Then, for $c > 10$ and m with $\log m > 1$, we have*

$$\mathbb{P}[X > c \log m(pa + 1)] \leq m^{-c/10}.$$

Proof. Take any subset $A \subset S$ of size $a+1/p$. Then, $|T \cap A|$ is the sum of $a+1/p$ independent random variables that are 1 with probability p, so $\mathbb{E}[|T \cap A|] = pa + 1$. By Lemma 2.5 we have that

$$\mathbb{P}[|T \cap A| \geq (1+\delta)(pa+1)] \leq e^{-\frac{\delta^2}{2+\delta}(pa+1)}.$$

All subsets A_i have size smaller than $a + 1/p$, so

$$\mathbb{P}[|T \cap A_i| \geq (1+\delta)(pa+1)] \leq \mathbb{P}[|T \cap A| \geq (1+\delta)(pa+1)] \leq e^{-\frac{\delta^2}{2+\delta}(pa+1)}.$$

Taking a union bound over i, we have

$$\mathbb{P}[X \geq (1+\delta)(pa+1)] \leq me^{-\frac{\delta^2}{2+\delta}(pa+1)}.$$

Finally take $\delta = c \log m/2$ which gives

$$\mathbb{P}[X \geq (1 + c\log m/2)(pa+1)] \leq me^{-c/2 \log m \frac{\delta}{2+\delta}(pa+1)} \leq m^{1-\frac{c\delta}{4+2\delta}}.$$

Given our assumptions that $c > 10$ and $\log m > 1$, we have $\delta > 5$, so

$$\mathbb{P}[X \geq c \log m(pa+1)] \leq m^{-c/10}.$$ □

2.2 Initial Construction

We first state the construction for Theorem 2.1.

Construction for Theorem 2.1. Let n be a positive integer and $0 < c < 1$ be a positive real. We construct $\mathcal{L} \subset \mathbb{C}^n$ as follows. Take N sufficiently large. Consider the set of all lines $\mathbb{L}$ of the form $\ell = (\mathbf{x}, 0) + \lambda(\mathbf{v}, 1)$ where

$$\mathbf{x} \in \{z : z \in \mathbb{Z},\, |z| \leq N\}^{n-1} \quad \text{and} \quad \mathbf{v} \in \{z : z \in \mathbb{Z},\, |z| \leq N^{1-c}\}^{n-1}.$$

Then, each line $\ell \in \mathbb{L}$ is included in $\mathcal{L}$ with probability $p = 10N^{-(1-c)(n-1)}$, independently of every other line. □

We first give a brief discussion of the construction before proving its properties. In finite fields $\mathbb{F}_p$, the construction of picking all lines independently with equal probability is good, as all lines pass through the same number of points. However, when we take a cuboid of lattice points in $\mathbb{C}^n$ and randomly take lines, they are unlikely to pass through many points.

Even when we take 2 points $(\mathbf{x}, 0), (\mathbf{y}, 1)$ such that $\mathbf{x}, \mathbf{y}$ are randomly chosen lattice points, the coordinates of $\mathbf{x}$ and $\mathbf{y}$ are expected to have a difference on the order of the side length of the cuboid, leading the line to have few intersections with the lattice points in the cuboid.

By picking $\mathbf{v}$ to have small coordinates, we force the lines of $\mathbb{L}$ to have many intersections with points in the cuboid. This increases the number of incidences with points in the cuboid, and hence the number of 2-rich points.

We now prove the remaining parts of Theorem 2.1. This is done by combining Proposition 2.7, Proposition 2.8, Proposition 2.9, which proves that the construction given above satisfies the conditions of Theorem 2.1 with at least positive constant probability.

Proposition 2.7. *With high probability, the number of lines chosen satisfies* $L = |\mathcal{L}| \sim_n N^{n-1}$.

Proof. Recall that $\mathcal{L}$ is chosen as a random subset of $\mathbb{L}$. The total number of lines picked can be modelled as a sum of indicator variables for each line $l \in \mathbb{L}$.

Now, we apply the Chernoff bound in Lemma 2.5 with a total of $|\mathbb{L}| \sim_n N^{n-1} \cdot N^{(1-c)(n-1)}$ identical indicator variables for the X_i, $p \sim_n N^{-(1-c)(n-1)}$. Note that $\mathbb{E}[|\mathcal{L}|] \sim_n N^{n-1}$. Hence, (taking δ large),

$$\mathbb{P}[|\mathcal{L}| \gtrsim_n N^{n-1}] \leq e^{-\mu}$$

with $\mu \sim_n N^{n-1}$, which clearly goes to 0.

Similarly, taking $\delta \approx 1$,

$$\mathbb{P}[\mathcal{L} \lesssim_n N^{n-1}] \leq e^{-\mu/3},$$

so with probability at least $1 - 2e^{\mu/3}$, we have $\mathcal{L} \sim_n N^{n-1}$. □

Proposition 2.8. *The construction satisfies* $P_2(\mathcal{L}) \gtrsim_n L^{1+\frac{c}{n-1}}$ *with probability* $> 1/2$.

Proof. Consider the set of integer points $\mathcal{P}$ consisting of points $(\mathbf{s}, t)$ where

$$\mathbf{s} \in \{z \mid z \in \mathbb{Z}, |z| \leq N\}^{n-1}, \quad t \in \{z \mid z \in \mathbb{Z}, |z| \leq N^c\}.$$

For each point $x = (\mathbf{s}, t) \in \mathcal{P}$, there is a subset of lines in $\mathbb{L}$ that passes through it. Denote this subset $\mathbb{L}_x$. We want a lower bound on $|\mathbb{L}_x|$ that gives each point a constant (possibly depending on n) probability that it is a 2-rich point.

For each line $\ell = \mathbf{x} + \lambda\mathbf{v} \in \mathbb{L}$, we have $\ell \in \mathbb{L}_x$ if and only if

$$\mathbf{x} + t\mathbf{v} = \mathbf{s}.$$

Looking at each coordinate separately, we now lower bound the number of tuples (x_i, v_i) such that $x_i + tv_i = s_i$, under the constraints that $|x_i|, |s_i| \leq N, |v_i| \leq N^{1-c}, |t| \leq N^c$.

If v_i is picked to have the same (positive or negative) sign as $s_i t$, then $x_i = s_i - tv_i$ is the difference of two numbers of the same sign that have absolute value less than N, and so will be in the acceptable range. Thus there are at least N^{1-c} choices in each coordinate, giving $|\mathbb{L}_x| \geq N^{(1-c)(n-1)}$.

The number of lines selected in $\mathcal{L}$ passing through x can be modeled as selecting lines in $\mathbb{L}_x$ with probability $p = 10N^{(1-c)(n-1)}$. It is clear that the probability x is 2-rich is minimized when $|\mathbb{L}_x|$ is minimized, and we have a lower bound of $|\mathbb{L}_x| \geq 10/p$.

We give a constant lower bound for the probability of being 2-rich by giving an upper bound of the complement, which occurs when zero or one line is chosen. For small p, this probability is

$$(1-p)^{10/p} + \frac{10}{p} \cdot p(1-p)^{(10/p)-1} \leq \frac{11}{e^{10}(1-p)} < \frac{1}{10}.$$

Hence, each point in $\mathcal{P}$ is 2-rich with probability at least $9/10$. By linearity of expectation, the expected number of 2-rich points in $\mathcal{P}$ is at least $9/10\,|\mathcal{P}|$. If the probability of picking at most $|\mathcal{P}|/2$ points exceeds $1/2$, then we get an upper bound of

$$1/2(|\mathcal{P}|/2 + |\mathcal{P}|) < 3/4\,|\mathcal{P}| < 9/10\,|\mathcal{P}|,$$

for the expected value, which contradicts the lower bound. Therefore, we have $\mathbb{P}\left[P_2(\mathcal{L}) \geq |P|/2 \sim_n L^{1+\frac{c}{n-1}}\right] > 1/2$ as desired. □

Proposition 2.9. *With high probability, we have no B lines of $\mathcal{L}$ on any k-flat, where $B \sim_{n,k} \left(1 + L^{\frac{(k-1)-(1-c)(n-k)}{n-1}}\right) \cdot \log L$.*

Proof. We want to use Lemma 2.6 with $T = \mathcal{L}$ and $S = \mathbb{L}$, each A_i being a subset of lines that lie in a k-flat and $X = B$. We have $p = 10N^{-(1-c)(n-1)}$. We set a to be an upper bound on the number of lines of $\mathbb{L}$ on any k-flat, and m to be an upper bound on the number of possible subsets of lines of $\mathbb{L}$ are the intersection of the set of lines of a k-flat with $\mathbb{L}$.

Let X be the maximum number of lines in $\mathcal{L} = T$ that lie in some k-flat. With these, for $C > 10$ we would get that

$$\mathbb{P}[X > C \log m(pa+1)] \leq m^{-C/10}.$$

It suffices to show that $\log m \sim_{n,k} \log L$ and $a \lesssim_{n,k} L^{\frac{k-1+(1-c)(k-1)}{n-1}}$. Then for sufficiently large C, we have

$$\mathbb{P}\left[X \gtrsim_{n,k} \log L \left(L^{\frac{(k-1)+(1-c)(k-1)-(1-c)(n-1)}{n-1}} + 1\right)\right] \leq L^{-C/10}$$

which gives the desired bound.

We first bound on $\log m$. We only need to take collection of subsets $\{A_i\}$ such that any subset of lines in $\mathbb{L}$ that lies on the same k-flat is contained in some subset A_i. Take any $k+1$ points in $\mathcal{P}$, and take a $\leq k$-flat by taking the affine subspace spanned by these points. Then define subset $A_i \subset \mathbb{L}$ by taking all lines in $\mathbb{L}$ that lie in this $\leq k$-flat. The number of such subsets is bounded above by $|\mathcal{P}|^{k+1} \lesssim_{n,k} N^{nk+n}$.

On the other hand, if we have some subset of lines in $\mathbb{L}$ that lies on the same k-flat, we can take some $\leq k+1$ points that are affinely independent such that all points of $\mathcal{P}$ that lie in this k-flat also lie in the affine subspace spanned by these points. We do this by greedily adding independent points, and the dimension means only $k+1$ points can ever be found.

This will be contained in one of the subsets previously chosen, thus proving that we can take m subsets where $\log m \lesssim_{n,k} \log N \sim_n \log L$.

To finish, we find a bound on a. The subsets were chosen such that any subset of lines lie on a single k-flat. We count the number of lines $\ell = (\mathbf{x}, 0) + \lambda(\mathbf{v}, 1)$ such that $(\mathbf{x}, 0)$ and $(\mathbf{x}+\mathbf{v}, 1)$ lie in the k-flat. If a k-flat contains any such lines, it cannot be fully contained in $(-, 0)$. So, its intersection with the $(n-1)$-dimensional hyperplane $(-, 0)$ is a $(k-1)$-dimensional affine subspace.

We wish to prove that a $(k-1)$-flat can only pass through at most $(2N+1)^{k-1}$ points of the integer lattice $\{z : z \in \mathbb{Z}, |z| \leq N\}^{n-1}$. We do this by induction on the value $n+k$. If $k-1 = 1$ then it is clear that it intersects at most $\leq 2N+1$ points as some coordinate must vary linearly and be non-constant.

Otherwise partition the n dimensional lattice into $n-1$ dimensional lattices on axis-aligned hyperplanes. If some hyperplane contains the $(k-1)$-flat then induct down to the $n-2$ dimensional lattice it is contained in. Otherwise the intersection of the hyperplane and the $(k-1)$-flat is a $(k-2)$-flat, and by the inductive hypothesis we find that it intersects at most $(2N+1) \cdot (2N+1)^{k-2} = (2N+1)^{k-1}$ lattice points.

Hence, there are at most $\lesssim_k N^{k-1}$ values of $\mathbf{x}$ such that $(\mathbf{x}, 0)$ lies on the k-flat. Similarly, we can consider the smaller lattice formed by $(\mathbf{x}+\mathbf{v}, 1)$for fixed $\mathbf{x}$. The same argument shows that there are $\lesssim_k N^{(1-c)(k-1)}$ possible choices of $\mathbf{v}$. Combined, we get a upper bound on the number of lines as the product:

$$a \lesssim_n N^{k-1} \cdot N^{(k-1)(1-c)}.$$

With these bounds on a and $\log m$ we use Lemma 2.6 as discussed above to finish the proof. □

With that, we have found and proven a consruction for Theorem 2.1. Finally, to show that this gives a construction with $|P_2(\mathcal{L})| = \omega(LB)$ we take explicit constants.

Corollary 2.10. *Fix a positive integer* $k \geq 2$. *There exists a set of lines* $\mathcal{L} \subset \mathbb{C}^{3k-2}$ *such that* $|P_2(\mathcal{L})| \gtrsim_k L^{1+\frac{1}{6(k-1)}}$ *and no* B *lines lie on a* k*-flat, where* $B \sim_k \log L$.

Proof. Substitute $n = 3k - 2, c = 1/2$ into Theorem 2.1. □

2.3 Projection to 3 Dimensions

We project the construction in the above section for $\mathbb{C}^4$ to $\mathbb{C}^3$ and aim to show it satisfies the claimed properties in Theorem 2.2.

Let $\mathcal{L}' \subset \mathbb{C}^4$ be the set of lines taken in $\mathbb{C}^4$ based on the above construction. Take three pairs of complex numbers $(a_1, b_1), (a_2, b_2), (a_3, b_3)$ such that the a_i and b_i are all algebraically independent over $\mathbb{Q}$. Define the linear transformation $f : \mathbb{C}^4 \to \mathbb{C}^3$ such that

$$f : (x_1, x_2, x_3, x_4) \mapsto x_1(a_1, b_1, 0) + x_2(a_2, b_2, 0) + x_3(a_3, b_3, 0) + x_4(0, 0, 1),$$

and let $\mathcal{L}$ be the image of $\mathcal{L}'$ under f.

Remark. The algebraically independent condition is to avoid "coincidences" between the points that result in a degree 2 surface after projection – we are essentially taking a "random" projection.

Most properties we want to show come directly from the initial construction.

Proposition 2.11. *With high probability, the number of lines chosen satisfies* $|\mathcal{L}| \sim N^3$.

Proof. No two points of $\mathcal{P} \subset \mathbb{C}^4$ map to the same point under f as a_i, b_i are linearly independent over $\mathbb{Q}$. Thus, no two pairs of points $(\mathbf{x}, 0), (\mathbf{x} + \mathbf{v}, 1)$ will map to the same pair of points. Hence, $\mathcal{L}' \to \mathcal{L}$ is injective, and thus have the same cardinality. By Proposition 2.7, with high probability, we have

$$|\mathcal{L}| = |\mathcal{L}'| \sim N^3.$$

□

Proposition 2.12. *With probability* $> 1/2$, *we have* $|P_2(\mathcal{L})| \gtrsim L^{1+c/3}$

Proof. If a point $x \in \mathcal{P}$ lies on two lines of $\mathcal{L}'$, then its image $f(x) \in \mathbb{C}^3$ lies on the projections of both lines, which are distinct. The points of $\mathcal{P}$ do not map to the same point under f, so $P_2(\mathcal{L}) \geq P_2(\mathcal{L}')$. Thus by Proposition 2.8, we have

$$\mathbb{P}\left[P_2(\mathcal{L}) \gtrsim L^{1+\frac{c}{n-1}}\right] \geq \mathbb{P}\left[P_2(\mathcal{L}') \gtrsim L^{1+\frac{c}{n-1}}\right] > 1/2.$$

□

To show that we do not find too many lines on a surface of degree at most 2 is more involved. We will use a version of Bézout's theorem given in [Gut16].

Lemma 2.13 ([Gut16, Theorem 6.7]). *For polynomials $P, Q \in \mathrm{Poly}(\mathbb{C}^3)$ with no common factor, the number of lines contained in $Z(P) \cap Z(Q)$ is at most $(\deg P)(\deg Q)$.*

Proposition 2.14. *With high probability, we have no B lines on any degree ≤ 2 surface where $B \sim L^{\frac{2c-1}{3}} \cdot \log L$.*

Proof. We assume these surfaces are irreducible. To account for degree 2 surfaces that reduce to two degree 1 surfaces, we may simply increase the bound by a factor of 2.

Following the proof of Proposition 2.9, we want to use Lemma 2.6. As before, this requires bounds on $\log m$ and r, where m is the number of subsets such that any set of lines contained in a degree ≤ 2 surface is contained in some subset, and r is the maximum size of these subsets.

We construct the subsets R_i by first considering all collections of 5 lines in $f(\mathbb{L})$. If these lines do not lie in an irreducible degree ≤ 2 surface, reject it. Otherwise there is a unique irreducible degree ≤ 2 surface that passes through it.

It is impossible for these 5 distinct lines to lie in the intersection of two irreducible degree ≤ 2 surfaces by Lemma 2.13, since $2 \cdot 2 < 5$, and so one irreducible surface would have to contain the other. Then, take R_i as the set of all lines in $f(\mathbb{L})$ that lie in this surface. For completeness, also include all subsets of $f(\mathbb{L})$ of size ≤ 4.

With this, we have $m \lesssim |\mathbb{L}|^5 \lesssim N^{6 \cdot 5}$. Consider any subset of lines in $f(\mathbb{L})$ that is contained in a degree ≤ 2 surface. If there are less than 5 lines we are done. Otherwise take any 5 lines and take the subset R_i generated above. Therefore we have $\log m \lesssim \log N \sim \log L$.

Now it suffices to bound r, which is bounded by the number of lines in $f(\mathbb{L})$ that lie on a degree ≤ 2 surface. To contain any lines of $\mathbb{L}$, this surface must not be fully contained in the planes $z = 0$ or $z = 1$. Then, the intersection of the surface with these planes is at most 1-dimensional. It is described by the same equation with the substitution $z = 0$ or $z = 1$.

We claim that the number of integer points in $z = 0$ that lie on this surface is $\lesssim N$. Again we assume the intersection is irreducible, otherwise it suffices to double the bounds. Take the image of the integer points on an axis-parallel plane under f, such as the points $(-, -, h, 0)$, for fixed $h \in \{z : z \in \mathbb{Z}, |z| \leq N\}$. The image is a $(2N + 1) \times (2N + 1)$ integer lattice in the $z = 0$ plane of $\mathbb{C}^3$.

If there are less than 5 points for each value of h that when projected lie on the surface, then the number of points on the surface is bounded by $(2N+1) \cdot 4 \lesssim N$. Otherwise, there is some value of h such that at least 5 points lie on the lattice corresponding to that h. If some 3 points lie on a straight line then the curve is that line. If not, the 5 points are in general linear position, and determine a unique conic that passes through them.

Given this, we claim no points from other values of h can lie on these lines or conics. These points $(x, y, 0)$ are such that $x - ha_3 \in \mathbb{Q}(a_1, a_2)$ and $y - hb_3 \in \mathbb{Q}(b_1, b_2)$. The coefficients of the equation can be determined by linear elimination, giving a nonzero polynomial of the form $P(x - ha_3, y - hb_3) \in \mathrm{Poly}_2(\mathbb{Q}(a_1, a_2, b_1, b_2))$.

For points with other values of h, this polynomial equation gives a non-trivial polynomial equation involving $\mathbb{Q}$,a_i,b_i, which is impossible, as the a_i and b_i are chosen to be algebraically independent. Therefore, if 5 points can be found for some value for h, no points with other values of h can lie on this surface.

Hence, it suffices to check the number of points for this fixed value of h that can lie on the surface. Each line can only intersect the surface at at most 2 points and we can split the lattice into $(2N + 1)$ axis-parallel lines, giving a bound of $2(2N + 1) \lesssim N$ points on the surface. Combining these cases, we find that the number of integer points that lie on any surface and $z = 0$ is bounded by $\lesssim N$.

This gives us $\lesssim N$ choices for $\mathbf{x}$. Similarly, for each $\mathbf{x}$, we can consider how many choices we have for $\mathbf{v}$ such that $(\mathbf{x} + \mathbf{v}, 1)$ lies on the same surface. This is also the projection of a 3 dimensional integer lattice into a plane, and using the argument above we find that we have $\lesssim N^c$ choices for $\mathbf{v}$. Thus, we have $r \lesssim N^{1+c}$, which aligns with Proposition 2.9 when $k = 2$.

Given these bounds on $\log m$ and r we use Lemma 2.6 as mentioned above. This finishes similarly to Proposition 2.9, proving the proposition. A minor difference with Theorem 2.1 is that $c \geq 1/2$ was specified to guarantee that $L^{\frac{2c-1}{3}} \gtrsim 1$. □

3 Bounds in General Vector Spaces

We now consider sets of lines in $\mathbb{F}^n$, for any field $\mathbb{F}$, while restricting the number of lines in any 3-flat.

The main result is the following.

Theorem 3.1. *Let $\mathcal{L}$ be a set of lines in $\mathbb{F}^n$ with at most B lines in any 3-flat. Then,*

$$|P_2(\mathcal{L})| \lesssim \min(B^{1/2}L^{3/2}, BL + L^{5/3}).$$

We can visualize this bound as in Fig. 2.

Remark. Note that if $\mathbb{F}$ is an infinite field, a construction of $|P_2(\mathcal{L})| \sim LB$ is achieved by simply taking L/B 2-flats, each containing B lines, similar to the situation in $\mathbb{C}^3$. Furthermore, in this case, the bound in Theorem 3.1 is tight (up to a constant factor) when $B \gtrsim L^{2/3}$.

We first prove the following lemma:

Lemma 3.2. *For any two lines ℓ_1, ℓ_2, there exists a 3-flat containing all lines ℓ that intersect ℓ_1 and ℓ_2 at distinct points.*

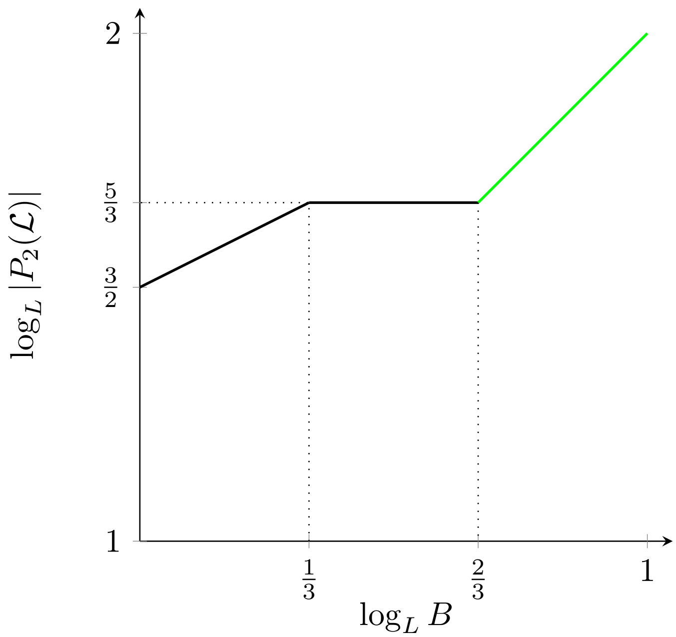

Figure 2: The curve shows the upper bound given by Theorem 3.1. The green segment, where $\log_L B \geq \frac{2}{3}$ shows the interval where the upper bound is known to be sharp.

Proof. If ℓ_1 and ℓ_2 lie in the same 2-flat, any line intersecting both of them must also lie in this 2-flat, and we can take any 3-flat containing it.

Consider now the case when ℓ_1 and ℓ_2 do not lie in the same 2-flat. Suppose the lines are given by $\ell_i = \{a_i + tb_i\}$. Then, the desired 3-flat is

$$\{\alpha_1 a_1 + \alpha_2 a_2 + \beta_1 b_1 + \beta_2 b_2 \mid \alpha_1 + \alpha_2 = 1\}.$$

In particular, any line ℓ intersecting ℓ_1 and ℓ_2 at distinct points contains two points of this form (which are the intersection points), so all points of ℓ have this form. $\square$

To prove Theorem 3.1, we interpret the intersections of the lines as a graph. In particular, consider a graph G whose vertices are the lines of $\mathcal{L}$. For each point P of $P_2(\mathcal{L})$, take two lines ℓ_1 and ℓ_2 that intersect at P, and add an edge corresponding to P between ℓ_1 and ℓ_2 in G. (If more than two lines meet at a point, pick two of those lines arbitrarily to add an edge.)

We prove the two parts of the bound separately.

Proposition 3.3. *If $\mathcal{L}$ is a set of lines in $\mathbb{F}^n$ with at most B lines in any 3-flat, then $|P_2(\mathcal{L})| \lesssim B^{1/2} L^{3/2}$.*

Proof. Define the graph G as above. Let d_ℓ be the degree of ℓ in G. The number of edges of G is then $|P_2(\mathcal{L})|$, so $\sum_{\ell\in\mathcal{L}} d_\ell = 2\,|P_2(\mathcal{L})|$.

By Lemma 3.2, for any two distinct vertices ℓ_1, ℓ_2, there exist at most B vertices ℓ adjacent to both of them, as any such ℓ intersects ℓ_1 and ℓ_2 at distinct points, and hence lies in a fixed 3-flat.

We now double-count triples (ℓ, ℓ_1, ℓ_2) such that ℓ is adjacent to both ℓ_1 and ℓ_2 in G. On one hand, counting by ℓ, the number of such triples is

$$\sum_{\ell\in\mathcal{L}} \binom{d_\ell}{2} \geq L\binom{\sum_{\ell\in\mathcal{L}} d_\ell/L}{2} \sim L\left(\frac{|P_2(\mathcal{L})|}{L}\right)^2 = \frac{|P_2(\mathcal{L})|^2}{L}.$$

On the other hand, counting by pairs (ℓ_1, ℓ_2) upper bounds the number of lines by $B\binom{L}{2} \sim BL^2$.

Hence,

$$\frac{|P_2(\mathcal{L})|^2}{L} \lesssim BL^2,$$

which gives the desired bound. □

Proposition 3.4. *If $\mathcal{L}$ is a set of lines in $\mathbb{F}^n$ with at most $B \gtrsim L^{2/3}$ lines in any 3-flat, then $|P_2(\mathcal{L})| \lesssim BL$.*

Proof. The following argument is rather similar to parts of the "contagious vanishing argument" made by Guth in [Gut16, Chapter 11.3].

Define the graph $G(V, E)$ as above. Take $A = P_2(\mathcal{L})L^{-1}$ and split into two cases: when there is some vertex with degree less than A, and when all vertices have degree at least A.

The easier case is if there is some vertex ℓ with degree less than A. Take $\mathcal{L}' = \mathcal{L} \setminus \{\ell\}$. By the inductive hypothesis, we know that $P_2(\mathcal{L}') \leq KB(L-1)$ for some constant K. Then,

$$P_2(\mathcal{L}) \leq A + KB(L-1) \leq \frac{P_2(\mathcal{L})}{L} + KB(L-1)$$

which implies $P_2(\mathcal{L}) \leq KBL$ as desired.

In the other case, all vertices have degree at least $A = P_2(\mathcal{L})L^{-1}$. Suppose $A \geq KL^{2/3}$ for sufficiently large constant K. We claim that there are then $\geq A/2$ lines that lie on a 3-flat.

In Proposition 3.3, we prove that the number of triples (ℓ, ℓ_1, ℓ_2) such that ℓ is adjacent to both ℓ_1 and ℓ_2 in G is at least $|P_2(\mathcal{L})|^2/L$. We also know that the number of unique pairs (ℓ_1, ℓ_2) is $\sim L^2$.

Hence there is some pair (ℓ_1, ℓ_2) such that there are $\gtrsim |P_2(\mathcal{L})|^2/L^3 = A^2/L$ lines ℓ that intersect both lines. In particular since $A \gtrsim L^{2/3}$, we have $\gtrsim L^{1/3}$

lines that intersect both ℓ_1 and ℓ_2. By Lemma 3.2, we know that ℓ_1, ℓ_2 and the $\gtrsim L^{1/3}$ lines that intersect both all lie in a 3-flat. We use this as a base case, and inductively find lines of $\mathcal{L}$ that are in the 3-flat until the number of lines in it is $\geq A/2$.

It suffices to show that if there is a subset $S \subset \mathcal{L}$ such that $L^{1/3} \lesssim |S| < A/2$ that is contained in a 3-flat, then there is always another line $\ell \in \mathcal{L} \setminus S$ that intersects two lines in S, and hence is contained in the same 3-flat.

Consider the number of edges between S and $\mathcal{L} \setminus S$. Each vertex in S has degree at least A, and at most $|S|$ of these edges are to other lines in S. This gives a lower bound on the number of edges between S and $\mathcal{L} \setminus S$ in G of

$$|S|\,(A - |S|) \geq |S|\,(A - A/2) \gtrsim AL^{1/3}.$$

When $K = A/L^{2/3}$ is chosen to be sufficiently large, the number of edges between S and $\mathcal{L} \setminus S$ is more than L. By the Pigeonhole Principle, there is some line in $\mathcal{L} \setminus S$ that intersects two lines in S, finishing the inductive step. This gives $\geq A/2$ lines in a 3-flat, and thus $B \geq A/2$. Hence, when $A \geq KL^{2/3}$,

$$B \gtrsim \frac{P_2(\mathcal{L})}{L}. \qquad \square$$

4 Further Questions and Discussion

The construction in Section 2.2 can be done in general dimension n, and a random projection can be performed onto the space $\mathbb{C}^3$ to get a set of lines with many 2-rich points. However, performing the construction with fixed $k = 2$ and dimension $n > 4$ result in worse constructions.

Using the result of Theorem 2.1, we see that as c varies from $(n-2)/(n-1)$ to 1, $\log_L B$ varies linearly from 0 to $1/(n-1)$ and $\log_L(|P_2(\mathcal{L})|)$ varies linearly from $1 + (n-2)/(n-1)^2$ to $1 + 1/(n-1)$. Since these two endpoints for any $n > 4$ are contained inside the convex region formed by $(0, 1), (1/3, 4/3), (0, 7/6)$ in Fig. 1, they give weaker constructions.

If we are to instead consider lines in space $\mathbb{C}^m$ for $m > 3$ and restrict the number of lines on a k-flat with $m > k > 2$, we are still able to project a construction on $\mathbb{C}^n$ into $\mathbb{C}^m$. The strength of our construction does not depend on m, since the projection will not increase the upper bounds on lines in a k-flat.

As an example, take $k = 3$. We see that the optimal dimension n to pick can differ based on c. Consider small values of dimension n.

- $n = 4, 5$ do not give a better construction than the standard LB does.
- $n = 6$ has $\log_L B$ vary linearly from 0 to $2/5$ with $\log_L |P_2(\mathcal{L})|$ varying linearly from $1 + 1/15$ to $1 + 1/5$. In particular, this bound is weaker than the standard LB construction when $\log_L B \geq 1/10$.

- $n = 7$ has $\log_L B$ vary linearly from 0 to $1/3$ with $\log_L |P_2(\mathcal{L})|$ varying linearly from $1 + 1/12$ to $1 + 1/6$. This bound is weaker than the LB construction when $\log_L B \geq 1/9$.
- $n = 8$ has $\log_L B$ vary linearly from 0 to $2/7$ with $\log_L |P_2(\mathcal{L})|$ varying linearly from $1 + 3/35$ to $1 + 1/7$. This bound is weaker than the LB construction when $\log_L B \geq 3/28$.

We see that when we fix $k = 3$, among $n = 4, 5, 6, 7, 8$, $n = 8$ gives the best bound when $\log_L B$ is close to 0 since $3/35 > 1/12 > 1/15$, but only $n = 7$ gives a better construction than the LB construction in the regime $3/28 \leq \log_L B \leq 1/9$.

In $\mathbb{C}^3$, this construction and the Guth-Katz result of Theorem 1.4 leaves a gap $1/3 \leq \log_L B < 1/2$ where the best known construction is $\sim LB$ but the best known bound is $\sim L^{3/2}$. (See Fig. 1). Given the limitations of this cubic lattice structure, better constructions seems difficult. It may be possible to use a different pseudo-regular structure, such as a fractal, to achieve a better bound.

It is also interesting to consider whether there are better constructions in the setting of Section 3, in which we restrict the number of lines on a 3-flat, where we again have a significant gap between the best constructions and best upper bounds, particularly for small values of B.

References

[Dvi09] Zeev Dvir. On the size of Kakeya sets in finite fields. *J. Amer. Math. Soc.*, 22(4):1093–1097, 2009.

[ES11] György Elekes and Micha Sharir. Incidences in three dimensions and distinct distances in the plane. *Combin. Probab. Comput.*, 20(4):571–608, 2011.

[GK10] Larry Guth and Nets Hawk Katz. Algebraic methods in discrete analogs of the Kakeya problem. *Adv. Math.*, 225(5):2828–2839, 2010.

[GK15] Larry Guth and Nets Hawk Katz. On the Erdős distinct distances problem in the plane. *Ann. of Math. (2)*, 181(1):155–190, 2015.

[Gut16] Larry Guth. *Polynomial methods in combinatorics*, volume 64 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2016.

[ST83] Endre Szemerédi and William T. Trotter, Jr. Extremal problems in discrete geometry. *Combinatorica*, 3(3-4):381–392, 1983.

[ST12] József Solymosi and Terence Tao. An incidence theorem in higher dimensions. *Discrete Comput. Geom.*, 48(2):255–280, 2012.

[Wal23] Miguel N. Walsh. Concentration estimates for algebraic intersections. *Amer. J. Math.*, 145(2):435–475, 2023.