On the number of ordinary lines determined by sets in complex space

Abdul Basit[*] Zeev Dvir[†] Shubhangi Saraf[‡] Charles Wolf[§]

November 26, 2016

Abstract

Kelly's theorem states that a set of n points affinely spanning $\mathbb{C}^3$ must determine at least one ordinary complex line (a line passing through exactly two of the points). Our main theorem shows that such sets determine 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 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 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 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 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*.

[*]Department of Computer Science, Rutgers University. Email: abasit@cs.rutgers.edu.

[†]Princeton University, Department of Mathematics and Department of Computer Science. Research supported by NSF CAREER award DMS-1451191 and NSF grant CCF-1523816. Email: zeev.dvir@gmail.com.

[‡]Department of Computer Science and Department of Mathematics, Rutgers University. Research supported in part by NSF grant CCF-1350572. Email: shubhangi.saraf@gmail.com.

[§]Department of Mathematics, Rutgers University. Research supported in part by NSF grant CCF-1350572. Email: ciw13@math.rutgers.edu.

Conjecture 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 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, 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 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 $\mathcal{V}$.*

Kelly's proof of Theorem 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 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 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 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.

[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.

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 (the constant 2/3 is arbitrary and can be replaced by any number smaller than one):

Theorem 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*

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

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 get that there are at least $k(n-2k)$ ordinary lines determined by $\mathcal{V}$. Then if $k = \epsilon n$, for $0 < \epsilon < 1/2$, we get that $\mathcal{V}$ has $\Omega_\epsilon(n^2)$ ordinary lines, where the hidden constant depends on ϵ. Therefore, the bound in Theorem 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 1.6 and 1.7 give lower bounds on the number of ordinary lines, i.e. $t_2(\mathcal{V})$. Another important contribution of Theorem 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 1.7. As a consequence, we get that if a non-planar configuration over $\mathbb{C}$ has many 4-rich lines, 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 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 1.6 and 1.7 over $\mathbb{C}^3$ and Theorem 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 $\mathbb{R}$.

Organization: In Section 2 we give a short overview of the new ideas in our proof (which builds upon [DSW14]). In Section 3 we develop the necessary machinery on matrix scaling and Latin squares. In Section 4, we prove some key lemmas that will be used in the proofs of our main results. Section 5 gives the proof of Theorems 1.6 and 1.8, which are considerably simpler than Theorem 1.7. In Section 6, we develop additional machinery needed for the proof of Theorem 1.7. The proof of Theorem 1.7 is presented in Section 7.

2 Proof overview

The starting point for the proofs of Theorems 1.6, 1.7 and 1.8 is the method developed in [BDWY13, DSW14] which uses rank bounds for *design matrices* – matrices in which the supports of different columns do not intersect in too many positions. We augment the techniques in these papers in several ways which give us more flexibility in analyzing the number of ordinary lines. We devote this short section to an overview of the general framework (starting with [DSW14]) outlining the places where new ideas come into play.

Let $\mathcal{V} = \{v_1, \ldots, v_n\}$ be points in $\mathbb{C}^d$ and denote by V the $n \times (d+1)$ matrix whose i^{th} row is the vector $(v_i, 1) \in \mathbb{C}^{d+1}$, i.e. the vector obtained by appending a 1 to the vector v_i. The dimension of the (affine) space spanned by the point set can be seen to be equal to $\mathrm{rank}(V) - 1$. We would now like to argue that too many collinearities in $\mathcal{V}$ (or too few ordinary lines) imply that all (or almost all) points of $\mathcal{V}$ must be contained in a low dimensional affine subspace, i.e. $\mathrm{rank}(V)$ is small. To do this, we construct a matrix A, encoding the dependencies in $\mathcal{V}$, such that $AV = 0$. Then we must have

$$\mathrm{rank}(V) \leq n - \mathrm{rank}(A),$$

and so it suffices to lower bound the rank of A.

We construct the matrix A in the following manner so that each row of A corresponds to a collinear triple in $\mathcal{V}$. For any collinear triple $\{v_i, v_j, v_k\}$, there exist coefficients a_i, a_j, a_k such that $a_i v_i + a_j v_j + a_k v_k = 0$. We can thus form a row of A by taking these coefficients as the nonzero entries in the appropriate columns. By carefully selecting the triples using constructions of Latin squares (see Lemma 3.12), we can ensure that A is a *design matrix*. Roughly speaking, this means that the supports of every two columns in A intersect in a small number of positions. Equivalently, every pair of points appears together only in a small number of triples.

The proof in [DSW14] now proceeds to prove a general rank lower bound on any such design-matrix. To understand the new ideas in our proof, we need to 'open the box' and see how the rank bound from [DSW14] is actually proved. To get some intuition, suppose that A is a matrix with 0/1 entries. To bound the rank of A, we can consider the matrix $M = A^*A$ and note that $\mathrm{rank}(M) = \mathrm{rank}(A)$. Since A is a design matrix, M has the property that the diagonal entries are very large (since we can show that each point is in many collinear triples) and that the off-diagonal elements are very small (since columns have small intersections). Matrices with this property are called *diagonal dominant* matrices, and it is easy to lower bound their rank using trace inequalities (see Lemma 3.5).

However, the matrix A that we construct could have entries of arbitrary magnitude and so bounding the rank requires more work. To do this, [DSW14] relies on *matrix scaling* techniques. We are allowed to multiply each row and each column of A by a nonzero scalar and would like to reduce to the case where the entries of A are 'mostly balanced' (see Theorem 3.3 and Corollary 3.4). Once scaled, we can consider $M = A^*A$ as before and use the bound for diagonal dominant matrices.

Our proof introduces two new main ideas into this picture. The first idea has to do with the conditions needed to scale A. It is known (see Corollary 3.4) that a matrix A has a good scaling if it does not contain a 'too large' zero submatrix. This is referred to as having Property-S (see Definition 3.2). The proof of [DSW14] uses A to construct a new matrix B, whose rows are the same as those of A but with some rows repeating more than once. Then one shows that B has Property-S and continues to scale B (which has rank at most that of A) instead of A. This loses the control on the exact number of rows in A which is crucial for bounding the number of ordinary lines. We instead perform a more careful case analysis: If A has Property-S then we scale A directly and gain more information about the number of ordinary lines. If A does not have Property-S, then we carefully examine the large zero submatrix that violates Property-S. Such a zero submatrix corresponds to a set of points and a set of lines such that no line passes through any of the points. We argue in Lemma 4.4 that such a submatrix implies the existence of many ordinary lines. In fact, the conclusion is slightly more delicate: We either get many ordinary lines (in which case we are done) or we get a point with many ordinary lines through it (but not enough to complete the proof). In the second case, we need to perform an iterative argument which removes the point we found and applies the same argument again to the remaining points.

The second new ingredient in our proof comes into play only in the proof of Theorem 1.7. Here, our goal is to improve on the rank bound of [DSW14] using the existence of lines with four or more points. Recall that our goal is to give a good upper bound on the off-diagonal entries of $M = A^*A$. Consider the (i,j)'th entry of M, obtained by taking the inner product of columns i and j in A. The i'th column of A contains the coefficients of v_i in a set of collinear triples containing v_i (we might not use all collinear triples). In [DSW14] this inner product is bounded using the Cauchy-Schwartz inequality, and uses the fact that we picked our triple family carefully so that v_i and v_j appear together in a small number of collinear triples. This does not use any information about possible cancellations that may occur in the inner product (considering different signs over the reals or angles of complex numbers). One of the key insights of our proof is to notice that having more than 3 points on a line, gives rise to such cancellations (which increase the more points we have on a single line).

To get a rough idea, let us focus on a real set of points. Consider two points v_1, v_2 on a line that has two more points v_3, v_4 on it. Suppose that v_3 is 'between' v_1 and v_2 and that v_4 is outside the interval v_1, v_2. Then, in the collinearity equation for the triple v_1, v_2, v_3 the signs of the coefficients of v_1, v_2 will both be positive. On the other hand, in the collinearity equation for v_1, v_2, v_4 the signs of the coefficients of v_1, v_2 will be different (one will be positive and the other negative). Thus, if both of these triples appear as rows of A, we will have non trivial cancellations! Of course, we need to also worry about the magnitudes of the coefficients but, luckily, this is possible since, if the coefficients are of magnitudes that differ from each other too much, we can 'win' in another Cauchy-Schwartz (which again translates into a better rank bound, see Lemma 6.6). To formalize the previous example, let v_1, v_2, v_3, v_4 be collinear points in $\mathbb{R}^d$. Then there exist coefficients such that

$$\begin{aligned} & r \cdot v_1 + (1-r) \cdot v_2 - v_3 = 0, \\ & s \cdot v_1 + (1-s) \cdot v_2 - v_4 = 0, \\ \text{and} \quad & t \cdot v_1 + (1-t) \cdot v_3 - v_4 = 0. \end{aligned}$$

Now at least one of $r(1-r)$, $s(1-s)$ and $t(1-t)$ must be negative, and at least one must be positive. Without loss of generality, say $r(1-r)$ is positive and $s(1-s)$ is negative. In order for the Cauchy-Schwarz inequality to be tight, we need that $r(1-r) = s(1-s)$, which cannot happen because one is positive and the other is negative. This phenomena is captured in Lemma 6.3 which generalizes this idea to the complex numbers. The lemma only analyzes the case of four points since we can bootstrap the lemma for lines with more points by applying it to a random four tuple (see Item 4 of Lemma 6.4).

3 Preliminaries

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

Definition 3.1 (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 3.2 (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 3.3 (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 3.3 appeared in [BDWY13].

Corollary 3.4 (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 3.4 is obtained by applying Theorem 3.3 to the matrix obtained by squaring the absolute values of the entries of the matrix A. Once again, we may assume that $\sum_{j \in [n]} \left|A'_{ij}\right|^2 = 1 + \epsilon$.

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 3.4, 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 3.5. *Let A be an $n \times n$ complex hermitian matrix, such that $|A_{ii}| \geq L$ for all $i \in n$. Then*

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

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 3.6. *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 3.7. *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 gives us 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 gives us 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 3.8. *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*

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

3.2 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 3.9 (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 3.10 ([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 3.11 ([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 $L_{ii} = L^T_{ii}$, 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 3.12. *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 3.10 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 3.11 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 $r \geq 4$. □

4 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 here that the construction we give here is preliminary, but suffices to prove Theorems 1.6 and 1.8. A refined construction is given in Section 6, where we select the triples more carefully. The rows of the matrix will consist of linear dependency coefficients, which we define below.

Definition 4.1 (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 4.2 (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 3.12. 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 3.12, 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 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 4.3. *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*

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

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

Proof. Fix $\epsilon > 0$. Since A satisfies Property-S, by Lemma 3.4 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 3.8 gives us that $\sum_{i \neq j} |M_{ij}|^2 \leq 4m(1+\epsilon)^2$. By applying Lemma 3.5 to M we get,

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

Taking ϵ to 0 we get

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

Recall that affine-dim$(\mathcal{V}) = d = \text{rank}(V) - 1$. Since $AV = 0$, we have that $\text{rank}(V) \leq n - \text{rank}(A)$. 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 4.4. *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 $\{v_i, w\}$, with $v_i \in \mathcal{V}_1$ and $w \in \mathcal{V} \setminus \mathcal{V}_1$, that lie on special lines. 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 $\{v_j, w\}$, where

$v_i, v_j \in \mathcal{V}_1$ and $w \in \mathcal{V} \setminus \mathcal{V}_1$ that lie on special lines. Each row of X_3 has all zero entries in the submatrix supported on $X \times Y$, so does not contribute any pairs. Recall, from Lemma 6.4, 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 gives 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 have 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 $A_{ij}, A_{ij'}$ with $j \neq j'$, that appear in the submatrix $U \times Y$. There are $\binom{n-b}{2}$ pairs of points in $\mathcal{V} \setminus \mathcal{V}_1$, 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, i.e. the number of pairs of entries equals $3a$. It follows that

$$3a \leq 6\binom{n-b}{2}. \tag{1}$$

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

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

Combining (1) and (2), we get

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

Solving for $t_2(\mathcal{V})$ gives us

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

□

5 Proofs of Theorems 1.6 and 1.8

The proofs of both Theorems 1.6 and 1.8 rely on Lemmas 4.3 and 4.4. Together, these lemmas imply that there must be a point with many ordinary lines containing it, or there are many ordinary lines in total. As mentioned in the proof overview, the theorems are then obtained by using an iterative argument removing a point with many ordinary lines through it, and then applying the same argument to the remaining points.

5.1 Proof of Theorem 1.6

We get the following easy corollary from Lemma 4.3 and Lemma 4.4.

Corollary 5.1. *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 4.3. Otherwise, let $b^* = 3$, and note that Lemma 4.4 gives us the statement of the corollary when $n \geq 5$. □

We are now ready to prove Theorem 1.6. For convenience, we state the theorem again.

Theorem 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 5.1, 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 5.1, 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 gives 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 $n \geq 9$, we get that $t_2(\mathcal{V}) \geq \frac{3}{2}n$.

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 5.1 gives us two cases. If there are $\frac{3}{2}(n-2)$ ordinary lines in $\mathcal{V}_2$, then we get that 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 get that

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

When $n \geq 24$, we get that $t_2(\mathcal{V}) \geq \frac{3}{2}n$. □

5.2 Proof of Theorem 1.8

We get the following easy corollary from Lemma 4.3 and Lemma 4.4.

Corollary 5.2. *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.*
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 4.3. Otherwise, let $b^* = n/6$. Now by Lemma 4.4, 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 1.8. For convenience, we state the theorem again.

Theorem 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 5.2 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 5.2 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 5.2, 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 point will the algorithm stop because the configuration becomes 3 dimensional. That is, we can use Corollary 5.2 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 5.2 for some $1 \leq j < n/3$. From case (2) of Corollary 5.2, we have that

$$t_2(\mathcal{V}_j) \geq \frac{(n-j)^2}{12}. \tag{3}$$

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 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, $1 \le i \le j$, 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

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

Combining (3) and (4), we get that

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

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

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

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 determined by $\mathcal{V}_{i-1}$ through it. 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 $j = n/3$, we get that the number of ordinary lines determined by $\mathcal{V}$ is at least

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

□

6 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 4.2 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 6.7) we will give a construction of a dependency matrix that will have an additional property (captured in Item 4 of Lemma 6.4) 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 6.1 (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 6.2 (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 to quantify the cancellations in lines with more than four points by applying it to random four tuples inside the line.

Lemma 6.3. *Let v_1, v_2, v_3, v_4 be 4 collinear points in $\mathbb{C}^d$. Then at least one of the following 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

$$a_1 v'_1 + a_2 v'_2 + a_3 v'_3 = 0 \tag{5}$$

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

$$b_1 v'_1 + b_2 v'_2 + b_4 v'_4 = 0. \tag{6}$$

We may assume, without loss of generality, that $a_3 = b_4 = 1$. Now equations (5) and (6) give us 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 (5) and (6), we get the following linear equation:

$$(b_2a_1 - b_1a_2)v'_1 + b_2v'_3 - a_2v'_4 = 0. \tag{7}$$

From (7), we get $C_{(1,3)}(4) = \frac{(b_2a_1-b_1a_2)\overline{b_2}}{|b_2a_1-b_1a_2||b_2|}$ and $C_{(1,4)}(3) = -\frac{(b_2a_1-b_1a_2)\overline{a_2}}{|b_2a_1-b_1a_2||a_2|}$.

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

$$\begin{aligned} &\left|\arg\left(\frac{a_1\overline{a_2}}{|a_1||a_2|}\frac{\overline{b_1}b_2}{|b_1||b_2|}\right)\right| \\ =\ &\left|\arg\left(a_1\overline{a_2}\overline{b_1}b_2\right)\right|. \end{aligned} \tag{8}$$

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

$$\begin{aligned} &\left|\arg\left(\frac{(b_2a_1-b_1a_2)\overline{b_2}}{|b_2a_1-b_1a_2||b_2|}\frac{\overline{a_1}}{|a_1|}\right)\right| \\ =\ &\left|\arg\left(\overline{a_1}\overline{b_2}(b_2a_1-b_1a_2)\right)\right|. \end{aligned} \tag{9}$$

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

$$\begin{aligned} &\left|\arg\left(-\frac{b_1}{|b_1|}\frac{\overline{(b_2a_1-b_1a_2)}a_2}{|b_2a_1-b_1a_2||a_2|}\right)\right| \\ =\ &\left|\arg\left(-b_1a_2\overline{(b_2a_1-b_1a_2)}\right)\right|. \end{aligned} \tag{10}$$

Note that the product of expressions inside the arg functions in (8), (9) and (10) is a negative real number, and so the sum of (8), (9) and (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 6.4. *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 3.12 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 3.12.

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 6.3, 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 6.5 (η-balanced row)**.** *Given an $m \times n$ matrix A, we say a row R_k is η-balanced for some constant η if $\left||A_{ki}|^2 - |A_{kj}|^2\right| \leq \eta$, 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 6.6. *There exists an absolute constant $c_0 > 0$ such that the following holds. Let l be a line in $\mathbb{C}^d$ and $\mathcal{V}_l = \{v_1, \dots v_r\}$ be points on l with $r \geq 4$. Let $A = A(l)$ be the dependency matrix for l, defined in Lemma 6.4, 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 3.7 to A' we get that

$$\sum_{i\neq j}|M_{ij}|^2 = 4(r^2-r)\alpha^4 - (D(A)+2E(A))\,. \tag{11}$$

We are able to give a lower bound on $D(A)+2E(A)$ using Property 4 of Lemma 6.4. 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 $i\in \mathrm{supp}(R_k)$. 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. $\cos\theta \le 1/2$. 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

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

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

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

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 (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 1.7.

Definition 6.7 (Dependency Matrix, second construction)**.** *Let $\mathcal{V} = \{v_1, \ldots v_n\}$ 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 4.2 and so satisfies all the properties mentioned there. In particular, we have $AV = 0$ and the number of rows in A equals $n^2 - n - 2t_2(\mathcal{V})$.

7 Proof of Theorem 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 4.4. The following lemma deals with the former case.

Lemma 7.1. *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 3.4 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 $j \in [n]$

$$\sum_{i \in [m]} \left|A'_{ij}\right|^2 \geq \frac{m}{n} - \epsilon. \quad (12)$$

Let C_i be denote the i^{th} column of A', and let $M = A'^*A'$. From (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 6.6, 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 3.5 gives us that

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

Taking ϵ to 0, we get

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

Note that

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

It follows that if

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

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

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

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 7.2. *There exists a positive integer n_0 such that the following holds. Let c_1 be the constant from Lemma 7.1 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 7.1. Otherwise, let b^* be an integer such that

$$\frac{n}{2}(b^*-1) < \frac{3n}{2} + c_1 \sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}) \leq \frac{n}{2}b^*. \tag{13}$$

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,

$$b^* < 4 + \frac{2c_1}{n}\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}) < \frac{1}{6}n. \tag{14}$$

Now by Lemma 4.4 and (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 (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 1.7.

Lemma 7.3. *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 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 1.8, i.e. we use Corollary 7.2 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 7.2 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 7.2, 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 7.2 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 7.2 for some $1 \leq j < n/3$. From case (2) of Lemma 7.2 and Lemma 7.3, we have that

$$\begin{aligned} t_2(\mathcal{V}_j) &\geq \frac{3(n-j)}{2} + c_1\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}_j) \\ &\geq \frac{3(n-j)}{2} + c_1\left(\sum_{r\geq 4}(r^2-r)t_r(\mathcal{V}) - 4\sum_{i=1}^{j}(n-i)\right). \end{aligned} \tag{15}$$

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

$$\sum_{i=1}^{j}\left(\frac{n-i}{2} - \frac{n-i}{4}\right) = \frac{1}{4}\sum_{i=1}^{j}(n-i)\,. \tag{16}$$

Combining (15) and (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, $1 \le i \le j$, 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

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

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

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

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

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

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

□

References

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

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

[BDWY13] B. Barak, Z. Dvir, A. Wigderson, and A. Yehudayoff. Fractional Sylvester–Gallai theorems. *Proceedings of the National Academy of Sciences*, 110(48):19213–19219, 2013.

[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.

[DSW14] Z. Dvir, S. Saraf, and A. 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.

[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.

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

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

[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.

[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.

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