BILINEAR AND QUADRATIC VARIANTS ON THE LITTLEWOOD-OFFORD PROBLEM

KEVIN P. COSTELLO

Abstract. If $f(x_1, \dots, x_n)$ is a polynomial dependent on a large number of independent Bernoulli random variables, what can be said about the maximum concentration of f on any single value? For linear polynomials, this reduces to one version of the classical Littlewood-Offord problem: Given nonzero constants $a_1, \dots, a_n$, what is the maximum number of sums of the form $\pm a_1 \pm a_2 \pm \dots \pm a_n$ which take on any single value? Here we consider the case where f is either a bilinear form or a quadratic form. For the bilinear case, we show that the only forms having concentration significantly larger than n^{-1} are those which are in a certain sense very close to being degenerate. For the quadratic case, we show that no form having many nonzero coefficients has concentration significantly larger than $n^{-1/2}$. In both cases the results are nearly tight.

1. Introduction: The Linear Littlewood-Offord Problem

In their study of the distribution of the number of real roots of random polynomials, Littlewood and Offord [12] encountered the following problem:

Question 1. *Let a_1, ... , a_n be real numbers such that $|a_i| > 1$ for every i. What is the largest number of the 2^n sums of the form*

$$\pm a_1 \pm a_2 \pm \dots \pm a_n$$

that can lie in any interval of length 1?

Littlewood and Offord showed an upper bound of $O(2^n \frac{\log n}{\sqrt{n}})$ on the number of such sums. Erdős [5] later removed the $\log n$ factor from this result, giving an exact bound of $\binom{n}{\lfloor n/2 \rfloor}$ via Sperner's Lemma, which is tight in the case where all of the a_i are equal. The same bound was later shown by Kleitman [10] in the case where the a_i are complex numbers. Rescaling Kleitman's result and using Stirling's approximation gives the following probabilistic variant of the lemma:

Theorem 1. *Let $n > 0$, and let $a_1, \dots, a_n$ be arbitrary complex numbers, at least $m \geq 1$ of which are nonzero. Let $x_1, \dots, x_n$ be independent ± 1 symmetric random variables (meaning each variable is drawn uniformly from $\{1, -1\}$). Then*

$$\sup_{c \in \mathbb{C}} \mathbf{P}(\sum_{i=1}^{n} a_i x_i = c) \leq \min\{\frac{1}{2}, \frac{1}{\sqrt{m}}\}.$$

This research was supported by NSF Grants DMS-0635607 and DMS-0902968.

In a sense Theorem 1 can be thought of as a quantitative description of the dispersion of a random walk: No matter what step sizes the walk takes, as the number of steps increases the walk becomes less and less concentrated on any particular value. In this interpretation the $\sqrt{m}$ in the bound is also unsurprising; if the step sizes are small integers, we would expect the walk to typically be at an integer about $O(\sqrt{m})$ distance from 0 at time m, so the concentration at individual points near 0 should be roughly $m^{-1/2}$.

In 1977 Halász [6] gave several far-reaching generalizations of Theorem 1, both to higher dimensions and to more general classes of random variables. One (rescaled) result of his is

Theorem 2. *Let $a_1, \dots, a_n$ be vectors in $\mathbb{C}^d$ such that no proper subspace of $\mathbb{C}^d$ contains more than $n-m$ of the a_i, where $m \geq 1$. Let $x_1, \dots, x_n$ be independent complex-valued random variables such that for some $\rho < 1$,*

$$\sup_{i,c} \mathbf{P}(x_i = c) \leq \rho.$$

Then

$$\sup_{c \in \mathbb{C}} \mathbf{P}(\sum_{i=1}^{n} a_i x_i = c) = O_{\rho,d}(m^{-d/2}).$$

The $O_{\rho,d}$ here means that the constant implicit in the $O()$ notation is allowed to depend on ρ and d.

The original Littlewood-Offord lemma corresponds to the special case where $d = 1$ and the x_i are iid Bernoulli variables. Again this can be thought of as a dispersion result: a linear polynomial which depends on a large number of independent, moderately dispersed random variables will itself be very dispersed. Furthermore, the dispersion will be greater if the coefficients of the polynomial are in some sense truly $d-$dimensional.

One application of these results is in the study of random matrices, since several key parameters of a matrix (e.g. the determinant, or the distance from one row to the span of the remaining rows) are linear forms in the entry of a single row or column of the matrix. Komlós [11] used Theorem 1 in 1967 to show that a random Bernoulli matrix (one whose entries are independently either 1 or -1) is almost surely non-singular. Later, Kahn, Komlós and Szemerédi [9] used the ideas of Halász to show that the singularity probability was exponentially small of the size of the matrix. The current best bound for this probability due to Bourgain, Vu, and Wood, $(\frac{1}{\sqrt{2}} + o(1))^n$ for an $n \times n$ matrix [1] (see also Tao and Vu [18], whose ideas this paper build on), comes from a detailed analysis of the *inverse* of the Littlewood-Offord problem, which can be thought of as

Question 2. *If $\sum a_i x_i$ is highly concentrated on one value, what can be said about the a_i?*

The intuition here is that the sum takes on a single value with probability close to $n^{-1/2}$, then the a_i should be very highly structured. Tao and Vu [19] and Rudelson and Vershynin [14] showed that this was in fact the case: If the sum takes on a single value with probability at least n^{-c} for some fixed c, then the coefficients must have been drawn from a short generalized arithmetic progression. One special case of this result can be expressed more quantitatively in the following theorem from [19]

Theorem 3. *Let $\epsilon < \frac{1}{2}$ and $\alpha > 0$ be fixed. Let $a_1, \dots, a_n$ be nonzero complex numbers, where n is at least some constant N_0 dependent only on ϵ and α. Suppose furthermore that for independent ± 1 symmetric random variables $x_1, \dots, x_n$ we have*

$$\sup_c \mathbf{P}(\sum_{i=1}^n a_i x_i = c) \geq n^{-1/2-\epsilon}.$$

Then there is a $d \in \mathbb{C}$ such that all but $n^{1-\alpha}$ of the a_i have the form

$$a_i = db_i,$$

where the b_i are integers such that $|b_i| \leq n^{\epsilon+\alpha}$.

The same holds true if the x_i are independent and identically distributed "lazy walker" variables satisfying $\mathbf{P}(x_i = 0) = 2\rho$, $\mathbf{P}(x_i = 1) = \mathbf{P}(x_i = -1) = 1/2 - \rho$ for some $0 < \rho < 1/2$ (N_0 is now also dependent on ρ).

2. Statement of Main Results

Our goal here will be to develop and strengthen extensions of Theorem 1 and related results to polynomials of higher degree, in particular bilinear and quadratic forms. To begin, let us consider the following result (implicit in [2]), which we reprove here for convenience:

Theorem 4. *Let $A = a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n$ be an array of complex numbers, and suppose that at least r distinct rows of A each contain at least r nonzero entries. Let $x^T = (x_1, \dots, x_m)$ and $y^T = (y_1, \dots, y_n)$ be two vectors whose $m+n$ entries are independent ± 1 symmetric random variables. Then*

$$\sup_{c \in \mathbb{C}} \mathbf{P}(x^T A y = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j = c) = O(r^{-1/2}).$$

Proof : Without loss of generality we may assume that the rows in question correspond to the variables x_1 through x_r.

Let $W_i = \sum_j a_{ij} y_j$, and let W denote the number of i between 1 and r for which W_i is equal to 0. We have

$$\mathbf{P}(x^T A y = c) \leq \mathbf{P}(W \geq \frac{r}{2}) + \mathbf{P}(x^T A y = c \wedge W < \frac{r}{2}).$$

We bound each term separately. For the first term, we view W as a sum of the indicator function of the events that each W_i is equal to 0. Since each W_i is a linear polynomial with at least r nonzero coefficients, it follows from Theorem 1 that each

W_i is equal to 0 with probability $O(r^{-1/2})$, it follows from linearity of expectation that $\mathbf{E}(W) = O(r^{1/2})$, and therefore from Markov's inequality that

$$\mathbf{P}(W \geq \frac{r}{2}) = O(r^{-1/2}).$$

For the second term, we treat y as fixed and write

$$x^T Ay = \sum_i W_i x_i.$$

If W is at most $\frac{r}{2}$, then the right hand side is a linear form in the x_i with at least $\frac{r}{2}$ nonzero coefficients. It follows from Theorem 1 and taking expectations over y that this term is $O(r^{-1/2})$. ■

In a certain sense this is a weaker result than we might expect. If A is an $n \times n$ matrix of small nonzero integers, then the magnitude of $x^T Ay$ will typically be around n, so we might expect a concentration probability of n^{-1} instead of $n^{-1/2}$. However, Theorem 4 is tight, as the case where A is the all ones matrix (corresponding to the polynomial $(x_1 + ... + x_n)(y_1 + ... + y_n)$) shows. What our first main result shows is that every bilinear form with sufficiently large concentration probability is in some sense close to this degenerate example.

Theorem 5. *Fix $\epsilon > 0$. There is an R_0 (dependent on ϵ) such that the following holds. Let A be an $m \times n$ coefficient matrix such that every row of A has at least r nonzero entries, where $R_0 \leq r \leq m$. Let $x = (x_1, \dots, x_m)$ and $y = (y_1, \dots, y_n)$ be vectors of independent ± 1 symmetric random variables. Suppose furthermore that there is a function $f : \mathbb{R}^n \to \mathbb{C}$ such that*

$$\mathbf{P}(x^T Ay = f(y)) \geq r^{-1+\epsilon}. \tag{1}$$

Then A contains a rank one submatrix of size at least $(m - O_\epsilon(\frac{r}{\log^6 r})) \times (n - O_\epsilon(\frac{r}{\log^6 r}))$

The same holds true if (1) holds when the entries of y are independently set equal to 0 (with probability 2ρ) or ± 1 (with probability $\frac{1}{2} - \rho$ each). In this case the constant R_0 and the constants implicit in the O notation depend on ρ as well as ϵ.

In particular, this holds for the case where $f(y) = c$ is constant.

Remark 1. Note that we now require the stronger condition that *every* row have many nonzero entries. If this does not hold, we can first expose the x_i corresponding to rows with few nonzero entries, then apply Theorem 5 to the bilinear form on the remaining variables. It follows that the rows of A having many nonzero entries must correspond almost entirely to a rank one submatrix.

Remark 2. Although for most of our applications we will be taking $f(y)$ to be either a fixed constant or a linear form in the y_i, the above theorem holds for arbitrary f. Taking transposes, we see that the same conclusion holds if the probability $x^T Ay = g(x)$ is large, only now we require that A have many nonzero entries in each *column*.

Remark 3. The -1 in the exponent is sharp. If A is a small integer matrix, then $x^T Ay$ will typically be on the order of n in absolute value, so by the pigeonhole principle some value is taken on with probability $\Omega(n^{-1})$. However, a randomly chosen such A will with high probability not have rank one submatrices of size larger than $O(\log n)$.

In terms of the original bilinear form, a rank one submatrix corresponds to a form which factors completely as $x^T Ay = g(x)h(y)$. Theorem 5 states that any bilinear form with sufficiently large concentration probability is highly structured in the sense that it can be made into a bilinear form which factors by setting only a small portion of the variables equal to 0 and considering the resultant form on the remaining variables (setting a variable equal to 0 is equivalent to deleting a row or column from the coefficient matrix)..

We next turn our attention to quadratic forms $x^T Ax$, where x is again random. Here we first aim to show

Theorem 6. *Let A be an $n \times n$ symmetric matrix of complex numbers such that every row of A has at least r nonzero entries, where $r \geq \exp((\log n)^{1/4})$. Let x be a vector of n independent ± 1 symmetric random variables. Let $\delta > 0$ be fixed. Then*

$$\sup_{L,c} \mathbf{P}(x^T Ax = L(x) + c) = O_\delta(r^{-1/2+\delta}),$$

where the supremum is taken over all linear forms $L(x)$ and all c (that is to say all affine functions). In particular, the above bound holds for the case where $L(x)$ is identically 0.

We will then remove the assumption that *every* row of A have many nonzero entries, obtaining the following corollary which may be easier to apply in practice

Corollary 1. *Let A be an $n \times n$ symmetric matrix of complex numbers such that at least mn of the entries of A are nonzero, where $m \geq 3\exp((\log n)^{1/4})$. Let L and x be as above. Then for any $\delta > 0$,*

$$\sup_{c} \mathbf{P}(x^T Ax = L(x) + c) = O_\delta(m^{-1/2+\delta}).$$

Remark 4. Again the 1/2 is sharp, as can be seen from the form
$(x_1 + \cdots + x_n)(x_1 + \cdots + x_m)$.

The original motivation for studying this problem came in the author's study with Tao and Vu [2] of the singularity probability of random symmetric matrices, whose determinant can be viewed as a quadratic form in the entries of any particular row or column of the matrix in question. By replacing Corollary 4.5 in that paper with Theorem 6, we immediately obtain the following improvement to the results in that paper ([2] had the below corollary with 1/2 replaced by 1/8):

Corollary 2. *Let Q_n be an $n \times n$ symmetric matrix whose entries are independent ± 1 symmetric random variables. Then the probability that Q_n is singular is $O(n^{-1/2+\epsilon})$ for any $\epsilon > 0$.*

Remark 5. Recent preprints of Nguyen [13] and Vershynin [20] have improved this bound further via more involved analysis of the inverse Quadratic Littlewood-Offord problem, with the best current bound being $2\exp(-n^c)$ (for a fixed $c < 1$) from

[20]. However, these bounds remain far weaker than the corresponding results in the non-symmetric case.

Weaker versions of Theorem 6 (with $\frac{1}{2}$ replaced by $\frac{1}{8}$ and $\frac{1}{4}$, respectively) were proved as a consequence of Theorem 4 in [2] and [3]. The improvement in the bound will come from a combination of Theorem 5 and the use of a probabilistic variant of the Szemerédi-Trotter theorem.

We will prove Theorem 5 in the next section, and the proof of Theorem 6 and Corollary 1 will come in the following section. The remainder of the paper will be devoted to conjectured extensions of both results.

3. The Proof of Theorem 5

As in the proof of Theorem 4, we begin by dividing the vectors y into two classes (implicitly depending on the value of r) based on how many coordinates of Ay are equal to 0.

Definition 1. A vector y is **typical** if at least $r^{1-\frac{\epsilon}{4}}$ entries of Ay are nonzero. Otherwise it is atypical.

Theorem 5 is an immediate consequence of the following two lemmas.

Lemma 1. *If A is a matrix satisfying the hypotheses of Theorem 5 and $\mathbf{P}(x^T Ay = f(y) \wedge y \text{ is typical }) \geq \frac{1}{2}r^{-1+\epsilon}$, then the conclusions of Theorem 5 hold.*

Lemma 2. *If A is a matrix satisfying the hypotheses of Theorem 5 and $\mathbf{P}(y \text{ is atypical }) \geq \frac{1}{2}r^{-1+\epsilon}$, then the conclusions of Theorem 5 hold.*

Remark 6. If we consider a form which factors perfectly as $x^T Ay = g(x)h(y)$, then the hypothesis of Lemma 1 corresponds to the case where $g(x)$ is very structured (concentrated on a single value with probability close to $r^{-1/2}$), while that of Lemma 2 corresponds with the same property holding for $h(y)$.

We will examine each lemma in turn.

3.1. The proof of Lemma 1.

We will assume throughout this section that A is a matrix such that

$$\mathbf{P}(x^T Ay = f(y) \wedge y \text{ is typical }) \geq \frac{1}{2}r^{-1+\epsilon}.$$

It follows from Theorem 1 that for any y_0 which is typical we have

$$\mathbf{P}_x(x^T Ay_0 = f(y_0)) \leq r^{-\frac{1}{2}+\frac{\epsilon}{8}}. \tag{2}$$

Our argument will go roughly as follows: Under our assumptions, we know that there must be many typical y_0 for which (2) is not too far from equality. By

Theorem 3, we know that for such y_0 the coordinates of Ay_0 must be very highly structured, in the sense that all of them except for a small exceptional set must lie in not too long an arithmetic progression.

The difficulty is that the exceptional sets in Theorem 3 may be different for different y_0. However, there will still be many "small" (of size much smaller than n) sets of coordinates which will lie entirely outside the exceptional set for most y. We will show that such sets correspond to small collections of rows in A which are very close to being multiples of each other, and then aggregate those collections to find our rank one submatrix. We now turn to the details.

We will make use of the following (truncated) quantitative description of how embeddable a small group of complex numbers is in a short arithmetic progression, which can be thought of as a variant of the essential LCD used in [14].

Definition 2. The **commensurability** of a $k-$tuple $(a_1, \dots, a_k)$ of complex numbers is defined by

$$\mathrm{Comm}(a_1, \dots, a_k) = \max\{r^{-\frac{1}{2}+\frac{\epsilon}{4}}, \frac{1}{R}\},$$

where R is the length of the shortest arithmetic progression containing 0 and every a_i simultaneously.

As with the typical/atypical classification before, the commensurability here depends on the particular value of r.

For example, if $a \leq b$ are positive integers, then, up to the truncation at $r^{-\frac{1}{2}+\frac{\epsilon}{4}}$, $\mathrm{Comm}(a,b) = \frac{b}{GCD(a,b)}$. Also, if $(a_1, a_2, \dots, a_k)$ are all drawn from an arithmetic progression of length q containing 0, we are trivially guaranteed that $\mathrm{Comm}(a_1, \dots, a_k)$ is at least $\frac{1}{q}$. We next characterize the "small sets" of coordinates mentioned above in terms of this commensurability.

Definition 3. A $k-$tuple $(v_1, v_2, \dots, v_k)$ of vectors in $\mathbb{C}^n$ is **neighborly** if

$$\mathbf{E}_y \, \mathrm{Comm}(v_1^T y, v_2^T y, \dots, v_k^T y) \geq \frac{1}{6} r^{-\frac{1}{2}+\frac{5\epsilon}{8}}.$$

Fix $k_0 := \log^7 r$. Our next lemma states that the number of neighborly tuples is quite large:

Lemma 3. *Let A satisfy the hypotheses of Theorem 5 and Lemma 1. Then for $k \leq k_0$, there are at least $m^k(1 - \frac{r^{1-\frac{\epsilon}{8}}}{m})$ neighborly $k-$tuples such that each v_i^T is a row of A.*

The proof of this lemma will be deferred to section 3.4. Our next goal will be to translate the neighborliness of a tuple into structural information about the corresponding rows of A. One natural way in which a tuple can be neighborly is if the rows in A are themselves small multiples of each other, in which case the corresponding coordinates of Ay will *always* be small multiples of each other. Our next lemma states that every neighborly tuple is in some sense close to this example.

Lemma 4. *Let $k \leq k_0$, and let $(v_1, v_2, \dots, v_k)$ be neighborly. Then there are unique complex numbers $d_2, \dots, d_k$ and sets $S_2, \dots, S_k$ of coordinates such that*

- *For each j, $v_1 = d_j v_j$ on all coordinates outside of S_j.*
- *$\prod_{j=2}^{k} |S_j \backslash \bigcup_{i=2}^{j-1} S_i|_1 = O_\epsilon(r^{1-\frac{5\epsilon}{4}})$, where $|S|_1 = 1$ if S is empty and $|S|_1 = \max\{|S|, 4\}$ otherwise.*

Remark 7. This product is dependent on the ordering of the v_i, even though the hypotheses of the lemma are independent of that ordering. The upper bound on the product holds for any ordering of the vectors.

What is important here is that not only does each row differ in only a few places from being a multiple of the first row in the tuple (the exceptional sets are of size $o(r)$), but also that the exceptions will tend to occur in the same columns. This latter fact will help keep the exceptional sets from growing too quickly when we attempt to examine many neighborly tuples at once. We will defer the proof of this lemma to sections 3.5-3.8.

Together, the above two lemmas state that the matrix A must have a great deal of local structure, in the sense that many not-too-large collections of rows are very close to being multiples of each other. Our goal will now be to combine these into a single global structure. Using Lemmas 3 and 4, we will be able to prove the following weakened version of Theorem 5, which allows the number of rows which are not in the rank one submatrix to be proportional to m instead of r.

Lemma 5. *If A satisfies the hypotheses of Theorem 5, then A contains a rank one submatrix of size $(m - O_\epsilon(\frac{m \log\log r}{\log^6 r})) \times (n - O_\epsilon(r^{1-\frac{5\epsilon}{4}}))$.*

In the following sections we will first prove Lemma 5 assuming the validity of Lemmas 3 and 4, then leverage that result into the stronger bound required by Theorem 5. We will finish the proof of Lemma 1 by proving Lemmas 3 and 4.

3.2. **The proof of Lemma 5 assuming Lemmas 3 and 4.**

Motivated by the conclusion of Lemma 4, we make the following definition:

Definition 4. Let $V = \{v_1, \dots, v_k\}$ be an (ordered) neighborly $k-$tuple. The **score** of V is given by

$$\text{Score}(V) = \sum_{j=2}^{k} \chi(S_j \not\subseteq \bigcup_{i=2}^{j-1} S_i),$$

where the S_j are as in Lemma 4 and $\chi(E)$ is the indicator function of the event E.

The score is well defined, since the d_j and S_j are unique in that lemma. It also has the following useful properties

- $\text{Score}(v_1, \dots, v_k) \leq \text{Score}(v_1, \dots, v_{k+1})$. Equality holds if and only if $S_{k+1} \subseteq \bigcup_{i=1}^{k} S_i$.

- If $(v_1, \dots, v_k)$ is neighborly, then there can be at most $\log_4(O_\epsilon(r^{1-\frac{5\epsilon}{4}})) < \log r - 1$ different j for which the score increases from $(v_1, \dots, v_j)$ to $(v_1, \dots, v_{j+1})$ (here the stated inequality follows from assumption $r > R_0$) .

For a given (ordered) neighborly k-tuple $V = (v_1, \dots, v_k)$ of rows of A with $k < k_0$, let $S(V)$ be the collection of all rows v of A such that $(v_1, \dots, v_k, v)$ is a neighborly tuple with the same score as V. Note that for any V, all of the rows in $S(V)$ are multiples of v_1 (and thus of each other) except in the coordinates where a prior $d_j v_j$ differed from v_1, and the number of such coordinates is at most

$$|\bigcup_{j=2}^{k} S_j| = \sum_{j=2}^{k} |S_j \backslash \bigcup_{i=2}^{j-1} S_i| = O_\epsilon(r^{1-\frac{5\epsilon}{4}}).$$

The second inequality here follows from Lemma 4 (which implies that the product of the nonzero summands is $O(r^{1-5\epsilon/4})$) and concavity. It follows that we have a rank one submatrix of dimensions $|S(V)| \times (n - O_\epsilon(r^{1-\frac{5\epsilon}{4}}))$. It therefore suffices to show some $S(V)$ is large. Let b be the maximal value of $|S(V)|$ over all neighborly tuples of rows of A of size at most $k_0 - 1$. We count the number of neighborly k_0-tuples in two ways.

Method 1: By Lemma 3, there are at least $m^{k_0}(1 - \frac{r^{1-\frac{\epsilon}{8}}}{m})$ such tuples.

Method 2: We can bound the number of such tuples by first choosing a set J of size $\log r - 1$ of places in which the score is allowed to increase, then restricting our attention only to those tuples whose scores increase only on J. For each j where the score fails to increase from $(v_1, \dots, v_j)$ to $(v_1, \dots, v_{j+1})$, there are at most b choices for v_{j+1}. For each other j, there are at most m choices. It follows that the number of tuples is at most

$$\binom{k_0 - 1}{\log r - 1} m^{\log r} b^{k_0 - \log r} \leq k_0^{\log r} m^{\log r} b^{k_0 - \log r}.$$

Comparing our methods, we have

$$1 - \frac{r^{1-\frac{\epsilon}{8}}}{m} \leq \left(\frac{b}{m}\right)^{k_0 - \log r} k_0^{\log r}.$$

Using the relationship $e^{(-1+o_x(1))x} \leq 1 - x \leq e^{-x}$, we have

$$e^{-(1+o(1))\frac{r^{1-\frac{\epsilon}{8}}}{m}} \leq e^{-(k_0 - \log r)\frac{m-b}{m} + \log r \log k_0}.$$

Taking logs and using the definition of k_0 gives

$$\frac{m-b}{m} \leq \frac{\log r \log k_0 + (1+o(1))\frac{r^{1-\frac{\epsilon}{8}}}{m}}{k_0 - \log r} = O(\frac{\log\log r}{\log^6 r}).$$

It follows that $b \geq m - O(\frac{m \log\log r}{log^6 r})$, so we are done.

3.3. **The proof of Lemma 1, from Lemma 5.** We construct our rank one submatrix using the following procedure. Let A_0 be a rank one submatrix of A of size $(m - O(\frac{m \log\log r}{log^6 r})) \times (n - O(r^{1-\frac{5\epsilon}{4}}))$ (such a matrix is guaranteed to exist by Lemma 5). We initialize $X_1 \subseteq \{x_1, \dots, x_m\}$ to be the variables corresponding to the rows of A_0, and X_2 to be the remaining variables, and X_3 to initially be empty. We also initially set Y_1 to be the variables corresponding to the columns of A_0. We now repeatedly follow the following procedure:

If the matrix corresponding to $(X_1 \cup X_2) \times Y_1$ has rank one, stop. If this is not the case, choose $x_i \in X_1, x_j \in X_2$, and $y_k, y_l \in Y_1$ such that $a_{ik}a_{jl} \neq a_{il}a_{kj}$. Move x_j from X_2 to X_3, and remove y_k and y_l from Y_1.

We can always find the necessary x_i and x_j since the matrix on $X_1 \times Y_1$ will always be a rank one matrix due to our choice of A_0. It remains to check that this procedure in fact terminates after at most $O(\frac{r}{\log^5 r})$ steps, so that the final rank one matrix is sufficiently large. Let us assume to the contrary that this does not occur, meaning that at some point $|X_3| > \frac{r}{\log^5 r}$.

Let S be a set of size r formed by taking $\frac{r}{\log^5 r}$ variables from X_3 and $r - \frac{r}{\log^5 r}$ variables from X_1, and let U be the remaining variables in X. Let $\widetilde{A}$ be the submatrix of A consisting of the rows corresponding to S. We can write

$$x^T A y - f(y) = x_S^T \widetilde{A} y - g(y, x_U),$$

where x_S (resp. x_T) is the vector of variables in S (resp. T). By assumption we have

$$\begin{aligned} \frac{1}{2} r^{-1+\epsilon} &\leq \mathbf{P}(x^T A y = f(y)) \\ &= \mathbf{E}_U(\mathbf{P}_S(x_S^T \widetilde{A} y = g(y, x_U))) \\ &\leq \sup_{x_U} \mathbf{P}_S(x_S^T \widetilde{A} y = g(y, x_U)). \end{aligned}$$

It follows from Lemma 5 that $\widetilde{A}$ must contain a rank one submatrix of size $(r - O(\frac{r \log\log r}{\log^6 r})) \times (n - O(r^{1-\frac{5\epsilon}{4}}))$. Since the number of excluded variables is much smaller than $\frac{r}{\log^5 r}$ (here we again use that r is sufficiently large), there must be a variable $x_j \in X_3$ such that both x_j and the corresponding y_k and y_l are contained in this submatrix, as well as some variable $x_{i'} \in X_1$. However, this is a contradiction, as $a_{i'k}a_{jl} \neq a_{i'l}a_{kj}$. This completes the proof of Lemma 1 modulo the proofs of Lemmas 3 and 4. We next turn to the proof of those two lemmas.

3.4. **The proof of Lemma 3.** We define g_y and D_y as follows:

- If y is atypical, then $g_y = 0$ and $D_y = \{1, \dots, m\}$.
- If y is typical and no arithmetic progression of length at most $r^{\frac{1}{2}-\frac{\epsilon}{4}}$ contains at least $m - r^{1-\frac{\epsilon}{4}}$ of the elements of Ay, then $g_y = r^{-\frac{1}{2}+\frac{\epsilon}{4}}$ and $D_y = \{1, \dots, m\}$.

- Otherwise, let R be an arithmetic progression of minimal length containing 0 and at least $m - r^{1-\frac{\epsilon}{4}}$ elements of Ay. We define $g_y = |R|^{-1}$, and D_y to be those i such that the i^{th} coordinate of Ay is in R.

Note that in this definition the D_y are not uniquely determined. We choose one arbitrarily for each y. Furthermore, by construction, for any y and for any $k-$tuple $\{a_1, \dots, a_k\} \subseteq D_y$ we have $\mathrm{Comm}(a_1, \dots, a_k) \geq g_y$.

By viewing the Inverse Littlewood-Offord Theorem 3 in the "forward" direction we can now obtain the following:

Lemma 6. *For every fixed $\epsilon < \frac{1}{2}$ there is an $r_0 > 0$ such that for all matrices A with $r > r_0$ and all typical y^* we have*

$$\mathbf{P}(x^T A y^* = f(y^*)) \leq r^{-\frac{1}{2}+\frac{3\epsilon}{8}} g_{y^*}.$$

Proof (of Lemma 6): Since by construction $g_{y^*} \geq r^{-\frac{1}{2}+\frac{\epsilon}{4}}$, there is nothing to prove unless the probability in question is at least $r^{-1+\frac{5\epsilon}{8}}$, which we will assume to be the case. Let r_1 be the number of nonzero coefficients of $x^T A y^*$, viewed as a linear form in x, and let $\mathbf{P}(x^T A y^* = f(y^*)) = r_1^{-\frac{1}{2}-\epsilon_0}$. Since y^* is typical, $r_1 \geq r^{1-\frac{\epsilon}{4}}$. In particular, this implies that $\epsilon_0 < \frac{1}{2}$.

Applying Theorem 3 to this form with $\alpha = \frac{\epsilon}{4}$, we see there is an arithmetic progression containing all but $r_1^{1-\frac{\epsilon}{4}}$ coefficients and of length

$$\begin{aligned} r_1^{\epsilon_0+\alpha} &= \frac{r_1^{-\frac{1}{2}+\frac{\epsilon}{4}}}{\mathbf{P}(x^T A y^* = f(y^*))} \\ &\leq \frac{r^{(-\frac{1}{2}+\frac{\epsilon}{4})(1-\frac{\epsilon}{4})}}{\mathbf{P}(x^T A y^* = f(y^*))} \\ &= \frac{r^{-\frac{1}{2}+\frac{3\epsilon}{8}}}{\mathbf{P}(x^T A y^* = f(y^*))} r^{-\frac{\epsilon^2}{16}}. \end{aligned}$$

If follows that $g_{y^*} \geq r^{\frac{1}{2}-\frac{3\epsilon}{8}} \mathbf{P}(x^T A y^* = f(y^*))$ as desired.

■

Taking expectations over all y, we see that

$$\mathbf{P}(x^T A y = f(y) \wedge y \text{ is typical }) \leq r^{-\frac{1}{2}+\frac{3\epsilon}{8}} \mathbf{E}_y(g_y),$$

which combined with the hypothesis of Lemma 1 in turn implies that

$$\mathbf{E}_y(g_y) \geq r^{-\frac{1}{2}+\frac{5\epsilon}{8}} \tag{3}$$

Let Z be the collection of $k-$tuples $\{a_1, \dots, a_k\}$ satisfying

$$\mathbf{E}_y(g_y \chi(\{a_1, \dots, a_k\} \subseteq D_y)) \geq \frac{1}{3} \mathbf{E}_y(g_y) \geq \frac{1}{3}.$$

For any $k-$tuple in Z, we have

$$\begin{aligned}\mathbf{E}_y(\mathrm{Comm}(a_1,\dots,a_k)) &\geq \mathbf{E}_y(\mathrm{Comm}(a_1,\dots,a_k)\chi(\{a_1,\dots,a_k\}\in D_y))\\ &\geq \mathbf{E}_y(g_y\chi(\{a_1,\dots,a_k\}\in D_y))\\ &\geq \frac{1}{3}\mathbf{E}_y(g_y)\geq\frac{1}{3}r^{-\frac{1}{2}+\frac{5\epsilon}{8}}.\end{aligned}$$

Here the first inequality follows from the observation that g_y is always a lower bound for commensurability (since g_y comes from a specific embedding into an arithmetic progression), and the last from (3). In particular, we know that any $k-$tuple in Z is neighborly. It remains to check that $|Z|$ is large.

Since by construction $|D_y|\geq m-r^{1-\frac{\epsilon}{4}}$ for every y, we have

$$\begin{aligned}\mathbf{E}_{a_1,\dots,a_k}\mathbf{E}_y(g_y\chi(\{a_1,\dots,a_k\}\subseteq D_y)) &= \mathbf{E}_y(g_y\mathbf{P}(\{a_1,\dots,a_k\}\in D_y))\\ &\geq (\frac{m-r^{1-\frac{\epsilon}{4}}}{m})^k\mathbf{E}_y(g_y),\end{aligned}$$

where the expectation here is taken with each a_i independent and uniform on $\{1,\dots,m\}$. Combining this with the definition of Z, we have

$$\begin{aligned}|Z|\mathbf{E}_y(g_y)+\frac{\mathbf{E}_y(g_y)}{3}(m^k-|Z|) &\geq (m-r^{1-\frac{\epsilon}{4}})^k\mathbf{E}_y(g_y)\\ &\geq m^k(1-\frac{kr^{1-\frac{\epsilon}{4}}}{m})\mathbf{E}_y(g_y).\end{aligned}$$

Solving the above inequality, we obtain

$$|Z|\geq m^k(1-\frac{3k}{2}\frac{r^{1-\frac{\epsilon}{4}}}{m})\geq m^k(1-\frac{r^{-\frac{\epsilon}{8}}}{m})$$

and we are done.

3.5. **The proof of Lemma 4 for $k=2$.**

We will first prove Lemma 4 in the special case where $k=2$. The case of larger k will then be much simpler, as we can take advantage of the observation that subsets of neighborly vectors are themselves neighborly to in a sense reduce to the $k=2$ case. Let (a,b) be a pair of neighborly vectors. Our goal will be to show that they are very close to being multiples of each other.

We make use of the general fact that for any random variable X taking values between 0 and 1

$$\mathbf{E}(X)=\int_0^1\mathbf{P}(X>u)du=\int_1^\infty\frac{\mathbf{P}(X>\frac{1}{t})}{t^2}dt \tag{4}$$

In our case X will be $\mathrm{Comm}(a^Ty,b^Ty)$, so bounding the right hand side becomes a question of how likely it is for a^Ty and b^Ty to be embeddable in a progression of a given length. We make the following further definitions:

Definition 5. A pair (l_1,l_2) of integers is **degenerate** for the vector pair (a,b) if l_1a and l_2b agree in at least $n-\frac{r}{5}$ positions and at least one of l_1 and l_2 is nonzero.

Note that there is (up to multiples) at most one degenerate pair for (a, b).

We further define for a real number q

$$\begin{aligned} p_{ab}(q) \quad := \quad & \mathbf{P}(\exists \text{ a non-degenerate } (l_1, l_2) \neq (0,0) \in \mathbb{Z} \times \mathbb{Z} \\ & \quad \text{with } l_1 a^T y = l_2 b^T y \text{ and } |l_1|, |l_2| \leq q). \end{aligned}$$

Using these definitions and the definition of Comm(a, b), we have

$$\begin{aligned} r^{-\frac{1}{2}+\frac{5\epsilon}{8}} \quad & \leq \quad \mathbf{E}_y(\text{Comm}(a^T y, b^T y)) \\ & \leq \quad (\int_1^{r^{\frac{1}{2}-\frac{\epsilon}{4}}} \frac{p_{ab}(q)}{q^2} dq) + r^{-\frac{1}{2}+\frac{\epsilon}{4}} \\ & \qquad + \mathbf{P}(\text{ there exists a degenerate } (k_0, l_0) \text{ and } k_0 a^T y = l_0 b^T y). \end{aligned}$$

The middle term on the right hand side is negligible, and in the next subsection we will show that the first term is also small by showing

Lemma 7. *For any positive $\alpha > 0$, any $q < \sqrt{r}$ and any a and b, there is a constant C_α dependent only on α such that $p_{ab}(q) \leq \frac{C_\alpha q}{r^{1/2-\alpha}}$.*

We may without loss of generality assume $C_\alpha > 1$. It follows that for any $0 < \alpha < \frac{1}{2}$, assuming Lemma 7, we have

$$\begin{aligned} \int_1^\infty \frac{p_{ab}(q)}{q^2} dq \quad & \leq \quad \int_1^{r^{1/2-\alpha}} \frac{C_\alpha}{q r^{1/2-\alpha}} dq + \int_{r^{1/2-\alpha}}^\infty \frac{dq}{q^2} \\ & = \quad O_\alpha(r^{-1/2+\alpha} \log r). \end{aligned}$$

By taking α sufficiently close to 0, we see that for large r the contribution from the first term is also $o(r^{-\frac{1}{2}+\frac{5\epsilon}{8}})$.

It follows that the dominant contribution to the expectation must come from the third term. This implies that a degenerate pair $(k_0, l_0) \neq (0, 0)$ exists, and that we furthermore must have

$$\mathbf{P}(k_0 a^T y = l_0 b^T y) \geq \frac{1}{12} r^{-\frac{1}{2}+\frac{5\epsilon}{8}}.$$

It follows by the Theorem 1 that the linear form $(k_0 a^T - l_0 b^T)y$ must have $O(r^{1-\frac{5\epsilon}{4}})$ nonzero coefficients. Note that k_0 cannot equal 0, since otherwise the form would have at least r nonzero coefficients (recall that b has by assumption at least r nonzero entries). It follows that there are at most $(r^{1-\frac{5\epsilon}{4}})$ places where a differs from $\frac{l_0}{k_0} b$, meaning that we can take $d_2 = \frac{l_0}{k_0}$ in Lemma 4.

This choice of d_2 is unique, for if $d_2 \neq d_2'$ then $d_2 v_2$ and $d_2' v_2$ disagree in every coordinate in which $v_2 \neq 0$. There are at least r such coordinates by assumption, which would make it impossible for both $d_2 v_2$ and $d_2' v_2$ to agree with v_1 in all but $O(r^{1-5\epsilon/4})$ places.

3.6. **The proof of Lemma 7.** It suffices to prove the following:

Lemma 8. *Let $a_1, \dots, a_n$ and $b_1, \dots, b_n$ be fixed (real or complex) constants such that for each i at least one of a_i and b_i is non-zero. Let $x_1, \dots, x_n$ be independent ± 1 symmetric random variables. Let E_q be the event that there exist u and v satisfying*

- $|u|, |v| \leq q$.
- *There are at least $\frac{n}{10}$ different i for which $va_i \neq ub_i$.*
- $v \sum a_i x_i = u \sum b_i x_i$.

Then for any $\alpha > 0$ and any $1 \leq q < \sqrt{n}$,

$$\mathbf{P}(E_q) = O(\frac{q}{\sqrt{n}} n^{\alpha}),$$

where the constant implicit in the O notation is as n tends to infinity and may depend on α.

We will throughout assume that both q and n are tending to infinity. By utilizing a Freiman isomorphism of order $2n^2$ (see for example [16], Lemma 5.25), we may assume that the a_i and the b_i are all real integers. We may furthermore without loss of generality assume for every i either b_i is positive or $b_i = 0$ and a_i is positive.

Let ℓ be a positive integer satisfying $\ell > \frac{1}{\alpha}$. We define $L_0 = 1$ and for $1 \leq j \leq \ell$, we define

$$L_j = \sup_{(c,d) \in \mathbb{Z}^2} |\{(i_1, \dots, i_j) : a_{i_1} + \dots + a_{i_j} = c \wedge b_{i_1} + \dots + b_{i_j} = d\}|.$$

Here the tuples in the definition of L_j may contain repeated elements. Clearly $1 \leq L_j \leq n^j$, and by treating i_j as fixed we furthermore see that
$L_{j-1} \leq L_j \leq nL_{j-1}$. This implies that one of the following two cases must hold

- There is a j between 1 and ℓ for which $L_j \geq nq^{-\frac{2}{2\ell+1}} L_{j-1}$.
- $L_\ell \leq n^\ell q^{-\frac{2\ell}{2\ell+1}}$.

We handle each case separately.

Case 1: $L_\ell \leq n^\ell q^{-\frac{2\ell}{2\ell+1}}$. Here we will make use of the following result of Halász (implicit in [6], see also [16]):

Theorem 7. *Let $\ell > 0$ be fixed, and let $a_1, \dots, a_n$ be nonzero (real or complex) coefficients. Let R_ℓ be the number of $2\ell-$tuples $(i_1, \dots, i_\ell, j_1, \dots, j_\ell)$ for which $a_{i_1} + \dots + a_{i_\ell} = a_{j_1} + \dots + a_{j_\ell}$. Then for x_i independent ± 1 symmetric variables,*

$$\mathbf{P}(\sum_{i=1}^{n} a_i x_i = c) = O(n^{-2\ell - 1/2} R_\ell).$$

Given two integers u and v and a such a $2\ell-$tuple, denote by $R(u, v, i_1, \dots, i_\ell, j_1, \dots, j_\ell)$ the property that

$$v(a_{i_1} + \dots + a_{i_\ell} - a_{j_1} - \dots - a_{j_\ell}) = u(b_{i_1} + \dots + b_{i_\ell} - b_{j_1} - \dots - b_{j_\ell}).$$

Combining the Halász bound in Theorem 7 above and the union bound, we can write

$$\begin{aligned}\mathbf{P}(E_q) &\leq \sum_{(u,v)} \mathbf{P}(\sum_{i=1}^{n}(a_i v - b_i u)x_i = 0)\\ &= O(n^{-2\ell-1/2}) \sum_{(u,v)} \sum_{\substack{(i_1,\dots,i_\ell)\\(j_1,\dots,j_\ell)}} \chi(R(u,v,i_1,\dots,i_\ell,j_1,\dots,j_\ell))\\ &= O(n^{-2\ell-1/2}) \sum_{\substack{(i_1,\dots,i_\ell)\\(j_1,\dots,j_\ell)}} \sum_{(u,v)} \chi(R(u,v,i_1,\dots,i_\ell,j_1,\dots,j_\ell)),\end{aligned}$$

where the sum is taken over all pairs (u,v) such that $|u|,|v| \leq q$, either $u > 0$ or $u = 0$ and $v > 0$, $GCD(u,v) = 1$, and at least $\frac{n}{10}$ different i satisfy $ub_i \neq va_i$. This last assumption guarantees that the linear form in the first inequality has at least $0.1n$ nonzero coefficients for every (u,v) we are summing over, so that the Halász bound in Theorem 7 above will be sufficiently strong.

In the final term in the above bound, the inner summand is at most 1 unless

$$(a_{i_1}, b_{i_1}) + (a_{i_2}, b_{i_2}) + \cdots + (a_{i_\ell}, b_{i_\ell}) = (a_{j_1}, b_{j_1}) + (a_{j_2}, b_{j_2}) + \cdots + (a_{j_\ell}, b_{j_\ell}), \tag{5}$$

since otherwise u and v are uniquely determined by the values of the a and b (recall that u and v are assumed relatively prime). If (5) holds then any relatively prime (u,v) contributes 1 to the inner summand, so we can bound the inner sum by q^2. Since the outer sum contains $O(n^{2\ell})$ terms and by assumption (5) has at most $L_\ell n^\ell$ solutions, it follows that

$$\mathbf{P}(E_q) = O(q^2 n^{-\ell-1/2} L_\ell + n^{-1/2}),$$

which by our assumptions on L_ℓ and ℓ is $O(\frac{q^{1+\frac{1}{2\ell+1}}}{\sqrt{n}}) = O(\frac{qn^\alpha}{\sqrt{n}})$.

Case 2: $L_j \geq \frac{n}{q^{\frac{2}{2\ell+1}}} L_{j-1}$. Let (c,d) be a (fixed and non-random) pair which is the sum of L_j different $j-$tuples. By our assumption on the b_i, and a_i, we know that either d is positive or $d = 0$ and c is positive. In particular, at least one of c and d must be nonzero.

We know that each variable (a_i, b_i) can be involved in at most jL_{j-1} different $j-$tuples which sum to (c,d), since there are j locations for (a_i, b_i) in the tuple and at most L_{j-1} choices for the remaining $j-1$ elements that add to $(c - a_i, c - b_i)$. Since each $j-$tuple adding to (c,d) intersects at most $j^2 L_{j-1}$ other such $j-$tuples, it follows that we can form a collection S of *disjoint* $j-$tuples summing to (c,d) satisfying

$$|S| \geq \frac{L_j}{j^2 L_{j-1}} \geq \frac{n}{j^2 q^{\frac{2}{2\ell+1}}}.$$

Define a $j-$tuple $(i_1, \dots, i_j)$ to be **agreeable** if $x_{i_1} = x_{i_2} = \cdots = x_{i_j}$. Note that each tuple has a constant probability 2^{1-j} of being agreeable. Let S' be the collection of tuples in S which are agreeable, and let B be the event that $|S'| \geq 2^{-j}|S|$. We have

$$\mathbf{P}(E_q) \leq \mathbf{P}(\neg B) + \mathbf{P}(E_q \wedge B).$$

Note that the agreeability of each tuple in S is an independent event due to our assumption that the tuples are disjoint. Since the expected number of agreeable tuples is $2^{1-j}|S|$, it follows from Hoeffding's inequality [7] that

$$\mathbf{P}(\big||S'| - 2^{1-j}|S|\big| > \lambda|S|) \leq e^{-2\lambda^2|S|}.$$

Taking $\lambda = 2^{-j}$, we see that $\mathbf{P}(\neg B) = o(n^{-1/2})$. We therefore focus on the second term.

To bound $\mathbf{P}(E_q \wedge B)$, we will expose the variables by first exposing S', then exposing the value of all the variables not involved in a tuple in S'. We will then finally expose the values of the variables in S'.

We have for any tuple that

$$\mathbf{P}(\sum_{k=1}^{j}(a_{i_k}, b_{i_k}) = (c, d)|(i_1, \dots, i_j) \text{ agreeable }) = 1/2,$$

and the same for $(-c, -d)$. It follows that, treating the set S' and the value of x_j for variables not in S' as fixed,

$$\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} b_i x_i} \sim \frac{c\sum_{j=1}^{|S'|} y_j + z_1}{d\sum_{j=1}^{|S'|} y_j + z_2},$$

where z_1 and z_2 are fixed constants and the y_i are independent symmetric ± 1 variables. Let y equal the sum of the y_i. By Hoeffding's inequality again, we know that

$$\mathbf{P}(|y| > \sqrt{n}\log n) \leq n^{-\log n/2},$$

so it suffices to estimate the contribution to $\mathbf{P}(E_q)$ from those y which are at most $\sqrt{n}\log n$ in absolute value. By the linear Littlewood-Offord Theorem 1, we know that each such y occurs with probability at most $n^{-1/2}$, and we will soon show:

Lemma 9. *Let $\alpha > 0$ be any fixed parameter. Let $n \geq N_0(\alpha)$ be an integer, and let w_1, w_2, w_3, w_4 be real numbers such that $w_1 w_4 \neq w_2 w_3$. Then for any $1 \leq q \leq n$, there are at most qn^{α} integers $z \in \{-n, \dots, n\}$ such that*

$$h(z) := \frac{w_1 z + w_2}{w_3 z + w_4}$$

has height at most q (where the height of a rational number h is defined to be the minimum of the absolute values of the numerator and denominator when h is written in lowest terms).

Assuming Lemma 9 to be true, we know that for fixed z_1, z_2 the probability that this fraction can be written as $\frac{u}{v} \neq \frac{c}{d}$ is at most $\frac{qn^{\frac{1}{3\ell}}}{\sqrt{|S'|}}$. Taking expectations over all z_1, z_2, S' and using our bounds on $|S'|$ under the assumption that B holds gives that

$$\mathbf{P}(E_q \wedge B) \leq \frac{C_j q^{1+\frac{1}{2\ell+1}} n^{\frac{1}{3\ell}}}{\sqrt{n}} + \mathbf{P}\Big(\frac{\sum a_i}{\sum b_i} = \frac{c}{d} \wedge da_i - ca_i \neq 0 \text{ for} \frac{n}{10} \text{ different } i\Big).$$

The second term on the right side corresponds to a linear form with $\frac{n}{10}$ nonzero coefficients, so is $O(n^{-1/2})$. The first term on the right side is at most $\frac{qn^{1/\ell}}{\sqrt{n}}$, since q is assumed to be at most n. Since $1/\ell$ is by construction at most α, this is at most $\frac{qn^\alpha}{\sqrt{n}}$. Again the result follows.

It remains to prove Lemma 9.

3.7. **The proof of Lemma 9.** [1]

We may without loss of generality assume that $|w_1| \geq |w_3|$. We will further assume without loss of generality that no prime divides all of the w_i.

Let $\Delta = |w_1w_4 - w_2w_3| > 0$. Note that any common divisor of $w_1z + w_2$ and $w_3z + w_4$ is also a common divisor of $|w_1(w_3z + w_4) - w_3(w_1z + w_2)| = \Delta$. Let $\tau(\Delta)$ be the number of divisors of Δ. We will consider two cases.

Case 1: $\tau(\Delta) < n^{\alpha/2}$. For $0 \leq i \leq (1 + \alpha)\log_2 n$, let S_i denote the set of $z \in \{-n, \dots, n\}$ such that $|w_1z + w_2| \in [2^i, 2^{i+1}]$. It is clear that each S_i lies in the union of two intervals, each of which has size at most 2^i. For any $z \in S_i$ such that $h(z)$ has height at most q, it must be the case that w_1z+w_2 shares a divisor v with $w_3z + w_4$ and Δ such that $v > 2^i/q$. We next claim that for any given v, there are not many v for which this can occur, as:

Claim 1. *If $v|GCD(w_1z_1 + w_2, w_3z_1 + w_4)$ and $v|GCD(w_1z_2 + w_2, w_3z_2 + w_4)$, then $v|(z_1 - z_2)$.*

Proof Let p be a (fixed) prime dividing v, and let p^m be the largest power of p dividing v. If p does not divide w_1, then p^m must divide $z_1 - z_2$, since v divides $(w_1z_1 + w_2) - (w_1z_2 + w_2) = w_1(z_1 - z_2)$. Similarly, either p^m divides $z_1 - z_2$ or p also divides w_3. However, p cannot divide both w_1 and w_3, for it would then follow that p also divided $(w_1z_1 + w_2) - w_1z_1 = w_2$ and, similarly, w_4, violating our assumption that the w_i share no common factor. Therefore it must be the case that $p^m|(z_1 - z_2)$. But this is true for any prime, so we are done. ■

It follows that for a given v, there are at most $2^{i+1}/v$ choices of z for which v provides the required cancellation. Adding up over all v, the number of $z \in S_i$ which lead to a height of at most q is at most

$$\sum_{\substack{v|\Delta \\ v>2^i/q}} \left(\frac{2^i}{v} + 1\right) \leq (q+1)\tau(\Delta) \leq (q+1)n^{\alpha/2}.$$

[1] Many of the key ideas in the proof of Lemma 9 are due to Ernie Croot

Adding up over all S_i, we see that the lemma holds in this case.

Case 2: $\tau(\Delta) \geq n^{\alpha/2}$. In this case it follows from classical number theoretic bounds on the number of divisors of an integer that $\Delta > n^{\omega(n)}$ for some $\omega(n)$ tending to infinity with n. We will show that if Δ is this large it is impossible to have even 3 different z in $[-n, n]$ with sufficiently small height.

Suppose to the contrary that there are z_1, z_2, and z_3 such that each $\frac{w_1 z_i + w_2}{w_3 z_i + z_4}$ has height at most q. It then follows that

$$\frac{w_1 z_1 + w_2}{w_3 z_1 + w_4} - \frac{w_1 z_3 + w_2}{w_3 z_3 + w_4} = \frac{\Delta(z_1 - z_3)}{(w_3 z_1 + w_4)(w_3 z_3 + w_4)} \tag{6}$$

and

$$\frac{w_1 z_2 + w_2}{w_3 z_2 + w_4} - \frac{w_1 z_3 + w_2}{w_3 z_3 + w_4} = \frac{\Delta(z_2 - z_3)}{(w_3 z_2 + w_4)(w_3 z_3 + w_4)} \tag{7}$$

have height at most $2q^2$. Taking their ratio, we see that the height of $\frac{(z_1 - z_3)(w_3 z_2 + w_4)}{(z_2 - z_3)(w_3 z_2 + w_4)}$ is at most $4q^4$. Since $|z_1 - z_3|$ and $|z_2 - z_3|$ are each at most $2n$, there must be a solution to

$$r_1(w_3 z_2 + w_4) = r_2(w_3 z_1 + w_4)$$

with $|r_1|, |r_2| \leq 8q^4 n$. Rearranging this equation as

$$\frac{w_3}{w_4} = \frac{r_2 - r_1}{r_1 z_2 - r_2 z_1},$$

we see that w_3/w_4 has height at most $16q^4 n^2$. An identical argument (replacing each fraction on the left hand side of (6) and (7) by its reciprocal) shows the same holds for w_1/w_2. In other words, we can write

$$\frac{w_1 z + w_2}{w_3 z + w_4} = \frac{W}{W'} \left(\frac{w_1' z + w_2'}{w_3' z + w_4'} \right),$$

where each $|w_i'| \leq 16q^4 n^2$, and $W = \frac{w_1}{w_1'}$. Here $W' = \frac{w_3}{w_3'}$ are relatively prime due to our assumption that the w_i share no common factor. For this fraction to ever have height at most q, it must be the case that $|W| \leq 32q^5 n^3$, and the same for W'. But this contradicts our lower bound on Δ.

3.8. **The proof of Lemma 4 for** $k > 2$**.** Let $(v_1, \dots, v_k)$ be a neighborly tuple. We first modify the definition of Commensurability slightly, writing

$$\text{Comm}^*(a_1, \dots, a_k) = \text{Comm}(a_1, \dots, a_k)\chi(\prod_{i=1}^{k} a_i \neq 0).$$

We have by Theorem 1 and the fact the Commensurability is always at most 1 that

$$\begin{aligned} \mathbf{E}_y(\text{Comm}^*(v_1^T y, \dots, v_k^T y)) &\geq \mathbf{E}_y(\text{Comm}(v_1^T y, \dots, v_k^T y)) - \mathbf{P}(\text{ some } v_i^T y = 0) \\ &\geq \mathbf{E}_y(\text{Comm}(v_1^T y, \dots, v_k^T y)) - kr^{-1/2} \\ &\geq \frac{1}{12} r^{-\frac{1}{2} + \frac{5\epsilon}{8}}. \end{aligned}$$

The advantage to this modified commensurability is that we have the relationship

$$\mathrm{Comm}^*(a_1,\dots,a_k) \geq \frac{1}{R} \Leftrightarrow \frac{a_1}{z_1} = \frac{a_2}{z_2} \cdots = \frac{a_k}{z_k}$$

for some integers $z_1,\dots,z_k$ which are at most R in absolute value.

As in the $k=2$ case, we have

$$\begin{aligned} \mathbf{E}_y(\mathrm{Comm}^*(v_1^T y,\dots,v_k^T y)) &\leq (\int_1^{r^{\frac{1}{2}-\frac{\epsilon}{4}}} \frac{p_v(q)}{q^2} dq) + r^{-\frac{1}{2}+\frac{\epsilon}{4}} \qquad (8) \\ &\quad +\mathbf{P}(\frac{v_1^T y}{l_1} = \frac{v_2^T y}{l_2} = \cdots = \frac{v_k^T y}{l_k} \text{ for a degenerate } l), \end{aligned}$$

where

$$p_v(q) := \mathbf{P}(\exists l = (l_1,\dots,l_k) : l \text{ is non-degenerate } \wedge \frac{a_i^T y}{l_i} \text{all equal } \wedge |l_i| \leq q),$$

and a $k-$tuple $(l_1,\dots,l_k)$ is degenerate if (l_i,l_j) is degenerate for (v_i,v_j) for every i and j. Note that a given $(v_1,\dots,v_k)$ again has (up to multiples) only one degenerate l.

It follows from Lemma 7 that for any particular (i,j), the contribution to $p_v(q)$ from those tuples where (l_i,l_j) is nondegenerate is $O(\frac{q}{r^{1/2-\alpha}})$ for any α. Adding up over all pairs, it follows that $p_v(q) = O(\frac{k^2 q}{r^{1/2-\alpha}})$. As in the $k=2$ case, we now have

$$\int_1^{r^{\frac{1}{2}-\frac{\epsilon}{4}}} \frac{p_v(q)}{q^2} dq = O(k^2 r^{-1/2+\alpha}\log r) = o(r^{-\frac{1}{2}+\frac{5\epsilon}{8}}).$$

by taking α to be sufficiently small. Again the contributions from the first two terms on the right hand side of (8) are small, so the last term must be large, that is to say

$$\mathbf{P}(\frac{v_1^T y}{l_1} = \frac{v_2^T y}{l_2} = \cdots = \frac{v_k^T y}{l_k}) \geq \frac{1}{14} r^{-\frac{1}{2}+\frac{5\epsilon}{8}}. \qquad (9)$$

Let $d_j = \frac{l_1}{l_j}$, and S_j be the places where v_1 differs from $d_j v_j$. We can rewrite the event on the left hand side of (9) as the system

$$\begin{aligned} \sum_{i\in S_2} (d_2 v_2(i) - v_1(i))x_i &= 0 \\ \sum_{i\in S_3\backslash S_2} (d_3 v_3(i) - v_1(i))x_i &= -\sum_{i\in S_2} (d_3 v_3(i) - v_1(i))x_i \\ \vdots \quad & \quad \vdots \\ \sum_{\substack{i\in S_k \\ i\notin S_2\cup\cdots\cup S_{k-1}}} (d_k v_k(i) - v_1(i))x_i &= -\sum_{i\in S_2\cup\cdots\cup S_{k-1}} (d_k v_k(i) - v_1(i))x_i. \end{aligned}$$

We now successively expose the variables in $S_j\backslash(S_2\cup\dots S_{j-1})$ for each j and examine each equation in turn.

After we expose the variables in S_2, the probability that the first equation above holds is at most $|S_2|_1^{-1/2}$ by the Linear Littlewood-Offord Theorem 1. We now treat the variables in S_2 as fixed, meaning that the right hand side of the second equation above is constant, and expose those in $S_3 \backslash S_2$. For any particular value of the variables in S_2, it again follows from Theorem 1 that the probability that the second equation holds is at most $|S_3 \backslash S_2|_1^{-1/2}$. Continuing onwards through the entire system, we have that the probability that the above system holds is at most

$$\prod_{j=2}^{k} |S_j \backslash \bigcup_{i=1}^{j-1} S_i|_1^{-1/2}.$$

The lemma follows by combining this with (9). Uniqueness follows as in the $k = 2$ case: Since the conclusion of the Lemma implies that each S_i is still $O_\epsilon(r^{1-5\epsilon/4})$, and it is impossible for $d_j v_j$ and $d'_j v_j$ to each disagree with v_1 in so few places unless $d_j = d'_j$.

3.9. The proof of Lemma 2.

This proof will follow along very similar lines to that of Lemma 1.

Again we let $k_0 := \lfloor \log^7 r \rfloor$, and the argument will make use of the following analogue of neighborliness:

Definition 6. A tuple $(v_1, \dots, v_k)$ of vectors is **friendly** if

$$\mathbf{P}(v_1^T y = v_2^T y = \dots = v_k^T y = 0) \geq \frac{1}{6} r^{-1+\epsilon}.$$

We again have that there are many friendly $k-$tuples.

Lemma 10. *Let $k \leq k_0$. Under the hypotheses of Lemma 2, there are at least $m^k(1 - \frac{r^{1-\frac{\epsilon}{4}}}{m})$ friendly $k-$tuples such that each v_i^T is a row of A.*

We also claim that friendly tuples exhibit a similar structure as neighborly ones:

Lemma 11. *Let $k \leq k_0$, and let $(v_1, \dots, v_k)$ be friendly. Then there are unique complex numbers d_j such that if S_j denotes the places where v_1 differs from $d_j v_j$, then*

$$\prod_{j=2}^{k} |S_j \backslash \bigcup_{i=2}^{j-1} S_i|_1 \leq 2r^{1-2\epsilon}.$$

(Recall that $|S|_1$ is defined to be 1 if S is empty, and $\max\{|S|, 4\}$ otherwise).

The proof of Lemma 2 from these two lemmas is exactly the same as that of Lemma 1 from Lemmas 3 and 4. We will therefore focus on the proofs of the two lemmas, which will again turn out to be similar to the proofs of the corresponding lemmas for neighborly tuples.

3.10. **The proof of Lemma 10.** Let A be such that y is atypical with probability at least $\frac{1}{2}r^{-1+\epsilon}$. We define Z to be those $k-$tuples satisfying

$$\mathbf{P}(v_1^T y = v_2^T y = \cdots = v_k^T y = 0 \wedge y \text{ atypical}) \geq \frac{1}{3}\mathbf{P}(y \text{ atypical}).$$

The hypotheses of Lemma 2 mean that the above equation guarantees that every tuple in Z is friendly. Now consider a tuple $(v_1, \dots, v_k; y)$ where the v_i^T are chosen randomly (with repetition allowed) from the rows of A and y is uniform and random. We estimate the probability that y is atypical and $v_j^T y = 0$ for every j in two different ways.

Method 1: For any atypical y, there are at least $(m - r^{1-\frac{\epsilon}{4}})^k$ choices for the tuple. It follows that the probability is at least

$$\frac{(m - r^{1-\frac{\epsilon}{4}})^k}{m^k}\mathbf{P}(y \text{ atypical }).$$

Method 2: We first choose the $k-$tuple, then bound the probability that y works based on whether or not the tuple is in Z. Doing this gives that the probability is at most

$$\frac{1}{m^k}(|Z| + \frac{1}{3}(m^k - |Z|))\mathbf{P}(y \text{ atypical }).$$

The result follows by comparing the bounds from the two methods, along with the bound

$$(m - r^{1-\frac{\epsilon}{4}})^k \geq m^k(1 - \frac{kr^{1-\frac{\epsilon}{4}}}{m}) \geq m^k(1 - \frac{r^{1-\frac{\epsilon}{4}+o(1)}}{m}).$$

Remark 8. The lower bound obtained on $|Z|$ above is independent of the probability that y is atypical. However, we need the hypotheses of Lemma 2 to be able to say that the tuples in Z are friendly.

3.11. **The proof of Lemma 2.** We first note that for any j, we can view the system $v_1^T y = v_j^T y$ as a single vector equation $\sum_i w_i y_i = 0$ in $\mathbb{C}^2$, where $w_i = \langle v_1(i), v_j(i)\rangle$. Since by assumption this equation is satisfied with probability $\frac{1}{6}r^{-1+\epsilon}$, it follows from the 2-dimensional Theorem 2 of Halász that there must be a 1-dimensional subspace containing all but $O(r^{1-\epsilon})$ of the w_i. In terms of the v_j, this says that for each j there is a multiple of v_j differing from v_1 in at most $r^{1-\epsilon}$ places. We will take those multiples to be our d_j, and S_j to be the places they differ.

The relationship $v_1^T y = v_2^T y = \cdots = v_k^T y = 0$ is equivalent to the system

$$\begin{aligned}
\sum_{i \in S_2} (d_2 v_2(i) - v_1(i)) x_i &= 0 \\
\sum_{i \in S_3 \setminus S_2} (d_3 v_3(i) - v_1(i)) x_i &= -\sum_{i \in S_2} (d_3 v_3(i) - v_1(i)) x_i \\
\vdots \quad &\quad \vdots \\
\sum_{\substack{i \in S_k \\ i \notin S_2 \cup \cdots \cup S_{k-1}}} (d_k(i) - v_1(i)) x_i &= -\sum_{i \in S_2 \cup \cdots \cup S_{k-1}} (d_k(i) - v_1(i)) x_i \\
\sum_{i \notin S_2 \cup \cdots \cup S_{k-1}} v_1(i) x_i &= -\sum_{i \in S_2 \cup \ldots S_{k-1}} v_1(i) x_i,
\end{aligned}$$

since the first $k-1$ equations each represent $d_j v_j^T y = v_1^T y$ for some j and the last equation represents $v_1^T y = 0$. As in the proof of Lemma 4, we expose each variable in S_2, then the remainder of S_3, then the remainder of S_4, and so forth. After all the variables in S_2 through S_j have been exposed, the probability that the remaining variables in S_{j+1} cause the next equation to be satisfied is at by Theorem 1 most

$$|S_{j+1} \setminus \bigcup_{i=2}^{j} S_i|_1^{-1/2}.$$

Since each S_j contains at most $r^{1-\epsilon}$ elements, it follows that for sufficiently large r there must be at least $r/2$ variables satisfying $v_1(i) \neq 0$ still unexposed by the time we expose S_k and arrive at the last equation. Therefore by Lemma 1 the probability this last equation holds is at most $2r^{-1/2}$, so

$$\mathbf{P}(v_1^T y = \cdots = v_k^T y = 0) \leq 2r^{-1/2} \prod_{j=2}^{k} |S_{j+1} \setminus \bigcup_{i=2}^{j} S_i|_1^{-1/2}.$$

The lemma follows.

4. The proof of Theorem 6

We first note that for any θ,

$$\mathbf{P}(x^T A x = L(x) + c) \leq \mathbf{P}(x^T \operatorname{Re}(e^{i\theta} A) x = \operatorname{Re}(e^{i\theta}(L(x) + c))).$$

Since we can always choose a θ such that $e^{i\theta} a_{ij}$ has non-zero real part for every i and j for which a_{ij} is nonzero, it suffices to prove the result for the case where the entries of A, as well as the coefficients of L and c, are real. We will now assume this to be the case.

The proof will proceed by contradiction. Let us assume that for some δ and all r_0 there is an $r > r_0$ and a matrix A such that for some c we have $\mathbf{P}(x^T A x = c) > r^{-1/2+\delta}$ and every row of A has at least r nonzero entries.

We will use a decoupling argument to relate probabilities involving the quadratic form defined by A to a probability involving a suitable bilinear form x^TBy. We will then combine those bounds with Theorem 5 to obtain

Lemma 12. *Let A be a matrix satisfying the hypotheses of Theorem 6 such that there is a c such that $\mathbf{P}(x^TAx = c) > r^{-1/2+\delta}$. Then there is a principal minor A' of A of size at least $n - O(\frac{r \log n}{\log^5 r})$ and a rank one matrix A'' such that $A' = A''$ everywhere off the main diagonal.*

This allows us to essentially reduce to the case where A is rank one. Let us (for now) assume that this lemma is true.

Without loss of generality we may assume that A' consists of the first m rows and columns of A. Let $z = (x_1, \dots, x_m)^T$. For any particular values of $x_{m+1}, \dots, x_n$, we have the relationship

$$x^TAx = z^TA'z + \widetilde{L}(z) + c',$$

where $\widetilde{L}$ and c' are dependent on the exposed variables. Because $x_i^2 = 1$ for every i, we can further replace A' by A'' by changing c'. It follows that

$$\mathbf{P}(x^TAx = L(x) + c) \leq \sup_{\widetilde{L}, c'} \mathbf{P}(z^TA''z = \widetilde{L}(z) + c').$$

Since A'' has rank one, the quadratic form $z^TA''z$ factors as the square of a linear form. Since we only removed $O(\frac{r \log n}{\log^5 r})$ columns in going from A to A', it follows that for sufficiently large r and n every coefficient of that linear form must be nonzero (as A'' still has at least $\frac{r}{2}$ nonzero entries per row). We will soon show

Lemma 13. *Let $b_1, \dots, b_m, c_1, \dots, c_m, d$ be real numbers such that all of the b_i are nonzero, and let $\alpha > 0$. Then*

$$\mathbf{P}((\sum_{i=1}^m b_i x_i)^2 = \sum_{i=1}^m c_i x_i + d) = O_\alpha(n^{-1/2+\alpha}). \tag{10}$$

Combining Lemma 13 with Lemma 12, we see that if for sufficiently large n we have $\mathbf{P}(x^TAx = c) > r^{-1/2+\delta}$, then we also have $\mathbf{P}(x^TAx = c) = O(r^{-1/2+\delta/2})$, which is a contradiction. We now turn to the proofs of the lemmas.

4.1. **The proof of Lemma 13.** We define

$$t_1 = \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} b_i x_i, \qquad s_1 = \sum_{i=1}^{\lfloor \frac{m}{2} \rfloor} c_i x_i,$$
$$t_2 = \sum_{i=\lfloor \frac{m}{2} \rfloor + 1}^{m} b_i x_i, \qquad s_2 = \sum_{i=\lfloor \frac{m}{2} \rfloor + 1}^{m} c_i x_i.$$

In terms of these new variables, we are attempting to show

$$\mathbf{P}(2t_1t_2 + t_1^2 + t_2^2 = s_1 + s_2 + d) = O(m^{-1/2+\alpha}). \tag{11}$$

The left hand side of (11) can be thought of as the probability that the point p and the line l are incident, where

$$p = (t_2, s_2 - t_2^2),\ l = \{y = 2t_1 x + t_1^2 - s_1 - d\}.$$

Note that p and l are independent, as p depends only on $\{x_1, \dots, x_{\lfloor m/2 \rfloor}\}$, while l depends only on $\{x_{\lfloor m/2 \rfloor} + 1, \dots, x_m\}$. We now make use of the following probabilistic variant of the Szemerédi-Trotter theorem, which is a rescaled version of the weighted Szemerédi-Trotter result of Iosevich, Konyagin, Rudnev, and Ten [8]:

Theorem 8. *Let (p, l) be a point and line independently chosen in $\mathbb{R}^2$. Let*

$$q_p := \sup_{p_0} \mathbf{P}(p = p_0) \quad q_l := \sup_{l_0} \mathbf{P}(l = l_0).$$

Then the probability that p and l are incident is bounded by

$$\mathbf{P}(p \in l) = O((q_p q_l)^{1/3} + q_p + q_l).$$

Since p uniquely determines t_2 and l uniquely determines t_1, it follows from Theorem 1 that q_p and q_l are at most $O(m^{-1/2})$. We are therefore done unless

$$q_p q_l \geq m^{-3/2+\alpha}. \tag{12}$$

If (12) holds, it follows that there is some point p_0 which is chosen with probability at least $m^{-1+\alpha}$. From the definition of p, we know that there are real numbers t_0 and s_0 such that

$$\mathbf{P}(t_2 = t_0 \wedge s_2 = s_0) \geq m^{-1+\alpha}.$$

If follows from the $d = 2$ case of Halász's Theorem 2 that the coefficient vectors of t_2 and s_2 must be close to being multiples of each other, that is to say there is an $|S| \subseteq \{\lfloor \frac{m}{2} \rfloor + 1, \dots, m\}$ with $|S| > \frac{m}{4}$ and a real number c_0 such that $c_j = b_j c_0$ for every $j \in S$.

We now expose every variable not in S. Once we have done so, (10) reduces to an equation of the form

$$(\sum_{j \in S} b_j x_j + d_1)^2 = c_0(\sum_{j \in S} b_j x_j) + d_2, \tag{13}$$

where d_1 and d_2 are constants depending on the exposed variables. For any given d_1 and d_2, there at most 2 values of $\sum_{j \in S} b_j x_j$ for which (13) holds. It therefore follows from the Linear Littlewood-Offord Theorem 1 that for any given d_1 and d_2 the probability that (13) holds is $O(m^{-1/2})$. Lemma 13 follows from taking expectations over all x_i not in S.

4.2. **The proof of Lemma 12.**

We will make use of the following "decoupling" lemma (originally proved in [15]) to reduce from the quadratic case to the bilinear one.

Lemma 14. *Let Y and Z be independent variables, and let Z' be a disjoint copy of Z. Let $E(Y, Z)$ be an event depending on Y and Z. Then*

$$\mathbf{P}(E(Y, Z))^2 \leq \mathbf{P}(E(Y, Z) \wedge E(Y, Z')).$$

In our case this implies that if $X = \{x_1, \dots, x_n\}$ is a collection of independent symmetric ±1 variables variables partitioned into two disjoint subsets Y and Z, then

$$\begin{aligned} &\mathbf{P}(x^T Ax = L(x) + c)^2 = \mathbf{P}(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = L(x) + c)^2 \\ \leq\ & \mathbf{P}(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = L_1(y) + L_2(z) + c \wedge \sum_{i=1}^n \sum_{j=1}^n a_{ij} \widetilde{x_i} \widetilde{x_j} = L_1(y) + L_2(z') + c) \\ \leq\ & \mathbf{P}(\sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j - L_1(y) - L_2(z) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} \widetilde{x_i} \widetilde{x_j} - L_1(y) - L_2(z')), \end{aligned}$$

where $\widetilde{x_j} = x_j$ if $j \in Y$ and $\widetilde{x_j} = x'_j$ if $j \in Z$. Here $L(x) = L_1(y) + L_2(z)$ is the natural decomposition of L into the sum of linear forms on y and z.

All terms only involving variables in Y disappear from this last inequality, and we have

$$\mathbf{P}(x^T Ax = L(x) + c)^2 \leq \mathbf{P}(2 \sum_{x_i \in Y} \sum_{x_j \in Z} a_{ij} x_i (x_j - x'_j) = L_1(z) - L_1(z') + Q(z, z')),$$

where Q is another quadratic form. By assumption the left hand side of this equation is at least $r^{-1+2\delta}$, while the right hand side has the form $y^T Bz = f(z)$.

If we further knew that for every $i \in Y$ there were at least $\frac{r}{4}$ different $j \in Z$ such that $a_{ij} \neq 0$, it would follow from Theorem 5 that the matrix B must contain a rank 1 square submatrix of size $n - O_\delta(\frac{r}{\log^6 r})$. With this observation in mind, we make the following definition:

Definition 7. Given a quadratic form A, a partition $\{x_1, \dots, x_n\} = Y \cup Z$ of the n variables into two disjoint subsets is **balanced** if for every $x_i \in Y$ there are at least $r/4$ different $x_j \in Z$ for which $a_{ij} \neq 0$.

In terms of our original A, we know that for any balanced decomposition of the variables into two equal parts Y and Z, the submatrix corresponding to $Y \times Z$ is equal to a rank one matrix except for a few rogue variables. Our next goal will be to play many such decompositions off of each other.

Since the reduction to a bilinear form only gives us information about the entries in $Y \times Z$, we will want to choose a collection of balanced decompositions such that most entries appear in this submatrix for some element of the decomposition. Motivated by this, we make the following definition:

Definition 8. Let $\mathcal{F} = (Y_1, Z_1), \dots, (Y_m, Z_m)$ be a collection of balanced partitions of a set $X = \{x_1, \dots, x_n\}$ into pairs of disjoint subsets of equal size. We say $\mathcal{F}$ **shatters** X if for every $i \neq j \neq k \neq l$ there is a $r = r(i, j, k, l)$ such that $i, j \in Y_r$ and $k, l \in Z_r$.

In terms of our decoupling, a shattering collection of partitions means that every pair of off-diagonal entries a_{ik} and a_{jl} will appear simultaneously in the bilinear form for some element of $\mathcal{F}$. We next show that we do not have to consider too many partitions at once.

Lemma 15. *If $|X| = n$, there is an $\mathcal{F}$ of size at most $\lceil \frac{5\log n}{\log(17/16)} \rceil < 83\log n$ which shatters X.*

Proof Let an $\mathcal{F}$ of size $\lceil \frac{5\log n}{\log(17/16)} \rceil$ be formed by independently and uniformly choosing (Y_s, Z_s) from the set of all partitions of X into two parts of equal size (or size differing by 1, if n is odd). For any given quadruple (i, j, k, l), the probability that Y_r contains $\{i, j\}$ while Z_r contains $\{k, l\}$ is at least $\frac{1}{17}$, and these events are independent over all r. It therefore follows from the union bound that the probability that X fails to be shattered by this collection is at most

$$n^4(\frac{16}{17})^{|\mathcal{F}|} + \mathbf{P}(\text{some } (Y_s, Z_s) \text{ is not balanced}).$$

The first term is $O(\frac{1}{n})$ by our choice of $|\mathcal{F}|$. For the second term, we first note that for a random partition and a given $x_i \in Y$, it follows from a "without replacement" variant of the Chernoff-Hoeffding bound (see Prop 2.1 in [4] for a precise statement) that the probability that x_i has too few nonzero a_{ij} across the partition is $O(e^{-r/8})$. Taking the union bound over all partitions in our collection and all variables in each Y, we see the second term is $o(1)$. Since a random collection almost surely shatters X, there must be at least one shattering collection. ■

We now fix some $\mathcal{F}_0$ which shatters our original set of variables and has size at most $83\log n$. For each r, we know from Theorem 5 that we can find exceptional sets $Y'_s \subseteq Y_s, Z'_s \subseteq Z_s$ with $|Y'_s|, |Z'_s| = O(\frac{r}{\log^6 r})$ such that the submatrix of A corresponding to $(Y_s \backslash Y'_s) \times (Z_s \backslash Z'_s)$ has rank one. Let

$$W = \bigcup_{(Y_s, Z_s)\in\mathcal{F}_0} (Y'_s \cup Z'_s).$$

Without loss of generality we may assume that $W = \{x_{n-t+1}, \dots, x_n\}$. By assumption $t = O(\frac{r\log n}{\log^5 r})$. Since r is by assumption at least $\exp((\log n)^{1/4})$, it follows that $t = o(r)$, meaning each row still contains many nonzero entries outside the columns of W. In particular, we may without loss of generality assume $a_{12} \neq 0$.

For any 4 distinct elements (i, j, k, l) disjoint from W, we know from the definition of $\mathcal{F}_0$ and W that for some s the 2×2 submatrix of A on $\{i, j\} \times \{k, l\}$ appeared in a rank one submatrix of $Y_s \times Z_s$. It follows that for every such set of distinct (i, j, k, l), we have $a_{ik}a_{jl} = a_{jk}a_{il}$. In particular, for every pair (j, l) with $3 \leq k \neq l \leq n - t$, we have

$$a_{jl} = a_{1l}\frac{a_{j2}}{a_{12}}. \tag{14}$$

We can therefore take A' to be the principal minor of A on $\{x_3, \dots, x_{n-t}\}$, and A'' to be the matrix for which the right hand side of (14) also holds for $j = l$.

4.3. **The proof of Corollary 1.** Construct a graph whose vertices are the variables x_i, with x_i adjacent to x_j for $i \neq j$ if and only if a_{ij} are nonzero. By assumption, this graph has average degree at least $m - 1$. It follows that it must contain a subgraph of minimum degree at least $\frac{m-1}{2}$ (e.g., by greedily removing the vertex of lowest degree, since removing vertices of degree less than $\frac{m-1}{2}$ never decreases the average degree). In matrix terms, this implies that A contains a principal minor A' such that every row of A' has at least $\frac{m-1}{2}$ nonzero entries. Without loss of generality we may assume that the minor corresponds to the variables $\widetilde{x} = \{x_1, \dots, x_k\}$. For any fixed value of $x_{k+1}, \dots, x_n$, the equation $x^T Ax = L(x) + c$ becomes

$$\widetilde{x}^T A\widetilde{x} = \widetilde{L}(\widetilde{x}) + \widetilde{c},$$

an equation which holds with probability $O_\delta(m^{-\frac{1}{2}+\delta})$ by Theorem 6. The result follows from taking expectations over all values of $x_{k+1}, \dots, x_n$.

5. Extensions of the Main Results and Conjectures

5.1. **Inverse results for more weakly concentrated Bilinear Forms:** It is an interesting problem to consider whether there are similar inverse results holding in general for when a bilinear form has polynomially large concentration on one value $P(x^T Ay = c) \geq n^{-b}$ for some b.

There are at least two different types of structure that lead to sufficient conditions for this to occur. One possibility is algebraic: If the coefficient matrix has low rank, then $x^T Ay$ will be equal to 0 whenever a small number of linear forms are simultaneously equal to 0, which may not be too unlikely an event if some of those forms are themselves structured (in the sense of the results of [19]). For example, if A is chosen to satisfy $a_{ij} = f(i) + g(j)$ (for arbitrary f and g), then $x^T Ay$ can be expressed as

$$(x_1 + x_2 + \dots + x_n)(g(1)y_1 + \dots + g(n)y_n) + (f(1)x_1 + \dots + f(n)x_n)(y_1 + \dots + y_n)$$

and is 0 whenever $x_1 + \dots + x_n = y_1 + \dots + y_n = 0$, an event which occurs with probability approximately $\frac{1}{n}$. More generally, if A has rank a_1, then $x^T Ay$ can equal 0 whenever a_1 linear forms are simultaneously 0, an event which can occur with probability on the order of $n^{-a_1/2}$.

Another structure that can cause polynomially large concentration is arithmetic: If the entries of the coefficient matrix A are all drawn from a short generalized arithmetic progression of rank a_2 and volume n^{a_3}, then the output of $x^T Ay$ will also lie in such a progression, and with high probability will take on one of only $n^{a_2+a_3}$ values. So by the pigeonhole principle some value is again taken on with polynomial probability. We conjecture that these two structures (and combinations thereof) are the *only* ways a form can have polynomial concentration

Conjecture 1. *Fix $a > 0$. There is an N_0 (dependent on a) such that the following holds. Let $n > N_0$ and let A be an $n \times n$ matrix of nonzero entries such that for x and y consisting of n independent ± 1 symmetric random variables we have*

$$\sup_c \mathbf{P}(x^T Ay = c) > n^{-a}.$$

Then there exist integers $a_1, a_2, a_3 \geq 0$ satisfying

$$\frac{a_1}{2} + a_2 + a_3 \leq a. \tag{15}$$

such that A can be written as $A_1+A_2+A_3$, where A_1 has rank at most a_1, the entries of A_2 are drawn from a generalized arithmetic progression of rank at most a_2 and volume at most n^{a_3}, and A_3 contains at most $\frac{n^2}{\log n}$ nonzero entries.

Note that if $a < 1$ we must have $a_1 = 1$ and $a_2 = a_3 = 0$, corresponding to Theorem 5. Nguyen [13] has recently proven a version of this conjecture [2] with (15) replaced by the weaker bound $a_1, a_2, a_3 = O_a(1)$.

5.2. **Higher degrees.** In this section we give several conjectured extentions of the main results to this paper to multilinear and polynomial forms. We begin with the following (simplified) analogue of Theorem 4, which can be proved by the same method.

Theorem 9. *Let k be a fixed positive integer. Let $y_1 = (x_{1,1}, \dots, x_{n,1}), \dots, y_k = (x_{1,k}, \dots, x_{n,k})$ be k independent vectors of independent ± 1 symmetric random variables, and let*

$$A(y_1, y_2, \dots, y_k) := \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} a_{i_1 i_2 \dots i_k} x_{i_1,1} \dots x_{i_k,k}$$

be a $k-$multilinear form whose coefficients $a_{i_1 \dots i_k}$ are all nonzero. Then for any function f of $k-1$ vectors of length n,

$$\mathbf{P}(A(y_1, \dots, y_k) = f(y_2, \dots, y_k)) = O_k(n^{-1/2}). \tag{16}$$

Again, this is tight for degenerate forms which contain a linear factor. A natural conjecture would be that non-degenerate forms are significantly less concentrated.

Conjecture 2. *Let k, A, y, and f be as in Theorem 9, and let $\epsilon > 0$ be fixed. Then there is a constant N_0 depending only on ϵ and k such that if*

$$\mathbf{P}(A(y_1, \dots, y_k) = f(y_2, \dots, y_k)) \geq n^{-\frac{k}{2}+\epsilon},$$

and $n > N_0$, then there is a partition of $\{y_1, \dots, y_k\}$ into disjoint sets S and T and functions f_1 and f_2 such that f_1 depends only the variables in S, f_2 only on the variables in T, and A differs from $f_1 f_2$ in $o(n^2)$ coefficients.

The $k/2$ in this conjecture comes from how $n^{k/2}$ is the typical magnitude of f in the case where the coefficients of A are random (small) integers.

We can also conjecture a polynomial analogue to Theorem 6, including an analogous inverse theorem to the above multilinear one.

Conjecture 3. *Let $x_1, \dots, x_n$ be independent ± 1 symmetric random variables, and let*

$$f(x_1, \dots, x_n) = \sum_{1 \leq i_1 \dots \leq i_k \leq n} a_{i_1 \dots i_k} x_{i_1} \dots x_{i_k}$$

[2] Nguyen actually gives a stronger characterization of the matrix A_1, for full details see his paper

be a degree k homogeneous polynomial with at least mn^{k-1} nonzero coefficients. Then

$$\sup_c \mathbf{P}(f(x_1,\dots,x_n)=c)=O(m^{-1/2}).$$

If the above concentration is at least $\Omega_k(m^{-k/2+\epsilon})$, then f differs in only a few coefficients from a polynomial which factors.

In [2], a proof of the first half of this conjecture was given with $m^{-1/2}$ replaced by m^{-c_k}, where $c_k = 2^{-(k^2+k)/2}$. For the second half, we do not have a proof of this conjecture even in the case $k=2$.

Acknowledgements: The author wishes to thank Ernie Croot, Tali Kaufman, Endre Szemerédi, Prasad Tetali, Van Vu, and Philip Matchett Wood for enlightening conversations and helpful comments on earlier versions of this paper. He is particularly indebted to Croot for providing much of the argument for Lemma 9.

References

[1] J. Bourgain, V. Vu, and P. Matchett Wood, On the singularity probability of discrete random matrices, *J. Funct. Anal* **258** (2010), 559-603

[2] K. Costello, T. Tao and V. Vu, Random symmetric matrices are almost surely non-singular, *Duke Math J.* **135** (2006), no. 2, 395-413

[3] K. Costello and V. Vu, The rank of random graphs, *Random Structures and Algorithms*, **33** (2008), 269-285

[4] W. de Launey and D. Levin, $(1,-1)$ matrices with near-extremal properties, *SIAM J. Disc. Math.*, **23** (2009), 1422-1440.

[5] P. Erdős, On a lemma of Littlewood and Offord, *Bull. Amer. Math. Soc.* **51** (1945), 898-902.

[6] G. Halász, Estimates for the concentration function of combinatorial number theory and probability, *Period. Math. Hungar.* **8** (1977), no. 3-4, 197-211

[7] W. Hoeffding, Probability inequalities for sums of bounded variables, *J. Amer. Statist. Assoc.* **58** (1963) 13-30

[8] A. Iosevich, S. Konyagin, M. Rudnev, and V. Ten, Combinatorial complexity of convex sequences, *Discrete Comput. Geom.* **35** (2006), 143-158

[9] J. Kahn, J. Komlós, E. Szemerédi, On the probability a random ±1 matrix is singular, *J. Amer. Math Soc.* **8** (1995), 223-240

[10] D. Kleitman, On a lemma of Littlewood and Offord on the distributions of certain sums, *Math. Z.*, **90** (1965) 251-259.

[11] J. Komlós, On the determinant of $(0,1)$ matrices, *Studia Sci. Math. Hungar.* **2** (1967) 7-22

[12] J. Littlewood and C. Offord, On the number of real roots of a random algebraic equation III *Rec. Math. [Mat. Sbornik] N.S.* **12** (1943), 277-286

[13] H. Nguyen, Inverse Littlewood-Offord problems and the singularity of random symmetric matrices, *Submitted*, preprint at arXiv:1101.3074

[14] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and invertibility of random matrices *Adv. Math.* **218** (2008) 600-633.

[15] A. Sidorenko, A correlation inequality for bipartite graphs, *Graphs Combin.* **9** (1991), no. 2, 201-204

[16] T. Tao and V. Vu, Additive Combinatorics, Cambridge Studies in Advanced Math **105**, Cambridge University Press, Cambridge, 2006

[17] T. Tao and V. Vu, Inverse Littlewood-Offord theorems and the condition number of random discrete matrices, *Annal. Math* **169** (2009), 595-632

[18] T. Tao and V. Vu, On the singularity probability of random Bernoulli matrices, *J. Amer. Math. Soc.* **20** (2007), 603-628

[19] T. Tao and V. Vu, A sharp inverse Littlewood-Offord theorem, *Random Structures and Algorithms*, **37** (2010), 525-539

[20] R. Vershynin, Invertibility of random symmetric matrices, *Submitted*, preprint at arXiv:1102:0300

Department of Mathematics, Georgia Institute of Technology, Atlanta, GA 30308

E-mail address: kcostell@math.gatech.edu