INCIDENCE PROBLEMS IN DISCRETE GEOMETRY

BY CHARLES WOLF

A dissertation submitted to the
Graduate School—New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy
Graduate Program in Mathematics
Written under the direction of
Shubhangi Saraf
and approved by

____

____

____

____

New Brunswick, New Jersey
May, 2017

ABSTRACT OF THE DISSERTATION

Incidence Problems in Discrete Geometry

by CHARLES WOLF

Dissertation Director: Shubhangi Saraf

Over the past decade, discrete geometry research has flourished with clever uses of algebraic methods. The polynomial method has had a deep impact on a wide collection of results in combinatorics, such as tight asymptotic lower bounds on finite field Kakeya and Nikodym sets, near optimal lower bound for Erdos' distinct distances problem, and improved bounds for cap sets. Spectral methods and rank bounds for matrices have shed new light on improved bounds for point-line incidences, subspace intersections and graph rigidity. This thesis is focused on developing new ways to improve on these techniques, and applying them in a few discrete geometry settings:

- Techniques such as matrix scaling and rank bounds for design matrices have recently found beautiful applications for understanding configurations of points and lines over the complex numbers ([BDWY13],[DSW14]). In particular, they give a different proof of Kelly's Theorem, which says that any configuration of points in complex space must either be contained in a plane or have a line passing through exactly 2 of those points, called an ordinary line. We expand on these techniques to prove the first quantitative bounds for the number of ordinary lines in a non planar configuration of points in complex space.

- In 2008, Dvir [Dvi09] showed in a breakthrough result that a Kakeya set over finite fields has an asymptotically tight lower bound using the polynomial method.

Later in 2008, Saraf and Sudan [SS08] improved on the polynomial method by interpolating a polynomial that vanishes with high multiplicity on points of the Kakeya set. We further enhance the polynomial method by introducing the notion of “fractional multiplicity,” and use this improvement to obtain a better lower bound for finite field Kakeya sets in 3 dimensions.

- While studying these 3-dimensional finite field Kakeya sets, we considered the related 3-dimensional finite field Nikodym sets. Previously, the lower bound for a 3-dimensional finite field Nikdoym sets was also obtained using the polynomial method, and had the same lower bound as for a Kakeya set. We achieve a better lower bound for 3-dimensional finite field Nikodym sets, thus separating it from the Kakeya set lower bound.

Acknowledgements

First, I would like to express my tremendous appreciation to my advisor Shubhangi Saraf for developing me as a researcher. Thank you for leading me in the right direction throughout our mathematical discussions in a way that encourages my confidence and curiosity. Also not overlooked was your maintaining kindness and a smile throughout our meetings. I would also like to thank Professors Swastik Kopparty, Jeffry Kahn, and Michael Saks for wonderful courses and discussions in combinatorics. Special thanks to Mike for his additional career and life advice.

I must commend my fellow graduate students at Rutgers for creating a wonderful learning environment. Sjuvon, Ed, Howie, Pat and Tim, thank you for making our office a welcoming space of inquiry and positivity. Ben and Basit, I enjoyed discussing and writing our results together, and I look forward to more joint work in the future! To the rest, including the frequent office visitors, combinatorics group, coffee hour regulars, and aerobie players, you have made this an amazing experience!

Finally, I attribute much of my graduate school success to my family. Abba, I look forward to all of our conversations, and I continue to value each bit of advice you provide. Tzvi, thanks for staying your fun self, yet assuming major responsibilities during tough times. Thank you Mom and Dad Leven, plus Saul, Jacob and Chaya, for your support and hospitality. Last and most precious, Rachel has been so much for me throughout this process: encouraging, consoling, entertaining, nurturing, and always available. Thank you for being by my side in this journey, and I can't wait to see what our future entails!

Dedication

To my mother, of blessed memory

Table of Contents

Chapter 1

Introduction

Discrete geometry problems are attributed to mathematicians as early as the ancient Greeks. Recently, discrete geometry research has been reinvigorated via algebraic applications. Many of these applications take advantage of the already present geometric correlations. In 2008, Dvir [Dvi09] showed in a breakthrough result that a Kakeya set over finite fields has an asymptotically tight lower bound using a technique called the polynomial method. This technique has been applied to several other problems, such as achieving near optimal lower bound for Erdos' distinct distances problem and improving bounds for cap sets. Around the same time, aspects of matrices, such as spectral methods and rank bounds, have shed new light on improved bounds for point-line incidences, subspace intersections and graph rigidity. In this thesis, we present a few applications of polynomial and matrix techniques to problems in discrete geometry. We will briefly summarize the results of this thesis in the following 3 sections:

1.1 Counting Ordinary Lines

Given n points in $\mathbb{R}^2$, define an *ordinary* line to be a line passing through exactly 2 of those points. Sylvester asked if there is a configuration of points with no ordinary lines, with the exception of all n points lying on a line. In 1944, Gallai proved [Gal44] that any configuration of n points not all on a line must contain at least one ordinary line. More recently in 2013, Green and Tao showed [GT13] that such a configuration contains at least $n/2$ ordinary lines.

While there exists an ordinary line in noncollinear configurations over $\mathbb{R}^2$, this is not the case over $\mathbb{C}^2$. Kelly's theorem states that a set of n points affinely spanning $\mathbb{C}^3$ must determine at least one ordinary line. Recently, Kelly's theorem was reproved using matrix scaling and rank bound techniques ([DSW14]). We expand on these techniques to show that a noncoplanar set in $\mathbb{C}^3$ determines at least $3n/2$ ordinary lines, unless the configuration has $n-1$ points in a plane and one point outside the plane (in which case there are at least $n-1$ ordinary lines). In addition, when at most $2n/3$ points are contained in any plane, we prove a theorem giving stronger bounds that take advantage of the existence of lines with 4 and more points (in the spirit of Melchior's and Hirzebruch's inequalities). Furthermore, when the points span 4 or more dimensions, with at most $2n/3$ points contained in any three dimensional affine subspace, we show that there must be a quadratic number of ordinary lines.

1.2 Finite Field Kakeya Sets in 3 Dimensions

Let $\mathbb{F}_q$ denote the finite field of q elements. A *Kakeya set* $K \subseteq \mathbb{F}_q^n$ is a set of points which contains 'a line in every direction'. More precisely, for all $x \in \mathbb{F}_q^n$ there is a $y \in \mathbb{F}_q^n$ such that the line[1] $\{xt + y, t \in \mathbb{F}_q\} \subseteq K$. In 2008, Dvir [Dvi09] showed in a breakthrough result that a Kakeya set in $\mathbb{F}_q^n$ has size at least $\frac{q^n}{n!}$ using the polynomial method. Later in 2008, Saraf and Sudan [SS08] improved the lower bound to the form $c^n q^n$, for some fixed constant $c < \frac{1}{2}$. In particular, for $n = 3$ they achieve a lower bound of $(0.208)q^3$. They refined the polynomial method by interpolating a polynomial that vanishes with high multiplicity on points of the Kakeya set. Our work [LSW16] introduces a notion that allows the polynomial to vanish with different multiplicities at different points of the Kakeya set. This technique is referred to as "fractional multiplicity," as the average multiplicity at each point is not necessarily an integer. This allows us to achieve an improved lower bound on Kakeya sets in $\mathbb{F}_q^3$.

1.3 Finite Field Nikodym Sets in 3 Dimensions

Related to a Kakeya set, a *Nikodym set* $\mathcal{N} \subseteq \mathbb{F}_q^n$ is a set of points such that, through each point $p \in \mathbb{F}_q^n$, there is a line ℓ such that $\ell \setminus \{p\} \subseteq \mathcal{N}$. We use spectral techniques to give improved lower bounds on the size of Nikodym sets over $\mathbb{F}_q^3$. We also propose a

[1] A *line* is an affine subspace of dimension 1.

natural conjecture on the minimum number of points in the union of a not-too-flat set of lines in $\mathbb{F}_q^3$, and show that this conjecture implies an optimal bound on the size of a Nikodym set. Finally, we study the notion of a weak Nikodym set and give improved, and in some special cases optimal, bounds for weak Nikodym sets in $\mathbb{F}_q^2$ and $\mathbb{F}_q^3$.

1.4 Organization of this Thesis

In Chapter 2, we show how to obtain lower bounds on the number of ordinary lines in complex space. In Section 2.2 we develop the necessary machinery on matrix scaling and Latin squares. In Sections 2.3 and 2.5, we prove some key lemmas that will be used in the proofs of our main results. Sections 2.4 and 2.6 give the proof of the main results. In Chapter 3, we give an improved lower bound for finite field Kakeya sets in 3 dimensions. In Section 3.2 we give preliminary lemmas and results to set up for the proof, and we give the proof in Section 3.3.

In Chapter 4, we discuss lower bounds for finite field Nikodym sets in 3 dimensions. In Section 4.2.1, we prove an improved lower bound for finite field Nikodym sets in 3 dimensions; as mentioned in the introduction, this is the first separation demonstrated between the minimum size of a Nikodym set and the minimum size of a Kakeya set in $\mathbb{F}_q^3$ that is valid for an arbitrary finite field $\mathbb{F}_q$. In Section 4.2.2, we show that the this improved lower bound immediately implies a lower bound on the number of points incident to a large set of lines, and that this bound is nearly tight. This implies that any substantial improvement to the lower bound will need to use some property of Nikodym sets that is not exploited by the proof given in Section 4.2.1. In Section 4.2.3, we observe that a weak Nikodym set has the property that not too many of the lines given by its definition can lie in any single plane. We make a conjecture about the size of point sets with not too many lines in any plane, and show that the proof of this conjecture would imply the main conjecture of finite field Nikodym sets in 3 dimensions.

Chapter 2

Counting Ordinary Lines in Complex Space

2.1 Introduction

Let $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ be a set of n points in $\mathbb{C}^d$. We denote by $\mathcal{L}(\mathcal{V})$ the set of lines determined by points in $\mathcal{V}$, and by $\mathcal{L}_r(\mathcal{V})$ (resp. $\mathcal{L}_{\geq r}(\mathcal{V})$) the set of lines in $\mathcal{L}(\mathcal{V})$ that contain exactly (resp. at least) r points. Let $t_r(\mathcal{V})$ denote the size of $\mathcal{L}_r(\mathcal{V})$. Throughout the write-up we omit the argument $\mathcal{V}$ when the context makes it clear. We refer to $\mathcal{L}_2$ as the set of *ordinary lines*, and $\mathcal{L}_{\geq 3}$ as the set of *special lines.*

A well known result in combinatorial geometry is the Sylvester-Gallai theorem.

Theorem 2.1.1 (Sylvester-Gallai theorem). *Let $\mathcal{V}$ be a set of n points in $\mathbb{R}^2$ not all on a line. Then there exists an ordinary line determined by points of $\mathcal{V}$.*

The statement was conjectured by Sylvester in 1893 [Syl93] and first proved by Melchior [Mel40]. It was later reproved by Gallai in 1944 [Gal44], and there are now several different proofs of the theorem. Of particular interest is the following result by Melchior [Mel40].

Theorem 2.1.2 (Melchior's inequality). *Let $\mathcal{V}$ be a set of n points in $\mathbb{R}^2$ that are not collinear. Then*

$$t_2(\mathcal{V}) \geq 3 + \sum_{r \geq 4} (r-3) t_r(\mathcal{V}).$$

Theorem 2.1.2 in fact proves something stronger than the Sylvester-Gallai theorem, i.e. there are at least three ordinary lines. A natural question to ask is how many ordinary lines must a set of n points, not all on a line, determine. This led to what is known as the *Dirac-Motzkin conjecture.*

Conjecture 2.1.1 (Dirac-Motzkin conjecture)**.** *Let $\mathcal{V}$ be a set of n points in $\mathbb{R}^2$, not all on a line. Suppose that $n \geq n_0$ for a sufficiently large absolute constant n_0. Then $\mathcal{V}$ determines at least $n/2$ ordinary lines.*

There were several results on this question (see [Mot51, KM58, CS93]), before it was completely resolved by Green and Tao [GT13].

Theorem 2.1.3 (Green-Tao)**.** *Let $\mathcal{V}$ be a set of n points in $\mathbb{R}^2$, not all on a line. Suppose that $n \geq n_0$ for a sufficiently large absolute constant n_0. Then $t_2(\mathcal{V}) \geq \frac{n}{2}$ for even n and $t_2(\mathcal{V}) \geq \lfloor \frac{3n}{4} \rfloor$ for odd n.*

[GT13] provides a nice history of the problem, and there are several survey articles on the topic, see for example [BM90].

The Sylvester-Gallai theorem is not true when the field $\mathbb{R}$ is replaced by $\mathbb{C}$. In particular, the well known Hesse configuration, realized by the 9 inflection points of a non-degenerate cubic, provides a counter example. A more general example is the following:

Example 1 (Fermat configuration)**.** *For any positive integer $k \geq 3$, let $\mathcal{V}$ be inflection points of the Fermat Curve $X^k + Y^k + Z^k = 0$ in $\mathbb{PC}^2$. Then $\mathcal{V}$ has $n = 3k$ points, and in particular*

$$\mathcal{V} = \bigcup_{i=1}^{k} \{[1 : \omega^i : 0]\} \cup \{[\omega^i : 0 : 1]\} \cup \{[0 : 1 : \omega^i]\},$$

where ω is the k^{th} root of -1.

It is easy to check that $\mathcal{V}$ determines 3 lines containing k points each, while every other line contains exactly 3 points. In particular, $\mathcal{V}$ determines no ordinary lines.[1]

In response to a question of Serre [Ser66], Kelly [Kel86] showed that when the points span more than 2 dimensions, the point set must determine at least one ordinary line.

Theorem 2.1.4 (Kelly's theorem)**.** *Let $\mathcal{V}$ be a set of n points in $\mathbb{C}^3$ that are not contained in a plane. Then there exists an ordinary line determined by points of $\mathcal{V}$.*

[1] *We note that the while Fermat configuration as stated lives in the projective plane, it can be made affine by any projective transformation that moves a line with no points to the line at infinity.*

Kelly's proof of Theorem 2.1.4 used a deep result of Hirzebruch [Hir83] from algebraic geometry. In particular, it used the following result, known as Hirzebruch's inequality.

Theorem 2.1.5 (Hirzebruch's inequality)**.** *Let $\mathcal{V}$ be a set of n points in $\mathbb{C}^2$, such that $t_n(\mathcal{V}) = t_{n-1}(\mathcal{V}) = t_{n-2}(\mathcal{V}) = 0$. Then*

$$t_2(\mathcal{V}) + \frac{3}{4}t_3(\mathcal{V}) \geq n + \sum_{r \geq 5}(2r-9)t_r(\mathcal{V}).$$

More elementary proofs of Theorem 2.1.4 were given in [EPS06] and [DSW14]. To the best of our knowledge, no lower bound greater than 1 is known for the number of ordinary lines determined by point sets spanning $\mathbb{C}^3$. Improving on the techniques of [DSW14], we make the first progress in this direction.

Theorem 2.1.6. *Let $\mathcal{V}$ be a set of $n \geq 24$ points in $\mathbb{C}^3$ not contained in a plane. Then $\mathcal{V}$ determines at least $\frac{3}{2}n$ ordinary lines, unless $n-1$ points are on a plane in which case there are at least $n-1$ ordinary lines.*

Clearly if $n-1$ points are coplanar, it is possible to have only $n-1$ ordinary lines. In particular, let $\mathcal{V}$ consist of the Fermat Configuration, for some $k \geq 3$, on a plane and one point v not on the plane. Then $\mathcal{V}$ has $3k+1$ points, and the only ordinary lines determined by $\mathcal{V}$ are lines that contain v, so there are exactly $3k$ ordinary lines. We are not aware of any examples that achieve the $\frac{3}{2}n$ bound when at most $n-2$ points are contained in any plane.

When $\mathcal{V}$ is sufficiently non-degenerate, i.e. no plane contains too many points, we are able to give a more refined bound in the spirit of Melchior's and Hirzebruch's inequalities, taking into account the existence of lines with more than three points. In particular, we show the following:

Theorem 2.1.7. *There exists an absolute constant $c > 0$ and a positive integer n_0 such that the following holds. Let $\mathcal{V}$ be a set of $n \geq n_0$ points in $\mathbb{C}^3$ with at most $\frac{2}{3}n$ points contained in any plane. Then*

$$t_2(\mathcal{V}) \geq \frac{3}{2}n + c\sum_{r \geq 4} r^2 t_r(\mathcal{V}).$$

the constant $2/3$ is arbitrary and can be replaced by any number smaller than 1. Suppose that $\mathcal{V}$ consists of $n-k$ points on a plane, and k points not on the plane. There are at least $n-k$ lines through each point not on the plane, at most $k-1$ of which could contain 3 or more points. So we see that there are at least $k(n-2k)$ ordinary lines determined by $\mathcal{V}$. Then if $k = \epsilon n$, for $0 < \epsilon < 1/2$, $\mathcal{V}$ has $\Omega_\epsilon(n^2)$ ordinary lines, where the hidden constant depends on ϵ. Therefore, the bound in Theorem 2.1.7 is only interesting when no plane contains too many points.

On the other hand, we note that having at most a constant fraction of the points on any plane is necessary to obtain a bound of this form. Indeed, let $\mathcal{V}$ consist of the Fermat Configuration for some $k \geq 3$ on a plane and $o(k)$ points not on the plane. Then $\mathcal{V}$ has $O(k)$ points and determines $o(k^2)$ ordinary lines. On the other hand, $\sum_{r\geq 4} r^2 t_r(\mathcal{V}) = \Omega(k^2)$.

Hirzebruch's inequality (which also gives a bound in $\mathbb{C}^3$, though without requiring that every plane contains not-too-many points) only gives a lower bound on $t_2(\mathcal{V}) + \frac{3}{4}t_3(\mathcal{V})$, whereas both Theorems 2.1.6 and 2.1.7 give lower bounds on the number of ordinary lines, i.e. $t_2(\mathcal{V})$. Another important contribution of Theorem 2.1.7 is replacing the linear $(2r-9)$ in Hirzebruch's inequality with a term quadratic in r. We also note that lines with 4 points do not play any role in Hirzebruch's inequality, where the summation starts at $r = 5$. This is not the case for Theorem 2.1.7. As a consequence, if a non-planar configuration over $\mathbb{C}$ has many lines with 4 points each, then it must have many ordinary lines.

Finally, when a point set $\mathcal{V}$ spans 4 or more dimensions in a sufficiently non-degenerate manner, i.e. no 3 dimensional affine subspace contains too many points, we prove that there must be quadratic number of ordinary lines.

Theorem 2.1.8. *There exists a positive integer n_0 such that the following holds. Let $\mathcal{V}$ be a set of $n \geq n_0$ points in $\mathbb{C}^4$ with at most $\frac{2}{3}n$ points contained in any 3 dimensional affine subspace. Then*

$$t_2(\mathcal{V}) \geq \frac{1}{12}n^2.$$

Here also the constant $2/3$ is arbitrary and can be replaced by any positive constant less than 1. However, increasing this constant will shrink the constant 1/12 in front of n^2. Also, a quadratic lower bound may be possible if at most $\frac{2}{3}n$ points are contained in any 2 dimensional space, but we have no proof or counterexample.

Note that while we state Theorems 2.1.6 and 2.1.7 over $\mathbb{C}^3$ and Theorem 2.1.8 over $\mathbb{C}^4$, the same bounds hold in higher dimensions as well since we may project a point set in $\mathbb{C}^d$ onto a generic lower dimensional subspace, preserving the incidence structures. In addition, while these theorems are proved over $\mathbb{C}$, these results are also new and interesting over $\mathbb{R}$.

2.2 Preliminaries

2.2.1 Matrix Scaling and Rank Bounds

One of the main ingredients in our proof is rank bounds for design matrices. These techniques were first used for incidence type problems in [BDWY13] and improved upon in [DSW14]. We first set up some notation. For a complex matrix A, let A^* denote the matrix conjugated and transposed. Let A_{ij} denote the entry in the i^{th} row and j^{th} column of A. For two complex vectors $u, v \in \mathbb{C}^d$, we denote their inner product by $\langle u, v\rangle = \sum_{i=1}^{d} u_i \cdot \overline{v_i}$.

Central to obtaining rank bounds for matrices is the notion of matrix scaling. We now introduce this notion and provide some definitions and lemmas.

Definition 2.2.1 (Matrix Scaling)**.** *Let A be an $m \times n$ matrix over some field $\mathbb{F}$. For every $\rho \in \mathbb{F}^m, \gamma \in \mathbb{F}^n$ with all entries nonzero, the matrix A' with $A'_{ij} = A_{ij} \cdot \rho_i \cdot \gamma_j$ is referred to as a scaling of A. Note that two matrices that are scalings of each other have the same rank.*

We will be interested in scalings of matrices that control the row and column sums. The following property provides a sufficient condition under which such scalings exist.

Definition 2.2.2 (Property-S)**.** *Let A be an $m \times n$ matrix over some field. We say that A satisfies Property-S if for every zero submatrix of size $a \times b$, we have*

$$\frac{a}{m} + \frac{b}{n} \leq 1.$$

The following theorem is given in [RS89].

Theorem 2.2.3 (Matrix Scaling theorem)**.** *Let A be an $m \times n$ real matrix with non-negative entries satisfying Property-S. Then, for every $\epsilon > 0$, there exists a scaling A' of A such that the sum of every row of A' is at most $1+\epsilon$, and the sum of every column of A' is at least $m/n-\epsilon$. Moreover, the scaling coefficients are all positive real numbers.*

We may assume that the sum of every row of the scaling A' is exactly $1 + \epsilon$. Otherwise, we may scale the rows to make the sum $1 + \epsilon$, and note that the column sums can only increase.

The following Corollary to Theorem 2.2.3 appeared in [BDWY13].

Corollary 2.2.4 (ℓ_2 scaling)**.** *Let A be an $m \times n$ complex matrix satisfying Property-S. Then, for every $\epsilon > 0$, there exists a scaling A' of A such that for every $i \in [m]$*

$$\sum_{j\in[n]} \left|A'_{ij}\right|^2 \leq 1+\epsilon,$$

and for every $j \in [n]$

$$\sum_{i\in[m]} \left|A'_{ij}\right|^2 \geq \frac{m}{n} - \epsilon$$

Moreover, the scaling coefficients are all positive real numbers.

Corollary 2.2.4 is obtained by applying Theorem 2.2.3 to the matrix obtained by squaring the absolute values of the entries of the matrix A. Once again, we may assume that $\sum_{j\in[n]} |A'_{ij}|^2 = 1 + \epsilon$.

To bound the rank of a matrix A, we will bound the rank of the matrix $M = A'^{*}A'$, where A' is some scaling of A. Then we have that $\text{rank}(A) = \text{rank}(A') = \text{rank}(M)$. We use Corollary 2.2.4, along with rank bounds for diagonal dominant matrices. The following lemma is a variant of a folklore lemma on the rank of diagonal dominant matrices (see [Alo09]) and appeared in this form in [DSW14].

Lemma 2.2.5. *Let A be an $n \times n$ complex hermitian matrix, such that $|A_{ii}| \geq L$ for all $i \in n$. Then*

$$\mathit{rank}(A) \geq \frac{n^2L^2}{nL^2 + \sum_{i\neq j} |A_{ij}|^2}.$$

The matrix scaling theorem allows us to control the ℓ_2 norms of the columns and rows of A, which in turn allow us to bound the sums of squares of entries of M. For this, we use a variation of a lemma from [DSW14]. While the proof idea is the same, our proof requires a somewhat more careful analysis. Before we provide the lemma, we need some definitions.

Definition 2.2.6. *Let A be an $m \times n$ matrix over $\mathbb{C}$. Then we define:*

$$D(A) := \sum_{i \neq j} \sum_{k<k'} \left| A_{ki}\overline{A_{kj}} - A_{k'i}\overline{A_{k'j}} \right|^2 ,$$

and

$$E(A) := \sum_{k=1}^{m} \sum_{i<j} \left(|A_{ki}|^2 - |A_{kj}|^2 \right)^2 .$$

Note that both $D(A)$ and $E(A)$ are non-negative real numbers.

Lemma 2.2.7. *Let A be an $m \times n$ matrix over $\mathbb{C}$. Suppose that each row of A has ℓ_2 norm α, the supports of every two columns of A intersect in exactly t locations, and the size of the support of every row is q. Let $M = A^*A$. Then*

$$\sum_{i \neq j} |M_{ij}|^2 = \left(1 - \frac{1}{q}\right) tm\alpha^4 - \left(D(A) + \frac{t}{q}E(A)\right).$$

Proof. Note that

$$\begin{aligned}\sum_{i \neq j} |M_{ij}|^2 &= \sum_{i \neq j} |\langle C_i, C_j \rangle|^2 \\ &= \sum_{i \neq j} \left| \sum_{k=1}^{m} A_{ki}\overline{A_{kj}} \right|^2 .\end{aligned}$$

Since the supports of any two columns of A intersect in exactly t locations, the Cauchy-Schwarz inequality shows that $\left|\sum_{k=1}^{m} A_{ki}\overline{A_{kj}}\right|^2 \leq t\sum_{k=1}^{m} |A_{ki}|^2|A_{kj}|^2$. Our approach requires somewhat more careful analysis, so we use the following equality:

$$\begin{aligned}\sum_{i \neq j} \left| \sum_{k=1}^{m} A_{ki}\overline{A_{kj}} \right|^2 &= \sum_{i \neq j} \left(t\sum_{k=1}^{m} |A_{ki}|^2|A_{kj}|^2 - \sum_{k<k'} \left| A_{ki}\overline{A_{kj}} - A_{k'i}\overline{A_{k'j}} \right|^2 \right) \\ &= t\sum_{i \neq j}\sum_{k=1}^{m} |A_{ki}|^2|A_{kj}|^2 - D(A) \\ &= t\sum_{k=1}^{m}\left(\sum_{i=1}^{n} |A_{ki}|^2\right)^2 - t\sum_{k=1}^{m}\left(\sum_{i=1}^{n} |A_{ki}|^4\right) - D(A).\end{aligned}$$

Since there are q nonzero entries for every row of A, the Cauchy-Schwarz inequality shows that $\sum_{i=1}^n |A_{ki}|^4 \geq \frac{1}{q}\left(\sum_{i=1}^n |A_{ki}|^2\right)^2$. Again, this turns out to be insufficient for our purpose and we consider the equality:

$$\begin{aligned}\sum_{i\neq j}|M_{ij}|^2 &= t\sum_{k=1}^m\left(\sum_{i=1}^n |A_{ki}|^2\right)^2 - t\sum_{k=1}^m \frac{1}{q}\left(\left(\sum_{i=1}^n |A_{ki}|^2\right)^2 + \sum_{i<j}\left(|A_{ki}|^2 - |A_{kj}|^2\right)^2\right) - D(A)\\ &= \left(1-\frac{1}{q}\right)t\sum_{k=1}^m\left(\sum_{i=1}^n |A_{ki}|^2\right)^2 - \frac{t}{q}\sum_{k=1}^m\sum_{i<j}\left(|A_{ki}|^2 - |A_{kj}|^2\right)^2 - D(A)\\ &= \left(1-\frac{1}{q}\right)t\sum_{k=1}^m\left(\sum_{i=1}^n |A_{ki}|^2\right)^2 - \frac{t}{q}E(A) - D(A)\\ &= \left(1-\frac{1}{q}\right)tm\alpha^4 - \left(D(A) + \frac{t}{q}E(A)\right).\end{aligned}$$

□

From this, we get the following easy corollary.

Corollary 2.2.8. *Let A be an $m \times n$ matrix over $\mathbb{C}$. Suppose that each row of A has ℓ_2 norm α, the supports of every two columns of A intersect in at most t locations, and the size of the support of every row is q. Let $M = A^*A$. Then*

$$\sum_{i\neq j}|M_{ij}|^2 \leq \left(1-\frac{1}{q}\right)tm\alpha^4.$$

2.2.2 Latin squares

Latin squares play a central role in our proof. While Latin squares play a role in both [DSW14] and [BDWY13], our proof exploits their design properties more strongly.

Definition 2.2.9 (Latin square). *An $r \times r$ Latin square is an $r \times r$ matrix L such that $L_{ij} \in [r]$ for all i, j and every number in $[r]$ appears exactly once in each row and exactly once in each column.*

If L is a Latin square and $L_{ii} = i$ for all $i \in [r]$, we call it a *diagonal* Latin square.

Theorem 2.2.10 ([Hil73]). *For every $r \geq 3$, there exists an $r \times r$ diagonal Latin square.*

Two Latin squares L and L' are called *orthogonal* if every ordered pair $(k, l) \in [r]^2$ occurs uniquely as (L_{ij}, L'_{ij}) for some $i, j \in [r]$. A Latin square is called *self-orthogonal* if it is orthogonal to its transpose, denoted by L^T.

Theorem 2.2.11 ([BCH74]). *For every $r \in \mathbb{N}$, $r \neq 2, 3, 6$, there exist an $r \times r$ self-orthogonal Latin square.*

Let L be a self-orthogonal Latin square. Since $L_{ii} = L_{ii}^T$, the diagonal entries give all pairs of the form (i, i) for every $i \in [r]$, i.e. the diagonal entries must be a permutation of $[r]$. Without loss of generality, we may assume that $L_{ii} = i$ and so L is also a diagonal Latin square.

The following lemma is a strengthening of a lemma from [BDWY13].

Lemma 2.2.12. *Let $r \geq 3$. Then there exists a set $T \subseteq [r]^3$ of $r^2 - r$ triples that satisfies the following properties:*

1. *Each triple consists of three distinct elements.*
2. *For every pair $i, j \in [r]$, $i \neq j$, there are exactly 6 triples containing both i and j.*
3. *If $r \geq 4$, for every $i, j \in [r]$, $i \neq j$, there are at least 2 triples containing i and j such that the remaining elements are distinct.*

Proof. Theorem 2.2.10 guarantees the existence of an $r \times r$ diagonal Latin square. Let L be such a Latin square. Let T be the set of triples $(i, j, k) \subseteq [r]^3$ with $i \neq j$ and $k = L_{ij}$. Clearly the number of such triples is $r^2 - r$. We verify that the properties mentioned hold.

Recall that we have $L_{ii} = i$ for all $i \in [r]$, and every value appears once in each row and column. So for $i \neq j \in [r]$, it can not happen that $L_{ij} = i$ or $L_{ij} = j$ and we get Property 1, i.e. all elements of a triple must be distinct.

For Property 2, note that a pair i, j appears once as (i, j, L_{ij}) and once as (j, i, L_{ji}). And since every element appears exactly once in every row and column, we have that i must appear once in the j^{th} row, j must appear once in the i^{th} row and the same for the columns. It follows that each of $(*, j, i), (j, *, i), (*, i, j)$ and $(i, *, j)$ appears exactly once, where $*$ is some other element of $[r]$. This gives us that every pair appears in exactly 6 triples.

If $r \geq 4$ and $r \neq 6$, Theorem 2.2.11 gives us the existence of an $r \times r$ self-orthogonal Latin square L. Since L can be assumed to be diagonal, we may use a self-orthogonal

Latin square and preserve Properties 1 and 2. Now note that for a self-orthogonal Latin square $L_{ij} \neq L_{ji}$ if $i \neq j$, and so the triples (i, j, L_{ij}) and (j, i, L_{ji}) have distinct third elements, i.e. Property 3 is satisfied.

The case $r = 6$ requires separate treatment. It is known that 6×6 self-orthogonal Latin squares do not exist. Fortunately, the property we require is weaker and we are able to give an explicit construction of a matrix that is sufficient for our needs. Let L be the matrix

$$\begin{bmatrix} 1 & 4 & 5 & 3 & 6 & 2 \\ 3 & 2 & 6 & 5 & 1 & 4 \\ 2 & 5 & 3 & 6 & 4 & 1 \\ 6 & 1 & 2 & 4 & 3 & 5 \\ 4 & 6 & 1 & 2 & 5 & 3 \\ 5 & 3 & 4 & 1 & 2 & 6 \end{bmatrix}.$$

Clearly L is diagonal, and it is straightforward to check that $L_{ij} \neq L_{ji}$ for $i \neq j$. This gives that (i, j, L_{ij}) and (j, i, L_{ji}) have distinct third elements. It follows that we have Property 3 for all $r \geq 4$.

□

2.3 The dependency matrix

Let $\mathcal{V} = \{v_1, \ldots, v_n\}$ be a set of n points in $\mathbb{C}^d$. We will use $\dim(\mathcal{V})$ to denote the dimension of the linear span of $\mathcal{V}$ and by affine-$\dim(\mathcal{V})$ the dimension of the affine span of $\mathcal{V}$ (i.e., the minimum r such that points of $\mathcal{V}$ are contained in a shift of a linear subspace of dimension r). We projectivize $\mathbb{C}^d$ and consider the set of vectors $\mathcal{V}' = \{v_1', \ldots, v_n'\}$, where $v_i' = (v_i, 1)$ is the vector in $\mathbb{C}^{d+1}$ obtained by appending a 1 to the vector v_i. Let V be the $n \times (d+1)$ matrix whose i^{th} row is the vector v_i'. Now note that

$$\text{affine-dim}(\mathcal{V}) = \dim(\mathcal{V}') - 1 = \text{rank}(V) - 1.$$

We now construct a matrix A, which we refer to as the dependency matrix of $\mathcal{V}$. Note that the construction we give here is preliminary, but suffices to prove Theorems 2.1.6 and 2.1.8. A refined construction is given in Section 2.5, where we select the triples more carefully. The rows of the matrix will consist of linear dependency coefficients, which we define below.

Definition 2.3.1 (Linear dependency coefficients)**.** *Let v_1, v_2 and v_3 be three distinct collinear points in $\mathbb{C}^d$, and let $v'_i = (v_i, 1)$, $i \in \{1, 2, 3\}$, be vectors in $\mathbb{C}^{d+1}$. Recall that v_1, v_2, v_3 are collinear if and only if there exist nonzero coefficients $a_1, a_2, a_3 \in \mathbb{C}$ such that*

$$a_1 v'_1 + a_2 v'_2 + a_3 v'_3 = 0.$$

We refer to the a_1, a_2 and a_3 as the linear dependency coefficients between v_1, v_2, v_3. Note that the coefficients are determined up to scaling by a complex number. Throughout our proof, the specific choice of coefficients does not matter, so we fix a canonical choice by setting $a_3 = 1$.

Definition 2.3.2 (Dependency Matrix)**.** *For every line $l \in \mathcal{L}_{\geq 3}(\mathcal{V})$, let $\mathcal{V}_l$ denote the points lying on l. Then $|\mathcal{V}_l| \geq 3$ and we assign each line a triple system $T_l \subseteq \mathcal{V}_l^3$, the existence of which is guaranteed by Lemma 2.2.12. Let A be the $m \times n$ matrix obtained by going over every line $l \in \mathcal{L}_{\geq 3}$ and for each triple $(i, j, k) \in T_l$, adding as a row of A a vector with three nonzero coefficients in positions i, j, k corresponding to the linear dependency coefficients among the points v_i, v_j, v_k.*

Note that we have $AV = 0$. Every row of A has exactly 3 nonzero entries. By Property 2 of Lemma 2.2.12, the supports of any distinct two columns intersect in exactly 6 entries when the two corresponding points lie on a special line[2]*, and 0 otherwise; that is, the supports of any two distinct columns intersect in at most 6 entries.*

We say a pair of points v_i, v_j, $i \neq j$, appears *in the dependency matrix A if there exists a row with nonzero entries in columns i and j. The number of times a pair appears is the number of rows with nonzero entries in both columns i and j.*

Every pair of points that lies on a special line appears exactly 6 times. The only pairs not appearing in the matrix are pairs of points that determine ordinary lines. There are $\binom{n}{2}$ pairs of points, $t_2(\mathcal{V})$ of which determine ordinary lines. So the number

[2] *Note that while the triple system T_l consists of ordered triples, the supports of the rows of A are unordered.*

of pairs appearing in A is $\binom{n}{2} - t_2$. The total number of times these pairs appear is then $6\left(\binom{n}{2} - t_2\right)$. Every row gives 3 distinct pairs of points, so it follows that the number of rows of A is $m = 6\left(\binom{n}{2} - t_2\right)/3 = n^2 - n - 2t_2$. Note that $m > 0$, unless $t_2 = \binom{n}{2}$, i.e. all lines are ordinary.

As mentioned in the proof overview, we will consider two cases: when A satisfies Property-S and when it does not. We now prove lemmas dealing with the two cases. The following lemma deals with the former case.

Lemma 2.3.3. *Let $\mathcal{V}$ be a set of n points affinely spanning $\mathbb{C}^d$, $d \geq 3$, and let A be the dependency matrix for $\mathcal{V}$. Suppose that A satisfies Property-S. Then*

$$t_2(\mathcal{V}) \geq \frac{(d-3)}{2(d+1)}n^2 + \frac{3}{2}n$$

Proof. Fix $\epsilon > 0$. Since A satisfies Property-S, by Lemma 2.2.4 there is a scaling A' such that the ℓ_2 norm of each row is at most $\sqrt{1+\epsilon}$ and the ℓ_2 norm of each column is at least $\sqrt{\frac{m}{n} - \epsilon}$. Let $M := A'^*A'$. Then $M_{ii} \geq \frac{m}{n} - \epsilon$ for all i. Since every row in A has support 3, and the supports of any two columns intersect in at most 6 locations, Corollary 2.2.8 implies that $\sum_{i \neq j} |M_{ij}|^2 \leq 4m(1+\epsilon)^2$. By applying Lemma 2.2.5 to M,

$$\operatorname{rank}(M) \geq \frac{n^2(\frac{m}{n} - \epsilon)^2}{n(\frac{m}{n} - \epsilon)^2 + 4m(1+\epsilon)^2}.$$

Taking ϵ to 0 we see that

$$\begin{aligned}
\operatorname{rank}(A) = \operatorname{rank}(A') = \operatorname{rank}(M) &\geq \frac{n^2\frac{m^2}{n^2}}{n\frac{m^2}{n^2} + 4m} = \frac{mn}{m+4n} \\
&= n - \frac{4n^2}{m+4n} = n - \frac{4n^2}{n^2 - n - 2t_2(\mathcal{V}) + 4n} \\
&= n - \frac{4n^2}{n^2 + 3n - 2t_2(\mathcal{V})}.
\end{aligned}$$

Recall that affine-dim$(\mathcal{V}) = d = \operatorname{rank}(V) - 1$. Since $AV = 0$, we have $\operatorname{rank}(V) \leq n - \operatorname{rank}(A)$. It follows that

$$\begin{aligned}
d + 1 &\leq \frac{4n^2}{n^2 + 3n - 2t_2(\mathcal{V})} \\
\text{i.e.} \quad t_2(\mathcal{V}) &\geq \frac{(d-3)}{2(d+1)}n^2 + \frac{3}{2}n.
\end{aligned}$$

□

We now consider the case when Property-S is not satisfied.

Lemma 2.3.4. *Let $\mathcal{V}$ be a set of n points in $\mathbb{C}^d$, and let A be the dependency matrix for $\mathcal{V}$. Suppose that A does not satisfy Property-S. Then, for every integer b^*, $1 < b^* < 2n/3$, one of the following holds:*

1. *There exists a point $v \in \mathcal{V}$ contained in at least $\frac{2}{3}(n+1) - b^*$ ordinary lines;*

2. *$t_2(\mathcal{V}) \geq nb^*/2$.*

Proof. Since A violates Property-S, there exists a zero submatrix supported on rows $U \subseteq [m]$ and columns $W \subseteq [n]$ of the matrix A, where $|U| = a$ and $|W| = b$, such that

$$\frac{a}{m} + \frac{b}{n} > 1.$$

Let $X = [m] \setminus U$ and $Y = [n] \setminus W$ and note that $|X| = m - a$ and $|Y| = n - b$. Let the violating columns correspond to the set $\mathcal{V}_1 = \{v_1, \ldots, v_b\} \subset \mathcal{V}$. We consider two cases: when $b < b^*$, and when $b \geq b^*$.

Case 1 ($b < b^*$). We may assume that U is maximal, so every row in the submatrix $X \times W$ has at least one nonzero entry. Partition the rows of X into 3 parts: Let X_1, X_2 and X_3 be rows with one, two and three nonzero entries in columns of W respectively. We will get a lower bound on the number of ordinary lines containing exactly one point in $\mathcal{V}_1$ and one point in $\mathcal{V} \setminus \mathcal{V}_1$ by bounding the number of pairs $\{v_i, w\}$ that lie on special lines, with $v_i \in \mathcal{V}_1$ and $w \in \mathcal{V} \setminus \mathcal{V}_1$. Note that there are at most $b(n-b)$ such pairs, and each pair that does not lie on a special line determines an ordinary line.

Each row of X_1 gives two pairs of points $\{v_i, w_1\}$ and $\{v_i, w_2\}$ that lie on a special line, where $v_i \in \mathcal{V}_1$ and $w_1, w_2 \in \mathcal{V} \setminus \mathcal{V}_1$. Each row of X_2 gives 2 pairs of points $\{v_i, w\}$ and $\{v_j, w\}$ that lie on special lines, where $v_i, v_j \in \mathcal{V}_1$ and $w \in \mathcal{V} \setminus \mathcal{V}_1$. Each row of X_3 has all zero entries in the submatrix supported on $X \times Y$, so it does not contribute any pairs. Recall, from Lemma 2.5.4, that each pair of points on a special line appears exactly 6 times in the matrix. This implies that the number of pairs that lie on special lines with at least one point in $\mathcal{V}_1$ and one point in $\mathcal{V} \setminus \mathcal{V}_1$ is $\frac{2|X_1|+2|X_2|}{6} \leq \frac{2|X|}{6}$. Hence, the number of ordinary lines containing exactly one of $v_1, \ldots, v_b$ is then at least $b(n-b) - \frac{|X|}{3}$.

Recall that

$$1 < \frac{a}{m} + \frac{b}{n} = \left(1 - \frac{|X|}{m}\right) + \frac{b}{n}.$$

Substituting $m \leq n^2 - n$, we get

$$|X| < \frac{bm}{n} \leq b(n-1).$$

This shows that the number of ordinary lines containing exactly one point in $\mathcal{V}_1$ is at least

$$b(n-b) - \frac{|X|}{3} > \frac{2b}{3}n - \frac{3b^2 - b}{3}.$$

We now see that there exists $v \in \mathcal{V}_1$ such that the number of ordinary lines containing v is at least

$$\left\lfloor \frac{2}{3}n - \frac{3b-1}{3} \right\rfloor \geq \left\lfloor \frac{2}{3}n - b^* + \frac{4}{3} \right\rfloor \geq \frac{2}{3}(n+1) - b^*.$$

Case 2 ($b \geq b^*$). We will determine a lower bound for $t_2(\mathcal{V})$ by counting the number of nonzero pairs of entries $A_{ij}, A_{ij'}$ that appear in the submatrix $U \times Y$, with $j \neq j'$. There are $\binom{n-b}{2}$ pairs of points in $\mathcal{V}\backslash\mathcal{V}_1$, each of which appears at most 6 times, therefore the number of pairs of such entries is at most $6\binom{n-b}{2}$. Each row of U has 3 pairs of nonzero entries, so the number of pairs of entries equals $3a$, and it follows that

$$3a \leq 6\binom{n-b}{2}. \tag{2.1}$$

Recall that $\frac{a}{m} + \frac{b}{n} > 1$, which gives us

$$a > m\left(1 - \frac{b}{n}\right) = \left(n^2 - n - 2t_2(\mathcal{V})\right)\left(1 - \frac{b}{n}\right). \tag{2.2}$$

Combining (2.1) and (2.2), we get

$$\left(n^2 - n - 2t_2(\mathcal{V})\right)\left(1 - \frac{b}{n}\right) < 2\binom{n-b}{2}, and$$

solving for $t_2(\mathcal{V})$ finally shows

$$t_2(\mathcal{V}) > \frac{nb}{2} \geq \frac{nb^*}{2}.$$

□

2.4 Proofs of Theorems 2.1.6 and 2.1.8

The proofs of both Theorems 2.1.6 and 2.1.8 rely on Lemmas 2.3.3 and 2.3.4. Together, these lemmas imply that there must be a point with many ordinary lines containing it, or else there are many ordinary lines in total. As mentioned in the proof overview, the theorems are then obtained by using an iterative argument that removes a point with many ordinary lines through it, and then applying the same argument to the remaining points.

2.4.1 Proof of Theorem 2.1.6

We get the following easy corollary from Lemma 2.3.3 and Lemma 2.3.4.

Corollary 2.4.1. *Let $\mathcal{V}$ be a set of $n \geq 5$ points in $\mathbb{C}^d$ not contained in a plane. Then one of the following holds:*

1. *There exists a point $v \in \mathcal{V}$ contained in at least $\frac{2}{3}n - \frac{7}{3}$ ordinary lines.*
2. *$t_2(\mathcal{V}) \geq \frac{3}{2}n$.*

Proof. Let A be the dependency matrix for $\mathcal{V}$. If A satisfies Property-S, then we are done by Lemma 2.3.3. Otherwise, let $b^* = 3$, and note that Lemma 2.3.4 gives us the statement of the corollary when $n \geq 5$. □

We are now ready to prove Theorem 2.1.6. For convenience, we state the theorem again.

Theorem 2.1.6. *Let $\mathcal{V}$ be a set of $n \geq 24$ points in $\mathbb{C}^3$, not contained in a plane. Then $\mathcal{V}$ determines at least $\frac{3}{2}n$ ordinary lines unless $n-1$ points are on a plane, in which case there are at least $n-1$ ordinary lines.*

Proof. If $t_2(\mathcal{V}) \geq \frac{3}{2}n$ then we are done. Else, by Corollary 2.4.1, we may assume there exists a point v_1 with at least $\frac{1}{3}(2n-7)$ ordinary lines and hence at most $\frac{1}{6}(n+4)$ special lines through it. Let $\mathcal{V}_1 = \mathcal{V} \setminus \{v_1\}$. If $\mathcal{V}_1$ is planar, then there are exactly $n-1$ ordinary lines through v_1. We note here that this is the only case where there exists fewer then $\frac{3}{2}n$ ordinary lines.

Suppose now that $\mathcal{V}_1$ is not planar. Again, by Corollary 2.4.1, there are either $\frac{3}{2}(n-1)$ ordinary lines in $\mathcal{V}_1$ or there exists a point $v_2 \in \mathcal{V}_1$ with at least $\frac{2}{3}(n-1)-\frac{7}{3} = \frac{1}{3}(2n-9)$ ordinary lines through it. In the former case, we get $\frac{3}{2}(n-1)$ ordinary lines in $\mathcal{V}_1$, at most $\frac{1}{6}(n+4)$ of which could contain v_1. This shows that the total number of ordinary lines in $\mathcal{V}$ is

$$t_2(\mathcal{V}) \geq \frac{3}{2}(n-1) - \frac{1}{6}(n+4) + \frac{1}{3}(2n-7) = \frac{1}{2}(4n-9).$$

When $n \geq 9$, $t_2(\mathcal{V}) \geq \frac{3}{2}n$.

In the latter case there exists a point $v_2 \in \mathcal{V}_1$ with at least $\frac{1}{3}(2n-9)$ ordinary lines in $\mathcal{V}_1$ through it. Note that at most one of these could contain v_1, so we get at least $\frac{1}{3}(2n-7)+\frac{1}{3}(2n-9)-1 = \frac{1}{3}(4n-19)$ ordinary lines through one of v_1 or v_2. Note also that the number of special lines through one of v_1 or v_2 is at most $\frac{1}{6}(n+4)+\frac{1}{6}(n+3) = \frac{1}{6}(2n+7)$.

Let $\mathcal{V}_2 = \mathcal{V}_1 \setminus \{v_2\}$. If $\mathcal{V}_2$ is contained in a plane, we get at least $n-3$ ordinary lines from each of v_1 and v_2 giving a total of $2n-6$ ordinary lines in $\mathcal{V}$. It follows that when $n \geq 12$, $t_2(\mathcal{V}) \geq \frac{3}{2}n$.

Otherwise $\mathcal{V}_2$ is not contained in a plane, and again Corollary 2.4.1 gives us two cases. If there are $\frac{3}{2}(n-2)$ ordinary lines in $\mathcal{V}_2$, then the total number of ordinary lines is

$$t_2(\mathcal{V}) = \frac{3}{2}(n-2) - \frac{1}{6}(2n+7) + \frac{1}{3}(4n-19) = \frac{1}{2}(5n-21).$$

When $n \geq 11$, we get that $t_2(\mathcal{V}) \geq \frac{3}{2}n$.

Otherwise there exists a point v_3 with at least $\frac{2}{3}(n-2)-\frac{7}{3}$ ordinary lines through it. At most 2 of these could pass through one of v_1 or v_2, so we get $\frac{2}{3}(n-2)-\frac{7}{3}-2 = \frac{1}{3}(2n-17)$ ordinary lines through v_3 in $\mathcal{V}$. Summing up the number of lines through one of v_1, v_2 and v_3, we have that

$$t_2(\mathcal{V}) \geq \frac{1}{3}(2n-17) + \frac{1}{3}(4n-19) = 2n-12,$$

and when $n \geq 24$, $t_2(\mathcal{V}) \geq \frac{3}{2}n$. □

2.4.2 Proof of Theorem 2.1.8

We get the following easy corollary from Lemma 2.3.3 and Lemma 2.3.4.

Corollary 2.4.2. *There exists a positive integer n_0 such that the following holds. Let $\mathcal{V}$ be a set of $n \geq n_0$ points in $\mathbb{C}^d$ not contained in a three dimensional affine subspace. Then either:*

1. *There exists a point with at least $\frac{n}{2}$ ordinary lines through it, or*
2. *$t_2(\mathcal{V}) \geq \frac{1}{12}n^2$.*

Proof. Let A be the dependency matrix of $\mathcal{V}$. If A satisfies Property-S, then we are done by Lemma 2.3.3. Otherwise, let $b^* = n/6$. Now by Lemma 2.3.4, either the number of ordinary lines

$$t_2(\mathcal{V}) \geq \frac{n}{2}b^* \geq \frac{1}{12}n^2$$

or there exists a point $v \in \mathcal{V}$, such that the number of ordinary lines containing v is at least

$$\frac{2}{3}(n+1) - b^* > \frac{1}{2}n.$$

□

We are now ready to prove Theorem 2.1.8. For convenience, we state the theorem again.

Theorem 2.1.8. *There exists a positive integer n_0 such that the following holds. Let $\mathcal{V}$ be a set of $n \geq n_0$ points in $\mathbb{C}^4$ with at most $\frac{2}{3}n$ points contained in any 3 dimensional affine subspace. Then*

$$t_2(\mathcal{V}) \geq \frac{1}{12}n^2.$$

Proof. The basic idea of the proof uses the following algorithm: We use Corollary 2.4.2 to find a point with a large number of ordinary lines, "prune" this point, and then repeat this process on the smaller set of points. We stop when either we can not find such a point, in which case Corollary 2.4.2 guarantees a large number of ordinary lines, or when we have accumulated enough ordinary lines.

Consider the following algorithm:

Let $\mathcal{V}_0 := \mathcal{V}$ and $j = 0$.

1. If $\mathcal{V}_j$ satisfies case (2) of Corollary 2.4.2, then stop.

2. Otherwise, there must exist a point $v_{j+1} \in \mathcal{V}_j$ with at least $\frac{n-j}{2}$ ordinary lines through it. Let $\mathcal{V}_{j+1} = \mathcal{V}_j \setminus \{v_{j+1}\}$.

3. Set $j = j + 1$. If $j = n/3$, then stop. Otherwise go to Step 1.

Note that since no 3 dimensional plane contains more than $2n/3$ points, at no step will the algorithm stop because the configuration becomes 3 dimensional. That is, we can use Corollary 2.4.2 in all steps of the algorithm.

We now analyze the two stopping conditions for the algorithm, and show that we can always find enough ordinary lines by the time the algorithm stops.

Suppose that we stop because $\mathcal{V}_j$ satisfies case (2) of Corollary 2.4.2 for some $1 \leq j < n/3$. From case (2) of Corollary 2.4.2, we have that

$$t_2(\mathcal{V}_j) \geq \frac{(n-j)^2}{12}. \tag{2.3}$$

On the other hand, each pruned point v_i, $1 \leq i \leq j$, has at least $\frac{n-i+1}{2} > \frac{n-i}{2}$ ordinary lines through it that are determined by $\mathcal{V}_{i-1}$, and hence at most $(n - i - \frac{n-i+1}{2})/2 < \frac{n-i}{4}$ special lines through it. Note that an ordinary line in $\mathcal{V}_i$ might not be ordinary in $\mathcal{V}_{i-1}$ if contains v_i. Thus, in order to lower bound the total number of ordinary lines in $\mathcal{V}$, we sum over the number of ordinary lines contributed by each of the pruned points v_i, $1 \leq i \leq j$, and subtract from the count the number of potential lines that could contain v_i.

Then the number of ordinary lines in $\mathcal{V}$ contributed by the pruned points is at least

$$\sum_{i=1}^{j} \left(\frac{n-i}{2} - \frac{n-i}{4} \right) = \frac{1}{4} \sum_{i=1}^{j} (n-i) = \frac{jn}{4} - \frac{j^2+j}{8}. \tag{2.4}$$

Combining (2.3) and (2.4), we have that

$$\begin{aligned} t_2(\mathcal{V}) &\geq \frac{1}{12}(n-j)^2 + \frac{jn}{4} - \frac{j^2+j}{8} \\ &= \frac{n^2}{12} + \frac{-j^2 + j(2n-3)}{24}. \end{aligned}$$

This is an increasing function for $j < n - 1$, implying that

$$t_2(\mathcal{V}) \geq \frac{n^2}{12}.$$

We now consider the case when the algorithm stops because $j = n/3$. Note that at this point, we will have pruned exactly j points. Each pruned point v_i, $1 \le i \le j$, has $\frac{n-i+1}{2}$ ordinary lines through it that are determined by $\mathcal{V}_{i-1}$. The only way such an ordinary line is not ordinary in $\mathcal{V}$ is that it contains one of the previously pruned points. At most $i-1$ of the ordinary lines through v_i contain other pruned points v_k, $k < i$. Therefore the total number of ordinary lines determined by $\mathcal{V}$ satisfies

$$t_2(\mathcal{V}) \ge \sum_{i=1}^{j} \frac{n-i+1}{2} - \sum_{i=1}^{j}(i-1) = \frac{jn}{2} - \frac{3}{4}(j^2 - j).$$

Since $j = n/3$, the number of ordinary lines determined by $\mathcal{V}$ is at least

$$t_2(\mathcal{V}) \ge \frac{n^2}{12}.$$

□

2.5 A dependency matrix for a more refined bound

In this section we give a more careful construction for the dependency matrix of a point set $\mathcal{V}$. Recall that we defined the dependency matrix in Definition 2.3.2 to contain a row for each collinear triple from a triple system constructed on each special line. The goal was to not have too many triples containing the same pair (as can happen when there are many points on a single line). At the end of this section (Definition 2.5.7) we will give a construction of a dependency matrix that will have an additional property (captured in Item 4 of Lemma 2.5.4) which is used to obtain cancellation in the diagonal dominant argument, as outlined in the proof overview.

We denote the argument of a complex number z by $\arg(z)$. We use the convention that for every complex number z, $\arg(z) \in (-\pi, \pi]$.

Definition 2.5.1 (angle between two complex numbers)**.** *We define the* angle between two complex numbers a and b *to be the the absolute value of the argument of $a\bar{b}$, denoted by $\left|arg\left(a\bar{b}\right)\right|$. Note that the angle between a and b equals the angle between b and a.*

Definition 2.5.2 (co-factor)**.** *Let v_1, v_2 and v_3 be three distinct collinear points in $\mathbb{C}^d$, and let a_1, a_2 and a_3 be the linear dependency coefficients among the three points. Define* *the* co-factor *of v_3 with respect to (v_1, v_2), denoted by $C_{(1,2)}(3)$, to be $\frac{a_1\overline{a_2}}{|a_1||a_2|}$. Notice that this is well defined with respect to the points, and does not depend on the choice of coefficients.*

The next lemma will be used to show that "cancellations" must arise in a line containing four points (as mentioned earlier in the proof overview). We will later use this lemma as a black box, in order to quantify the cancellations in lines with more than four points by applying it to random four tuples inside the line.

Lemma 2.5.3. *Let v_1, v_2, v_3, v_4 be 4 collinear points in $\mathbb{C}^d$. Then at least one of the following conditions hold:*

1. *The angle between $C_{(1,2)}(3)$ and $C_{(1,2)}(4)$ is at least $\pi/3$.*
2. *The angle between $C_{(1,3)}(4)$ and $C_{(1,3)}(2)$ is at least $\pi/3$.*
3. *The angle between $C_{(1,4)}(2)$ and $C_{(1,4)}(3)$ is at least $\pi/3$.*

Proof. For $i \in \{1, 2, 3, 4\}$, let $v'_i = (v_i, 1)$, i.e. the vector obtained by appending 1 to v_i. Since v_1, v_2, v_3, v_4 are collinear, there exist $a_1, a_2, a_3 \in \mathbb{C}$ such that

$$a_1 v'_1 + a_2 v'_2 + a_3 v'_3 = 0 \tag{2.5}$$

and $b_1, b_2, b_4 \in \mathbb{C}$ such that

$$b_1 v'_1 + b_2 v'_2 + b_4 v'_4 = 0. \tag{2.6}$$

We may assume, without loss of generality, that $a_3 = b_4 = 1$. Now equations (2.5) and (2.6) imply that $C_{(1,2)}(3) = \frac{a_1\overline{a_2}}{|a_1||a_2|}$, $C_{(1,2)}(4) = \frac{b_1\overline{b_2}}{|b_1||b_2|}$, $C_{(1,3)}(2) = \frac{a_1}{|a_1|}$ and $C_{(1,4)}(2) = \frac{b_1}{|b_1|}$.

Combining equations (2.5) and (2.6), we get the following linear equation:

$$(b_2 a_1 - b_1 a_2) v'_1 + b_2 v'_3 - a_2 v'_4 = 0. \tag{2.7}$$

Using (2.7), we see $C_{(1,3)}(4) = \frac{(b_2 a_1 - b_1 a_2)\overline{b_2}}{|b_2 a_1 - b_1 a_2||b_2|}$ and $C_{(1,4)}(3) = -\frac{(b_2 a_1 - b_1 a_2)\overline{a_2}}{|b_2 a_1 - b_1 a_2||a_2|}$.

Then the angle between $C_{(1,2)}(3)$ and $C_{(1,2)}(4)$ is

$$\begin{aligned} & \left|\arg\left(\frac{a_1\overline{a_2}}{|a_1||a_2|}\frac{\overline{b_1}b_2}{|b_1||b_2|}\right)\right| \\ = & \left|\arg\left(a_1\overline{a_2}\overline{b_1}b_2\right)\right|. \end{aligned} \tag{2.8}$$

The angle between $C_{(1,3)}(4)$ and $C_{(1,3)}(2)$ is

$$\begin{aligned} & \left|\arg\left(\frac{(b_2a_1-b_1a_2)\overline{b_2}}{|b_2a_1-b_1a_2||b_2|}\frac{\overline{a_1}}{|a_1|}\right)\right| \\ = & \left|\arg\left(\overline{a_1}\overline{b_2}(b_2a_1-b_1a_2)\right)\right|. \end{aligned} \tag{2.9}$$

The angle between $C_{(1,4)}(2)$ and $C_{(1,4)}(3)$ is

$$\begin{aligned} & \left|\arg\left(-\frac{b_1}{|b_1|}\frac{\overline{(b_2a_1-b_1a_2)}a_2}{|b_2a_1-b_1a_2||a_2|}\right)\right| \\ = & \left|\arg\left(-b_1a_2\overline{(b_2a_1-b_1a_2)}\right)\right|. \end{aligned} \tag{2.10}$$

Note that the product of expressions inside the arg functions in (2.8), (2.9) and (2.10) is a negative real number, and so the sum of (2.8), (2.9) and (2.10) must be π. It follows that one of the angles must be at least $\pi/3$. □

Our final dependency matrix will be composed of blocks, each given by the following lemma. Roughly speaking, we construct a block of rows $A(l)$ for each special line l. The rows in $A(l)$ will be chosen carefully and will correspond to triples that will eventually give non trivial cancellations.

Lemma 2.5.4. *Let l be a line in $\mathbb{C}^d$ and $\mathcal{V}_l = \{v_1, \dots v_r\}$ be points on l with $r \geq 3$. Let V_l be the $r \times (d+1)$ matrix whose i^{th} row is the vector $(v_i, 1)$. Then there exists an $(r^2 - r) \times r$ matrix $A = A(l)$, which we refer to as the* dependency matrix *of l, such that the following hold:*

1. *$AV_l = 0$;*

2. *Every row of A has support of size 3;*

3. *The support of every two columns of A intersects in exactly 6 locations;*

4. *If $r \geq 4$ then for at least $1/3$ of choices of $k \in [r^2-r]$, there exists $k' \in [r^2-r]$ such that following holds: For $k \in [r^2-r]$, let R_k denote the rth row of A. Suppose supp$(R_k) = \{i,j,s\}$. Then supp$(R_{k'}) = \{i,j,t\}$ (for some $t \neq s$) and the angle between the co-factors $C_{(i,j)}(s)$ and $C_{(i,j)}(t)$ is at least $\pi/3$.*

Proof. Recall that Lemma 2.2.12 gives us a family of triples T_r on the set $[r]^3$. For every bijective map $\sigma : \mathcal{V}_l \to [r]$, construct a matrix A_σ in the following manner: Let T_l be the triple system on $\mathcal{V}_l^3$ induced by composing σ and T_r. For each triple $(v_i, v_j, v_k) \in T_l$, add a row with three non-zero entries in positions i, j, k corresponding to the linear dependency coefficients between v_i, v_j and v_k.

Note that for every σ, A_σ has $r^2 - r$ rows and r columns. Since the rows correspond to linear dependency coefficients, clearly we have $A_\sigma V_l = 0$ satisfying Property 1. Properties 2 and 3 follow from properties of the triple system from Lemma 2.2.12.

We will use a probabilistic argument to show that there exists a matrix A that has Property 4. Let Σ be the collection of all bijective maps from $[r]$ to the points $\mathcal{V}_l$, and let $\sigma \in \Sigma$ be a uniformly random element. Consider A_σ. Since every pair of points occurs in at least 2 distinct triples, for every row R_k of A_σ, there exists a row $R_{k'}$ such that the supports of R_k and $R_{k'}$ intersect in 2 entries. Suppose that R_k and $R_{k'}$ have supports contained in $\{i,j,s,t\}$. Suppose that σ maps $\{v_i, v_j, v_s, v_t\}$ to $\{1,2,3,4\}$ and that $(1,2,3)$ and $(1,2,4)$ are triples in T_r. Without loss of generality, assume v_i maps to 1. Then by Lemma 2.5.3, the angle between at least one of the pairs $\{C_{(i,j)}(s), C_{(i,j)}(t)\}$, $\{C_{(i,s)}(j), C_{(i,s)}(t)\}$, $\{C_{(i,t)}(j), C_{(i,t)}(s)\}$ must be at least $\pi/3$. That is, given that v_i maps to 1, we have that the probability that R_k satisfies Property 4 is at least 1/3. Then it is easy to see that

$$\Pr(R_k \text{ satisfies Property 4}) \geq 1/3.$$

Define the random variable X to be the number of rows satisfying Property 4, and note that we have

$$\mathbb{E}[X] \geq (r^2 - r)\frac{1}{3}.$$

It follows that there exists a matrix A in which at least 1/3 of the rows satisfy Property 4. □

To argue about the off diagonal entries of $M = A^*A$ (where $A = A(l)$), we will use the following notion of balanced rows. The main idea here is that, if there are many rows that are not balanced then we win in one of the Cauchy-Schwartz applications and, if many rows are balanced then we win from cancellations that show up via the different angles.

Definition 2.5.5 (η-balanced row)**.** *Given an $m \times n$ matrix A, we say a row R_k is η-balanced for some constant η if $\big||A_{ki}|^2 - |A_{kj}|^2\big| \leq \eta$, for every $i, j \in supp(R_k)$. Otherwise we say that R_k is η-unbalanced. When η is clear from the context, we say that the row is balanced/unbalanced.*

Lemma 2.5.6. *There exists an absolute constant $c_0 > 0$ such that the following holds. Let l be a line in $\mathbb{C}^d$ and $\mathcal{V}_l = \{v_1, \ldots v_r\}$ be points on l with $r \geq 4$. Let $A = A(l)$ be the dependency matrix for l, defined in Lemma 2.5.4, and A' a scaling of A such that the ℓ_2 norm of every row is α. Let $M = A'^* A'$.*

$$\sum_{i \neq j} |M_{ij}|^2 \leq 4(r^2 - r)\alpha^4 - c_0(r^2 - r)\alpha^2.$$

Proof. Recall that A is an $(r^2 - r) \times r$ matrix, that the support of every row has size exactly 3, and that the supports of any two distinct columns of A intersects in 6 locations. Clearly, any scaling A' of A will also satisfy these properties. Applying Lemma 2.2.7 to A' we get that

$$\sum_{i \neq j} |M_{ij}|^2 = 4(r^2 - r)\alpha^4 - (D(A) + 2E(A))\,. \tag{2.11}$$

We are able to give a lower bound on $D(A)+2E(A)$ using Property 4 of Lemma 2.5.4. From here on, we focus on the rows mentioned in Property 4. Recall that there are at least $(r^2 - r)/3$ such rows. For some η to be determined later, suppose that β fraction of these rows is η-unbalanced. We will show each such row contributes to either $D(A)$ or $E(A)$.

If a row R_k is η-imbalanced, we get that

$$\sum_{i<j} \left(|A_{ki}|^2 - |A_{kj}|^2\right)^2 > \eta^2.$$

Alternatively suppose that R_k is η-balanced. Recall that $\sum_{i=1}^{n} |A_{ki}|^2 = \alpha$ and note that we must have that $|A_{ki}|^2 \in [\frac{\alpha}{3} - \frac{2\eta}{3}, \frac{\alpha}{3} + \frac{2\eta}{3}]$ for all $i \in \text{supp}(R_k)$. Suppose that both R_k and $R_{k'}$ have non-zero entries in columns i and j, but R_k has a third nonzero entry in column s and $R_{k'}$ has a third nonzero entry in column t, where $s \neq t$. Suppose further that the angle θ between the co-factors $C_{(i,j)}(s)$ and $C_{(i,j)}(t)$ is at least $\pi/3$, i.e. $\cos\theta \leq 1/2$. Then

$$\begin{aligned} &\left|A_{ki}\overline{A_{kj}} - A_{k'i}\overline{A_{k'j}}\right|^2 \\ &= |A_{ki}\overline{A_{kj}}|^2 + |A_{k'i}\overline{A_{k'j}}|^2 - 2|A_{ki}\overline{A_{kj}}||A_{k'i}\overline{A_{k'j}}|\cos\theta \\ &\geq |A_{ki}\overline{A_{kj}}|^2 + |A_{k'i}\overline{A_{k'j}}|^2 - |A_{ki}\overline{A_{kj}}||A_{k'i}\overline{A_{k'j}}|. \end{aligned}$$

For any positive real numbers a, b, we have that

$$a^2 + b^2 - ab = \left(\frac{a}{2} - b\right)^2 + \frac{3}{4}a^2 \geq \frac{3}{4}a^2.$$

Substituting $a = |A_{ki}\overline{A_{kj}}|$ and $b = |A_{k'i}\overline{A_{k'j}}|$, we get that

$$\begin{aligned} &|A_{ki}\overline{A_{kj}}|^2 + |A_{k'i}\overline{A_{k'j}}|^2 - |A_{ki}\overline{A_{kj}}||A_{k'i}\overline{A_{k'j}}| \\ &\geq \frac{3}{4}|A_{ki}\overline{A_{kj}}|^2 \\ &\geq \frac{3}{4}\left(\frac{\alpha}{3} - \frac{2\eta}{3}\right)^2 \\ &= \frac{1}{12}\left(\alpha - 2\eta\right)^2. \end{aligned}$$

Summing over the η-unbalanced rows, we get that

$$E(A) \geq \beta\frac{(r^2 - r)}{3}\eta^2.$$

Summing over all the η-balanced rows, we get that

$$\begin{aligned} D(A) &= \sum_{i\neq j}\sum_{k<k'}\left|A_{ki}\overline{A_{kj}} - A_{k'i}\overline{A_{k'j}}\right|^2 \\ &= \frac{1}{2}\sum_{k\neq k'}\sum_{i\neq j}\left|A_{ki}\overline{A_{kj}} - A_{k'i}\overline{A_{k'j}}\right|^2 \\ &\geq \frac{1}{2}\cdot(1-\beta)\frac{(r^2 - r)}{3}\cdot\frac{1}{12}\left(\alpha - 2\eta\right)^2. \\ &= (1-\beta)\frac{(r^2 - r)}{72}\left(\alpha - 2\eta\right)^2. \end{aligned}$$

Setting $\eta = \alpha/10$, we get that

$$\begin{aligned} D(A) + 2E(A) &\geq (1-\beta)\frac{(r^2-r)}{72}(\alpha - 2\eta)^2 + 2\beta\frac{(r^2-r)}{3}\eta^2 \\ &= (r^2-r)\left((1-\beta)\frac{1}{72}\left(\frac{4}{5}\alpha\right)^2 + \beta\frac{2}{3}\left(\frac{1}{10}\alpha\right)^2\right) \\ &\geq c_0(r^2-r)\alpha^2 \end{aligned}$$

for some absolute constant c_0. Combining the above with equation (2.11), we get

$$\sum_{i \neq j} |M_{ij}|^2 \leq 4(r^2-r)\alpha^4 - c_0(r^2-r)\alpha^2.$$

□

We are now ready to define the full dependency matrix that we will use in the proof of Theorem 2.1.7.

Definition 2.5.7 (Dependency Matrix, second construction). *Let $\mathcal{V} = \{v_1, \dots v_n\}$ be a set of n points in $\mathbb{C}^d$ and let V be the $n \times (d+1)$ matrix whose i^{th} row is the vector $(v_i, 1)$. For each matrix $A(l)$, where $l \in \mathcal{L}_{\geq 3}(\mathcal{V})$, add $n - r$ column vectors of all zeroes, with length $r^2 - r$, in the column locations corresponding to points not in l, giving an $(r^2 - r) \times n$ matrix. Let A be the matrix obtained by taking the union of rows of these matrices for every $l \in \mathcal{L}_{\geq 3}(\mathcal{V})$. We refer to A as the* dependency matrix *of $\mathcal{V}$.*

Note that this construction is a special case of the one given in Definition 2.3.2 and so satisfies all the properties mentioned there. In particular, we have $AV = 0$ and the number of rows in A equals $n^2 - n - 2t_2(\mathcal{V})$.

2.6 Proof of Theorem 2.1.7

Before we prove the theorem, we give some key lemmas. As before, we consider two cases: When the dependency matrix A satisfies Property-S and when it does not. In the latter case, we rely on Lemma 2.3.4. The following lemma deals with the former case.

Lemma 2.6.1. *There exists an absolute constant $c_1 > 0$ such that the following holds. Let $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ be a set of points in $\mathbb{C}^d$ not contained in a plane. Let A be the $m \times n$ dependency matrix for $\mathcal{V}$, and suppose that A satisfies Property-S. Then*

$$t_2(\mathcal{V}) \geq \frac{3}{2}n + c_1 \sum_{r \geq 4} (r^2 - r) t_r(\mathcal{V}).$$

Proof. Since A satisfies Property-S, by Corollary 2.2.4 for every $\epsilon > 0$, there exists a scaling A' of A such that for every $i \in [m]$

$$\sum_{j \in [n]} \left|A'_{ij}\right|^2 = 1 + \epsilon,$$

and for every $j \in [n]$

$$\sum_{i \in [m]} \left|A'_{ij}\right|^2 \geq \frac{m}{n} - \epsilon. \tag{2.12}$$

Let C_i be denote the i^{th} column of A', and let $M = A'^{*}A'$. From (2.12), we get that $|M_{ii}| = \langle C_i, C_i \rangle \geq \left(\frac{m}{n} - \epsilon\right)$.

To bound the sum of squares of the off-diagonal entries, we go back to the construction of the dependency matrix. Recall that the matrix A was obtained by taking the union of rows of matrices $A(l)$, for each $l \in \mathcal{L}_{\geq 3}$. Then we have that A' is the union of scalings of the rows of the matrices $A(l)$, for each $l \in \mathcal{L}_{\geq 3}$. Note that $|M_{ij}| = \langle C_i, C_j \rangle$ and that the intersection of the supports of any two distinct columns in contained within a scaling of $A(l)$, for some $l \in \mathcal{L}_{\geq 3}$. Therefore, to get a bound on $\sum_{i \neq j} |M_{ij}|^2$, it suffices to consider these component matrices. Combining the bounds obtained from Lemma 2.5.6, for $\alpha = 1 + \epsilon$, we get that

$$\begin{aligned}
\sum_{i \neq j} |M_{ij}|^2 &\leq \sum_{l \in \mathcal{L}_3} 4(r^2 - r)\alpha^4 + \sum_{l \in \mathcal{L}_{\geq 4}} \left(4(r^2 - r)\alpha^4 - c_0(r^2 - r)\alpha^2\right) \\
&= \sum_{l \in \mathcal{L}_{\geq 3}} 4(r^2 - r)\alpha^4 - \sum_{l \in \mathcal{L}_{\geq 4}} c_0(r^2 - r)\alpha^2 \\
&= 4m(1 + \epsilon)^4 - (1 + \epsilon)^2 c_0 \sum_{r \geq 4} (r^2 - r) t_r.
\end{aligned}$$

Let $F = c_0 \sum_{r \geq 4}^{n} (r^2 - r) t_r$. Lemma 2.2.5 gives us that

$$\begin{aligned}
\operatorname{rank}(M) &\geq \frac{n^2 L^2}{nL^2 + \sum_{i \neq j} |M_{ij}|^2} \\
&\geq \frac{n^2 \left(\frac{m}{n} - \epsilon\right)^2}{n\left(\frac{m}{n} - \epsilon\right)^2 + 4m(1 + \epsilon)^4 - (1 + \epsilon)^2 F}.
\end{aligned}$$

Taking ϵ to 0, we get

$$\begin{aligned}\operatorname{rank}(M) &\geq \frac{n^2\left(\frac{m}{n}\right)^2}{n\left(\frac{m}{n}\right)^2 + 4m - F} \\ &= n - \frac{4n^2m - n^2F}{m^2 + 4mn - nF}.\end{aligned}$$

Note that

$$\text{affine-dim}(\mathcal{V}) = \operatorname{rank}(V) - 1 \leq \frac{4n^2m - n^2F}{m^2 + 4mn - nF} - 1.$$

It follows that if

$$\frac{4n^2m - n^2F}{m^2 + 4mn - nF} < 4,$$

we get that $\mathcal{V}$ must be contained in a plane, contradicting the assumption of the theorem. Substituting $m = n^2 - n - 2t_2(\mathcal{V})$ and simplifying, we get

$$4t_2^2 - (2n^2 + 4n)t_2 + 3n^3 - 3n^2 + \frac{n^2F}{4} - nF > 0.$$

This holds when

$$\begin{aligned}t_2(\mathcal{V}) &< \frac{3n}{2} + \frac{F}{8} \\ &= \frac{3n}{2} + \frac{c_0}{8}\sum_{r=4}^{n}(r^2 - r)t_r(\mathcal{V})\end{aligned}$$

which completes the proof. □

We now have the following easy corollary.

Corollary 2.6.2. *There exists a positive integer n_0 such that the following holds. Let c_1 be the constant from Lemma 2.6.1 and let $\mathcal{V}$ be a set of $n \geq n_0$ points in $\mathbb{C}^d$ not contained in a plane. Then one of the following must hold:*

1. *There exists a point $v \in \mathcal{V}$ contained in at least $\frac{n}{2}$ ordinary lines.*

2. *$t_2(\mathcal{V}) \geq \frac{3}{2}n + c_1\sum_{r\geq 4}(r^2 - r)t_r(\mathcal{V})$.*

Proof. If A satisfies Property-S, then we are done by Lemma 2.6.1. Otherwise, let b^* be an integer such that

$$\frac{n}{2}(b^* - 1) < \frac{3n}{2} + c_1\sum_{r\geq 4}(r^2 - r)t_r(\mathcal{V}) \leq \frac{n}{2}b^*. \tag{2.13}$$

Clearly we have $b^* > 1$. Recall that $\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}) < n^2$, implying that for c_1 small enough and n large enough,

$$b^* < 4 + \frac{2c_1}{n}\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}) < \frac{1}{6}n. \tag{2.14}$$

Now by Lemma 2.3.4 and (2.13), either the number of ordinary lines

$$t_2(\mathcal{V}) \geq \frac{n}{2}b^* \geq \frac{3n}{2} + c_1\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}),$$

or, using (2.14), there exists a point $v \in \mathcal{V}$, such that the number of ordinary lines containing v is at least

$$\frac{2}{3}(n+1) - b^* > \frac{1}{2}n.$$

□

The following lemma will be crucially used in the proof of Theorem 2.1.7.

Lemma 2.6.3. *Let $\mathcal{V}$ be a set of n points in $\mathbb{C}^d$, and $\mathcal{V}' = \mathcal{V}\setminus\{v\}$ for some $v \in \mathcal{V}$. Then*

$$\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}') \geq \sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}) - 4(n-1).$$

Proof. Note that when we remove v from the set $\mathcal{V}$, we only affect lines that go through v. In particular, ordinary lines through v are removed and the number of points on every special line through v goes down by 1. Every other line remains unchanged and so it suffices to consider only lines that contain the point v.

We consider the difference

$$K = \sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}) - \sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}').$$

We will consider the contribution of a line l determined by $\mathcal{V}$ to the difference K.

Each line $l \in \mathcal{L}_{\geq 5}(\mathcal{V})$, i.e. a line that has $r \geq 5$ points, that contains v contributes r^2-r to the summation $\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V})$. In $\mathcal{V}'$, l has $r-1$ points, and contributes $(r-1)^2-(r-1)$ to the summation $\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}')$. Therefore, l contributes $2(r-1)$ to the difference K. We may charge this contribution to the points on l that are not v. There are $r-1$ other points on l, so each point contributes 2 to K.

Each line $l \in \mathcal{L}_4(\mathcal{V})$ that contains v contributes $r^2 - r = 12$ to the summation $\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V})$. These lines contain 3 points in $\mathcal{V}'$, and so do not contribute anything in the $\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}')$ term. Once again, we charge this contribution to the points lying on l that are not v. Each such line has 3 points on it other than v, so each point contributes $12/3 = 4$ to K.

There is a unique line through v and any other point, and each point either contributes 0, 2 or 4 to K. This gives us that

$$\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}) - \sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}') \leq 4(n-1).$$

Rearranging completes the proof. □

We are now ready to prove the main theorem. For convenience, we restate the theorem here.

Theorem 2.1.7. *There exists an absolute constant $c > 0$ and a positive integer n_0 such that the following holds. Let $\mathcal{V}$ be a set of $n \geq n_0$ points in $\mathbb{C}^3$ with at most $\frac{2}{3}n$ points contained in any plane. Then*

$$t_2(\mathcal{V}) \geq \frac{3}{2}n + c\sum_{r\geq 4} r^2 t_r(\mathcal{V}).$$

Proof. The remainder of the proof is similar to the proof of Theorem 2.1.8, i.e. we use Corollary 2.6.2 to find a point with a large number of ordinary lines, "prune" this point, and then repeat this on the smaller set of points. We stop when either we can not find such a point, in which case Corollary 2.6.2 guarantees a large number of ordinary lines, or when we have accumulated enough ordinary lines.

As before, consider the following algorithm: Let $\mathcal{V}_0 := \mathcal{V}$ and $j = 0$.

1. If $\mathcal{V}_j$ satisfies case (2) of Lemma 2.6.2, then stop.

2. Otherwise, there must exist a point v_{j+1} with at least $\frac{n-j}{2}$ ordinary lines through it. Let $\mathcal{V}_{j+1} = \mathcal{V}_j \setminus \{v_{j+1}\}$.

3. Set $j = j + 1$. If $j = n/3$, then stop. Otherwise go to Step 1.

Note that since no plane contains more than $2n/3$ points, at no point will the algorithm stop because the configuration becomes planar. That is, we can use Corollary 2.6.2 at every step of the algorithm. We now analyze the two stopping conditions for the algorithm, and show that we can always find enough ordinary lines by the time the algorithm stops.

Suppose that we stop because $\mathcal{V}_j$ satisfies case (2) of Corollary 2.6.2 for some $1 \leq j < n/3$. From case (2) of Lemma 2.6.2 and Lemma 2.6.3, we have that

$$\begin{aligned} t_2(\mathcal{V}_j) &\geq \frac{3(n-j)}{2} + c_1 \sum_{r\geq 4} (r^2 - r) t_r(\mathcal{V}_j) \\ &\geq \frac{3(n-j)}{2} + c_1 \left(\sum_{r\geq 4} (r^2 - r) t_r(\mathcal{V}) - 4 \sum_{i=1}^{j} (n-i) \right). \end{aligned} \tag{2.15}$$

On the other hand, each pruned point v_i, $1 \leq i \leq j$, has at least $\frac{n-i+1}{2} > \frac{n-i}{2}$ ordinary lines determined by $\mathcal{V}_{i-1}$ through it, and hence at most $(n - i - \frac{n-i+1}{2})/2 < \frac{n-i}{4}$ special lines through it. Note that an ordinary line in $\mathcal{V}_i$ might not be ordinary in $\mathcal{V}_{i-1}$ if contains v_i. Thus, in order to lower bound the total number of ordinary lines in $\mathcal{V}$, we sum over the number of ordinary lines contributed by each of the pruned points v_i, $1 \leq i \leq j$, and subtract from the count the number of potential lines that could contain v_i. Then the number of ordinary lines contributed by the pruned points is at least

$$\sum_{i=1}^{j} \left(\frac{n-i}{2} - \frac{n-i}{4} \right) = \frac{1}{4} \sum_{i=1}^{j} (n-i)\,. \tag{2.16}$$

Combining (2.15) and (2.16), we get that

$$\begin{aligned} t_2(\mathcal{V}) &\geq \frac{3}{2}(n-j) + c_1 \left(\sum_{r\geq 4} (r^2 - r) t_r(\mathcal{V}) - 4 \sum_{i=1}^{j} (n-i) \right) + \frac{1}{4} \sum_{i=1}^{j} (n-i) \\ &= \frac{3}{2}n + c_1 \sum_{r\geq 4} (r^2 - r) t_r(\mathcal{V}) + \left(\frac{1}{4} - 4c_1 \right) \sum_{i=1}^{j} (n-i) - \frac{3}{2}j. \end{aligned}$$

For c_1 small enough and n large, the term $\left(\frac{1}{4} - 4c_1\right) \sum_{i=1}^{j} (n-i) - \frac{3}{2}j$ is positive. Therefore, there exists some absolute constant $c > 0$ such that

$$t_2(\mathcal{V}) \geq \frac{3}{2}n + c \sum_{r\geq 4} r^2 t_r(\mathcal{V}).$$

We now consider the case when the algorithm stops because $j = n/3$. Note that at this point, we will have pruned exactly j points. Each pruned point v_i, $1 \leq i \leq j$, has $\frac{n-i+1}{2} > \frac{n-i}{2}$ ordinary lines determined by $\mathcal{V}_{i-1}$ through it. However, as many as $i-1 < i$ ordinary lines through v_i contain other pruned points v_k, $k < i$, i.e. lines that could be special in $\mathcal{V}$. Therefore the total number of ordinary lines determined by $\mathcal{V}$ is at least

$$t_2(\mathcal{V}) \geq \sum_{i=1}^{j} \frac{n-i}{2} - \sum_{i=1}^{j} i = \frac{1}{2}\sum_{i=1}^{j}(n-3i).$$

Since $j = \frac{n}{3}$, we get that the number of ordinary lines determined by $\mathcal{V}$ is at least

$$t_2(\mathcal{V}) \geq \frac{1}{2}\sum_{i=1}^{j}(n-3i) = \frac{5n^2-12n}{64}.$$

Recall that $n^2 \geq \sum_{r\geq 4}(r^2-r)t_r(\mathcal{V})$, which gives us that

$$t_2(\mathcal{V}) \geq \frac{3}{2}n + c\sum_{r\geq 4} r^2 t_r(\mathcal{V})$$

for some absolute constant $c > 0$ and n large enough.

□

Chapter 3

Finite Field Kakeya Sets in Three Dimensions

3.1 Finite Field Kakeya Sets

Let $\mathbb{F}_q$ denote the finite field of q elements. A *Kakeya set* $K \subseteq \mathbb{F}_q^n$ is a set of points which contains 'a line in every direction'. More precisely, for all $x \in \mathbb{F}_q^n$ there is a $y \in \mathbb{F}_q^n$ such that the *line* $\{xt + y, t \in \mathbb{F}_q\} \subseteq K$.

The question of establishing lower bounds for Kakeya sets over finite fields was asked by Wolff [Wol99]. In 2008, in a breakthrough result, Dvir [Dvi09] showed that for a Kakeya set K over a finite field $\mathbb{F}$ of size q, $|K| > \frac{q^n}{n!}$, thus exactly pinning down the exponent of q in the lower bound. Later in 2008, Saraf and Sudan [SS08] improved the lower bound to the form $1/2 \cdot \beta^n q^n$, where β is approximately $1/2.6$. Moreover, Dvir showed how to construct a Kakeya set of size $\frac{q^n}{2^{n-1}} + O(q^{n-1})$ (see [SS08]). In 2009, Dvir, Kopparty, Saraf and Sudan [DKSS09] proved a lower bound of $\frac{q^n}{2^n}$ for the size of Kakeya sets. Thus the gap between the lower bound and the upper bound given by the construction is only at most a factor of 2, and it is a very interesting question to close this gap. Though we now know extremely strong lower bounds, we still do not know an exact bound for any dimension other than 2. For $n = 2$, Blokhuis and Mazzocca gave exact bounds on the size of a Kakeya set of $q(q+1)/2 + (q-1)/2$ for odd q and $q(q+1)/2$ for even q. Here we give improved lower bounds for dimension $n = 3$, using an extension of the argument presented in [SS08].

In the rest of Chapters 3, all asymptotics will be in terms of q. We will use n to represent the dimension of the underlying space, but we will think of it as a fixed constant and the underlying field size q to be growing. Thus $o(1)$ will be a function that tends to 0 as q tends to ∞.

3.1.1 Kakeya sets: Background and our results

We prove the following improved lower bounds for Kakeya sets in dimensions $n = 3$.

Theorem 3.1.1. *There exists a constant $C > 0$, such that for any prime power $q > C$, if $K \subseteq \mathbb{F}_q^3$ is a Kakeya set, then*

$$|K| \geq 0.2107q^3.$$

Prior to this work, the best lower bound for $n = 3$ was obtained by Saraf and Sudan [SS08], and they achieved a lower bound of $(0.208)q^3$.

Though the quantitative improvement in the lower bound is small, we believe our proof method is interesting and might be of independent interest. The proof of Saraf and Sudan [SS08] extended the beautiful polynomials based lower bounds of Dvir [Dvi09] by using the notion of the multiplicity of roots of polynomials. Our work uses the notion of "fractional multiplicity" to obtain the improved result. We say a few more words about these proof methods.

Dvir [Dvi09] obtained his lower bound via the following argument using polynomials: If the size of K is small, then interpolate a nonzero low degree polynomial P vanishing on all the points of K. Then, use the properties of K to show that P must actually vanish at all points of the underlying space[1]. However this contradicts the low degreeness of P.

The work of Saraf and Sudan [SS08] extends this idea by taking a polynomial P that vanishes of each point of K with some higher multiplicity m. To enable this, they allow the degree of P to be somewhat higher, but they cap the individual degree of each variable of P. This idea somehow still enables them to get the same conclusion as Dvir, but now with stronger bounds. The novelty of the current work is that we allow the multiplicity m to take a non-integer value. We need to now specify what it means for a polynomial to vanish with multiplicity m, where m is a positive real number that is not an integer. For this we define a suitable random process which makes the expected

[1] Actually in this step Dvir uses a polynomial very closely related to P, but for simplicity we think of it to be P itself.

multiplicity of P at a point equal to m. By allowing m to take a non-integer value we are able to make finer optimizations.

3.2 Preliminary Results and Lemmas

Let $\mathbb{F}_q[x_1, ..., x_n] = \mathbb{F}_q[\mathbf{x}]$ be the ring of polynomials in $x_1, ..., x_n$ with coefficients in $\mathbb{F}_q$.

The following is a basic and well known fact about zeroes of polynomials.

Fact 1. *Let $P \in \mathbb{F}_q[\mathbf{x}]$ be a polynomial of degree at most $q-1$ in each variable. If $P(a) = 0$ for each $a \in \mathbb{F}_q^n$, then $P \equiv 0$.*

Let $N_q(n,m)$ be the number of monomials in $\mathbb{F}_q[x_1, ..., x_n]$ of individual degree $< q$ and total degree $< mq$. Note that m need not be a natural number to define $N_q(n,m)$, rather m can be any positive real number greater than or equal to 1.

Lemma 3.2.1.

$$N_q(n,m) = \sum_{i=0}^{n} (-1)^i \binom{n}{i} \binom{\lfloor (m-i)q + n - 1 \rfloor}{n},$$

where $\lfloor x \rfloor$ is the largest integer that is at most x.

Proof. The proof will be via inclusion-exclusion. Consider the total number of monomial terms of a polynomial of total degree strictly less than mq. This equals $\binom{\lfloor mq+n-1 \rfloor}{n}$. We only want to include those monomials in our count that have individual degree at most $q-1$. Let C_r be the total number of monomials of total degree strictly less than mq and some particular r of the variables having degree q or more. Then by inclusion-exclusion,

$$N_q(n,m) = \sum_{i=0}^{n} (-1)^i \binom{n}{i} C_i.$$

It is not hard to see that $C_i = \binom{\lfloor (m-i)q+n-1 \rfloor}{n}$ since if a particular set of i variables must have degree at least q, we can "peel off" degree q part from each of these variables to get a resulting monomial of total degree at most $\lfloor (m-i)q + n - 1 \rfloor$. C_i is then then number of such monomials which equals $\binom{\lfloor (m-i)q+n-1 \rfloor}{n}$.

□

Definition 3.2.2 (multiplicity)**.** *For a polynomial $g \in \mathbb{F}_q[\mathbf{x}]$, we say g vanishes at a point $\mathbf{a}$ with multiplicity m if $g(\mathbf{x} + \mathbf{a})$ has no monomial term of degree lower than m.*

The following lemma is a simple adaptation of a lemma from [SS08] (where instead of two sets S_1 and S_2 there was only one set).

Lemma 3.2.3. *Let $m_1 \geq 0$ and $m_2 \geq 0$ be integers and $m > 0$ be a real number. Let $S_1, S_2 \subset \mathbb{F}_q^n$ be disjoint sets such that $|S_1|\binom{m_1+n-1}{n} + |S_2|\binom{m_2+n-1}{n} < N_q(n, m)$. Then there exists a non-zero polynomial $g \in \mathbb{F}_q[\mathbf{x}]$ of total degree less than mq and individual degree at most $q - 1$ such that g vanishes on each point of S_1 with multiplicity m_1 and on S_2 with multiplicity m_2.*

Proof. The total number of possible monomials in g is $N_q(n, m)$. We consider the coefficients of these monomials to be free variables. For each point $\mathbf{a} \in \mathbb{F}_\mathbf{q}^\mathbf{n}$, requiring that the polynomial vanishes on $\mathbf{a}$ with multiplicity m_i adds $\binom{m_i+n-1}{n}$ homogeneous linear constraints on these coefficients. Requiring that g vanishes on each point of S_1 with multiplicity m_1 and on S_2 with multiplicity m_2 imposes a total of $|S_1|\binom{m_1+n-1}{n} + |S_2|\binom{m_2+n-1}{n}$ homogeneous linear constraints. Since $|S_1|\binom{m_1+n-1}{n} + |S_2|\binom{m_2+n-1}{n} < N_q(n, m)$, the total number of homogeneous linear constraints is strictly less than the number of variables and hence a nonzero solution exists. Thus there exists a non-zero polynomial $g \in \mathbb{F}_q[\mathbf{x}]$ of total degree less than mq and individual degree at most $q - 1$ such that g vanishes on each point of S_1 with multiplicity m_1 and on S_2 with multiplicity m_2. □

For $g \in \mathbb{F}_q[\mathbf{x}]$ let $g_{\mathbf{a},\mathbf{b}}(t) = g(\mathbf{a} + t\mathbf{b})$ denote its restriction to the "line" $\{\mathbf{a} + t\mathbf{b}, t \in \mathbb{F}_q\}$. The lemma below is a basic result that also appears in [SS08].

Lemma 3.2.4. *If $g \in \mathbb{F}_q[\mathbf{x}]$ vanishes with multiplicity m at some point $\boldsymbol{a} + t_0\boldsymbol{b}$ then $g_{a,b}$ vanishes with multiplicity m at t_0.*

Proof. By definition, the fact that g has a zero of multiplicity m at $\mathbf{a}+t_0\mathbf{b}$ implies that the polynomial $g(\mathbf{x}+\mathbf{a}+t_0\mathbf{b})$ has no support on monomials of degree less than m. Thus under the homogeneous substitution of $\mathbf{x} \to t\mathbf{b}$, we get no monomials of degree less than

m either, and thus we have t^m divides $g(t\mathbf{b}+\mathbf{a}+t_0\mathbf{b}) = g(\mathbf{a}+(t+t_0)\mathbf{b}) = g_{\mathbf{a},\mathbf{b}}(t+t_0)$. Hence $g_{\mathbf{a},\mathbf{b}}$ has a zero of multiplicity m at t_0.

□

The following theorem was the lower bound result from [SS08].

Theorem 3.2.5 (Kakeya lower bound from [SS08])**.** *If K is a Kakeya set in $\mathbb{F}_q^n$, then $|K| \geq \frac{1}{\binom{m+n-1}{n}} N_q(n,m)$.*

By setting $n = 3$ and $m = 2$, it is concluded in [SS08] that for a Kakeya set $K \subseteq \mathbb{F}_q^n$, $|K| \geq \frac{5}{24}q^3 \approx 0.2083q^3$. We manage to obtain our strengthened lower bound by allowing m to take values that are not necessarily integers. In particular, we introduce a notion of vanishing with fractional multiplicity and show that it can be used for an improved bound.

3.3 Proof of Theorem 3.1.1

Let $K \subseteq \mathbb{F}_q^3$ be a Kakeya set. As a first step in the proof, we will interpolate a nonzero polynomial vanishing on the points of K with some possibly fractional multiplicity m. If we wanted to interpolate a polynomial vanishing with multiplicity m where m is sandwiched between two positive integers u and $u+1$, for each point we could make it vanish with multiplicity u with some probability, say α, and with multiplicity $u+1$ with probability $1-\alpha$, so that in expectation the multiplicity of vanishing would be at least m. It turns out that the main property of the multiplicities of vanishing we will need is that on each *line* of the Kakeya set, almost the correct α fraction of points have multiplicity of vanishing being at least u and the rest have multiplicity of vanishing at least $u+1$. To do this we will first identify an appropriate subset S of the Kakeya set on which we will want the vanishing multiplicity to be u, and in the lemma below we show that such a set can be suitably picked.

Lemma 3.3.1. *Let $K \subseteq \mathbb{F}_q^3$ be a Kakeya set. Let $0 \leq \alpha \leq 1$, and $\delta = \frac{1}{\sqrt[3]{q}}$. Then there exists a constant $C > 0$ such that for $q > C$ we can pick a subset $S \subset K$ such that $||S| - \alpha|K|| < \delta\alpha|K|$, and such that for each line L contained in $|K|$, $||L \cap S| - \alpha q| < \delta\alpha q$.*

Proof. Consider a random subset $S \subset K$, where we choose each point in S independently with probability α. By the Chernoff Bound, $\mathbb{P}[||S| - \alpha|K|| \geq \delta\alpha|K|] \leq \exp(-\frac{\alpha|K|\delta^2}{3})$. Since $|K|$ is certainly larger than q, $\exp(-\frac{\alpha|K|\delta^2}{3}) \leq \exp(-\frac{\alpha q\delta^2}{3})$.

Note also that there are only $q^4 + q^3 + q^2$ distinct lines in $\mathbb{F}_q^3$, and thus at most $q^4 + q^3 + q^2$ lines in K. Let L be any line in K. Again, via the Chernoff Bound, we have $\mathbb{P}[||L \cap S| - \alpha q| \geq \delta\alpha q] \leq \exp(-\frac{\alpha q\delta^2}{3})$. By the union bound, the probability that any one of the lines in K has more than $(1 + \alpha\delta)q$ or fewer than $(1 - \alpha\delta)q$ points in S is at most $(q^4 + q^3 + q^2)\exp(-\frac{\alpha q\delta^2}{3})$.

Thus if we show that $\exp(-\frac{\alpha q\delta^2}{3}) + (q^4+q^3+q^2)\exp(-\frac{\alpha q\delta^2}{3}) < 1$, then by the probabilistic method, such a subset S with the desired properties exists. Since $\lim_{q\to\infty} \exp(-\frac{\alpha q\delta^2}{3}) + (q^4 + q^3 + q^2)\exp(-\frac{\alpha q\delta^2}{3}) = 0$ for$\delta = \frac{1}{\sqrt[3]{q}}$, there exists some constant $C > 0$ such that for $q > C$, there exists such a set S. □

Lemma 3.3.2. *Let $K \subseteq \mathbb{F}_q^3$ be a Kakeya set. Let $u \in \{1, 2\}$, let α be such that $0 \leq \alpha \leq 1$, $\delta = \frac{1}{\sqrt[3]{q}}$ and $m = (\alpha - \delta\alpha)u + (1 - \alpha - \delta\alpha)(u + 1)$. Then*

$$N_q(3, m) \leq (\alpha + \delta\alpha)\binom{2+u}{3}|K| + (1 - \alpha + \delta\alpha)\binom{3+u}{3}|K|.$$

Proof. Suppose for contradiction, $N_q(3, m) > (\alpha+\delta\alpha)\binom{2+u}{3}|K| + (1-\alpha+\delta\alpha)\binom{3+u}{3}|K|$. By Lemma 3.3.1, choose S such that each line in K has between $\alpha q - \delta\alpha q$ and $\alpha q + \delta\alpha q$ points in S and $||S| - \alpha|K|| < \delta\alpha|K|$. In particular $|S| < (\alpha+\delta\alpha)|K|$ and $|K \setminus S| < (1-\alpha + \delta\alpha)|K|$. Then by Lemma 3.2.3 there exists a nonzero polynomial $g \in \mathbb{F}_q[x_1, x_2, x_3]$ with total degree less than mq and individual degree less than q such that g vanishes on S with multiplicity at least u and on $K \setminus S$ with multiplicity at least $u + 1$. Let d denote the degree of g. Let $g = g_0 + g_1$, where g_0 denotes the homogeneous part of degree d and g_1 the part with degree less than d. Note that g_0 also has degree at most $q - 1$ in each of its variables.

Fix a "direction" $\mathbf{b} \in \mathbb{F}_q^3$. Since K is a Kakeya set, there exists $\mathbf{a} \in \mathbb{F}_q^3$ such that the line $\mathbf{a} + t\mathbf{b} \in K$ for all $t \in \mathbb{F}_q$. So consider $g_{a,b}(t)$, the univariate polynomial of g restricted to the line $\mathbf{a} + t\mathbf{b}$. By Lemma 3.3.1 and Lemma 3.2.4, there are at least $(1 - \delta)\alpha q$ choices of t where $g_{a,b}$ vanishes with multiplicity at least u and there are at least $q - \alpha q - \delta\alpha q$ choices of t, where $g_{a,b}$ vanishes with multiplicity at least $u + 1$. So in total, $g_{a,b}$ has at least $(\alpha-\delta\alpha)uq+(1-\alpha-\delta\alpha)(u+1)q = mq$ zeros, which is more than its degree. Therefore, $g_{a,b}$ must be identically zero. In particular, its leading coefficient must be 0. Since this leading coefficient equals $g_0(\mathbf{b})$, $g_0(\mathbf{b}) = 0$. Since b was chosen arbitrarily, this must happen for all $\mathbf{b} \in \mathbb{F}_q^3$. However, by Fact 1, this contradicts the fact that g_0 is a nonzero polynomial of degree at most $q-1$ in each of its variables. □

Proof of Theorem 3.1.1. Let $\delta = \frac{1}{\sqrt[3]{q}}$, let $u \in \{1,2\}$, let α be such that $0 \le \alpha \le 1$, and $m = (\alpha-\delta\alpha)u + (1-\alpha-\delta\alpha)(u+1)$. Note that once we set the value for u and m between 1 and 2, this will determine a value for α. For now suppose we have chosen some values for u, α and m.

By Lemma 3.3.2, $|K| \ge \frac{N_q(3,m)}{(\alpha+\delta\alpha)\binom{2+u}{3}+(1-\alpha+\delta\alpha)\binom{3+u}{3}}$. Since we are considering $|K|$ as q grows asymptotically, we only need to consider the leading term when $N_q(3,m)$ is expressed as a polynomial in q. Also, note that δ becomes small as q grows large.

The reason we only let u take value 1 or 2 is the following. Since we only care about polynomials with individual variable degree less than q, the total degree must be less than $3q$. Choosing a value of m that is greater than or equal to 3 will just end up being somewhat redundant and end up giving a worse bound. Thus we only consider $m < 3$. Given the relationship between u and m and given that u needs to be an integer, the only choices for u are hence 1 or 2 as in the statement of the above lemma.

When $u = 1$, this makes $m = 2-(1+o(1))\alpha$ for large q. By Lemma 3.2.1,

$$N_q(3,m) = \left(\frac{-2m^3+9m^2-9m+3}{6} + o(1)\right) q^3.$$

Substituting $u = 1$, by Lemma 3.3.2 we get that

$$|K| \ge \left(\frac{-2m^3+9m^2-9m+3}{6(4-3\alpha)} + o(1)\right) q^3 = \left(\frac{-2m^3+9m^2-9m+3}{6(3m-2)} + o(1)\right) q^3.$$

We maximize this for $1 \le m \le 2$. For m=1.84, this gives $|K| \ge (0.21076 + o(1))q^3$. When $u = 2$, the best lower bound achieved in this case is $|K| \ge (.2083 + o(1))q^3$. Thus overall the best lower bound we achieve is $(0.21076 + o(1))q^3$. □

Chapter 4

Finite Field Nikodym Sets in Three Dimensions

4.1 Introduction

A very closely related notion to Kakeya sets is that of Nikodym sets. A *Nikodym set* $\mathcal{N} \subseteq \mathbb{F}_q^n$ is a set of points such that, through each point $p \in \mathbb{F}_q^n$, there is a line ℓ such that $\ell \setminus \{p\} \subseteq \mathcal{N}$.

In fact, a lower bound for Kakeya sets implies a lower bound for Nikodym sets by the following argument: first observe that up to a multiplicative factor, lower bounds for Kakeya or Nikodym sets will not change regardless of whether the set is over affine or projective spaces over finite fields. Now take a Nikodym set over the finite projective space $PG(n, q)$. We will argue that it is also a Kakeya set. Consider the lines through points in the hyperplane at infinity. Each point determines a line pointing in each different affine "direction." An entire line pointing in the direction dictated by the point must be included in the Nikodym set. By definition, a set containing a line pointing in every direction is a Kakeya set.

Almost all lower bounds for Nikodym sets currently follow from a lower bound for Kakeya sets, although we believe that much stronger lower bounds should hold for Nikodym sets. Here we study Nikodym sets in 3 dimensions over finite fields and give improved bounds for this setting. We also study a related notion of weak Nikodym sets in 3 dimensions.

The main conjecture in the study of finite Nikodym sets we focus on is the following.

Conjecture 4.1.1. *Let $\mathcal{N}$ be a Nikodym set in $\mathbb{F}_q^n$. Then,*

$$|\mathcal{N}| \geq (1 - o(1))q^n.$$

Conjecture 4.1.1 is known in some special cases. Feng, Li, and Shen [FLS10] showed that the complement of a Nikodym set in $\mathbb{F}_q^2$ is at most $q^{3/2}+q$ points. Guo, Kopparty, and Sudan [GKS13] proved Conjecture 4.1.1 for all dimensions, but only over fields of constant characteristic. The only known lower bound on the size of a Nikodym set for general n and q matches the corresponding bound for Kakeya sets.

In Section 4.2, we prove the following theorem which gives the first separation between the minimum possible size of Kakeya and Nikodym sets in $\mathbb{F}_q^3$ for any sufficiently large prime power q.

Theorem 4.1.1. *Let $\mathcal{N}$ be a Nikodym set in $\mathbb{F}_q^3$. Then,*

$$|\mathcal{N}| \geq (0.38 - o(1))q^3.$$

While this falls short of proving the case $n = 3$ of Conjecture 4.1.1, it does show a separation between Kakeya and Nikodym sets in $\mathbb{F}_q^3$, since the construction in [SS08] gives a Kakeya set of size $(0.25 + o(1))q^3$.

A conjecture on the union of lines

For L a set of lines, we define $P(L)$ to be the collection of points contained in some line of L. More precisely,

$$P(L) = \bigcup_{\ell \in L} \{p \mid p \in \ell\}.$$

In Section 4.2.2, we show that a slight modification of the proof of Theorem 4.1.1 shows that if L is any set of $(0.62 + o(1))q^3$ lines in $\mathbb{F}_q^3$, then $|P(L)| \geq (0.38 - o(1))q^3$. Such a result is stronger than Theorem 4.1.1 since the definition of a Nikodym set guarantees the existence of a set L of lines, one for each point in the complement of the Nikodym set, such that all but one point of each line of L is contained in the Nikodym set. We also show that the $(0.38 - o(1))q^3$ bound is nearly tight.

The proof of Theorem 4.1.1 uses very little information about L (the set of lines corresponding to the complement of a Nikodym set), and there is actually a lot more structure that one might be able to exploit in order to get a stronger result. For example, we show in Section 4.2.3 that no more than $(1 + o(1))q^{3/2}$ lines of L can be contained in any plane. We believe that the approach of bounding the size of the set of lines associated to the complement of a Nikodym set could lead to a proof of Conjecture 4.1.1, if this additional structure of L is used.

To this end, we propose the following conjecture.

Conjecture 4.1.2. *If L is a set of lines in $\mathbb{F}_q^3$ such that $|L| = \Omega(q^3)$, and such that no plane contains $\omega(q)$ lines of L, then $|P(L)| \geq (1 - o(1))q^3$.*

In Section 4.2.3, we show that Conjecture 4.1.2 implies the three dimensional case of Conjecture 4.1.1. In addition to making it a very interesting conjecture for understanding Nikodym sets, the conjecture seems also very natural and worthwhile to study for its own sake.

Conjecture 4.1.2 resembles a recent result of Ellenberg and Hablisek [EH13]. A special case of Ellenberg and Hablisek's theorem states that, if p is a prime and L is a set of p^2 lines in $\mathbb{F}_p^3$ such that no more than p lines of L lie in any plane, then $|P(L)| = \Omega(p^3)$. The main differences between Conjecture 4.1.2 and the result of Ellenberg and Hablisek is that we take L to be much larger, we allow the underlying field to have composite order, and our desired conclusion is stronger.

For Ellenberg and Hablisek's result, the condition that the underlying field has prime order is necessary. Indeed, they observe that a nondegenerate Hermitian variety in $\mathbb{F}_q^3$ for q a perfect square (which we discuss further in Section 4.3.1) contains a set L of q^2 lines, no more than $(1 + o(1))q^{1/2}$ on any plane, such that $|P(L)| = (1 + o(1))q^{5/2}$ points.

Although Conjecture 4.1.2 would be sufficient for an application to Conjecture 4.1.1, we do not have a counterexample to the following, much stronger, conjecture.

Conjecture 4.1.3. *Let $\epsilon > 0$ be any constant and let q be a sufficiently large prime power. Let L be a set of at least $q^{5/2+\epsilon}$ lines in $\mathbb{F}_q^3$ such that no plane contains more than $(1/2)q^{3/2}$ lines of L. Then, $|P(L)| \geq (1 - o(1))q^3$.*

It may even be that the conclusion $|P(L)| \geq (1 - o(1))q^3$ in Conjecture 4.1.3 could be replaced by $|P(L)| \geq q^3 - 2q^{5/2}$ without admitting a counterexample.

There are reasons to be skeptical of Conjecture 4.1.3. Although the construction of Ellenberg and Hablisek mentioned above does not directly give a counterexample, it might be possible to construct a counterexample by taking the union of many, carefully chosen, copies of their construction. In fact, in Section 4.3.1 we use Hermitian varieties to construct a set of lines with the following parameters.

Proposition 4.1.1. *Let $q = p^2$ for a prime power p. There is a set L of $(1/2-o(1))q^{7/2}$ lines in $\mathbb{F}_q^3$ such that no plane contains more than $(1/2)q^{3/2}$ lines of L, and $|P(L)| = q^3 - (1/2 + o(1))q^{5/2}$.*

A proof of Conjecture 4.1.2 would be new and very interesting even in the case of prime order fields, for which the above constructions based on Hermitian varieties do not occur and it is thus even more likely that even Conjecture 4.1.3 might be true.

Weak Nikodym sets

All existing lower bounds on the size of a Nikodym set use only much weaker properties of Nikodym sets. To capture the part of the definition that is actually used by the existing proofs, we introduce and initiate the explicit study of *weak Nikodym sets.* A weak Nikodym set $\mathcal{N}$ in $\mathbb{F}_q^n$ is a set of points such that, through each point p in the complement $\mathcal{N}^c$ of $\mathcal{N}$, there is a line ℓ such that $\ell \setminus \{p\} \subseteq \mathcal{N}$.

In Section 4.3.1 we give improved constructions of weak Nikodym sets, and based on these we conjecture that, at least for fields of square order, there are weak Nikodym that contain significantly fewer points than any Nikodym set. Since current lower bound proofs for Nikodym sets only use the fact that Nikodym sets are also weak Nikodym sets, these proofs are inadequate to prove such a separation.

4.2 Nikodym sets in 3 dimensions and the union of lines

In this section, we investigate Nikodym sets in $\mathbb{F}_q^3$ and give improved lower bounds.

We will find it easier to work with the complement of a Nikodym set rather than the Nikodym set itself. We define

$$f(n, q) = \text{ the maximum size of the complement of a Nikodym set in } \mathbb{F}_q^n.$$

We additionally denote the complement of a set $\mathcal{N}$ by $\mathcal{N}^c$.

Using this notation, Conjecture 4.1.1 states that $f(n,q) = o(q^n)$, and Theorem 4.1.1 states that $f(3,q) \leq (0.62 + o(1))q^3$.

In Section 4.2.1, we prove Theorem 4.1.1; as mentioned in the introduction, this is the first separation demonstrated between the minimum size of a Nikodym set and the minimum size of a Kakeya set in $\mathbb{F}_q^3$ that is valid for an arbitrary finite field $\mathbb{F}_q$.

In Section 4.2.2, we show that the proof of Theorem 4.1.1 given in Section 4.2.1 immediately implies a lower bound on the number of points incident to a large set of lines, and that this bound is nearly tight. This implies that any substantial improvement to Theorem 4.1.1 will need to use some property of Nikodym sets that is not exploited by the proof given in Section 4.2.1.

In Section 4.2.3, we observe that a weak Nikodym set has the property that not too many of the lines given by its definition can lie in any single plane. We further suggest that exploiting this property might lead to a proof of Conjecture 4.1.1 in the three dimensional case. In particular, we show that a proof of Conjecture 4.1.2 would immediately imply the case $n = 3$ of Conjecture 4.1.1.

4.2.1 Proof of Theorem 4.1.1

Our bound on $f(3,q)$ will use a bound on the number of incidences between points and lines. The bound we will use was essentially proved by Lund and Saraf in [LS14], but is not explicitly stated there; a similar bound was obtained by Bennett, Iosevich, and Pakianathan [BIP14]. We show how to recover the bound from arguments given in [LS14].

Given a set P of points and a set L of lines, we denote the number of incidences between P and L as

$$I(P,L) = |\{(p,\ell) \in P \times L \mid p \in \ell\}|.$$

Theorem 4.2.1. *Let L be a set of lines and P a set of points in $\mathbb{F}_q^3$. Then,*

$$I(P,L) \leq (1 + o(1)) \left(|P||L|q^{-2} + q\sqrt{|P||L|(1 - |P|q^{-3})(1 - |L|q^{-4})}\right).$$

Proof. A (d_U, d_V)-biregular graph G is a bipartite graph such that each each left vertex has degree d_U and each right vertex has degree d_V. We denote by $e(G)$ the number of edges in a graph G, and by $e(S,T)$ the number of edges between two subsets S,T of the vertices of a graph. We will use the expander mixing lemma [AC88], specifically the following bipartite version. A proof of Lemma 4.2.2 is given in [LS14], and an equivalent result was proved much earlier by Haemers, e.g. [Hae95].

Lemma 4.2.2 (Bipartite expander mixing lemma, [LS14]). *Let G be a (d_U, d_V)-biregular graph with left vertices U and right vertices V. Let A be the (square) adjacency matrix of G, and let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{|U|+|V|}$ be the eigenvalues of A. Let $\lambda = \lambda_2/\lambda_1$. Let $S \subseteq U$ with $|S| = \alpha|U|$ and let $T \subseteq V$ with $|T| = \beta|V|$. Then,*

$$\left|\frac{e(S,T)}{e(G)} - \alpha\beta\right| \leq \lambda\sqrt{\alpha\beta(1-\alpha)(1-\beta)}.$$

Construct a bipartite graph G with left vertices U being the points of $\mathbb{F}_q^3$, and right vertices V being the lines of $\mathbb{F}_q^3$, with (p,ℓ) in the edge set of G if and only if $p \in \ell$. The number of points in $\mathbb{F}_q^3$ is $|U| = q^3$; the number of lines is $|V| = (1+o(1))q^4$; and the number of incidences between points and lines in $\mathbb{F}_q^3$ is $e(G) = (1+o(1))q^5$. It is shown in Section 4 of [LS14] that the largest eigenvalue of this graph is $(1+o(1))q^{3/2}$, and the second largest eigenvalue is $(1+o(1))q$. We are interested in the number of incidences between a set $P \subseteq U$ and $L \subseteq V$. This is exactly the number of edges between P and L in G, and hence we apply Lemma 4.2.2 with $\alpha = |P|q^{-3}$ (which is the density of P in U) and $\beta = (1-o(1))|L|q^{-4}$ (which is the density of L in V), to get that

$$\left|(1+o(1))(I(P,L)q^{-5} - |L||P|q^{-7})\right| \leq (1+o(1))q^{-4}\sqrt{|P||L|(1-|P|q^{-3})(1-|L|q^{-4})}.$$

Thus, simplifying we get

$$I(P,L) \leq (1+o(1))\left(|P||L|q^{-2} + q\sqrt{|P||L|(1-|P|q^{-3})(1-|L|q^{-4})}\right).$$

□

Now, we complete the proof of Theorem 4.1.1.

Proof of Theorem 4.1.1. Suppose that $\mathcal{N}^c$ is the complement of a weak Nikodym set in $\mathbb{F}_q^3$. Let L be a set of $|\mathcal{N}^c|$ lines such that each line has exactly one point in common with $\mathcal{N}^c$, and there is exactly one line of L through each point of $\mathcal{N}^c$; the existence of such a set is guaranteed by the definition of a weak Nikodym set. Let $P = \mathcal{N}$; by definition, $|P| = q^3 - |L|$. Then each line of L is incident to exactly $q-1$ points of P, so $I(P, L) = (q-1)|L|$. Applying Theorem 4.2.1, we get that

$$(q-1)|L| \leq (1+o(1))\left((q^3 - |L|)|L|q^{-2} + q\sqrt{(q^3 - |L|)|L|(|L|q^{-3})}\right).$$

Simplifying the above expression one can show (with a little bit of effort) that

$$|L| \leq \left((\sqrt{5}-1)/2 + o(1)\right) q^3 \leq (1+o(1))0.62q^3.$$

Simplifying the first inequality to get the second one is a messy calculation that we omit, but it can easily be seen that for instance setting $|L|/q^3$ to be any constant greater than 0.62 in the first inequality yields a contradiction, for q sufficiently large. □

4.2.2 The union of lines

The proof of Theorem 4.1.1 only uses the fact that the definition of a Nikodym set $\mathcal{N}$ guarantees the existence of $|\mathcal{N}^c|$ distinct lines, each of which are incident to at least $q-1$ points of $\mathcal{N}$. While we do not believe that Theorem 4.1.1 is anywhere near tight, the same proof gives a nearly tight lower bound on the size of the union of any set of at least $0.62q^3$ lines.

Recall from the introduction that, for any set L of lines,

$$P(L) = \{p \in \ell \mid \ell \in L\}.$$

Proposition 4.2.1. *If L is a set of $0.62q^3$ lines in $\mathbb{F}_q^3$, then $|P(L)| \geq (1-o(1))0.38q^3$.*

Proof. Since each point on any line in L is contained in $P = P(L)$, the number of incidences between L and P is $q|L| = 0.62q^4$. Applying Theorem 4.2.1,

$$0.62q^4 \leq (1+o(1))(0.62|P|q + q\sqrt{0.62|P|q^3(1-|P|q^{-3})}), \text{ so, simplifying as before,}$$
$$|P| > (1-o(1))0.38q^3.$$

□

We now show that without any further condition on the set of lines, Proposition 4.2.1 is nearly tight.

Proposition 4.2.2. *There is a set L of $(1-o(1))0.62q^3$ lines in $\mathbb{F}_q^3$ such that $|P(L)| < 0.43q^3$.*

Proof. Let p be an arbitrary point of $\mathbb{F}_q^3$. We show below that we can choose a set Π of $0.62q$ planes incident to p, such that no line is contained in 3 planes of Π. The set L will be the set of all lines contained in the union of the planes of Π. By inclusion-exclusion, the total number of lines chosen is $|L| \geq 0.62q^3 - \binom{0.62q}{2} = (1-o(1))0.62q^3$, and the total number of points on these lines is $(0.62q^3-1)-(q-1)\binom{0.62q}{2}+1 < 0.43q^3$, for q sufficiently large.

To choose the set Π, we first project from the point p; this is a map from the lines incident to p to points in $PG(2,q)$, the projective plane over $\mathbb{F}_q$. In this projection, each plane incident to p corresponds to a line in $PG(2,q)$. A conic in $PG(2,q)$ is a set of $q+1$ points, no three collinear; the projective dual to a conic is a set of $q+1$ lines, no three coincident. By choosing Π to be an arbitrary subset of size $0.62q$ among the planes associated to such a set of lines, we ensure that no three contain a common line. □

4.2.3 Coplanar lines and Conjecture 4.1.2

A consequence of the near tightness of Proposition 4.2.1 is that any substantial improvement to Theorem 4.1.1 must use some additional information about Nikodym sets, beyond the fact that the definition of a Nikodym set $\mathcal{N}$ guarantees the existence of $|\mathcal{N}^c|$ distinct lines, each incident to $q-1$ points of $\mathcal{N}$. One such property is that no plane can contain too many of the lines associated to the complement of a Nikodym set.

Proposition 4.2.3. *Let $\mathcal{N} \subseteq \mathbb{F}_q^3$ be a Nikodym set. Let L be a set of lines, such that each line of L is incident to exactly one point of $\mathcal{N}^c$, and each point of $\mathcal{N}^c$ is incident to exactly one line of L. Then any plane in $\mathbb{F}_q^3$ contains at most $(1+o(1))q^{3/2}$ lines of L.*

Note that the existence of a set satisfying the conditions on L in this proposition is guaranteed by the definition of a Nikodym set.

Proof. Let π be a plane, and let L' be the subset of lines of L that are contained in π. Let $P \subseteq \mathcal{N}^c$ be the set of points associated to lines in L'. From the definition, P is the complement of a planar weak Nikodym set in π. By the result of Feng, Li, and Shen [FLS10]), $|L'| = |P| \leq (1 + o(1))q^{3/2}$. □

The observation recorded in Proposition 4.2.3 enables us to show that Conjecture 4.1.2 implies the three dimensional case of Conjecture 4.1.1. Since Proposition 4.2.3 only gives an upper bound of $(1 + o(1))q^{3/2}$ lines contained in any plane, while Conjecture 4.1.2 requires a bound of any function in $\omega(q)$, we will need to use some additional incidence theory to bridge the gap. In particular, we will use the following lemma, which is a special case of Corollary 6 in [LS14].

Lemma 4.2.3 ([LS14])**.** *For $k > 1$, a set of kq planes in $\mathbb{F}_q^3$ is incident to at least $(1 - \frac{1}{k-1+k^{-1}})q^3$ points. A set of kq lines in $\mathbb{F}_q^2$ is incident to $(1 - \frac{1}{k-1+k^{-1}})q^2$ points.*

We now prove that Conjecture 4.1.2 implies the three dimensional case of Conjecture 4.1.1.

Theorem 4.2.4. *If Conjecture 4.1.2 holds, then the case $n = 3$ of Conjecture 4.1.1 holds.*

Proof. Suppose that Conjecture 4.1.2 holds.

Let $\mathcal{N}^c$ be the complement of a Nikodym set in $\mathbb{F}_q^3$. Let L be a set of lines such that each line of L is incident to exactly one point of $\mathcal{N}^c$, and each point of $\mathcal{N}^c$ is incident to exactly one line of L; the existence of such a set is guaranteed by the definition of a Nikodym set. Let $L_1 \subset L$ be an arbitrary subset of $\lfloor |L|/2 \rfloor$ lines of L, and let $P \subset \mathcal{N}^c$ be the set of points in $\mathcal{N}^c$ that are not incident to any line in L_1.

Let $\alpha(q) \in \omega(q)$, and let Π be the set of planes that contain more than $\alpha(q)$ lines of L_1. Let $L_2 \subseteq L_1$ be the subset of lines in L_1 that are each contained in some plane of Π.

Suppose that $|L_2| = \Omega(q^{5/2}\log(q))$. Since each plane $\pi \in \Pi$ contains at least $\alpha(q)$ lines of L_2, Lemma 4.2.3 implies that the probability that a uniformly chosen point of π is not on any line of L_2 is bounded above by $(1 + o(1))q/\alpha(q)$. By Proposition 4.2.3, no plane of Π contains more than $(1 + o(1))q^{3/2}$ lines of L_2; hence, $|\Pi| \geq (1 - o(1))q^{-3/2}|L_2| = \Omega(q \log q)$. By Lemma 4.2.3, the probability that a uniformly chosen point of $\mathbb{F}_q^3$ is not on any plane of Π is bounded above by $O(1/\log(q))$. By a union bound, all but $O(q^3/\log(q) + q^4/\alpha(q)) = o(q^3)$ points of $\mathbb{F}_q^3$ are contained in some line of L_2. By construction, half of the points of $\mathcal{N}^c$ are not in any line of L_1, and hence $|\mathcal{N}^c| = o(q^3)$.

Now, suppose that $|L_2| = O(q^{5/2}\log q) = o(q^3)$. By construction, no plane contains more than $\alpha(q)$ lines of $L_1 \setminus L_2$. Hence, Conjecture 4.1.2 implies that either $|L_1 \setminus L_2| = o(q^3)$, and hence $|\mathcal{N}^c| = o(q^3)$, or $|P(L_1 \setminus L_2)| = (1 - o(1))q^3$, and hence $|\mathcal{N}^c| = o(q^3)$. □

4.3 Weak Nikodym sets

In this section, we begin the investigation of *weak Nikodym sets*, with a particular focus on possible differences between weak Nikodym sets and Nikodym sets.

We will find it convenient to work in projective geometry; we denote the n dimensional projective geometry over $\mathbb{F}_q$ as $PG(n, q)$.

We define (weak) Nikodym sets in projective geometry the same way as in affine geometry. We say $\mathcal{N}$ is a Nikodym set if, through each point p in $PG(n, q)$ there is a line ℓ such that $\ell \setminus \{p\} \subseteq \mathcal{N}$, and $\mathcal{N}$ is a weak Nikodym set if, through each point $p \in \mathcal{N}^c$, there is a line ℓ such that $\ell \setminus \{p\} \subseteq \mathcal{N}$.

Let

$$\begin{aligned}
f(n, q) &= \text{the maximum size of the complement of a Nikodym set in } \mathbb{F}_q^n,\\
f_w(n, q) &= \text{the maximum size of the complement of a weak Nikodym set in } \mathbb{F}_q^n,\\
f^*(n, q) &= \text{the maximum size of the complement of a Nikodym set in } PG(n, q),\\
f_w^*(n, q) &= \text{the maximum size of the complement of a weak Nikodym set in } PG(n, q).
\end{aligned}$$

There are some easy relations among the above quantities. From the definitions, a

Nikodym set is also a weak Nikodym set. Hence,

$$f_w(n,q) \geq f(n,q), \tag{4.1}$$

$$f_w^*(n,q) \geq f^*(n,q). \tag{4.2}$$

Suppose that $\mathcal{N}^c$ is the complement of a (weak) Nikodym set in $\mathbb{F}_q^n$. Take the projective closure of $\mathbb{F}_q^n$ by adding a hyperplane, and include the new hyperplane in $\mathcal{N}$. This expanded $\mathcal{N}$ is still a (weak) Nikodym set, and hence

$$f_w^*(n,q) \geq f_w(n,q), \tag{4.3}$$

$$f^*(n,q) \geq f(n,q). \tag{4.4}$$

Suppose that $\mathcal{N}^c$ is the complement of a (weak) Nikodym set in $PG(n,q)$. The expected number of points of $\mathcal{N}^c$ contained in a hyperplane chosen uniformly at random is $E = (1+o(1))\mathcal{N}^c/q$; hence, there exists a hyperplane that contains at most E points of $\mathcal{N}^c$. We can obtain a (weak) Nikodym set in $\mathbb{F}_q^n$ by removing this hyperplane, and hence

$$f_w(n,q) \geq (1+o(1))(1-1/q)f_w^*(n,q), \tag{4.5}$$

$$f(n,q) \geq (1+o(1))(1-1/q)f^*(n,q). \tag{4.6}$$

We can do somewhat better when $n = 2$.

Suppose that $\mathcal{N}^c$ is the complement of a Nikodym set in $PG(2,q)$. By the definition of a Nikodym set, if we take a point $p \in \mathcal{N}$, there exists a line ℓ through p such that $\ell \in \mathcal{N}$. We can remove ℓ to obtain an affine plane, and $\mathcal{N}^c$ will be the complement of a Nikodym set in this affine plane. Hence, $f(2,q) \geq f^*(2,q)$, and so

$$f(2,q) = f^*(2,q). \tag{4.7}$$

Suppose that $\mathcal{N}^c$ is the complement of a weak Nikodym set in $PG(n,q)$. If we take a point $p \in \mathcal{N}^c$, there exists a line ℓ through p such that $\ell \setminus \{p\} \in \mathcal{N}$. We can remove ℓ to obtain an affine plane, and $\mathcal{N}^c \setminus \{p\}$ will be the complement of a weak Nikodym set in this affine plane. Hence, $f_w(2,q) + 1 \geq f_w^*(2,q)$, and so

$$f_w^*(2,q) - 1 \leq f_w(2,q) \leq f_w^*(2,q). \tag{4.8}$$

4.3.1 Constructions

In this section, we show how to construct two infinite families of point sets that form the complement of (weak) Nikodym sets in $PG(n, q)$; we also (in Section 4.3.1) give the proof of Proposition 4.1.1, which provides an extreme example for Conjecture 4.1.3.

It is easy to see that a hyperplane in $\mathbb{F}_q^n$ is the complement of a weak Nikodym set consisting of q^{n-1} points, and, to our knowledge, no better construction than this was known. Our first construction is a refinement of this idea, and gives the complement of a Nikodym set consisting of $(1 - o(1))nq^{n-1}$ points, or the complement of a weak Nikodym set consisting of $(1 - o(1))(n + 1)q^{n-1}$ points. Our second construction gives the complement of a weak Nikodym set consisting of $(1 - o(1))q^{n-1/2}$ points, but only works in fields of square order, and cannot be used to construct the complement of a standard Nikodym set.

Union of hyperplanes with a few points removed

In this section, we construct the complement of Nikodym sets consisting of $(1-o(1))nq^{n-1}$ points, and the complement of weak Nikodym sets consisting of $(1 - o(1))(n + 1)q^{n-1}$ points. These constructions work for any sufficiently large finite field.

Let q be a prime power; we will assume that q is sufficiently large relative to n.

Let S be the union of $n + 1$ hyperplanes $\Lambda_1, \ldots, \Lambda_{n+1}$ in $PG(n, q)$ that do not all pass through a single point. For each $I \subset [n + 1]$ with $1 \leq |I| \leq n$, remove a point p_I from S such that $p_I \in \Lambda_i$ for $i \in I$, and $p_I \notin \Lambda_j$ for $j \notin I$. By simple dimension counting arguments one can show that such a point always exists. Here is a sketch of the argument. For any k between 1 and n, the intersection of any k hyperplanes must be exactly an $n-k$ dimensional space, since if it was larger then there would be a point in common with all the hyperplanes. If we want a point on those k hyperplanes but not on any other plane, then it is easy to see that for q large enough, a random point on the $n - k$ dimensional intersection would not lie on any of the other hyperplanes.

We claim that the resulting set S (after deleting the points as mentioned above) is the complement of a weak Nikodym set.

Let q be an arbitrary point of S. Let $J \subseteq [n+1]$, such that $q \notin \Lambda_j$ for each $j \in J$, and $q \in \Lambda_i$ for $i \notin J$. Note that $1 \leq |J| \leq n$. Let ℓ be the line through q and p_J. Note that ℓ intersects each Λ_i at either q or p_J, and does not intersect any Λ_i at both points. Hence, q and p_J are the only points at which ℓ intersects any Λ_i. Since $p_J \notin S$, q is the unique point in the intersection of S and ℓ. Hence, S is the complement of a weak Nikodym set.

Consequently,

$$f^*_w(n,q) \geq (1-o(1))(n+1)q^{n-1}.$$

We can modify S to be the complement of a standard Nikodym set by removing Λ_{n+1} from the construction. Then, for any point $q \notin S$, the line through q and $p_{[n]}$ is disjoint from S. Hence,

$$f^*(n,q) \geq (1-o(1))nq^{n-1}.$$

Hermitian varieties

In this section, we give an improved construction of weak Nikodym sets in $\mathbb{F}_q^3$ for square q, and we prove Proposition 4.1.1, which describes the construction of an extreme example related to Conjecture 4.1.3. Both of these constructions are based on Hermitian varieties.

Let $q = p^2$, for p a prime power. For $v \in \mathbb{F}_q$, we define the conjugate $\overline{v} = v^p$. Since q has order p^2, we have $\overline{\overline{v}} = v$. We will use homogenous coordinates to represent a point $v \in PG(n,q)$ as a column vector $\mathbf{v} = (v_0, v_1, \ldots, v_n)^T$.

A square matrix $H = ((h_{ij}))$ for $i,j = 0,1,\ldots,n$ and $h_{ij} \in \mathbb{F}_q$ is *Hermitian* if $h_{ij} = \overline{h_{ji}}$ for all i,j. Let $\mathbf{x}^T = (x_0, x_1, \ldots, x_n)$ and $\overline{\mathbf{x}} = (\overline{x_0}, \overline{x_1}, \ldots, \overline{x_n})^T$. The set of points x in $PG(n,q)$ whose coordinates satisfy $\mathbf{x}^T H \overline{\mathbf{x}} = 0$ for a Hermitian matrix H is a *Hermitian variety*. The *rank* of the Hermitian variety V defined by $\mathbf{x}^T H \overline{\mathbf{x}} = 0$ is defined to be the rank of H. We say that V is *non-degenerate* if its rank is $n+1$.

Let V be a rank r Hermitian variety in $PG(n,q)$ defined by $\mathbf{x}^T H \overline{\mathbf{x}} = 0$. A point c of V is *singular* if $\mathbf{c}^T H = \mathbf{0}$. Clearly, if V is non-degenerate, it has no singular points. Otherwise, $\mathbf{c}^T H = 0$ has $n - r + 1$ independent solutions, and hence defines an

$(n-r)$-flat, which we term the *singular space* of V.

The set of points corresponding to row vectors $\mathbf{x}^T$ that satisfy the equation $\mathbf{x}^T H\overline{\mathbf{c}} = 0$ is the *tangent space* at $\mathbf{c}$. If $\mathbf{c}$ is singular, this is the entire space; otherwise, $H\overline{\mathbf{c}}$ is a non-zero vector, and hence the tangent space is a hyperplane.

We will use the following properties of Hermitian varieties, determined by Bose and Chakravarti [BC66].

Lemma 4.3.1 (Section 7 in [BC66])**.** *The intersection of a Hermitian variety with a flat space is a Hermitian variety. In particular, a line intersects a Hermitian variety in a single point, $q^{1/2}+1$ points, or is entirely contained in the variety.*

Given a Hermitian variety V, we define *tangent lines* to be those lines that intersect V in exactly 1 point.

Theorem 4.3.2 (Theorem 7.2 in [BC66])**.** *If V is a degenerate Hermitian variety of rank $r < n+1$, and c is a point belonging to the singular space of V, and d is an arbitrary point of V, then each point on the line cd belongs to V.*

Theorem 4.3.3 (Theorem 7.4 in [BC66])**.** *If V is a non-degenerate Hermitian variety, the tangent hyperplane at a point c of V intersects V in a degenerate Hermitian variety U of rank $n-1$. The singular space of U consists of the single point c.*

Theorem 4.3.4 (Theorem 8.1 in [BC66])**.** *The number of points on a non-degenerate Hermitian variety is*

$$\phi(n,q) = (q^{(n+1)/2} - (-1)^{n+1})(q^{n/2} - (-1)^n)(q-1)^{-1}.$$

The number of points on a degenerate Hermitian variety of rank r is

$$(q^{n-r+1}-1)\phi(r-1,q) + (q^{n-r+1}-1)(q-1)^{-1} + \phi(r-1,q).$$

Using the above definitions and properties, we can use Hermitian varieties to construct small weak Nikodym sets, as well as an extreme example for Conjecture 4.1.3.

Proposition 4.3.1. *Let $q = p^2$ for a prime power p, and let $n \geq 2$.*

$$f_w^* \geq \phi(n,q),$$

where $\phi(n,q) = \Omega(q^{n-1/2})$ is the function defined in Theorem 4.3.4.

Proof. Let V be a non-degenerate Hermitian variety in $PG(n, q)$, and let c be a point of V. By Theorem 4.3.3, the tangent hyperplane Σ at c intersects V in a Hermitian variety of rank $n-1$ in $PG(n-1, q)$. By the second part of Theorem 4.3.4, there is a point $d \in \Sigma$ that is not contained in V. By Theorem 4.3.2, the intersection of the line cd with V is only the point c itself. Since this holds for an arbitrary point $c \in V$, it holds for each point in V, and hence V is the complement of a weak Nikodym set. The proposition follows from the first part of Theorem 4.3.4. □

In Proposition 4.3.1, we use the fact that there is at least one line tangent to V at each point, together with the fact that a tangent line contains exactly one point of V. In fact, we know that there are many tangent lines at each point of V, and we use this fact to prove Proposition 4.1.1. Indeed, we prove a slightly stronger result.

Proposition 4.3.2. *Let $q = p^2$ for a prime power p, and let $0 < \alpha < 1$. Then, there is a set L of $(\alpha + o(1))q^{7/2}$ lines in $\mathbb{F}_q^3$ such that no plane contains more than $(\alpha + o(1))q^{3/2}$ lines of L, and $|P(L)| = q^3 - (1 - \alpha + o(1))q^{5/2}$.*

Proposition 4.1.1 follows immediately from Proposition 4.3.2 by taking $\alpha = 1/2$.

Proof. Let V be a non-degenerate Hermitian variety in $PG(3, q)$. By Theorem 4.3.4, we have $|V| = (1 + o(1))q^{5/2}$. Let P be a set of $\lfloor \alpha |V| \rfloor$ of the points of V, chosen uniformly at random. Let L be the set of tangent lines to V at points of P. Since the tangent lines intersect V only at their points of tangency, it is clear that the $\lceil (1 - \alpha)|V| \rceil = (1 - \alpha + o(1))q^{5/2}$ points of $V \setminus P$ are not incident to any line of L. It remains to show $|L| = (\alpha + o(1))q^{7/2}$, and that no plane contains more than $(\alpha + o(1))(q^{3/2})$ lines of L.

By Theorem 4.3.3, the tangent plane Σ to V at an arbitrary point $c \in P$ intersects V in a rank 2 Hermitian variety $U \subseteq \Sigma$, having the single singular point c. From the second part of Theorem 4.3.4, we have that U contains $q^{3/2} + q + 1$ points. Together with Theorem 4.3.2, this implies that U is the union of $q^{1/2} + 1$ lines coincident at c. The remaining $q - q^{1/2}$ lines contained in Σ and incident to c are tangent lines to V. Hence, L consists of $(q - q^{1/2})|P| = (\alpha + o(1))q^{7/2}$ distinct lines, and tangent planes to V each contain at most $q - q^{1/2}$ lines of L.

By Lemma 4.3.1, the intersection of a plane Σ with V is a Hermitian variety U; if Σ is not tangent to V, then U is non-degenerate. By Theorem 4.3.4, we have that $|U| = q^{3/2} + q + 1$, and there is a single tangent line at each of these points. In addition, a line of L will be contained in Σ only if it is tangent to one of the points of U. Hence, in order to show that no plane contains more than $(\alpha + o(1))q^{3/2}$ lines of L, it suffices to show that no plane contains more than $(\alpha + o(1))q^{3/2}$ points of P.

The expected number of points of P on Σ is $\alpha|U|$. Since the points of P are chosen uniformly at random, the Chernoff bound for Bernoulli random variables implies that, for any $0 < \delta < 1$, the probability that we have more than $(1 + \delta)\alpha|U|$ points of P on Σ is bounded above by $e^{-\delta^2\alpha|U|/3}$. Taking a union bound over the $(1 + o(1))q^3$ planes in $PG(3, q)$, we have that the probability that any plane has more than $(1 + \delta)\alpha|U|$ points of P is bounded above by $(1 + o(1))q^3e^{-(1+o(1))\delta^2\alpha q^{3/2}/3}$. Hence, taking $\delta > (1 + o(1))9\alpha^{-1}q^{-3/4}\log q = o(1)$ ensures that this happens with probability strictly less than 1, and hence there is a choice of P such that there are fewer than $(\alpha + o(1))q^{3/2}$ on any plane.

□

References

[AC88] Noga Alon and Fan RK Chung. Explicit construction of linear sized tolerant networks. *Annals of Discrete Mathematics*, 38:15–19, 1988.

[Alo09] N. Alon. Perturbed identity matrices have high rank: Proof and applications. *Combinatorics, Probability and Computing*, 18(1-2):3–15, 2009.

[BC66] RC Bose and IM Chakravarti. Hermitian varieties in a finite projective space $PG(N, q^2)$. *Canad. J. Math*, 18:1161–1182, 1966.

[BCH74] R. K. Brayton, D. Coppersmith, and A. J. Hoffman. Self-orthogonal latin squares of all orders $n \neq 2, 3, 6$. *Bulletin of the American Mathematical Society*, 80, 1974.

[BDWY13] Boaz Barak, Zeev Dvir, Avi Wigderson, and Amir Yehudayoff. Fractional Sylvester-Gallai theorems. *Proceedings of the National Academy of Sciences*, 110(48):19213–19219, 2013.

[BIP14] Mike Bennett, Alex Iosevich, and Jonathan Pakianathan. Three-point configurations determined by subsets of $\mathbb{F}_q^2$ via the Elekes-Sharir paradigm. *Combinatorica*, 34(6):689–706, 2014.

[BM90] P. Borwein and W. Moser. A survey of Sylvester's problem and its generalizations. *Aequationes Mathematicae*, 40(1):111–135, 1990.

[CS93] J. Csima and E. T. Sawyer. There exist $6n/13$ ordinary points. *Discrete & Computational Geometry*, 9(2):187–202, 1993.

[DKSS09] Zev Dvir, Swastik Kopparty, Shubhangi Saraf, and Madhu Sudan. Extensions to the method of multiplicities, with applications to kakeya sets and mergers. pages 181–190. IEEE Computer Society, 2009.

[DSW14] Zeev Dvir, Shubhangi Saraf, and Avi Wigderson. Improved rank bounds for design matrices and a new proof of Kelly's theorem. In *Forum of Mathematics, Sigma*, volume 2, page e4. Cambridge Univ Press, 2014.

[Dvi09] Zeev Dvir. On the size of kakeya sets in finite fields. *Journal of the American Mathematical Society*, 22:1093–1097, 2009.

[EH13] Jordan S Ellenberg and Marton Hablicsek. An incidence conjecture of Bourgain over fields of positive characteristic. *arXiv preprint arXiv:1311.1479*, 2013.

[EPS06] N. Elkies, L. M. Pretorius, and K. Swanepoel. Sylvester–Gallai theorems for complex numbers and quaternions. *Discrete & Computational Geometry*, 35(3):361–373, 2006.

[FLS10] Chunrong Feng, Liangpan Li, and Jian Shen. Some inequalities in functional analysis, combinatorics, and probability theory. *the electronic journal of combinatorics*, 17(1):R58, 2010.

[Gal44] Tibor Gallai. Solution of problem 4065. *American Mathematical Monthly*, 51:169–171, 1944.

[GKS13] Alan Guo, Swastik Kopparty, and Madhu Sudan. New affine-invariant codes from lifting. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 529–540. ACM, 2013.

[GT13] B. Green and T. Tao. On sets defining few ordinary lines. *Discrete & Computational Geometry*, 50(2):409–468, 2013.

[Hae95] Willem H Haemers. Interlacing eigenvalues and graphs. *Linear Algebra and its applications*, 226:593–616, 1995.

[Hil73] A. J. W. Hilton. On double diagonal and cross latin squares. *Journal of the London Mathematical Society*, 2(4):679–689, 1973.

[Hir83] F. Hirzebruch. Arrangements of lines and algebraic surfaces. In *Arithmetic and geometry*, pages 113–140. Springer, 1983.

[Kel86] L. Kelly. A resolution of the Sylvester-Gallai problem of J.-P. Serre. *Discrete & Computational Geometry*, 1(1):101–104, 1986.

[KM58] L. Kelly and W. Moser. On the number of ordinary lines determined by n points. *Canadian Journal of Mathematics*, 10:210–219, 1958.

[LS14] Ben Lund and Shubhangi Saraf. Incidence bounds for block designs. *arXiv preprint arXiv:1407.7513*, 2014.

[LSW16] Ben Lund, Shubhangi Saraf, and Charles Wolf. Finite field kakeya and nikodym sets in three dimensions. *arXiv preprint arXiv:1609.01048*, 2016.

[Mel40] E. Melchior. Über vielseite der projektiven ebene. *Deutsche Math*, 5:461–475, 1940.

[Mot51] Th. Motzkin. The lines and planes connecting the points of a finite set. *Transactions of the American Mathematical Society*, pages 451–464, 1951.

[RS89] U. Rothblum and H. Schneider. Scalings of matrices which have prespecified row sums and column sums via optimization. *Linear Algebra and its Applications*, 114:737–764, 1989.

[Ser66] J.-P. Serre. Advanced problem 5359. *American Mathematical Monthly*, 73(1):89, 1966.

[SS08] Shubhangi Saraf and Madhu Sudan. Improved lower bound on the size of kakeya sets over finite fields. *Analysis and PDE*, 1(3):375–379, 2008.

[Syl93] J. J. Sylvester. Mathematical question 11851. *Educational Times*, 59(98):256, 1893.

[Wol99] Thomas Wolff. Recent work connected with the kakeya problem. *Prospects in mathematics (Princeton, NJ, 1996)*, 2:129–162, 1999.