On the Number of Rich Lines in Truly High Dimensional Sets

Zeev Dvir[1] and Sivakanth Gopi[2]

1 Department of Computer Science and Department of Mathematics, Princeton University
35 Olden Street, Princeton, NJ 08540-5233, USA
zeev.dvir@gmail.com

2 Department of Computer Science, Princeton University
35 Olden Street, Princeton, NJ 08540-5233, USA
sgopi@cs.princeton.edu

Abstract

We prove a new upper bound on the number of r-rich lines (lines with at least r points) in a 'truly' d-dimensional configuration of points $v_1, \ldots, v_n \in \mathbb{C}^d$. More formally, we show that, if the number of r-rich lines is significantly larger than n^2/r^d then there must exist a large subset of the points contained in a hyperplane. We conjecture that the factor r^d can be replaced with a tight r^{d+1}. If true, this would generalize the classic Szemerédi-Trotter theorem which gives a bound of n^2/r^3 on the number of r-rich lines in a planar configuration. This conjecture was shown to hold in $\mathbb{R}^3$ in the seminal work of Guth and Katz [7] and was also recently proved over $\mathbb{R}^4$ (under some additional restrictions) [14]. For the special case of arithmetic progressions (r collinear points that are evenly distanced) we give a bound that is tight up to lower order terms, showing that a d-dimensional grid achieves the largest number of r-term progressions.

The main ingredient in the proof is a new method to find a low degree polynomial that vanishes on many of the rich lines. Unlike previous applications of the polynomial method, we do not find this polynomial by interpolation. The starting observation is that the degree $r-2$ Veronese embedding takes r-collinear points to r linearly dependent images. Hence, each collinear r-tuple of points, gives us a dependent r-tuple of images. We then use the design-matrix method of [1] to convert these 'local' linear dependencies into a global one, showing that all the images lie in a hyperplane. This then translates into a low degree polynomial vanishing on the original set.

1998 ACM Subject Classification G.2.1 Combinatorics: Counting Problems

Keywords and phrases Incidences, Combinatorial Geometry, Designs, Polynomial Method, Additive Combinatorics

Digital Object Identifier 10.4230/LIPIcs.SOCG.2015.584

1 Introduction

The Szemerédi-Trotter theorem gives a tight upper bound on the number of incidences between a collection of points and lines in the real plane. We write $A \lesssim B$ to denote $A \leq C \cdot B$ for some absolute constant C and $A \approx B$ if we have both $A \lesssim B$ and $B \lesssim A$. We use $A \gg B$ to mean $A \geq C \cdot B$ for some sufficiently large constant C and we sometimes use a subscript d to mean that the constant C in the inequalities can depend on d.

31st International Symposium on Computational Geometry (SoCG'15).
Editors: Lars Arge and János Pach; pp. 584–598
LIPICS Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

▶ **Theorem 1** ([17]). *Given a set of points V and a set of lines $\mathcal{L}$ in $\mathbb{R}^2$, let $I(V,\mathcal{L})$ be the set of incidences between V and $\mathcal{L}$. Then,*

$$I(V,\mathcal{L}) \lesssim |V|^{2/3}|\mathcal{L}|^{2/3} + |V| + |\mathcal{L}|.$$

This fundamental theorem has found many applications in various areas (see [4] for some examples) and is known to also hold in the complex plane $\mathbb{C}^2$ [18, 19]. In recent years there has been a growing interest in high dimensional variants of line-point incidence bounds [13, 9, 11, 14, 16, 3]. This is largely due to the breakthrough results of Guth and Katz [7] who proved the Erdős distinct distances conjecture. One of the main ingredients in their proof was an incidence theorem for configurations of lines in $\mathbb{R}^3$ satisfying some 'truly 3 dimensional' condition (e.g, not too many lines in a plane). The intuition is that, in high dimensions, it is 'harder' to create many incidences between points and lines. This intuition is of course false if our configuration happens to lie in some low-dimensional space. In this work we prove stronger line-point incidence bounds for sets of points that do not contain a large low-dimensional subset.

To state our main theorem we first restate the Szemerédi-Trotter theorem as a bound on the number of r-rich lines (lines containing at least r points) in a given set of points. Since our results will hold over the complex numbers we will switch now from $\mathbb{R}$ to $\mathbb{C}$. The complex version of Szemerédi-Trotter was first proved by Tóth [18] and then proved using different methods by Zahl [19]. For a finite set of points V, we denote by $\mathcal{L}_r(V)$ the set of r-rich lines in V. The following is equivalent to Theorem-1 (but stated over $\mathbb{C}$).

▶ **Theorem 2** ([18, 19]). *Given a set V of n points in $\mathbb{C}^2$, for $r \geq 2$,*

$$|\mathcal{L}_r(V)| \lesssim \frac{n^2}{r^3} + \frac{n}{r}.$$

Theorem 2 is tight since a two dimensional square grid of n points contains $\gtrsim n^2/r^3$ lines that are r-rich. We might then ask whether a d-dimensional grid $G_d = \{1, 2, \ldots, h\}^d$, with $h \approx n^{1/d}$, has asymptotically the maximal number of r-rich lines among all n-point configurations that do not have a large low-dimensional subset. In [15], it was shown that for $r \ll_d n^{1/d}$,

$$|\mathcal{L}_r(G_d)| \approx_d \frac{n^2}{r^{d+1}}.$$

Clearly, we can obtain a larger number of rich lines in $\mathbb{C}^d$ if V is a union of several low-dimensional grids. For example, for some $\alpha \gg_d 1$ and $d > \ell > 1$, we can take a disjoint union of $r^{d-\ell}/\alpha$ ℓ-dimensional grids G_ℓ of size $\alpha n/r^{d-\ell}$ each. Each of these grids will have $\gtrsim_d \alpha^2 n^2/r^{2d-\ell+1}$ r-rich lines and so, together we will get $\gtrsim_d \alpha n^2/r^{d+1}$ rich lines. We can also take a union of n/r lines containing r points each, to get more r-rich lines than in the d-dimensional grid G_d when $r \gg_d n^{1/d}$. We thus arrive at the following conjecture which, if true, would mean that the best one can do is to paste together a number of grids as above.

▶ **Conjecture 3.** *For $r \geq 2$, suppose $V \subset \mathbb{C}^d$ is a set of n points with*

$$|\mathcal{L}_r(V)| \gg_d \frac{n^2}{r^{d+1}} + \frac{n}{r}.$$

Then there exists $1 < \ell < d$ and a subset $V' \subset V$ of size $\gtrsim_d n/r^{d-\ell}$ which is contained in an ℓ-flat (i.e. an ℓ-dimensional affine subspace).

This conjecture holds in $\mathbb{R}^3$ [7] and, in a slightly weaker form, in $\mathbb{R}^4$ [14]. We compare these two results with ours later in the introduction. Our main result makes a step in the direction of this conjecture. First of all, our bound is off by a factor of r from the optimal bound (i.e. with n^2/r^d instead of n^2/r^{d+1}). Secondly, we are only able to detect a $(d-1)$-dimensional subset (instead of finding the correct ℓ which may be smaller).

▶ **Theorem 4.** *For all $d \geq 1$ there exists constants C_d, C'_d such that the following holds. Let $V \subset \mathbb{C}^d$ be a set of n points and let $r \geq 2$ be an integer. Suppose that for some $\alpha \geq 1$,*

$$|\mathcal{L}_r(V)| \geq C_d \cdot \alpha \cdot \frac{n^2}{r^d}.$$

Then, there exists a subset $\tilde{V} \subset V$ of size at least $C'_d \cdot \alpha \cdot \frac{n}{r^{d-2}}$ contained in a $(d-1)$-flat. We can take the constants C_d, C'_d to be $d^{cd}, d^{c'd}$ for absolute constants $c, c' > 0$.

Notice that the theorem is only meaningful when $r \gg d^c$ for some constant c (otherwise the factor r^d in the assumption will be swallowed by the constant C_d). On the other hand, if $r \gg n^{1/(d-1)}$ then the conclusion always holds. Hence, the theorem is meaningful when r is in a ‘middle’ range. Notice also that for $d = 2, 3$ and r sufficiently small, the condition of the theorem also cannot hold, by the Szemerédi-Trotter theorem. However, when d becomes larger, our theorem gives nontrivial results (and becomes closer to optimal for large d). The proof of Theorem 4 actually shows (Lemma 19) that, under the same hypothesis, most of the rich lines must be contained in a hypersurface of degree smaller than r. This in itself can be very useful, as we will see in the proof of Theorem 9 which uses this fact to prove certain sum-product estimates. The existence of such a low-degree hypersurface containing most of the curves can also be obtained when there are many r-rich curves of bounded degree with ‘two degrees of freedom’, i.e. through every pair of points there are at most $O(1)$ curves (see Remark 22).

Counting arithmetic progressions

An r-term arithmetic progression in $\mathbb{C}^d$ is simply a set of r points of the form $\{y, y+x, y+2x, \dots, y+(r-1)x\}$ with $x, y \in \mathbb{C}^d$. This is a special case of r collinear points and, for this case, we can derive a tighter bound than for the general case. In a nutshell, we can show that a d-dimensional grid contains the largest number of r-term progressions, among all sets that do not contain a large $d-1$ dimensional subset. The main extra property of arithmetic progressions we use in the proof is that they behave well under products. That is, if we take a Cartesian product of V with itself, the number of progressions of length r squares.

For a finite set $V \subset \mathbb{C}^d$, let us denote the number of r-term arithmetic progressions contained in V by $\mathbf{AP}_r(V)$. We first observe that, for all sufficiently small r, the grid G_d (defined above) contains at least $\gtrsim_d n^2/r^d$ r-term progressions. To see where the extra factor of r comes from, notice that the $2r$-rich lines in G_d will contain r arithmetic progressions of length r each. Our main theorem shows that this is optimal, as long as there is no large low-dimensional set.

▶ **Theorem 5.** *Let $0 < \epsilon < 1$ and $V \subset \mathbb{C}^d$ be a set of size n and suppose that for some $r \geq 4$ we have*

$$\mathbf{AP}_r(V) \gg_{d,\epsilon} \frac{n^2}{r^{d-\epsilon}}.$$

Then, there exists a subset $\tilde{V} \subset V$ of size $\gtrsim_{d,\epsilon} \frac{n}{r^{2d/\epsilon - 1}}$ contained in a hyperplane.

1.1 Related Work

To make the comparison with prior work easier, Theorem 4 can be stated equivalently as follows:

▶ **Theorem 6** (Equiv. to Theorem 4). *Given a set V of n points in $\mathbb{C}^d$, let s_{d-1} denote the maximum number of points of V contained in a hyperplane. Then for $r \geq 2$,*

$$|\mathcal{L}_r(V)| \lesssim_d \frac{n^2}{r^d} + \frac{ns_{d-1}}{r^2}.$$

Using the incidence bound between points and lines in $\mathbb{R}^3$ proved by Guth and Katz [7], one can prove the following theorem from which Conjecture 3 in $\mathbb{R}^3$ trivially follows.

▶ **Theorem 7** (Guth and Katz [7]). *Given a set V of n points in $\mathbb{R}^3$, let s_2 denote the maximum number of points of V contained in a 2-flat. Then for $r \geq 2$,*

$$|\mathcal{L}_r(V)| \lesssim \frac{n^2}{r^4} + \frac{ns_2}{r^3} + \frac{n}{r}.$$

Similarly, using the results of Sharir and Solomon [14], we can prove the following theorem from which a slightly weaker version of Conjecture 3 in $\mathbb{R}^4$ trivially follows.

▶ **Theorem 8** (Sharir and Solomon [14]). *Given a set V of n points in $\mathbb{R}^4$, let s_2 denote the maximum number of points of V contained in a 2-flat and s_3' denote the maximum number of points of V contained in a quadric hypersurface or a hyperplane. Then there is an absolute constant $c > 0$ such that for $r \geq 2$,*

$$|\mathcal{L}_r(V)| \lesssim 2^{c\sqrt{\log n}} \cdot \left(\frac{n^2}{r^5} + \frac{ns_3'}{r^4} + \frac{ns_2}{r^3} + \frac{n}{r}\right).$$

We are not aware of any examples where points arranged on a quadric hypersurface in $\mathbb{R}^4$ result in significantly more rich lines than in a four dimensional grid. It is, however, possible that one needs to weaken Conjecture 3 so that for some $1 < \ell < d$, an ℓ-dimensional hypersurface of constant degree (possibly depending on ℓ) contains $\gtrsim_d n/r^{d-\ell}$ points.

In [15], it was shown that $|\mathcal{L}_r(V)| \lesssim_d \frac{n^2}{r^{d+1}}$ when $V \subset \mathbb{R}^d$ is a *homogeneous* set. This roughly means that the point set is a perturbation of the grid G_d. In [10], the result was extended for pseudolines and homogeneous sets in $\mathbb{R}^d$ where pseudolines are a generalization of lines which include constant degree irreducible algebraic curves. Adding the homogeneous condition on a set is a much stronger condition (for sufficiently small r) than requiring that no large subset belongs to a hyperplane (however, we cannot derive these results from ours since our dependence on d is suboptimal).

1.1.1 Subsequent Work

Subsequent to our work, Hablicsek and Scherr [8] improved Theorem 4 in the case of $V \subset \mathbb{R}^d$. It was shown that if $\mathcal{L}_r(V) \gg_d \frac{n^2}{r^{d+1}}$, then $\gtrsim_d \frac{n}{r^{d-1}}$ points are contained in a $(d-1)$-flat. In a further improvement, Zahl [20] extended this result to $V \subset \mathbb{C}^d$ though with an ϵ loss in the exponent of n, i.e. if $\mathcal{L}_r(V) \gg_{d,\epsilon} \frac{n^{2+\epsilon}}{r^{d+1}}$ then $\gtrsim_{d,\epsilon} \frac{n^{1+\epsilon}}{r^{d-1}}$ points are contained in a $(d-1)$-flat. This brings us closer to Conjecture 3, although the conclusion about a large low-dimensional subset is still very weak.

1.2 Overview of the proof

The main tool used in the proof of Theorem 4 is a rank bound for design matrices. A *design matrix* is a matrix with entries in $\mathbb{C}$ and whose support (set of non-zero entries) forms a specific pattern. Namely, the supports of different columns have small intersections, the columns have large support and rows are sparse (see Definition 11). Design matrices were introduced in [1, 5] to study quantitative variants of the Sylvester-Gallai theorem. These works prove certain lower bounds on the rank of such matrices, depending only on the combinatorial properties of their support (see Section 2.1). Such rank bounds can be used to give *upper bounds* on the dimension of point configurations in which there are many 'local' linear dependencies. This is done by using the local dependencies to construct rows of a design matrix M, showing that its rank is high and then arguing that the dimension of the original set is small since it must lie in the kernel of M.

Suppose we have a configuration of points with many r-rich lines. Clearly, $r \geq 3$ collinear points are also linearly dependent. However, this conclusion does not use the fact that r may be larger than 3. To use this information, we observe that a certain map, called the Veronese embedding, takes r-collinear points to r linearly dependent points in a larger dimensional space (see Section 2.2). Thus we can create a design matrix using these linear dependencies similarly to the constructions of [2, 5] to get an upper bound on the dimension of the *image* of the original set, under the Veronese embedding. We use this upper bound to conclude that there is a polynomial of degree $r-2$ which contains all the points in our original configuration. We then proceed in a way similar to the proof of the Joints conjecture by Guth and Katz [6] to conclude that there is a hyperplane which contains many points of the configuration (by finding a 'flat' point of the surface).

1.3 Application: Sum-product estimates

Here, we show a simple application of our techniques to prove sum product estimates over $\mathbb{C}$. Though we can get slightly better estimates (i.e. without the log factor) using the Szemerédi-Trotter theorem in the complex plane, we include them only as an example of how to use a higher-dimensional theorem in this setting. We hope that future progress on proving Conjecture 3 will result in progress on sum product problems.

We begin with some notation. For two sets $A, B \subset \mathbb{C}$ we denote by $A+B = \{a+b \mid a, b \in A\}$ the sum set of A and B. For a set $A \subset \mathbb{C}$ and a complex number $t \in \mathbb{C}$ we denote by $tA = \{ta \mid a \in A\}$ the dilate of A by t. Hence we have that $A + tA = \{a + ta' \mid a, a' \in A\}$.

▶ **Theorem 9.** *Let $A \subset \mathbb{C}$ be a set of N complex numbers and let $1 \ll C \ll \sqrt{N}$. Define the set*

$$T_C = \left\{ t \in \mathbb{C} \;\middle|\; |A + tA| \leq \frac{N^{1.5}}{C\sqrt{\log N}} \right\}.$$

Then, $|T_C| \lesssim \frac{N}{C^2}$.

By taking C to be a large constant, an immediate corollary is:

▶ **Corollary 10.** *Let $A \subset \mathbb{C}$ be a finite set. Then*

$$|A + A \cdot A| = |\{a + bc \mid a, b, c \in A\}| \gtrsim \frac{|A|^{1.5}}{\sqrt{\log |A|}}.$$

1.4 Organization

In Section 2 we give some preliminaries, including on design matrices and the Veronese embedding. In Section 3 we prove Theorem 4. In Section 4 we prove Theorem 5. In Section 5 we prove Theorem 9.

2 Preliminaries

We begin with some notation. For a vector $v \in \mathbb{C}^n$ and a set $I \subset [n]$ we denote by $v_I \subset \mathbb{C}^I$ the restriction of v to indices in I. We denote the *support* of a vector $v \in \mathbb{C}^d$ by $\mathsf{supp}(v) = \{i \in [d] \,|\, v_i \neq 0\}$ (this notation is extended to matrices as well). For a set of n points $V \subset \mathbb{C}^d$ and an integer ℓ, we denote by $V^\ell \subset \mathbb{C}^{d\ell}$ its ℓ-fold Cartesian product i.e. $V^\ell = V \times V \times \cdots \times V$ (ℓ times) where we naturally identify $\mathbb{C}^d \times \mathbb{C}^d \times \cdots \times \mathbb{C}^d$ (ℓ times) with $\mathbb{C}^{d\ell}$.

2.1 Design matrices

Design matrices, defined in [1], are matrices that satisfy a certain condition on their support.

▶ **Definition 11** (Design matrix)**.** Let A be an $m \times n$ matrix over a field $\mathbb{F}$. Let $R_1, \ldots, R_m \in \mathbb{F}^n$ be the rows of A and let $C_1, \ldots, C_n \in \mathbb{F}^m$ be the columns of A. We say that A is a (q,k,t)-*design matrix* if

1. For all $i \in [m]$, $|\mathsf{supp}(R_i)| \leq q$.
2. For all $j \in [n]$, $|\mathsf{supp}(C_j)| \geq k$.
3. For all $j_1 \neq j_2 \in [n]$, $|\mathsf{supp}(C_{j_1}) \cap \mathsf{supp}(C_{j_2})| \leq t$.

Surprisingly, one can derive a general bound on the rank of complex design matrices, despite having no information on the values present at the non-zero positions of the matrix. The first bound of this form was given in [1] which was improved in [5].

▶ **Theorem 12** ([5])**.** *Let A be an $m \times n$ matrix with entries in $\mathbb{C}$. If A is a (q,k,t)-design matrix then the following bounds hold:*

$$\mathsf{rank}(A) \geq n - \frac{ntq^2}{k}, \tag{1}$$

$$\mathsf{rank}(A) \geq n - \frac{mtq^2}{k^2}. \tag{2}$$

2.2 The Veronese embedding

We denote by

$$\mathbf{m}(d,r) = \binom{d+r}{d}$$

the number of monomials of degree at most r in d variables. We will often use the lower bound $\mathbf{m}(d,r) \geq (r/d)^d$. The Veronese embedding $\phi_{d,r} : \mathbb{C}^d \mapsto \mathbb{C}^{\mathbf{m}(d,r)}$ sends a point $a = (a_1, \ldots, a_d) \in \mathbb{C}^d$ to the vector of evaluations of all monomials of degree at most r at the point a. For example, the map $\phi_{2,2}$ sends (a_1, a_2) to $(1, a_1, a_2, a_1^2, a_1a_2, a_2^2)$. We can identify each point $w \in \mathbb{C}^{\mathbf{m}(d,r)}$ with a polynomial $f_w \in \mathbb{C}[x_1, \ldots, x_d]$ of degree at most r in an obvious manner so that the value $f_w(a)$ at a point $a \in \mathbb{C}^d$ is given by the standard inner product $\langle w, \phi_{d,r}(a) \rangle$. We will use the following two easy claims.

▶ **Claim 13.** *Let $V \subset \mathbb{C}^d$ and let $U = \phi_{d,r}(V) \subset \mathbb{C}^{\mathbf{m}(d,r)}$. Then U is contained in a hyperplane iff there is a non-zero polynomial $f \in \mathbb{C}[x_1, \ldots, x_d]$ of degree at most r that vanishes on all points of V.*

Proof. Each hyperplane in $\mathbb{C}^{\mathbf{m}(d,r)}$ is given as the set of points having inner product zero with some $w \in \mathbb{C}^{\mathbf{m}(d,r)}$. If we take the corresponding polynomial $f_w \in \mathbb{C}[x_1, \ldots, x_d]$ we get that it vanishes on V iff $\phi_{d,r}(V)$ is contained in the hyperplane defined by w. ◀

▶ **Claim 14.** *Suppose the $r+2$ points $v_1, \ldots, v_{r+2} \in \mathbb{C}^d$ are collinear and let $\phi = \phi_{d,r} : \mathbb{C}^d \mapsto \mathbb{C}^{\mathbf{m}(d,r)}$. Then, the points $\phi(v_1), \ldots, \phi(v_{r+2})$ are linearly dependent. Moreover, every $r+1$ of the points $\phi(v_1), \ldots, \phi(v_{r+2})$ are linearly independent.*

Proof. Denote $u_i = \phi(v_i)$ for $i = 1, \ldots, r+2$. To show that the u_i's are linearly dependent it is enough to show that, for any $w \in \mathbb{C}^{\mathbf{m}(d,r)}$, if all the $r+1$ inner products $\langle w, u_1\rangle, \ldots, \langle w, u_{r+1}\rangle$ are zero, then the inner product $\langle w, u_{r+2}\rangle$ must also be zero. Suppose this is the case, and let $f_w \in \mathbb{C}[x_1, \ldots, x_d]$ be the polynomial of degree at most r associated with the point w so that $\langle w, u_i\rangle = f_w(v_i)$ for all $1 \le i \le r+1$. Since the points $v_1, \ldots, v_{r+2}$ are on a single line $L \subset \mathbb{C}^d$, and since the polynomial f_w vanishes on $r+1$ of them, we have that f_w must vanish identically on the line L and so $f_w(v_{r+2}) = \langle w, u_{r+2}\rangle = 0$ as well.

To show the 'moreover' part, suppose in contradiction that u_{r+1} is in the span of $u_1, \ldots, u_r$. We can find, by interpolation, a non-zero polynomial $f \in \mathbb{C}[x_1, \ldots, x_d]$ of degree at most r such that $f(v_1) = \ldots = f(v_r) = 0$ and $f(v_{r+1}) = 1$. More formally, we can transform the line containing the $r+1$ points to the x_1-axis by a linear transformation, and then interpolate a degree r polynomial in x_1 with the required properties using the invertibility of the Vandermonde matrix. Now, let $w \in \mathbb{C}^{\mathbf{m}(d,r)}$ be the point such that $f = f_w$. We know that $\langle w, u_i\rangle = 0$ for $i = 1 \ldots r$ and thus, since u_{r+1} is in the span of $u_1, \ldots, u_r$, we get that $f(v_{r+1}) = \langle w, u_{r+1}\rangle = 0$ in contradiction. This completes the proof. ◀

2.3 Polynomials vanishing on grids

We recall the Schwartz-Zippel lemma.

▶ **Lemma 15** ([12, 21]). *Let $S \subset \mathbb{F}$ be a finite subset of an arbitrary field $\mathbb{F}$ and let $f \in \mathbb{F}[x_1, \ldots, x_d]$ be a non-zero polynomial of degree at most r. Then*

$$|\{(a_1, \ldots, a_d) \in S^d \subset \mathbb{F}^d \mid f(a_1, \ldots, a_d) = 0\}| \le r \cdot |S|^{d-1}.$$

An easy corollary is the following claim about homogeneous polynomials.

▶ **Lemma 16.** *Let $S \subset \mathbb{F}$ be a finite subset of an arbitrary field $\mathbb{F}$ and let $f \in \mathbb{F}[x_1, \ldots, x_d]$ be a non-zero homogeneous polynomial of degree at most r. Then*

$$|\{(1, a_2, \ldots, a_d) \in \{1\} \times S^{d-1} \mid f(1, a_2, \ldots, a_d) = 0\}| \le r \cdot |S|^{d-2}.$$

Proof. Let $g(x_2, \ldots, x_d) = f(1, x_2, \ldots, x_d)$ be the polynomial one obtains from fixing $x_1 = 1$ in f. Then g is a polynomial of degree at most r in $d-1$ variables. If g was the zero polynomial then f would have been divisible by $1 - x_1$ which is impossible for a homogeneous polynomial. Hence, we can use Lemma 15 to bound the number of zeros of g in the set S^{d-1} by $r \cdot |S|^{d-2}$. This completes the proof. ◀

Another useful claim says that if a degree one polynomial (i.e. the equation of a hyperplane) vanishes on a large subset of the product set V^ℓ, then there is another degree one polynomial that vanishes on a large subset of V.

▶ **Lemma 17.** *Let $V \subset \mathbb{C}^d$ be a set of n points and let $V^\ell \subset \mathbb{C}^{d\ell}$ be its ℓ-fold Cartesian product. Let $H \subset \mathbb{C}^{d\ell}$ be an affine hyperplane such that $|H \cap V^\ell| \geq \delta \cdot n^\ell$. Then, there exists an affine hyperplane $H' \subset \mathbb{C}^d$ such that $|H' \cap V| \geq \delta \cdot n$.*

Proof. Let $h \in \mathbb{C}^{d\ell}$ be the vector perpendicular to H so that $x \in H$ iff $\langle x, h\rangle = b$ for some $b \in \mathbb{C}$. Observing the product structure of $\mathbb{C}^{d\ell} = (\mathbb{C}^d)^\ell$ we can write $h = (h_1, \ldots, h_\ell)$ with each $h_i \in \mathbb{C}^d$. W.l.o.g suppose that $h_1 \neq 0$. For each $a = (a_2, \ldots, a_\ell) \in V^{\ell-1}$ let $V_a^\ell = V \times \{a_2\} \times \ldots \{a_\ell\}$. Since there are $n^{\ell-1}$ different choices for $a \in V^{\ell-1}$, and since

$$|V^\ell \cap H| = \sum_{a \in V^{\ell-1}} |V_a^\ell \cap H|,$$

there must be some a with $|V_a^\ell \cap H| \geq \delta \cdot n$. Let $H' \subset \mathbb{C}^d$ be the hyperplane defined by the equation

$$x \in H' \text{ iff } \langle x, h_1\rangle + \langle a_2, h_2\rangle + \ldots + \langle a_\ell, h_\ell\rangle = b.$$

Then, $H' \cap V$ is in one-to-one correspondence with the set $V_a^\ell \cap H$ and so has the same size. ◀

2.4 A graph refinement lemma

We will need the following simple lemma, showing that any bipartite graph can be refined so that both vertex sets have high minimum degree (relative the to the original edge density).

▶ **Lemma 18.** *Let $G = (A \sqcup B, E)$ be a bipartite graph with $E \subset A \times B$ and edge set $E \neq \phi$. Then there exists non-empty sets $A' \subset A$ and $B' \subset B$ such that if we consider the induced subgraph $G' = (A' \sqcup B', E')$ then*

1. *The minimum degree in A' is at least $\frac{|E|}{4|A|}$*
2. *The minimum degree in B' is at least $\frac{|E|}{4|B|}$*
3. *$|E'| \geq |E|/2$.*

Proof. We will construct A' and B' using an iterative procedure. Initially let $A' = A$ and $B' = B$. Let $G' = (A' \sqcup B', E')$ be the induced subgraph of G. If there is a vertex in A' with degree (in the induced subgraph G') less than $\frac{|E|}{4|A|}$, remove it from A'. If there is a vertex in B' with degree (in the induced subgraph G') less than $\frac{|E|}{4|B|}$, remove it from B'. At the end of this procedure, we are left with sets A', B' with the required min-degrees. We can count the number of edges lost as we remove vertices in the procedure. Whenever a vertex in A' is removed we lose at most $\frac{|E|}{4|A|}$ edges and whenever a vertex from B' is removed we lose at most $\frac{|E|}{4|B|}$ edges. So

$$|E'| \geq |E| - |A|\frac{|E|}{4|A|} - |B|\frac{|E|}{4|B|} \geq |E|/2.$$

◀

3 Proof of Theorem 4

The main technical tool will be the following lemma, which shows that one can find a vanishing polynomial of low degree, assuming each point is in many rich lines.

▶ **Lemma 19.** *For each $d \geq 1$ there is a constant $K_d \leq 32(2d)^d$ such that the following holds. Let $V \subset \mathbb{C}^d$ be a set of n points and let $r \geq 4$ be an integer. Suppose that, through each point $v \in V$, there are at least k r-rich lines where*

$$k \geq K_d \cdot \frac{n}{r^{d-2}}.$$

Then, there exists a non-zero polynomial $f \in \mathbb{C}[x_1, \ldots, x_d]$ *of degree at most* $r-2$ *such that* $f(v) = 0$ *for all* $v \in V$.

If we have the stronger condition that the number of r*-rich lines through each point of* V *is between* k *and* $8k$ *then we can get the same conclusion (vanishing* f *of degree* $r-2$*) under the weaker inequality*

$$k \geq K_d \cdot \frac{n}{r^{d-1}}.$$

Proof. Let $V = \{v_1, \ldots, v_n\}$ and let $\phi = \phi_{d,r-2} : \mathbb{C}^d \mapsto \mathbb{C}^{\mathbf{m}(d,r-2)}$ be the Veronese embedding with degree bound $r-2$. Let us denote $U = \{u_1, \ldots, u_n\} \subset \mathbb{C}^{\mathbf{m}(d,r-2)}$ with $u_i = \phi(v_i)$ for all $i \in [n]$.

We will prove the lemma by showing that U is contained in a hyperplane and then using Claim 13 to deduce the existence of the vanishing polynomial. Let M be an $n \times \mathbf{m}(d, r-2)$ matrix whose i'th row is $u_i = \phi(v_i)$. To show that U is contained in a hyperplane, it is enough to show that $\mathsf{rank}(M) < \mathbf{m}(d, r-2)$. This will imply that the columns of M are linearly dependent, which means that all the rows lie in some hyperplane.

We will now construct a design matrix A such that $A \cdot M = 0$. Since $\mathsf{rank}(A) + \mathsf{rank}(M) \leq n$, we will be able to translate a lower bound on the rank of A (which will be given by Theorem 12) to the required upper bound on the rank of M. Each row in A will correspond to some collinear r-tuple in V. We will construct A in several stages. First, for each r-rich line $L \in \mathcal{L}_r(V)$ we will construct a set of r-tuples $R_L \subset \binom{V}{r}$ such that

1. Each r-tuple in R_L is contained in $L \cap V$.
2. Each point $v \in L \cap V$ is in at least one and at most two r-tuples from R_L.

If $|L \cap V|$ is a multiple of r, we can construct such a set R_L easily by taking a disjoint cover of r-tuples. If $|L \cap V|$ is not a multiple of r (but is still of size at least r) we can take a maximal set of disjoint r-tuples inside it and then add to it one more r-tuple that will cover the remaining elements and will otherwise intersect only one other r-tuple. This will guarantee that each point in $L \cap V$ is in at most two r-tuples from R_L. We define $R \subset \binom{V}{r}$ to be the union of all sets R_L over all r-rich lines L. We can now prove:

▶ **Claim 20.** *The set* $R \subset \binom{V}{r}$ *defined above has the following three properties.*

1. *Each point* $v \in V$ *is contained in at least* k r*-tuples from* R.
2. *Every pair of distinct points* $u, v \in V$ *is contained together in at most two* r*-tuples from* R.
3. *Let* $(v_{i_1}, \ldots, v_{i_r}) \in R$. *Then there exists* r *non-zero coefficients* $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$ *so that* $\alpha_1 \cdot u_{i_1} + \ldots + \alpha_r \cdot u_{i_r} = 0$.

If, in addition, we know that each point belongs to at most $8k$ *rich lines (as in the second part of the lemma) then we also have that* $|R| \leq 16nk/r$.

Proof. The first property follows from the fact that each v is in at least k r-rich lines and that each R_L with $v \in L$ has at least one r-tuple containing v. The second property follows from the fact that each pair u, v can belong to at most one r-rich line L and that each R_L can contain at most two r-tuples with both u and v. The fact that the r-tuple of point $u_{i_1}, \ldots, u_{i_r}$ is linearly dependent follows from Claim 14. The fact that all the coefficients α_j are non-zero holds since no proper subset of that r-tuple is linearly dependent (by the 'moreover' part of Claim 14). If each point is in at most $8k$ lines then each point is in at most $16k$ r-tuples (at most two on each line). This means that there could be at most $16nk/r$ tuples in R since otherwise, some point would be in too many tuples. ◀

We now construct the matrix A of size $m \times n$ where $m = |R|$. For each r-tuple $(v_{i_1}, \ldots, v_{i_r}) \in R$ we add a row to A (the order of the rows does not matter) that has zeros in all positions except $i_1, \ldots, i_r$ and has values $\alpha_1, \ldots, \alpha_r$ given by Claim 20 in those positions. Since the rows of M are the points $u_1, \ldots, u_n$, the third item of Claim 20 guarantees that $A \cdot M = 0$ as we wanted. The next claim asserts that A is a design matrix.

▶ **Claim 21.** *The matrix A constructed above is a $(r, k, 2)$-design matrix.*

Proof. Clearly, each row of A contains at most r non-zero coordinates. Since each point $v \in V$ is in at least k r-tuples from R we have that each column of A contains at least k non-zero coordinates. The size of the intersection of the supports of two distinct columns in A is at most two by item (2) of Claim 20. ◀

We now use Eq. (1) from Theorem 12 to get

$$\mathsf{rank}(A) \geq n - \frac{2nr^2}{k}.$$

This implies (using $r \geq 4$) that

$$\mathsf{rank}(M) \leq \frac{2nr^2}{k} \leq \left(\frac{r-2}{d}\right)^d < \mathbf{m}(d, r-2),$$

if

$$k \geq 2(2d)^d \cdot \frac{n}{r^{d-2}}.$$

If we have the additional assumption that each point is in at most $8k$ lines then, using the bound $m = |R| \leq 16nk/r$ in Eq. (2) of Theorem 12. We get

$$\mathsf{rank}(A) \geq n - \frac{2mr^2}{k^2} \geq n - \frac{32nr}{k}$$

which gives

$$\mathsf{rank}(M) \leq \frac{32nr}{k} < \mathbf{m}(d, r-2)$$

for

$$k \geq 32(2d)^d \frac{n}{r^{d-1}}.$$

Hence, the rows of M lie in a hyperplane. This completes the proof of the lemma. ◀

▶ Remark 22. Lemma 19 can be extended to the case where we have r-rich curves of bounded degree $D = O(1)$ with 'two degrees of freedom', i.e. through every pair of points there can be at most $C = O(1)$ distinct curves (e.g. unit circles). Under the Veronese embedding $\phi_{d,\lfloor \frac{r-2}{D} \rfloor}$, the images of r points on a degree D curve are linearly dependent. So we can still construct a design matrix as in the above proof where the design parameters depend on D, C. Once we get a hypersurface of degree $\lfloor \frac{r-2}{D} \rfloor$ vanishing on all the points, the hypersurface should also contain all the degree D r-rich curves.

We will now use Lemma 19 to prove Theorem 4. The reduction uses Lemma 18 to reduce to the case where each point has many rich lines through it. Once we find a vanishing low degree polynomial we analyze its singularities to find a point such that all lines though it are in some hyperplane.

Proof of Theorem 4. Since $\mathcal{L}_r(V) \leq n^2$ for all $r \geq 2$, by choosing $C_d > R_d^d$ we can assume that $r \geq R_d$ for any large constant R_d depending only on d.

Let $\mathcal{L} = \mathcal{L}_r(V)$ be the set of r-rich lines in V and let $I = I(\mathcal{L}, V)$ be the set of incidences between $\mathcal{L}$ and V. By the conditions of the theorem we have

$$|I| \geq r|\mathcal{L}| \geq C_d \cdot \frac{\alpha n^2}{r^{d-1}}. \tag{3}$$

Applying Lemma 18 to the incidence graph between V and $\mathcal{L}$, we obtain non-empty subsets $V' \subset V$ and $\mathcal{L}' \subset \mathcal{L}$ such that each $v \in V'$ is in at least $k = \frac{|I|}{4n}$ lines from $\mathcal{L}'$ and such that each line in $\mathcal{L}'$ is $r/4$-rich w.r.t to the set V' and

$$|I'| = |I(\mathcal{L}', V')| \geq |I|/2.$$

We would like to apply Lemma 19 with the stronger condition that each point is incident on approximately the same number of lines (which gives better dependence on r). To achieve this, we will further refine our set of points using dyadic pigeonholing.

Let $V' = V_1' \sqcup V_2' \sqcup \cdots$ be a partition of V' into disjoint subsets where V_j' is the set of points incident to at least $k_j = 2^{j-1}k$ and less than $2^j k$ lines from $\mathcal{L}'$. Let $I_j' = I(\mathcal{L}', V_j')$, so that

$$\sum_{j \geq 1} |I_j'| = |I'| \geq |I|/2.$$

Since $\sum_{j\geq 1} \frac{1}{2j^2} < 1$, there exists j such that $|I_j'| \geq \frac{|I|}{4j^2}$. Let us fix j to this value for the rest of the proof.

We will first upper bound j. Since $|I_j'| > 0$, V_j' is non-empty and let $p \in V_j'$. There are at least k_j $(r/4)$-rich lines through p and by choosing $R_d \geq 8$, there are at least $r/4 - 1 \geq r/8$ points other than p on each of these lines and they are all distinct. So,

$$n = |V| \geq 2^{j-1}k \cdot \frac{r}{8} = \frac{2^{j-6} r|I|}{n} \geq C_d \frac{2^{j-6}\alpha n}{r^{d-2}} \geq \frac{2^{j-6} n}{r^{d-2}}.$$

This implies $j \lesssim d \log r$ where we assumed above that $C_d \geq 1$.

Since the lines in $\mathcal{L}'$ need not be $r/4$-rich w.r.t V_j', we need further refinement. Apply Lemma 18 again on the incidence graph $I_j' = I(\mathcal{L}', V_j')$ to get non-empty $V'' \subset V_j'$ and $\mathcal{L}'' \subset \mathcal{L}'$ and

$$|I''| = |I(\mathcal{L}'', V'')| \geq \frac{|I_j'|}{2} \geq \frac{|I|}{8j^2} \geq \frac{r|\mathcal{L}|}{8j^2}.$$

Each line in $\mathcal{L}''$ is incident to at least

$$\frac{|I_j'|}{4|\mathcal{L}'|} \geq \frac{r}{16j^2} = r_0$$

points from V'' and so $\mathcal{L}''$ is r_0-rich w.r.t V''. And each point in V'' is incident to at least

$$\frac{|I_j'|}{4|V_j'|} \geq \frac{k_j}{4} = 2^{j-3}k = k_0$$

and at most $2^j k = 8k_0$ lines from $\mathcal{L}''$. Since $j \lesssim d \log r$, we can assume $r_0 = \frac{r}{16j^2} \geq 4$ by choosing $R_d \gg d^3$.

The following claim shows that we can apply Lemma 19 to V'' and $\mathcal{L}''$.

▶ **Claim 23.** *$k_0 \geq K_d \cdot \frac{|V''|}{r_0^{d-1}}$ where K_d is the constant in Lemma 19.*

Proof. We have

$$|V''| \le |V'_j| \le \frac{|I|}{2^{j-1}k} = \frac{n}{2^{j-3}}.$$

So it is enough to show that

$$k_0 \ge K_d \cdot \frac{n}{2^{j-3} r_0^{d-1}}.$$

Substituting the bounds we have for k_0 and r_0, this will follow from

$$|I| \ge 16K_d \cdot 2^{4d} \cdot \left(\frac{j^{2(d-1)}}{2^{2j}}\right) \frac{n^2}{r^{d-1}}$$

which follows from Eq. (3) by choosing $C_d > 16K_d \cdot 2^{4d} \cdot \max_j \left(\frac{j^{2(d-1)}}{2^{2j}}\right)$. ◀

Hence, by Lemma 19, there exists a non-zero polynomial $f \in \mathbb{C}[x_1, \ldots, x_d]$ of degree at most $r_0 - 2$, vanishing at all points of V''. W.l.o.g suppose f has minimal total degree among all polynomials vanishing on V''. Since f has degree at most $r_0 - 2$ it must vanish identically on all lines in $\mathcal{L}''$.

We say that a point $v \in V''$ is 'flat' if the set of lines from $\mathcal{L}''$ passing through v are contained in some affine hyperplane through v. Otherwise, we call the point v a 'joint'. We will show that there is at least one flat point in V''. Suppose towards a contradiction that all points in V'' are joints. Let $v \in V''$ be some point and let $\nabla f(v)$ be the gradient of f at v. Since f vanishes identically on all lines in $\mathcal{L}''$ we get that $\nabla f(v) = 0$ (v is a singular point of the hypersurface defined by f). We now get a contradiction since one of the coordinates of ∇f is a non-zero polynomial of degree smaller than the degree of f that vanishes on the entire set V''.

Hence, there exists a point $v \in V''$ and an affine hyperplane H passing through v such that all r_0-rich lines in $\mathcal{L}''$ passing through v are contained in H. Since there are at least k_0 such lines, and each line contain at least $r_0 - 1$ points in addition to v, we get that H contains at least

$$(r_0 - 1)k_0 \ge \frac{r}{32j^2} \cdot 2^{j-3}\frac{|I|}{4n} \ge C_d \left(\frac{2^{j-10}}{j^2}\right) \frac{\alpha n}{r^{d-2}} \ge C'_d \frac{\alpha n}{r^{d-2}}$$

points from V where $C'_d = C_d \cdot \min_j \left(\frac{2^{j-10}}{j^2}\right)$. We can take the constants to be $C_d = d^{\Theta(d)}$ and $C'_d = \frac{C_d}{2^{11}}$. ◀

▶ Remark 24. Observe that, we can take $\mathcal{L}$ to be any subset of $\mathcal{L}_r(V)$ of size $\ge C_d \frac{\alpha n^2}{r^d}$ and obtain the same conclusion. Moreover, the hyperplane H that we obtain at the end contains $k_0 \gtrsim \frac{\alpha n}{r^d}$ lines of $\mathcal{L}$.

4 Proof of Theorem 5

We will reduce the problem of bounding r-term arithmetic progressions to that of bounding r-rich lines using the following claim:

▶ **Claim 25.** *Let $V \subset \mathbb{C}^d$ then $\mathbf{AP}_r(V) \le |\mathcal{L}_r([r] \times V)|$ where $[r] = \{0, 1, \cdots, r-1\}$*

Proof. For $u, w \in \mathbb{C}^d, w \ne 0$, let $(u, u+w, \cdots, u+(r-1)w)$ be an r-term arithmetic progression in V. Then the line $\{(0,u)+z(1,w)\}_{z\in\mathbb{C}}$ is r-rich w.r.t the point set $[r]\times V \subset \mathbb{C}^{1+d}$; moreover this mapping is injective. ◀

We need the following claim regarding arithmetic progressions in product sets.

▶ **Claim 26.** *Let $V \subset \mathbb{C}^d$ be a set of n points and let $\ell \geq 1$ be an integer. Then, for all $r \geq 1$, the product set $V^\ell \subset \mathbb{C}^{d\ell}$ satisfies*

$$\mathbf{AP}_r(V^\ell) \geq \mathbf{AP}_r(V)^\ell.$$

Proof. Let $P(V)$ be the set of r-term arithmetic progressions in V and let $P(V^\ell)$ be the set of r-term progressions in V^ℓ. We will describe an injective mapping from $P(V)^\ell$ into $P(V^\ell)$. For $u, w \in \mathbb{C}^d$ let $L_{u,w} = \{u, u+w, \ldots, u+(r-1)w\}$ be the r-term progression starting at u with difference w. Let $u_1, \ldots, u_\ell, w_1, \ldots, w_\ell \in \mathbb{C}^d$ such that $L_{u_i,w_i} \in P(V)$ for each $i \in [\ell]$. We map them into the arithmetic progression $L_{u,w} \in P(V^\ell)$ with $u = (u_1, \ldots, u_\ell)$ and $w = (w_1, \ldots, w_\ell)$. Clearly, this map is injective (care should be taken to assign each progression a unique difference since these are determined up to a sign). ◀

Proof of Theorem 5. Let us assume $\mathbf{AP}_r(V) \gg_{d,\epsilon} \frac{n^2}{r^{d-\epsilon}}$. Let $\ell = \lceil \frac{1}{\epsilon} \rceil$. By Claim 26, $\mathbf{AP}_r(V^\ell) \geq \mathbf{AP}_r(V)^\ell$. Let $\mathcal{L}$ be the collection of r-rich lines w.r.t $[r] \times V^\ell \subset \mathbb{C}^{1+d\ell}$ corresponding to nontrivial r-term arithmetic progressions in V^ℓ, as given by Claim 25. So

$$|\mathcal{L}_r([r] \times V^\ell)| \geq |\mathcal{L}| = \mathbf{AP}_r(V^\ell) \geq \mathbf{AP}_r(V)^\ell \gg_{d,\epsilon} \frac{n^{2\ell}}{r^{d\ell-\epsilon\ell}} \geq \frac{n^{2\ell}}{r^{d\ell-1}} = \frac{(n^\ell r)^2}{r^{d\ell+1}}.$$

By Theorem 4 (choosing the constants appropriately), there is a hyperplane H in $\mathbb{C}^{1+d\ell}$ which contains $\gtrsim_{d,\epsilon} \frac{n^\ell r}{r^{d\ell-1}}$ points of $[r] \times V^\ell$. Moreover, by Remark 24, H contains some of the lines of $\mathcal{L}$. So H cannot be one of the hyperplanes $\{z_1 = i\}_{i \in [r]}$ because they do not contain any lines of $\mathcal{L}$. So the intersection of H with one of the r hyperplanes $\{z_1 = i\}_{i \in [r]}$ (say j) gives a $(d\ell - 1)$-flat which contains $\gtrsim_{d,\epsilon} \frac{n^\ell}{r^{d\ell-1}}$ points of $V^\ell \times \{j\}$. This gives a hyperplane H' in $\mathbb{C}^{d\ell}$ which contains $\gtrsim_{d,\epsilon} \frac{n^\ell}{r^{d\ell-1}}$ points of V^ℓ. Now by Lemma 17, we can conclude that there is a hyperplane in $\mathbb{C}^d$ which contains $\gtrsim_{d,\epsilon} \frac{n}{r^{d\ell-1}} \geq \frac{n}{r^{2d/\epsilon-1}}$ points of V. ◀

5 Proof of Theorem 9

Suppose in contradiction that $|T_C| > \lambda N/C^2$ for some large absolute constant λ which we will choose later. Let $Q \subset T_C$ be a set of size $|Q| = \lceil \lambda N/C^2 \rceil$ containing the zero element $0 \in Q$ (we have $0 \in T_C$ since the sum-set $|A + 0A| = |A|$ is small). Let

$$r = |Q|,\, m = \frac{N^{1.5}}{C\sqrt{\log N}},\, d = \lceil 100 \log N \rceil.$$

We will use our assumption on the size of Q to construct a configuration of points $V \subset \mathbb{C}^d$ with many r-rich lines. Then we will use Lemma 19 to derive a contradiction. The set V will be a union of the sets

$$V_t = \{t\} \times (A + tA)^{d-1} = \{(t, a_2 + tb_2, \ldots, a_d + tb_d) \,|\, a_i, b_j \in A\}$$

over all $t \in Q$, i.e. $V = \bigcup_{t \in Q} V_t$. Notice the special structure of the set $V_0 = \{0\} \times A^{d-1}$. We denote by

$$n = |V| \leq r \cdot m^{d-1} \tag{4}$$

Notice that, by construction, for every $a = (0, a_2, \ldots, a_d)$ and every $b = (1, b_2, \ldots, b_d)$ (with all the a_i, b_j in A), the line through the point $a \in V_0$ in direction b is r-rich w.r.t V.

Let us denote by $\mathcal{L} \subset \mathcal{L}_r(V)$ the set of all lines of this form. We thus have $|\mathcal{L}| = N^{2d-2}$. Let $I = I(V, \mathcal{L})$, then $|I| \geq r|\mathcal{L}|$. We now use Lemma 18 to find subsets $V' \subset V$ and $\mathcal{L}' \subset \mathcal{L}$ such that each point in V' is in at least

$$k = \frac{rN^{2d-2}}{4n}$$

lines from $\mathcal{L}'$, each line in $\mathcal{L}'$ is $r_0 = r/4$-rich w.r.t to the set V' and

$$|I(V', L')| \geq |I|/2.$$

Observe that, since each line in $\mathcal{L}'$ contains at most r points from V', we have

$$|\mathcal{L}'| \geq |I(V', \mathcal{L}')|/r \geq |\mathcal{L}|/2.$$

The following claim shows that we can apply Lemma 19 on the set V'.

▶ **Claim 27.**

$$k \geq K_d \frac{n}{r_0^{d-2}}.$$

where $K_d = 32(2d)^d$ is the constant in Lemma 19.

Proof. Plugging in the value of k, r_0 and using bound Eq. 4 to bound n, we need to show that

$$r^{d-3} \geq \frac{32(8d)^d N^{d-1}}{(C^2)^{d-1}(\log N)^{d-1}}.$$

We now raise both sides to the power $1/(d-3)$ and use the fact that, for $\ell > \log X$, we have $1 \leq X^{1/\ell} \leq 2$. Thus it is enough to show

$$r \geq \frac{K' dN}{C^2 \log N} = \frac{K'N \lceil 100 \log N \rceil}{C^2}$$

for some absolute constant K' which holds by choosing $\lambda = 100K'$. ◀

Since $C \ll \sqrt{N}$, $r_0 \geq 4$. Applying Lemma 19, we get a non-zero polynomial $f \in \mathbb{C}[x_1, \ldots, x_d]$ of degree at most $r_0 - 2$ that vanishes on all points in V'. This means that f must also vanish identically on all lines in $\mathcal{L}'$ (since these are all r_0-rich w.r.t V'). Since each line in $\mathcal{L}'$ intersects V_0 exactly once, and since $|V_0| = N^{d-1}$, we get that there must be at least one point $v \in V_0$ that is contained in at least $|\mathcal{L}'|/N^{d-1} \geq \frac{1}{2}N^{d-1}$ lines (in different directions) from $\mathcal{L}'$. Let $\tilde{f}$ denote the homogeneous part of f of highest degree. If f vanishes identically on a line in direction $b \in \mathbb{C}^d$, this implies that $\tilde{f}(b) = 0$ (to see this notice that the leading coefficient of $g(t) = f(a + tb)$ is $\tilde{f}(b)$). Hence, since all the directions of lines in $\mathcal{L}'$ are from the set $\{1\} \times A^{d-1}$, we get that $\tilde{f}$ has at least $\frac{1}{2}N^{d-1}$ zeros in the set $\{1\} \times A^{d-1}$. This contradicts Lemma 16 since the degree of $\tilde{f}$ is at most $r_0 - 2 = r/4 - 2 < N/2$ (since $r = \lceil \lambda N/C^2 \rceil$ and $C \gg 1$). This completes the proof of Theorem 9. ◀

Acknowledgements. We thank Ben Green and Noam Solomon for helpful comments. Research supported by NSF grant CCF-1217416 and by the Sloan fellowship. Some of the work on the paper was carried out during the special semester on 'Algebraic Techniques for Combinatorial and Computational Geometry', held at the Institute for Pure and Applied Mathematics (IPAM) during Spring 2014.

References

1 B. Barak, Z. Dvir, A. Wigderson, and A. Yehudayoff. Fractional Sylvester-Gallai theorems. *Proceedings of the National Academy of Sciences*, 2012.

2 B. Barak, Z. Dvir, A. Yehudayoff, and A. Wigderson. Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC '11, pages 519–528, New York, NY, USA, 2011. ACM.

3 Saugata Basu and Martin Sombra. Polynomial partitioning on varieties and point-hypersurface incidences in four dimensions. *arXiv preprint arXiv:1406.2144*, 2014.

4 Z. Dvir. Incidence Theorems and Their Applications. *Foundations and Trends in Theoretical Computer Science*, 6(4):257–393, 2012.

5 Z. Dvir, S. Saraf, and A. Wigderson. Improved rank bounds for design matrices and a new proof of Kelly's theorem. *Forum of Mathematics, Sigma*, 2, 10 2014.

6 L. Guth and N. Katz. Algebraic methods in discrete analogs of the Kakeya problem. *Advances in Mathematics*, 225(5):2828 – 2839, 2010.

7 Larry Guth and Nets Hawk Katz. On the Erdős distinct distances problem in the plane. *Annals of Mathematics*, 181(1):155–190, 2015.

8 Marton Hablicsek and Zachary Scherr. On the number of rich lines in high dimensional real vector spaces. *arXiv preprint arXiv:1412.7025*, 2014.

9 J. Kollar. Szemerédi–Trotter-type theorems in dimension 3. *arXiv:1405.2243*, 2014.

10 Izabella Laba and József Solymosi. Incidence theorems for pseudoflats. *Discrete & Computational Geometry*, 37(2):163–174, 2007.

11 M. Rudnev. On the number of incidences between planes and points in three dimensions. *arXiv:1407.0426v2*, 2014.

12 J. T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, 1980.

13 Micha Sharir, Adam Sheffer, and Joshua Zahl. Improved bounds for incidences between points and circles. In *Proceedings of the Twenty-ninth Annual Symposium on Computational Geometry*, SoCG '13, pages 97–106, New York, NY, USA, 2013. ACM.

14 Micha Sharir and Noam Solomon. Incidences between points and lines in $\mathbb{R}^4$. *arXiv preprint arXiv:1411.0777*, 2014.

15 József Solymosi and VH Vu. Distinct distances in high dimensional homogeneous sets. *Contemporary Mathematics*, 342:259–268, 2004.

16 József Solymosi and Terence Tao. An incidence theorem in higher dimensions. *Discrete and Computational Geometry*, 48(2):255–280, 2012.

17 E. Szemerédi and W. T. Trotter. Extremal problems in discrete geometry. *Combinatorica*, 3(3):381–392, 1983.

18 C. Toth. The Szemerédi-Trotter theorem in the complex plane. *arXiv:math/0305283v4*, 2003.

19 Joshua Zahl. A Szemerédi-Trotter type theorem in $\mathbb{R}^4$. *CoRR*, abs/1203.4600, 2012.

20 Joshua Zahl. A note on rich lines in truly high dimensional sets. *arXiv preprint arXiv:1503.01729*, 2015.

21 R. Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposiumon on Symbolic and Algebraic Computation*, pages 216–226. Springer-Verlag, 1979.