Annals of Mathematics, **169** (2009), 595–632

Inverse Littlewood-Offord theorems and the condition number of random discrete matrices

By Terence Tao and Van H. Vu*

Abstract

Consider a random sum $\eta_1 v_1 + \cdots + \eta_n v_n$, where $\eta_1, \ldots, \eta_n$ are independently and identically distributed (i.i.d.) random signs and $v_1, \ldots, v_n$ are integers. The Littlewood-Offord problem asks to maximize concentration probabilities such as $\mathbf{P}(\eta_1 v_1 + \cdots + \eta_n v_n = 0)$ subject to various hypotheses on $v_1, \ldots, v_n$. In this paper we develop an *inverse* Littlewood-Offord theory (somewhat in the spirit of Freiman's inverse theory in additive combinatorics), which starts with the hypothesis that a concentration probability is large, and concludes that almost all of the $v_1, \ldots, v_n$ are efficiently contained in a generalized arithmetic progression. As an application we give a new bound on the magnitude of the least singular value of a random Bernoulli matrix, which in turn provides upper tail estimates on the condition number.

1. Introduction

Let $\mathbf{v}$ be a multiset (allowing repetitions) of n integers $v_1, \ldots, v_n$. Consider a class of discrete random walks $Y_{\mu,\mathbf{v}}$ on the integers $\mathbf{Z}$, which start at the origin and consist of n steps, where at the i^{th} step one moves backwards or forwards with magnitude v_i and probability $\mu/2$, and stays at rest with probability $1-\mu$. More precisely:

Definition 1.1 (Random walks). For any $0 \le \mu \le 1$, let $\eta^\mu \in \{-1, 0, 1\}$ denote a random variable which equals 0 with probability $1-\mu$ and ± 1 with probability $\mu/2$ each. In particular, η^1 is a random sign ± 1, while η^0 is identically zero. Given $\mathbf{v}$, we define $Y_{\mu,\mathbf{v}}$ to be the random variable

$$Y_{\mu,\mathbf{v}} := \sum_{i=1}^n \eta_i^\mu v_i$$

*T. Tao is a Clay Prize Fellow and is supported by a grant from the Packard Foundation. V. Vu is an A. Sloan Fellow and is supported by an NSF Career Grant.

where the η_i^μ are i.i.d. copies of η^μ. Note that the exact enumeration $v_1, \dots, v_n$ of the multiset is irrelevant. The *concentration probability* $\mathbb{P}_\mu(\mathbf{v})$ of this random walk is defined to be the quantity

$$\mathbb{P}_\mu(\mathbf{v}) := \max_{a \in \mathbf{Z}} \mathbf{P}(Y_{\mu,\mathbf{v}} = a). \tag{1}$$

Thus we have $0 < \mathbb{P}_\mu(\mathbf{v}) \leq 1$ for any $\mu, \mathbf{v}$.

The concentration probability (and more generally, the concentration function) is a central notion in probability theory and has been studied extensively, especially by the Russian school (see [21], [19], [18] and the references therein).

The first goal of this paper is to establish a relation between the magnitude of $\mathbb{P}_\mu(\mathbf{v})$ and the arithmetic structure of the multiset $\mathbf{v} = \{v_1, \dots, v_n\}$. This gives an answer to the general question of finding conditions under which one can squeeze large probability inside a small interval. We will primarily be interested in the case $\mu = 1$, but for technical reasons it will be convenient to consider more general values of μ. Generally, however, we think of μ as fixed, while letting n become very large.

A classical result of Littlewood-Offord [16], found in their study of the number of real roots of random polynomials, asserts that if all of the v_i's are nonzero, then $\mathbb{P}_1(\mathbf{v}) = O(n^{-1/2} \log n)$. The log term was later removed by Erdős [5]. Erdős' bound is sharp, as shown by the case $v_1 = \cdots = v_n \neq 0$. However, if one forbids this special case and assumes that the v_i's are all distinct, then the bound can be improved significantly. Erdős and Moser [6] showed that under this stronger assumption, $\mathbb{P}_1(\mathbf{v}) = O(n^{-3/2} \ln n)$. They conjectured that the logarithmic term is not necessary and this was confirmed by Sárközy and Szemerédi [22]. Again, the bound is sharp (up to a constant factor), as can be seen by taking $v_1, \dots, v_n$ to be a proper arithmetic progression such as $1, \dots, n$. Later, Stanley [24], using algebraic methods, gave a very explicit bound for the probability in question.

The higher dimensional version of Littlewood-Offord's problem (where the v_i are nonzero vectors in $\mathbf{R}^d$, for some fixed d) also drew lots of attention. Without the assumption that the v_i's are different, the best result was obtained by Frankl and Füredi in [7], following earlier results by Katona [11], Kleitman [12], Griggs, Lagarias, Odlyzko and Shearer [8] and many others. However, the techniques used in these papers did not seem to yield the generalization of Sárközy and Szemerédi's result (the $O(n^{-3/2})$ bound under the assumption that the vectors are different).

The generalization of Sárközy and Szemerédi's result was obtained by Halász [9], using analytical methods (especially harmonic analysis). Halász' paper was one of our starting points in this study.

In the above two examples, we see that in order to make $\mathbb{P}_\mu(\mathbf{v})$ large, we have to impose a very strong additive structure on $\mathbf{v}$ (in one case we set the v_i's to be the same, while in the other we set them to be elements of an arithmetic progression). We are going to show that this is the only way to make $\mathbb{P}_\mu(\mathbf{v})$ large. More precisely, we propose the following phenomenon:

If $\mathbb{P}_\mu(\mathbf{v})$ is large, then $\mathbf{v}$ has a strong additive structure.

In the next section, we are going to present several theorems supporting this phenomenon. Let us mention here that there is an analogous phenomenon in combinatorial number theory. In particular, a famous theorem of Freiman asserts that if A is a finite set of integers and $A+A$ is small, then A is contained efficiently in a generalized arithmetic progression [28, Ch. 5]. However, the proofs of Freiman's theorem and those in this paper are quite different.

As an application, we are going to use these inverse theorems to study random matrices. Let M_n^μ be an n by n random matrix, whose entries are i.i.d. copies of η^μ. We are going to show that with very high probability, the condition number of M_n^μ is bounded from above by a polynomial in n (see Theorem 3.3 below). This result has high potential of applications in the theory of probability in Banach spaces, as well as in numerical analysis and theoretical computer science. A related result was recently established by Rudelson [20], with better upper bounds on the condition number but worse probabilities. We will discuss this application with more detail in Section 3.

To see the connection between this problem and inverse Littlewood-Offord theory, observe that for any $\mathbf{v} = (v_1, \dots, v_n)$ (which we interpret as a column vector), the entries of the product $M_n^\mu \mathbf{v}$ are independent copies of $Y_{\mu,\mathbf{v}}$. Thus we expect that $\mathbf{v}^T$ is unlikely to lie in the kernel of M_n^μ unless the concentration probability $\mathbb{P}_\mu(\mathbf{v})$ is large. These ideas are already enough to control the singularity probability of M_n^μ (see e.g. [10], [25], [26]). To obtain the more quantitative condition number estimates, we introduce a new discretization technique that allows one to estimate the probability that a certain random variable is small by the probability that a certain discretized analogue of that variable is zero.

The rest of the paper is organized as follows. In Section 2 we state our main inverse theorems. In Section 3 we state our main results on condition numbers, as well as the key lemmas used to prove these results. In Section 4, we give some brief applications of the inverse theorems. In Section 7 we prove the result on condition numbers, assuming the inverse theorems and two other key ingredients: a discretization of generalized progressions and an extension of the famous result of Kahn, Komlós and Szemerédi [10] on the probability that a random Bernoulli matrix is singular. The inverse theorems is proven in Section 6, after some preliminaries in Section 5 in which we establish basic properties of $\mathbb{P}_\mu(\mathbf{v})$. The result about discretization of progressions are proven in Section 8. Finally, in Section 9 we prove the extension of Kahn, Komlós and Szemerédi [10].

We conclude this section by setting out some basic notation. A set

$$P = \{c + m_1 a_1 + \cdots + m_d a_d | M_i \le m_i \le M_i'\}$$

is called a *generalized arithmetic progression* (GAP) of rank d. It is convenient to think of P as the image of an integer box

$$B := \{(m_1, \dots, m_d) | M_i \le m_i \le M_i'\}$$

in $\mathbf{Z}^d$ under the linear map

$$\Phi : (m_1, \dots, m_d) \mapsto c + m_1 a_1 + \cdots + m_d a_d.$$

The numbers a_i are the *generators* of P. In this paper, all GAPs have rational generators. A GAP is *proper* if Φ is one to one on B. The product $\prod_{i=1}^d (M_i' - M_i + 1)$ is the *volume* of P. If $M_i = -M_i'$ and $c = 0$ (so $P = -P$) then we say that P is *symmetric*.

For a set A of reals and a positive integer k, we define the iterated sumset

$$kA := \{a_1 + \cdots + a_k | a_i \in A\}.$$

One should take care to distinguish the sumset kA from the dilate $k \cdot A$, defined for any real k as

$$k \cdot A := \{ka | a \in A\}.$$

We always assume that n is sufficiently large. The asymptotic notation $O()$, $o()$, $\Omega()$, $\Theta()$ is used under the assumption that $n \to \infty$. Notation such as $O_d(f)$ means that the hidden constant in O depends only on d.

2. Inverse Littlewood-Offord theorems

Let us start by presenting an example when $\mathbb{P}_\mu(\mathbf{v})$ is large. This example is the motivation of our inverse theorems.

Example 2.1. Let P be a symmetric generalized arithmetic progression of rank d and volume V; we view d as being fixed independently of n, though V can grow with n. Let $v_1, \dots, v_n$ be (not necessarily different) elements of V. Then the random variable $Y_{\mu,\mathbf{v}} = \sum_{i=1}^n \eta_i v_i$ takes values in the GAP nP which has volume $n^d V$. From the pigeonhole principle it follows that

$$\mathbb{P}_\mu(\mathbf{v}) \ge n^{-d} V^{-1}.$$

In fact, the central limit theorem suggests that $\mathbb{P}_\mu(\mathbf{v})$ should typically be of the order of $n^{-d/2} V^{-1}$.

This example shows that if the elements of $\mathbf{v}$ belong to a GAP with small rank and small volume then $\mathbb{P}_\mu(\mathbf{v})$ is large. One might hope that the inverse also holds, namely,

If $\mathbb{P}_\mu(\mathbf{v})$ *is large, then (most of) the elements of* $\mathbf{v}$ *belong to a* GAP *with small rank and small volume.*

In the rest of this section, we present three theorems, which support this statement in a quantitative way.

Definition 2.2 (Dissociativity). Given a multiset $\mathbf{w} = \{w_1, \dots, w_r\}$ of real numbers and a positive number k, we define the GAP $Q(\mathbf{w}, k)$ and the cube $S(\mathbf{w})$ as follows:

$$Q(\mathbf{w}, k) := \{m_1 w_1 + \cdots + m_r w_r | -k \le m_i \le k\},$$
$$S(\mathbf{w}) := \{\epsilon_1 w_1 + \cdots + \epsilon_r w_r | \epsilon_i \in \{-1, 1\}\}.$$

We say that $\mathbf{w}$ is *dissociated* if $S(\mathbf{w})$ does not contain zero. Furthermore, $\mathbf{w}$ is k*-dissociated* if there do not exist integers $-k \le m_1, \dots, m_r \le k$, not all zero, such that $m_1 w_1 + \cdots + m_r w_r = 0$.

Our first result is the following simple proposition:

PROPOSITION 2.3 (Zeroth inverse theorem). *Let* $\mathbf{v} = \{v_1, \dots, v_n\}$ *be such that* $\mathbb{P}_1(\mathbf{v}) > 2^{-d-1}$ *for some integer* $d \ge 0$*. Then* $\mathbf{v}$ *contains a subset* $\mathbf{w}$ *of size* d *such that the cube* $S(\mathbf{w})$ *contains* $v_1, \dots, v_n$*.*

The next two theorems are more involved and also more useful. In these two theorems and their corollaries, we assume that k and n are sufficiently large, whenever needed.

THEOREM 2.4 (First inverse theorem). *Let* μ *be a positive constant at most* 1 *and let* d *be a positive integer. Then there is a constant* $C = C(\mu, d) \ge 1$ *such that the following holds. Let* $k \ge 2$ *be an integer and let* $\mathbf{v} = \{v_1, \dots, v_n\}$ *be a multiset such that*

$$\mathbb{P}_\mu(\mathbf{v}) \ge C(\mu, d) k^{-d}.$$

Then there exists a k*-dissociated multiset* $\mathbf{w} = \{w_1, \dots, w_r\}$ *such that*

(1) $r \le d - 1$ *and* $w_1, \dots, w_r$ *are elements of* $\mathbf{v}$*;*

(2) *The union* $\bigcup_{\tau \in \mathbf{Z}, 1 \le \tau \le k} \frac{1}{\tau} \cdot Q(\mathbf{w}, k)$ *contains all but* k^2 *of the integers* $v_1, \dots, v_n$ *(counting multiplicity).*

This theorem should be compared against the heuristics in Example 2.1 (setting k equal to a small multiple of $\sqrt{n}$). In particular, note that the GAP $Q(\mathbf{w}, k)$ has very small volume, only $O(k^{d-1})$.

The above theorem does not yet show that most of the elements of $\mathbf{v}$ belong to a single GAP. Instead, it shows that they belong to the union of a few dilates of a GAP. One could remove the unwanted $\frac{1}{\tau}$ factor by clearing denominators, but this costs us an exponential factor such as $k!$, which is often too large in applications. Fortunately, a more refined argument allows us to eliminate these denominators while losing only polynomial factors in k:

THEOREM 2.5 (Second inverse theorem). *Let μ be a positive constant at most one, ϵ be an arbitrary positive constant and d be a positive integer. Then there are constants $C = C(\mu, \epsilon, d) \geq 1$ and $k_0 = k_0(\mu, \epsilon, d) \geq 1$ such that the following holds. Let $k \geq k_0$ be an integer and let $\mathbf{v} = \{v_1, \dots, v_n\}$ be a multiset such that*

$$\mathbb{P}_\mu(\mathbf{v}) \geq Ck^{-d}.$$

Then there exists a GAP *Q with the following properties*:

(1) *The rank of Q is at most $d-1$;*
(2) *The volume of Q is at most $k^{2(d^2-1)+\epsilon}$;*
(3) *Q contains all but at most $\epsilon k^2 \log k$ elements of $\mathbf{v}$ (counting multiplicity);*
(4) *There exists a positive integer s at most $k^{d+\epsilon}$ such that $su \in \mathbf{v}$ for each generator u of Q.*

Remark 2.6. A small number of exceptional elements cannot be avoided. For instance, one can add $O(\log k)$ completely arbitrary elements to $\mathbf{v}$, and decrease $\mathbf{P}_\mu(\mathbf{v})$ by a factor of $k^{-O(1)}$ at worst.

For the applications in this paper, the following corollary of Theorem 2.5 is convenient.

COROLLARY 2.7. *For any positive constants A and α there is a positive constant A' such that the following holds. Let μ be a positive constant at most one and assume that $\mathbf{v} = \{v_1, \dots, v_n\}$ is a multiset of integers satisfying $\mathbb{P}_\mu(\mathbf{v}) \geq n^{-A}$. Then there is a* GAP *$Q$ of rank at most A' and volume at most $n^{A'}$ which contains all but at most n^α elements of $\mathbf{v}$ (counting multiplicity). Furthermore, there exists a positive integer $s \leq n^{A'}$ such that $su \in \mathbf{v}$ for each generator u of $\mathbf{Q}$.*

Remark 2.8. The assumption $\mathbb{P}_\mu(\mathbf{v}) \geq n^{-A}$ in all statements can be replaced by the following more technical, but somewhat weaker, assumption that

$$\int_0^1 \prod_{i=1} |(1-\mu) + \mu \cos 2\pi v_i \xi| \; d\xi \geq n^{-A}.$$

The right-hand side is an upper bound for $\mathbb{P}_\mu(\mathbf{v})$, provided that μ is sufficiently small. Assuming that $\mathbb{P}_\mu(\mathbf{v}) \geq n^{-A}$, what is actually used in the proofs is the consequence

$$\int_0^1 \prod_{i=1} |(1-\mu) + \mu \cos 2\pi v_i \xi| d\xi \geq n^{-A}.$$

(See §5 for more details.) This weaker assumption is useful in applications (see [27]).

The vector versions of all three theorems hold (when the v_i's are vectors in $\mathbf{R}^r$, for any positive integer r), thanks to Freiman's isomorphism principle (see, e.g., [28, Ch. 5]). This principle allows us to project the problem from $\mathbf{R}^r$ onto $\mathbf{Z}$. The value of r is irrelevant and does not appear in any quantitative bound. In fact, one can even replace $\mathbf{R}^r$ by any torsion free additive group.

In an earlier paper [26] we introduced another type of inverse Littlewood-Offord theorem. This result showed that if $\mathbb{P}_\mu(\mathbf{v})$ was comparable to $\mathbb{P}_1(\mathbf{v})$, then $\mathbf{v}$ could be efficiently contained inside a GAP of bounded rank (see [26, Th. 5.2] for details).

We shall prove these inverse theorems in Section 6, after some combinatorial and Fourier-analytic preliminaries in Section 5. For now, we take these results for granted and turn to an application of these inverse theorems to random matrices.

3. The condition number of random matrices

If M is an $n \times n$ matrix, we use

$$\sigma_1(M) := \sup_{x \in \mathbf{R}^n, \|x\|=1} \|Mx\|$$

to denote the largest singular value of M. Tthis parameter is also often called the operator norm of M. Here $\|x\|$ denotes the Euclidean magnitude of a vector $x \in \mathbf{R}^n$. If M is invertible, the *condition number* $c(M)$ is defined as

$$c(M) := \sigma_1(M)\sigma_1(M^{-1}).$$

We adopt the convention that $c(M)$ is infinite if M is not invertible.

The condition number plays a crucial role in applied linear algebra and computer science. In particular, the complexity of any algorithm which requires solving a system of linear equations usually involves the condition number of a matrix; see [1], [23]. Another area of mathematics where this parameter is important is the theory of probability in Banach spaces (e.g. see [15], [20]).

The condition number of a random matrix is a well-studied object (see [3] and the references therein). In the case when the entries of M are i.i.d. Gaussian random variables (with mean zero and variance one), Edelman [3], answering a question of Smale [23] showed

Theorem 3.1. *Let N_n be an $n \times n$ random matrix, whose entries are* i.i.d. *Gaussian random variables (with mean zero and variance one). Then* $\mathbf{E}(\ln c(N_n)) = \ln n + c + o(1)$, *where $c > 0$ is an explicit constant.*

In application, it is usually useful to have a tail estimate. It was shown by Edelman and Sutton [4] that

Theorem 3.2. *Let N_n be a n by n random matrix, whose entries are* i.i.d. *Gaussian random variables (with mean zero and variance one). Then for any constant $A > 0$,*

$$\mathbf{P}(c(N_n) \geq n^{A+1}) = O_A(n^{-A}).$$

On the other hand, for the other basic case when the entries are i.i.d. Bernoulli random variables (copies of η^1), the situation is far from being settled. Even to prove that the condition number is finite with high probability is a nontrivial task (see [13]). The techniques used to study Gaussian matrices rely heavily on the explicit joint distribution of the eigenvalues. This distribution is not available for discrete models.

Using our inverse theorems, we can prove the following result, which is comparable to Theorem 3.2, and is another main result of this paper. Let M_n^μ be the n by n random matrix whose entries are i.i.d. copies of η^μ. In particular, the Bernoulli matrix mentioned above is the case when $\mu = 1$.

Theorem 3.3. *For any positive constant A, there is a positive constant B such that the following holds. For any positive constant μ at most one and any sufficiently large n*

$$\mathbf{P}(c(M_n^\mu) \geq n^B) \leq n^{-A}.$$

Given an invertible matrix M of order n, we set $\sigma_n(M)$ to be the smallest singular value of M:

$$\sigma_n(M) := \min_{x \in \mathbf{R}^n, \|x\|=1} \|Mx\|.$$

Then

$$c(M) = \sigma_1(M)/\sigma_n(M).$$

It is well known that there is a constant C_μ such that the largest singular value of M_n^μ is at most $C_\mu n^{1/2}$ with exponential probability $1 - \exp(-\Omega_\mu(n))$ (see, e.g. [14]). Thus, Theorem 3.3 reduces to the following lower tail estimate for the smallest singular value of $\sigma_n(M)$:

Theorem 3.4. *For any positive constant A, there is a positive constant B such that the following holds. For any positive constant μ at most one and any sufficiently large n*

$$\mathbf{P}(\sigma_n(M_n^\mu) \leq n^{-B}) \leq n^{-A}.$$

Shortly prior to this paper, Rudelson [20] proved the following result.

THEOREM 3.5. *Let $0 < \mu \leq 1$. There are positive constants $c_1(\mu), c_2(\mu)$ such that the following holds. For any $\epsilon \geq c_1(\mu)n^{-1/2}$,*

$$\mathbf{P}(\sigma_n(M_n^\mu) \leq c_2(\mu)\epsilon n^{-3/2}) \leq \epsilon.$$

In fact, Rudelson's result holds for a larger class of matrices. The description of this class is, however, somewhat technical. We refer the reader to [20] for details.

It is useful to compare Theorems 3.4 and 3.5. Theorem 3.5 gives an explicit dependence between the bound on σ_n and the probability, while the dependence between A and B in Theorem 3.4 is implicit. Actually our proof does provide an explicit value for B, but it is rather large and we make no attempt to optimize it. On the other hand, Theorem 3.5 does not yield a probability better than $n^{-1/2}$. In many applications (especially those involving the union bound), it is important to have a probability bound of order n^{-A} with arbitrarily given A.

The proof of Theorem 3.4 relies on Corollary 2.7 and two other ingredients, which are of independent interest. In the rest of this section, we discuss these ingredients. These ingredients will then be combined in Section 7 to prove Theorem 3.4.

3.1. *Discretization of* GAP*s*. Let P be a GAP of integers of rank d and volume V. We show that given any specified scale parameter R_0, one can "discretize" P near the scale R_0. More precisely, one can cover P by the sum of a coarse progression and a small progression, where the diameter of the small progression is much smaller (by an arbitrarily specified factor of S) than the spacing of the coarse progression, and that both of these quantities are close to R_0 (up to a bounded power of SV).

THEOREM 3.6 (Discretization). *Let $P \subset \mathbf{Z}$ be a symmetric* GAP *of rank d and volume V. Let R_0, S be positive integers. Then there exists a scale $R \geq 1$ and two* GAP*s P_{small}, P_{sparse} of rational numbers with the following properties.*

- (*Scale*) $R = (SV)^{O_d(1)}R_0$.
- (*Smallness*) P_{small} *has rank at most d, volume at most V, and takes values in $[-R/S, R/S]$.*
- (*Sparseness*) P_{sparse} *has rank at most d, volume at most V, and any two distinct elements of SP_{sparse} are separated by at least RS.*
- (*Covering*) $P \subseteq P_{\mathrm{small}} + P_{\mathrm{sparse}}$.

This theorem is elementary but is somewhat involved. The detailed proof will appear in Section 8. Here, we give an informal explanation, appealing to the analogy between the combinatorics of progressions and linear algebra. Recall that a GAP of rank d is the image $\Phi(B)$ of a d-dimensional box under a linear map Φ. This can be viewed as a discretized, localized analogue of the object $\Phi(V)$, where Φ is a linear map from a d-dimensional vector space V to some other vector space. The analogue of a "small" progression would be an object $\Phi(V)$ in which Φ vanished. The analogue of a "sparse" progression would be an object $\Phi(V)$ in which the map Φ was injective. Theorem 3.6 is then a discretized, localized analogue of the obvious linear algebra fact that given any object of the form $\Phi(V)$, one can split $V = V_{\text{small}} + V_{\text{sparse}}$ for which $\Phi(V_{\text{small}})$ is small and $\Phi(V_{\text{sparse}})$ is sparse. Indeed one simply sets V_{small} to be the kernel of Φ, and V_{sparse} to be any complementary subspace to V_{small} in V. The proof of Theorem 3.6 follows these broad ideas, with P_{small} being essentially a "kernel" of the progression P, and P_{sparse} being a kind of "complementary progression" to this kernel.

To oversimplify, we shall exploit this discretization result (as well as the inverse Littlewood-Offord theorems) to control the event that the singular value is small, by the event that the singular value (of a slightly modified random matrix) is *zero*. The control of this latter quantity is the other ingredient of the proof, to which we now turn.

3.2. *Singularity of random matrices.* A famous result of Kahn, Komlós and Szemerédi [10] asserts that the probability that M_n^1 is singular (or equivalently, that $\sigma_n(M_n^1) = 0$) is exponentially small:

THEOREM 3.7. *There is a positive constant ε such that*

$$\mathbf{P}(\sigma_n(M_n^1) = 0) \leq (1-\varepsilon)^n.$$

In [10] it was shown that one can take $\varepsilon = .001$. Improvements on ε are obtained recently in [25], [26]. The value of ϵ does not play a critical role in this paper.

To prove Theorem 3.3, we need the following generalization of Theorem 3.7. Note that the row vectors of M_n^1 are i.i.d. copies of X^1, where $X^1 = (\eta_1^1, \dots, \eta_n^1)$ and η_i^1 are i.i.d. copies of η^1. By changing 1 to μ, we can define X^μ in the obvious manner. Now let Y be a set of l vectors $y_1, \dots, y_l$ in $\mathbf{R}^n$ and $M_n^{\mu,Y}$ be the random matrix whose rows are $X_1^\mu, \dots, X_{n-l}^\mu, y_1, \dots, y_l$, where X_i^μ are i.i.d. copies of X^μ.

THEOREM 3.8. *Let $0 < \mu \leq 1$, and let l be a nonnegative integer. Then there is a positive constant $\varepsilon = \varepsilon(\mu, l)$ such that the following holds. For any set Y of l independent vectors from $\mathbf{R}^n$,*

$$\mathbf{P}(\sigma_n(M_n^{\mu,Y}) = 0) \leq (1-\varepsilon)^n.$$

COROLLARY 3.9. *Let $0 < \mu \leq 1$. Then there is a positive constant $\varepsilon = \varepsilon(\mu)$ such that the following holds. For any vector $y \in \mathbf{R}^n$, the probability that there are $w_1, \ldots, w_{n-1}$, not all zeros, such that*

$$y = X_1^\mu w_1 + \ldots X_{n-1}^\mu w_{n-1}$$

is at most $(1-\varepsilon)^n$.

We will prove Theorem 3.10 in Section 9 by using the machinery from [25].

4. Some quick applications of the inverse theorems

The inverse theorems provide effective bounds for counting the number of "exceptional" collections $\mathbf{v}$ of numbers with high concentration probability; see [26] for a demonstration of how such bounds can be used in applications. In this section, we present two such bounds that can be obtained from the inverse theorems developed here. In the first example, let ϵ be a positive constant and M be a large integer, and consider the following question:

How many sets $\mathbf{v}$ of n integers with absolute values at most M are there such that $\mathbb{P}_1(\mathbf{v}) \geq \epsilon$?

By Erdős' result, all but at most $O(\epsilon^{-2})$ of the elements of $\mathbf{v}$ are nonzero. Thus we have the upper bound $\binom{n}{\epsilon^{-2}}(2M+1)^{O(\epsilon^{-2})}$ for the number in question. Using Proposition 2.3, we can obtain a better bound as follows. There are only $M^{O(\ln \epsilon^{-1})}$ ways to choose the generators of the cube. After the cube is fixed, we need to choose $O(\epsilon^{-2})$ nonzero elements inside it. As the cube has volume $O(\epsilon^{-1})$, the number of ways to do this is $(\frac{1}{\epsilon})^{O(\epsilon^{-2})}$. Thus, we end up with a bound

$$M^{O(\ln \epsilon^{-1})} \left(\frac{1}{\epsilon}\right)^{O(\epsilon^{-2})}$$

which is better than the previous bound if M is considerably larger than ϵ^{-1}.

For the second application, we return to the question of bounding the singularity probability $\mathbf{P}(\sigma_n(M_n^1) = 0)$ studied in Theorem 3.7. This probability is conjectured to equal $(1/2 + o(1))^n$, but this remains open (see [26] for the latest results and some further discussion). The event that M_n^1 is singular is the same as the event that there exists some nonzero vector $v \in \mathbf{R}^n$ such that $M_n^1 v = 0$. For simplicity, we use the notation M_n instead of M_n^1 in the rest of this section. It turns out that one can obtain the optimal bound $(1/2 + o(1))^n$ if one restricts v to some special set of vectors.

Let Ω_1 be the set of vectors in $\mathbf{R}^n$ with at least $3n/\log_2 n$ coordinates. Komlós proved the following:

THEOREM 4.1. *The probability that $M_n v = 0$ for some nonzero $v \in \Omega_1$ is $(1/2 + o(1))^n$.*

A proof of this theorem can be found in Bollobás' book [2].

We are going to consider another restricted class. Let C be an arbitrary positive constant and let Ω_2 be the set of integer vectors in $\mathbf{R}^n$ where the coordinates have absolute values at most n^C. Using Theorem 2.4, we can prove

THEOREM 4.2. *The probability that $M_n v = 0$ for some nonzero $v \in \Omega_2$ is $(1/2 + o(1))^n$.*

Proof. The lower bound is trivial so we focus on the upper bound. For each nonzero vector v, let $p(v)$ be the probability that $X \cdot v = 0$, where X is a random Bernoulli vector. From independence we have $\mathbf{P}(M_n v = 0) = p(v)^n$. Since a hyperplane can contain at most 2^{n-1} vectors from $\{-1, +1\}^n$, $p(v)$ is at most $1/2$. For $j = 1, 2, \dots$, let S_j be the number of nonzero vectors v in Ω_2 such that $2^{-j-1} < p(v) \leq 2^{-j}$. Then the probability that $M_n v = 0$ for some nonzero $v \in \Omega_2$ is at most

$$\sum_{j=1}^{n} (2^{-j})^n S_j.$$

Let us now restrict the range of j. Note that if $p(v) \geq n^{-1/3}$, then by Erdős's result (mentioned in the introduction) most of the coordinates of v are zero. In this case, by Theorem 4.1 the contribution from these v is at most $(1/2+o(1))^n$. Next, since the number of vectors in Ω_2 is at most $(2n^C + 1)^n \leq n^{(C+1)n}$, we can ignore those j where $2^{-j} \leq n^{-C-2}$. Now it suffices to show that

$$\sum_{n^{-C-2} \leq 2^{-j} \leq n^{-1/3}} (2^{-j})^n S_j = o((1/2)^n).$$

For any relevant j, we can find an integer $d = O(1)$ and a positive number $\epsilon = \Omega(1)$ such that

$$n^{-(d-1/3)\epsilon} \leq 2^{-j} < n^{-(d-2/3)\epsilon}.$$

Set $k := n^\epsilon$. Thus $2^{-j} \gg k^{-d}$ and we can use Theorem 2.4 to estimate S_j. Indeed, by invoking this theorem, we see that there are at most $\binom{n}{k^2}(2n^C + 1)^{k^2} = n^{O(k^2)} = n^{o(n)}$ ways to choose the positions and values of exceptional coordinates of v. Furthermore, there are only $(2n^C+1)^{d-1} = n^{O(1)}$ ways to fix the generalized progression $P := Q(\mathbf{w}, k)$.

Note that the elements of P are polynomially bounded in n. Such integers have only $n^{o(1}$ divisors. Thus, if P is fixed any (nonexceptional) coordinate of v has at most $|P|n^{o(1)}$ possible values. This means that once P is fixed, the number of ways to set the nonexceptional coordinates of v is at most $(n^{o(1)}|P|)^n = (2k+1)^{(d-1+o(1))n}$. Putting these together,

$$S_j \leq n^{O(k^2)} k^{(d-1+o(1))n}.$$

As $k = n^{\varepsilon}$ and $2^{-j} \leq n^{-(d-2/3)\epsilon}$, it follows that

$$2^{-jn} S_j \leq n^{o(n)} n^{-\epsilon n/3} = o\left(\frac{1}{\log n}\right) 2^{-n}.$$

Since there are only $O(\log n)$ relevant j, we can conclude the proof by summing the bound over j. □

5. Properties of $\mathbb{P}_\mu(\mathbf{v})$

In order to prove the inverse Littlewood-Offord theorems in Section 2, we shall first need to develop some useful tools for estimating the quantity $\mathbb{P}_\mu(\mathbf{v})$. Note that the tools here are only used for the proof of the inverse Littlewood-Offord theorems in Section 6 and are not required elsewhere in the paper.

It is convenient to think of $\mathbf{v}$ as a word, obtained by concatenating the numbers v_i:

$$\mathbf{v} = v_1 v_2 \dots v_n.$$

This allows us to perform several operations such as concatenating, truncating and repeating. For instance, if $\mathbf{v} = v_1 \dots v_n$ and $\mathbf{w} = w_1 \dots w_m$, then

$$\mathbb{P}_\mu(\mathbf{vw}) = \max_{a \in Z} \Big(\sum_{i=1}^{n} \eta_i^\mu v_i + \sum_{j=1}^{m} \eta_{n+j}^\mu w_j = a \Big)$$

where $\eta_k^\mu, 1 \leq k \leq n+m$ are i.i.d. copies of η^μ. Furthermore, we use $\mathbf{v}^k$ to denote the concatenation of k copies of $\mathbf{v}$.

It turns out that there is a nice calculus concerning the expressions $\mathbb{P}_\mu(\mathbf{v})$, especially when μ is small. The core properties are summarized in the next lemma.

Lemma 5.1. *The following properties hold.*

- $\mathbb{P}_\mu(\mathbf{v})$ *is invariant under permutations of* $\mathbf{v}$.
- *For any words* $\mathbf{v}, \mathbf{w}$

$$\mathbb{P}_\mu(\mathbf{v})\mathbb{P}_\mu(\mathbf{w}) \leq \mathbb{P}_\mu(\mathbf{vw}) \leq \mathbb{P}_\mu(\mathbf{v}). \tag{2}$$

- *For any* $0 < \mu \leq 1$, *any* $0 < \mu' \leq \mu/4$, *and any word* $\mathbf{v}$,

$$\mathbb{P}_\mu(\mathbf{v}) \leq \mathbb{P}_{\mu'}(\mathbf{v}). \tag{3}$$

- *For any number* $0 < \mu \leq 1/2$ *and any word* $\mathbf{v}$,

$$\mathbb{P}_\mu(\mathbf{v}) \leq \mathbb{P}_{\mu/k}(\mathbf{v}^k). \tag{4}$$

- *For any number* $0 < \mu \leq 1/2$ *and any words* $\mathbf{v}, \mathbf{w}_1, \dots, \mathbf{w}_m$,

$$\mathbb{P}_\mu(\mathbf{v}\mathbf{w}_1 \dots \mathbf{w}_m) \leq \Big(\prod_{j=1}^{m} \mathbb{P}_\mu(\mathbf{v}\mathbf{w}_j^m)\Big)^{1/m}. \tag{5}$$

- *For any number* $0 < \mu \leq 1/2$ *and any words* $\mathbf{v}, \mathbf{w}_1, \dots, \mathbf{w}_m$, *there is an index* $1 \leq j \leq m$ *such that*

$$\mathbb{P}_\mu(\mathbf{v}\mathbf{w}_1 \dots \mathbf{w}_m) \leq \mathbb{P}_\mu(\mathbf{v}\mathbf{w}_j^m). \tag{6}$$

Proof. The first two properties are trivial. To verify the rest, note that from Fourier analysis

$$\mathbf{P}(\eta_1^{(\mu)} v_1 + \dots + \eta_n^{(\mu)} v_n = a) = \int_0^1 e^{-2\pi i a \xi} \prod_{j=1}^n (1 - \mu + \mu \cos(2\pi v_j \xi))\, d\xi. \tag{7}$$

When $0 < \mu \leq 1/2$, the expression $1 - \mu + \mu\cos(2\pi v_j\xi))$ is positive, and thus

$$\mathbb{P}_\mu(\mathbf{v}) = \mathbf{P}(Y_{\mu,\mathbf{v}} = 0) = \int_0^1 \prod_{j=1}^n (1 - \mu + \mu \cos(2\pi v_j \xi))\, d\xi. \tag{8}$$

To prove (3), note that for any $0 < \mu \leq 1$, $0 < \mu' \leq \mu/4$ and any θ we have the elementary inequality

$$|(1-\mu) + \mu\cos\theta| \leq (1-\mu') + \mu'\cos 2\theta.$$

Using this,

$$\begin{aligned}
\mathbb{P}_\mu(\mathbf{v}) &\leq \int_0^1 \prod_{j=1}^n |(1 - \mu + \mu \cos(2\pi v_j \xi))|\, d\xi \\
&\leq \int_0^1 \prod_{j=1}^n (1 - \mu' + \mu' \cos(4\pi v_j \xi))\, d\xi \\
&= \int_0^1 \prod_{j=1}^n (1 - \mu' + \mu' \cos(4\pi v_j \xi)) d\xi \\
&= \mathbb{P}_{\mu'}(\mathbf{v})
\end{aligned}$$

where the next to last equality follows by changing ξ to 2ξ and considering the periodicity of cosine.

Similarly, observe that for $0 < \mu \leq 1/2$ and $k \geq 1$,

$$(1 - \mu + \mu\cos(2\pi v_j\xi)) \leq \Big(1 - \frac{\mu}{k} + \frac{\mu}{k}\cos(2\pi v_j \xi)\Big)^k.$$

From the concavity of $\log(1-t)$ when $0 < t < 1$, $\log(1-t) \leq k\log(1-\frac{t}{k})$. The claim follows by exponentiating this with $t := \mu(1 - \cos(2\pi v_j\xi)))$, which proves (4).

Finally, (5) is a consequence of (8) and Hölder's inequality, while (6) follows directly from (5). $\square$

Now we consider the distribution of the equal-steps random walk $\eta_1^\mu + \cdots + \eta_m^\mu = Y_{\mu,1^m}$. Intuitively, this random walk is concentrated in an interval of length $O((1+\mu m)^{1/2})$ and has a roughly uniform distribution in the integers in this interval (though when μ is close to 1, parity considerations may cause $Y_{\mu,1^m}$ to favor the even integers over the odd ones, or vice versa); compare with the discussion in Example 2.1. The following lemma is a quantitative version of this intuition.

LEMMA 5.2. *For any* $0 < \mu \leq 1$ *and* $m \geq 1$

$$\mathbb{P}_\mu(1^m) = \sup_a \mathbf{P}(\eta_1^\mu + \cdots + \eta_m^\mu = a) = O((\mu m)^{-1/2}). \tag{9}$$

In fact, we have the more general estimate

$$\mathbf{P}(\eta_1^\mu+\cdots+\eta_m^\mu = a) = O((\tau^{-1}+(\mu m)^{-1/2})\mathbf{P}(\eta_1^\mu+\cdots+\eta_m^\mu \in [a-\tau, a+\tau]) \tag{10}$$

for any $a \in \mathbf{Z}$ *and* $\tau \geq 1$.

Finally, if $\tau \geq 1$ *and if* S *is any* τ*-separated set of integers* (*i.e. any two distinct elements of* S *are at least* τ *apart*) *then*

$$\mathbf{P}(\eta_1^\mu + \cdots + \eta_m^\mu \in S) \leq O(\tau^{-1} + (\mu m)^{-1/2}). \tag{11}$$

Proof. We first prove (9). From (3) we may assume $\mu \leq 1/4$, and then by (8)

$$\mathbb{P}_\mu(1^m) = \int_0^1 |1 - \mu + \mu\cos(2\pi\xi)|^m \, d\xi.$$

Next we use the elementary estimate

$$1 - \mu + \mu\cos(2\pi\xi) \leq \exp(-\mu\|\xi\|^2/100),$$

where $\|\xi\|$ denotes the distance to the nearest integer. This implies that $\mathbb{P}_\mu(1^m)$ is bounded from above by $\int_0^1 \exp(-\mu m\|\xi\|^2/100)d\xi$, which is of order $O((\mu m)^{-1/2})$. To see this, note that for $\xi \geq 1000(\mu m)^{-1/2}$ the function $\exp(-\mu m\|\xi\|^2/100)$ is quite small and its integral is negligible.

Now we prove (10). We may assume that $\tau \leq (\mu m)^{1/2}$, since the claim for larger τ follows automatically. By symmetry we can take $a \geq 2$.

For each integer a, let c_a denote the probability

$$c_a := \mathbf{P}(\eta_1^{(\mu)} + \cdots + \eta_m^{(\mu)} = a).$$

Direct computation (letting i denote the number of $\eta^{(\mu)}$ variables which equal zero) yields the explicit formula

$$c_a = \sum_{j=0}^m \binom{m}{j}(1-\mu)^j(\mu/2)^{m-j}\binom{m-j}{(a+m-j)/2},$$

with the convention that the binomial coefficient $\binom{a}{b}$ is zero when b is not an integer between 0 and a. This in particular yields the monotonicity property

$c_a \geq c_{a+2}$ whenever $a \geq 0$. This is already enough to yield the claim when $a > \tau$, so it remains to verify the claim when $a \leq \tau$. Now the random variable $\eta_1^\mu + \cdots + \eta_m^\mu$ is symmetric around the origin and has variance μm, so from Chebyshev's inequality we know that

$$\sum_{0 \leq a \leq 2(\mu m)^{1/2}} c_a = \Theta(1).$$

From (9) we also have $c_a = O((\mu m)^{-1/2})$ for all a. From this and the monotonicity property $c_a \geq c_{a+2}$ and the pigeonhole principle we see that $c_a = \Theta((\mu m)^{-1/2})$ either for all even $0 \leq a \leq (\mu m)^{1/2}$, or for all odd $0 \leq a \leq (\mu m)^{1/2}$. In either case, the claim (10) is easily verified. The bound in (11) then follows by summing (10) over all $a \in S$ and noting that $\sum_a c_a = 1$. □

One can also use the formula for c_a to prove (9). The simple details are left as an exercise.

6. Proofs of the inverse theorems

We now have enough machinery to prove the inverse Littlewood-Offord theorems. We first give a quick proof of Proposition 2.3:

Proof of Proposition 2.3. Suppose that the conclusion failed. Then an easy greedy algorithm argument shows that $\mathbf{v}$ must contain a dissociated subword $\mathbf{w} = (w_1, \ldots, w_{d+1})$ of length $d+1$. By (2),

$$2^{-d-1} < \mathbb{P}_1(\mathbf{v}) \leq \mathbb{P}_1(\mathbf{w}).$$

On the other hand, since $\mathbf{w}$ is dissociated, all the sums of the form $\eta_1 w_1 + \cdots + \eta_{d+1} w_{d+1}$ are distinct and so $\mathbb{P}_1(\mathbf{w}) \leq 2^{-d-1}$, yielding the desired contradiction. □

To prove Theorem 2.4, we modify the above argument by replacing the notion of dissociativity by k-dissociativity. Unfortunately this makes the proof somewhat longer:

Proof of Theorem 2.4. We construct an k-dissociated tuple $(w_1, \ldots, w_r)$ for some $0 \leq r \leq d-1$ by the following algorithm:

- Step 0. Initialize $r = 0$. In particular, $(w_1, \ldots, w_r)$ is trivially k-dissociated. From (4) we have

$$\mathbb{P}_{\mu/4d}(\mathbf{v}^d) \geq \mathbb{P}_{\mu/4}(\mathbf{v}) \geq \mathbb{P}_\mu(\mathbf{v}). \tag{12}$$

- Step 1. Count how many $1 \leq j \leq n$ there are such that $(w_1, \ldots, w_r, v_j)$ is k-dissociated. If this number is less than k^2, halt the algorithm. Otherwise, move on to Step 2.

- Step 2. Applying the last property of Lemma 5.1, we can locate a v_j such that $(w_1, \dots, w_r, v_j)$ is k-dissociated, and

$$\mathbb{P}_{\mu/4d}(\mathbf{v}^{d-r} w_1^{k^2} \dots w_r^{k^2}) \le \mathbb{P}_{\mu/4d}(\mathbf{v}^{d-r-1} w_1^{k^2} \dots w_r^{k^2} v_j^{k^2}). \tag{13}$$

  Then set $w_{r+1} := v_j$ and increase r to $r+1$. Return to Step 1. Note that $(w_1, \dots, w_r)$ remains k-dissociated, and (12) remains true.

Suppose that we terminate at some step $r \le d-1$. Then we have an r-tuple $(w_1, \dots, w_r)$ which is k-dissociated, but such that $(w_1, \dots, w_r, v_j)$ is k-dissociated for at most k^2 values of v_j. Unwinding the definitions, this shows that for all but at most k^2 values of v_j, there exists $\tau \in [1, k]$ such that $\tau v_j \in Q(\mathbf{w}, k)$, proving the claim.

It remains to show that we must indeed terminate at some step $r \le d-1$. Assume (for a contradiction) that we have reached step d. Then there exists a k-dissociated tuple $(w_1, \dots, w_d)$, and by (12), (13),

$$\mathbb{P}_\mu(\mathbf{v}) \le \mathbb{P}_{\mu/4d}(w_1^{k^2} \dots w_d^{k^2}) = \mathbf{P}(Y_{\mu/4d, w_1^{k^2} \dots w_d^{k^2}} = 0).$$

Let $\Gamma \subset \mathbf{Z}^d$ be the lattice

$$\Gamma := \{(m_1, \dots, m_d) \in \mathbf{Z}^d : m_1 w_1 + \dots + m_d w_d = 0\}.$$

By using independence we can write

$$\mathbb{P}_\mu(\mathbf{v}) \le \mathbf{P}(Y_{\mu/4d, w_1^{k^2} \dots w_d^{k^2}} = 0) = \sum_{(m_1, \dots, m_d) \in \Gamma} \prod_{j=1}^d \mathbf{P}(Y_{\mu/4d, 1^{k^2}} = m_j). \tag{14}$$

Now we use a volume packing argument. From Lemma 5.2,

$$\mathbf{P}(Y_{\mu/4d, 1^{k^2}} = m) = O_{\mu,d}\left(\frac{1}{k} \sum_{m' \in m + (-k/2, k/2)} \mathbf{P}(Y_{\mu/4d, 1^{k^2}} = m')\right)$$

and hence from (14),

$$\mathbb{P}_\mu(\mathbf{v}) \le O_{\mu,d}\Bigg(k^{-d} \sum_{(m_1, \dots, m_d) \in \Gamma} \sum_{(m_1', \dots, m_d') \in (m_1, \dots, m_d) + (-k/2, k/2)^d} \prod_{j=1}^d \mathbf{P}(Y_{\mu/4d, 1^{k^2}} = m_j')\Bigg).$$

Since $(w_1, \dots, w_d)$ is k-dissociated, all the $(m_1', \dots, m_d')$ tuples in

$$\Gamma + (-k/2, k/2)^d$$

are different. Thus, we conclude

$$\mathbb{P}_\mu(\mathbf{v}) \le O_{\mu,d}\left(k^{-d} \sum_{(m_1, \dots, m_d) \in \mathbf{Z}^d} \prod_{j=1}^d \mathbf{P}(Y_{\mu/4d, 1^{k^2}} = m_j)\right).$$

But from the union bound

$$\sum_{(m_1,\dots,m_d)\in\mathbf{Z}^d}\prod_{j=1}^{d}\mathbf{P}(Y_{\mu/4d,1^{k^2}}=m_j)=1,$$

and so

$$\mathbb{P}_\mu(\mathbf{v})\leq O_{\mu,d}(k^{-d}).$$

To complete the proof, set the constant $C=C(\mu,d)$ in the theorem to be larger than the hidden constant in $O_{\mu,d}(k^{-d})$. □

Remark 6.1. One can also use the Chernoff bound and obtain a shorter proof (avoiding the volume packing argument) but with an extra logarithmic loss in the estimates.

Finally we perform some additional arguments to eliminate the $\frac{1}{\tau}$ dilations in Theorem 2.4 and obtain our final inverse Littlewood-Offord theorem. The key will be the following lemma.

Given a set S and a number v, the torsion of v with respect to S is the smallest positive integer τ such that $\tau v\in S$. If such τ does not exists, we say that v has infinite torsion with respect to S.

The key new ingredient will be the following lemma, which asserts that adding a high torsion element to a random walk reduces the concentration probability significantly.

LEMMA 6.2 (Torsion implies dispersion). *Let $0<\mu\leq 1$ and consider a* GAP $Q:=\{\sum_{i=1}^{d}x_iW_i|-L_i\leq x_i\leq L_i\}$. *Assume that W_{d+1} has finite torsion τ with respect to $2Q$. Then there is a constant C_μ depending only on μ such that*

$$\mathbb{P}_\mu(W_1^{L_1}\dots W_d^{L_d}W_{d+1}^{\tau^2})\leq C_\mu\tau^{-1}\mathbb{P}_\mu(W_1^{L_1}\dots W_d^{L_d}).$$

Proof. Let a be an integer such that

$$\mathbb{P}_\mu(W_1^{L_1}\dots W_d^{L_d}W_{d+1}^{\tau^2})=\mathbf{P}\left(\sum_{i=1}^{d}W_i\sum_{j=1}^{L_i}\eta_{j,i}^{\mu}+W_{d+1}\sum_{j=1}^{\tau^2}\eta_{j,d+1}^{\mu}=a\right),$$

where the $\eta_{j,i}^{\mu}$ are i.i.d. copies of η^μ. It suffices to show that

$$\mathbf{P}\left(\sum_{i=1}^{d}W_i\sum_{j=1}^{L_i}\eta_{j,i}^{\mu}+W_{d+1}\sum_{j=1}^{\tau^2}\eta_{j,d+1}^{\mu}=a\right)=O_\mu(\tau^{-1})\mathbb{P}_\mu(W_1^{L_1}\dots W_d^{L_d}).$$

Let S be the set of all $m\in[-\tau^2,\tau^2]$ such that $Q+mW_{d+1}$ contains a. Observe that in order for $\sum_{i=1}^{d}W_i\sum_{j=1}^{L_i}\eta_{j,i}^{\mu}+W_{d+1}\sum_{j=1}^{\tau^2}\eta_{j,d+1}^{\mu}$ to equal a, the quantity $\sum_{j=1}^{k} \eta_{j,d+1}^{\mu}$ must lie in S. By the definition of $\mathbb{P}_\mu(W_1^{L_1} \dots W_d^{L_d})$ and Bayes identity, we conclude

$$\mathbf{P}\left(\sum_{i=1}^{d} W_i \sum_{j=1}^{L_i} \eta_{j,i}^{\mu} + W_{d+1} \sum_{j=1}^{\tau^2} \eta_{j,d+1}^{\mu} = a\right) \le \mathbb{P}_\mu(W_1^{L_1} \dots W_d^{L_d}) \mathbf{P}\left(\sum_{j=1}^{\tau^2} \eta_{j,d+1}^{\mu} \in S\right).$$

Consider two elements $x, y \in S$. By the definition of S, $(x-y)v \in Q - Q = 2Q$. From the definition of τ, $|x-y|$ is either zero or at least τ. This implies that S is τ-separated and the claim now follows from Lemma 5.2. □

The following technical lemma is also needed.

LEMMA 6.3. *Consider a* GAP $Q(\mathbf{w}, L)$. *Assume that* v *is an element with (finite) torsion* τ *with respect to* $Q(\mathbf{w}, L)$. *Then*

$$Q(\mathbf{w}, L) + Q(v, L') \subset \frac{1}{\tau} \cdot Q(\mathbf{w}, L(L' + \tau)).$$

Proof. Assume $\mathbf{w} = w_1 \dots w_r$. We can write v as $\frac{1}{\tau} \sum_{i=1}^{r} a_i w_i$, where $|a_i| \le L$. An element y in $Q(\mathbf{w}, L) + Q(v, L')$ can be written as

$$y = \sum_{i=1}^{r} x_i w_i + xv$$

where $|x_i| \le L$ and $|x| \le L'$. Substituting v,

$$y = \sum_{i=1}^{r} x_i w_i + x \frac{1}{\tau} \sum_{i=1}^{r} a_i w_i = \frac{1}{\tau} \sum_{i=1}^{r} w_i (\tau x_i + x a_i),$$

where $|\tau x_i + x a_i| \le \tau L + L' L$. This concludes the proof. □

Proof of Theorem 2.5. We begin by running the algorithm in the proof of Theorem 2.4 to locate a word $\mathbf{w}$ of length at most $d-1$ such that the set $\bigcup_{1 \le \tau \le k} \frac{1}{\tau} \cdot Q(\mathbf{w}, k)$ covers all but at most k^2 elements of $\mathbf{v}$. Set $\mathbf{v}^{[0]}$ to be the word formed by removing the (at most k^2) exceptional elements from $\mathbf{v}$ which do not lie in $\bigcup_{1 \le \tau \le k} \frac{1}{\tau} \cdot Q(\mathbf{w}, k)$.

By increasing the constant k_0 in the assumption of the theorem, we can assume, in all arguments below, that k is sufficiently large, whenever needed.

By (2) and (3)

(15)
$$\mathbb{P}_{\mu/4d}(\mathbf{v}^{[0]} \mathbf{w}^{k^2}) \ge \mathbb{P}_{\mu/4d}(\mathbf{v} \mathbf{w}^{k^2}) \ge \mathbb{P}_{\mu/4d}(\mathbf{v}) \mathbb{P}_{\mu/4d}(\mathbf{v} \mathbf{w}^{k^2}) \ge k^{-d} \mathbb{P}_{\mu/4d}(\mathbf{v} \mathbf{w}^{k^2}).$$

In the following, assume that there is at least one nonzero entry in $\mathbf{w}$; otherwise the claim is trivial.

Now we perform an additional algorithm. Let $K = K(\mu, d, \epsilon) > 2$ be a large constant to be chosen later.

- Step 0. Initialize $i = 0$ and set $Q_0 := Q(\mathbf{w}, k^2)$ and $\mathbf{v}^{[0]}$ as above.
- Step 1. Count how many $v \in \mathbf{v}^{[i-1]}$ having torsion at least K with respect to $2Q_{i-1}$. (We need to have the factor 2 here in order to apply Lemma 6.2.) If this number is less than k^2, halt the algorithm. Otherwise, move on to Step 2.
- Step 2. Locate a multiset S of k^2 elements of $\mathbf{v}^{[i-1]}$ with torsion at least K with respect to $2Q_{i-1}$. Applying (6), we can find an element $v \in S$ such that

$$\mathbb{P}_{\mu/4d}(\mathbf{v}^{[i-1]}\mathbf{w}^{k^2} W_1^{\tau_1^2} \dots W_{i-1}^{\tau_{i-1}^2}) \leq \mathbb{P}_{\mu/4d}(\mathbf{v}^{[i]}\mathbf{w}^{k^2} W_1^{\tau_1^2} \dots W_{i-1}^{\tau_{i-1}^2} v^{k^2})$$

  where $\mathbf{v}^{[i]}$ is obtained from $\mathbf{v}^{[i-1]}$ by deleting S.
  Let τ_i be the torsion of v with respect to $2Q_{i-1}$. Since every element of $\mathbf{v}^{[0]}$ has torsion at most k with respect to Q_0, $K \leq \tau_i \leq k$. We then set $W_i := v$, $Q_i := Q_{i-1} + Q(W_i, \tau_i^2)$, increase i to $i+1$ and return to Step 1.

Consider a stage i of the algorithm. From construction and induction and (15), we have a word $W_1 \dots W_i$ with

$$\mathbb{P}_{\mu/4d}(\mathbf{v}^{[i]}\mathbf{w}^{k^2} W_1^{\tau_1^2} \dots W_i^{\tau_i^2}) \geq \mathbb{P}(\mathbf{v}^{[0]}\mathbf{w}^{k^2}) \geq k^{-d}\mathbb{P}(\mathbf{w}^{k^2}).$$

On the other hand, by applying Lemma 6.2 iteratively,

$$\mathbb{P}_{\mu/4d}(\mathbf{w}^{k^2} W_1^{\tau_1^2} \dots W_i^{\tau_i^2}) \leq \mathbb{P}_{\mu/4d}(\mathbf{w}^{k^2}) \prod_{j=1}^{i} (C_\mu \tau_j^{-1}).$$

It follows that $\prod_{j=1}^i (C_\mu \tau_j^{-1}) \geq k^{-d}$, or equivalently $\prod_{j=1}^i (C_\mu^{-1}\tau_j) \leq k^d$. Recall that $\tau_j \geq K$. Thus by setting K sufficiently large (compared to C_μ, d and $1/\epsilon$), we can guarantee that

$$\prod_{j=1}^{i} \tau_j \leq k^{d+\epsilon/2d} \tag{16}$$

where ϵ is the constant in the assumption of the theorem. It also follows that the algorithm must terminate at some stage $D \leq \log_K k^{d+\epsilon/2d} \leq (d+1)\log_K k$.

Now look at the final set Q_D. Applying Lemma 6.3 iteratively,

$$Q_D \subset \left(\prod_{j=1}^{D} \frac{1}{\tau_j}\right) \cdot Q(\mathbf{w}, L_D)$$

where $L_0 := k^2$ and

$$L_i := L_{i-1}(\tau_i + \tau_i^2) \leq (1 + 1/K) L_{i-1}\tau_i^2. \tag{17}$$

We now show that the GAP $Q := \frac{1}{K!} \cdot (2K!) Q(\mathbf{w}, L_D) = \frac{1}{K!} \cdot Q(\mathbf{w}, 2K! L_D)$ satisfies the claims of the theorem.

- (*Rank*) We have $\operatorname{rank}(Q) = \operatorname{rank}(Q(\mathbf{w}, L_D)) = \operatorname{rank}(Q_0) = r \le d-1$, as shown in the proof of the previous theorem.
- (*Volume*) We have $\operatorname{Vol}(Q) = (2K!)^r \operatorname{Vol}(Q(\mathbf{w}, L_D)) = O(\operatorname{Vol}(Q(\mathbf{w}, L_D)))$. On the other hand, by (16) and (17)

$$\operatorname{Vol}(Q(\mathbf{w}, L_D)) = (2L_D+1)^r \le (3L_D)^r = O\left(\left(k^2 \prod_{j=1}^{D} (1+1/K)\tau_j^2\right)^r\right)$$
$$= O\left(\left(k^{2+2(d+\epsilon/2d)}(1+K)^D\right)^r\right).$$

By definition, $D \le \log_K k^{d+\epsilon/2d} < \log k$, given that K is sufficiently large compared to d. Thus $(1+1/K)^D \le \exp(D/K) \le k^{1/K}$ which implies that

$$\operatorname{Vol}(Q(\mathbf{w}, L_D)) = O(k^{r(2+2(d+\epsilon/2d)+1/K)}) = o(k^{2(d^2-1)+\epsilon})$$

provided that $r \le d-1$ and K is sufficiently large compared to d and $1/\epsilon$. (The asymptotic notation here is used under the assumption that $k \to \infty$.)
- (*Number of exceptional elements*) At each stage in the second algorithm, we discard a set of k^2 elements; thus all but $Dk^2 \le (d+1)k^2 \log_K k)$ elements of $\mathbf{v}^{[0]}$ have torsion at most K with respect to $2Q_D$. As $Q_D \subset Q(\mathbf{w}, L_D)$ and $v \backslash v^{[0]} \le k^2$, it follows that all but at most

$$(d+1)k^2 \log_K k + k^2$$

elements of $\mathbf{v}$ have torsion at most K with respect to

$$2Q(\mathbf{w}, L_D) = Q(\mathbf{w}, 2L_D).$$

By setting K sufficiently large compared to d and $1/\epsilon$, we can guarantee that

$$(d+1)k^2 \log_K k + k^2 \le \epsilon k^2 \log k.$$

To conclude, note that any element with torsion at most K with respect to $Q(\mathbf{w}, 2L_D)$ belongs to $Q := \frac{1}{K!} \cdot Q(\mathbf{w}, 2K!L_D)$. Thus, Q contains all but at most $\epsilon k^2 \log k$ elements of $\mathbf{v}$.
- (*Generators*) The generators of $\frac{1}{K!} \cdot Q(\mathbf{w}, 2K!L_D)$ are $\frac{1}{K! \prod_{j=1}^{D} \tau_j} w_i$, $1 \le i \le r$. Since $w_i \in \mathbf{v}$ and $\prod_{j=1}^{D} \tau_j \le k^{d+\epsilon/2d} = o(k^{d+\epsilon})$, the claim about generators follows.

The proof is complete. □

7. The smallest singular value

In this section, we prove Theorem 3.4, modulo two key results, Theorem 3.6 and Corollary 3.9, which will be proved in later sections.

Let $B10$ be a large number (depending on A) to be chosen later. Suppose that $\sigma_n(M_n^\mu) < n^{-B}$. This means that there exists a unit vector v such that

$$\|M_n^\mu v\| < n^{-B}.$$

By rounding each coordinate v to the nearest multiple of n^{-B-2}, we can find a vector $\tilde{v} \in n^{-B-2} \cdot \mathbf{Z}^n$ of magnitude $0.9 \le \|\tilde{v}\| \le 1.1$ such that

$$\|M_n^\mu \tilde{v}\| \le 2n^{-B}.$$

Thus, writing $w := n^{B+2}\tilde{v}$, we can find an integer vector $w \in \mathbf{Z}^n$ of magnitude $0.9n^{B+2} \le \|w\| \le 1.1n^{B+2}$ such that

$$\|M_n^\mu w\| \le 2n^2.$$

Let Ω be the set of integer vectors $w \in \mathbf{Z}^n$ of magnitude $0.9n^{B+2} \le \|w\| \le 1.1n^{B+2}$. It suffices to show the probability bound

$$\mathbf{P}(\text{there is some } w \in \Omega \text{ such that } \|M_n^\mu w\| \le 2n^2) = O_{A,\mu}(n^{-A}).$$

We now partition the elements $w = (w_1, \ldots, w_n)$ of Ω into three sets:

- We say that w is *rich* if

  $$\mathbb{P}_\mu(w_1 \ldots w_n) \ge n^{-A-10}$$

  and *poor* otherwise. Let Ω_1 be the set of poor w's.

- A rich w is *singular* w if fewer than $n^{0.2}$ of its coordinates have absolute value n^{B-10} or greater. Let Ω_2 be the set of rich and singular w's.

- A rich w is *nonsingular* w, if at least $n^{0.2}$ of its coordinates have absolute value n^{B-10} or greater. Let Ω_3 be the set of rich and nonsingular w's.

The desired estimate follows directly from the following lemmas and the union bound.

Lemma 7.1 (Estimate for poor w).

$$\mathbf{P}(\textit{there is some } w \in \Omega_1 \textit{ such that } \|M_n^\mu w\| \le 2n^2) = o(n^{-A}).$$

Lemma 7.2 (Estimate for rich singular w).

$$\mathbf{P}(\textit{there is some } w \in \Omega_2 \textit{ such that } \|M_n^\mu w\| \le 2n^2) = o(n^{-A}).$$

Lemma 7.3 (Estimate for rich nonsingular w).

$$\mathbf{P}(\textit{there is some } w \in \Omega_3 \textit{ such that } \|M_n^\mu w\| \le 2n^2) = o(n^{-A}).$$

Remark 7.4. Our arguments will show that the probabilities in Lemmas 7.2 and 7.3 are exponentially small.

The proofs of Lemmas 7.1 and 7.2 are relatively simple and rely on well-known methods. We delay these proofs to the end of this section and focus on the proof of Lemma 7.3, which is the heart of the matter, and which uses all the major tools discussed in previous sections.

Proof of Lemma 7.3. Informally, the strategy is to use the inverse Littlewood-Offord theorem (Corollary 2.7) to place the integers $w_1, \dots, w_n$ in a progression, which we then discretize using Theorem 3.6. This allows us to replace the event $\|M_n^\mu w\| \leq 2n^2$ by the discretized event $M_n^{\mu,Y} = 0$ for a suitable Y, at which point we apply Corollary 3.9.

We turn to the details. Since w is rich, we see from Corollary 2.7 that there exists a symmetric GAP Q of integers of rank at most A' and volume at most $n^{A'}$ which contains all but $\lfloor n^{0.1} \rfloor$ of the integers $w_1, \dots, w_n$, where A' is a constant depending on μ and A. Also the generators of Q are of the form w_i/s for some $1 \leq i \leq n$ and $1 \leq s \leq n^{A'}$.

Using the description of Q and the fact that $w_1, \dots, w_n$ are polynomially bounded (in n), it is easy to derive that the total number of possible Q is $n^{O_{A'}(1)}$. Next, by paying a factor of

$$\binom{n}{\lfloor n^{0.1} \rfloor} \leq n^{\lfloor n^{0.1} \rfloor} = \exp(o(n))$$

we may assume that it is the last $\lfloor n^{0.1} \rfloor$ integers $w_{m+1}, \dots, w_n$ which possibly lie outside Q, where we set $m := n - \lfloor n^{0.1} \rfloor$. As each of the w_i has absolute value at most $1.1n^{B+2}$, the number of ways to fix these exceptional elements is at most $(2.2n^{B+2})^{n^{0.1}} = \exp(o(n))$. Overall, it costs a factor of at most $\exp(o(n))$ to fix Q and the positions and values of the exceptional elements of w.

Once we have fixed $w_{m+1}, \dots, w_n$, we can then write

$$M_n w = w_1 X_1^\mu + \cdots + w_m X_m^\mu + Y,$$

where Y is a random variable determined by X_i^μ and w_i, $m < i \leq n$. (In this proof we think of X_i^μ as the column vectors of the matrix.) For any number y, let F_y be the event that there exists $w_1, \dots, w_m$ in Q, where at least one of the w_i has absolute value larger or equal n^{B-10}, such that

$$|w_1 X_1^\mu + \cdots + w_m X_m^\mu + y| \leq 2n^2.$$

It suffices to prove that

$$\mathbf{P}(F_y) = o(n^{-A})$$

for any y. Our argument will in fact show that this probability is exponentially small.

We now apply Theorem 3.6 to the GAP Q with $R_0 := n^{B/2}$ and $S := n^{10}$ to find a scale $R = n^{B/2+O_A(1)}$ and symmetric GAPs Q_{sparse}, Q_{small} of rank at most A' and volume at most $n^{A'}$ such that:

- $Q \subseteq Q_{\mathrm{sparse}} + Q_{\mathrm{small}}$.
- $Q_{\mathrm{small}} \subseteq [-n^{-10}R, n^{-10}R]$.
- The elements of $n^{10}Q_{\mathrm{sparse}}$ are $n^{10}R$-separated.

Since Q (and hence $n^{10}Q$) contains $w_1, \dots, w_m$, we can write

$$w_j = w_j^{\mathrm{sparse}} + w_j^{\mathrm{small}}$$

for all $1 \leq j \leq m$, where $w_j^{\mathrm{sparse}} \in Q_{\mathrm{sparse}}$ and $w_j^{\mathrm{small}} \in Q_{\mathrm{small}}$. In fact, this decomposition is unique.

Suppose that the event F_y holds. Writing $X_i^\mu = (\eta_{i,1}^\mu, \dots, \eta_{i,n}^\mu)$ (where $\eta_{i,j}^\mu$ are i.i.d. copies of η^μ) and $y = (y_1, \dots, y_n)$,

$$w_1\eta_{i,1}^\mu + \cdots + w_m\eta_{i,m}^\mu = y_i + O(n^2)$$

for all $1 \leq i \leq n$. Splitting the w_j into sparse and small components and estimating the small components using the triangle inequality, we obtain

$$w_1^{\mathrm{sparse}}\eta_{i,1}^\mu + \cdots + w_m^{\mathrm{sparse}}\eta_{i,m}^\mu = y_i + O(n^{-9}R)$$

for all $1 \leq i \leq n$. Note that the left-hand side lies in $mQ_{\mathrm{sparse}} \subset n^{10}Q_{\mathrm{sparse}}$, which is known to be $n^{10}R$-separated. Thus there is a unique value for the right-hand side, denoted as y_i', which depends only on y and Q such that

$$w_1^{\mathrm{sparse}}\eta_{i,1} + \cdots + w_m^{\mathrm{sparse}}\eta_{i,m} = y_i'.$$

The point is that now we have eliminated the $O()$ errors, and thus have essentially converted the singular value problem to the zero determinant problem. Note also that since one of the $w_1, \dots, w_m$ is known to have magnitude at least n^{B-10} (which will be much larger than $n^{10}R$ if B is chosen large depending on A), we see that at least one of the $w_1^{\mathrm{sparse}}, \dots, w_n^{\mathrm{sparse}}$ is nonzero.

Consider the random matrix M' of order $m \times m+1$ whose entries are i.i.d. copies of η^μ and let $y' \in \mathbf{R}^{m+1}$ be the column vector $y' = (y_1', \dots, y_{m+1}')$. We conclude that if the event F_y holds, then there exists a nonzero vector $w \in \mathbf{R}^m$ such that $M'w = y'$. But from Corollary 3.9, this holds with the desired probability

$$\exp(-\Omega(m+1)) = \exp(-\Omega(n)) = o(n^{-A})$$

and we are done. □

Proof of Lemma 7.1. We use a conditioning argument, following [20]. (An argument of the same spirit was used by Komlós to prove the bound $O(n^{-1/2})$ for the singularity problem [2].)

Let M be a matrix such that there is $w \in \Omega_1$ satisfying $\|Mw\| \leq 2n^2$. Since M and its transpose have the same spectral norm, there is a vector w' which has the same norm as w such that $\|w'M\| \leq 2n^2$. Let $u = w'M$ and X_i be the row vectors of M. Then

$$u = \sum_{i=1}^{n} w'_i X_i$$

where w'_i are the coordinates of w'.

Now we think of M as a random matrix. By paying a factor of n, we can assume that w'_n has the largest absolute value among the w'_i. We expose the first $n-1$ rows $X_1, \ldots, X_{n-1}$ of M. If there is $w \in \Omega_1$ satisfying $\|Mw\| \leq 2n^2$, then there is a vector $y \in \Omega_1$, depending only on the first $n-1$ rows such that

$$\Big(\sum_{i=1}^{n-1} (X_i \cdot y)^2\Big)^{1/2} \leq 2n^2.$$

Now consider the inner product $X_n \cdot y$. We can write X_n as

$$X_n = \frac{1}{w'_n}\Big(u - \sum_{i=1}^{n-1} w'_i X_i\Big).$$

Thus,

$$|X_n \cdot y| = \frac{1}{\|w'_n\|} |u \cdot y - \sum_{i=1}^{n-1} w'_i X_i \cdot y|.$$

The right-hand side, by the triangle inequality, is at most

$$\frac{1}{\|w'_n\|}(\|u\|\|y\| + \|w'\|\Big(\sum_{i=1}^{n-1} (X_i \cdot y)^2)^{1/2}\Big).$$

By assumption $\|w'_n\| \geq n^{-1/2}\|w'\|$. Furthermore, as $\|u\| \leq 2n^2$, $\|u\|\|y\| \leq 2n^2\|y\| \leq 3n^2\|w'\|$ as $\|w'\| = \|w\|$, and both y and w belong to Ω_1. (Any two vectors in Ω_1 have roughly the same length.) Finally $(\sum_{i=1}^{n-1}(X_i \cdot y)^2)^{1/2} \leq 2n^2$. Putting all these together,

$$|X_n \cdot y| \leq 5n^{5/2}.$$

Recall that y is fixed (after we expose the first $n-1$ rows) and X_n is a copy of X^μ. The probability that $|X^\mu \cdot y| \leq 5n^{5/2}$ is at most $(10n^{5/2}+1)\mathbb{P}_\mu(y)$. On the other hand, y is poor, and so $\mathbb{P}_\mu(y) \leq n^{-A-10}$. Thus, it follows that

$$\begin{aligned}
&\mathbf{P}(\text{there is some } w \in \Omega_1 \text{ such that } \|M_n^\mu w\| \leq 2n^2) \\
&\quad \leq n^{-A-10}(10n^{5/2}+1)n = o(n^{-A}),
\end{aligned}$$

where the extra factor n comes from the assumption that w'_n has the largest absolute value. This completes the proof. □

Proof of Lemma 7.2. We use an argument from [15]. The key point will be that the set Ω_2 of rich nonsingular vectors has sufficiently low entropy so that one can proceed using the union bound.

A set N of vectors on the n-dimensional unit sphere S_{n-1} is said to be an *ϵ-net* if for any $x \in S_{n-1}$, there is $y \in N$ such that $\|x-y\| \leq \epsilon$. A standard greedy argument shows the following:

LEMMA 7.5. *For any n and $\epsilon \leq 1$, there exists an ϵ-net of cardinality at most $O(1/\varepsilon)^n$.*

Next, a simple concentration of measure argument shows

LEMMA 7.6. *For any fixed vector y of magnitude between* 0.9 *and* 1.1

$$\mathbf{P}(\|M_n^\mu y\| \leq n^{-2}) = \exp(-\Omega(n)).$$

It suffices to verify this statement for the case $|y|=1$. Note that

$$\|M_n^\mu y\|^2 = \sum_{i=1}^n (X_i \cdot y)^2 = \sum_{i=1}^n Z_i$$

where $Z_i = (X_i \cdot y)^2$. The Z_i are i.i.d. random variables with expectation μ and bounded variance. Thus $\sum_{i=1}^n Z_i$ has mean $\Omega(n)$ and the claimed bound follows from Chernoff's large deviation inequality (see, e.g., [28, Ch. 1]). (In fact, one can replace the n^{-2} by $cn^{1/2}$ for some small constant c, but this refinement is not necessary.)

For a vector $w \in \Omega_2$, let w' be its normalization $w' := w/\|w\|$. Thus, w' is a unit vector with at most $n^{0.2}$ coordinates with absolute values larger or equal n^{-10}. Let Ω_2' be the collection of those w' with this property.

If $\|Mw\| \leq 2n^2$ for some $w \in \Omega_2$, then $\|Mw'\| \leq 3n^{-B}$, as $\|w\| \geq 0.9n^{B+2}$. Thus, it suffices to give an exponential bound on the event that there is $w' \in \Omega_2'$ such that $\|M_n^\mu w'\| \leq 3n^{-B}$.

By paying a factor $\binom{n}{n^{0.2}} = \exp(o(n))$ in probability, we can assume that the large coordinates (with absolute value at least n^{-10}) are among the first $l := n^{0.2}$ coordinates. Consider an n^{-3}-net N in S_{l-1}. For each vector $y \in N$, let y' be the n-dimensional vector obtained from y by letting the last $n-l$ coordinates be zeros, and let N' be the set of all such vectors obtained. These vectors have magnitude between 0.9 and 1.1, and from Lemma 7.5, $|N'| \leq O(n^3)^l$.

Now consider a rich singular vector $w' \in \Omega_2$ and let w'' be the l-dimensional vector formed by the first l coordinates of this vector. Since the remaining coordinates are small, $\|w''\| = 1 + O(n^{-9.5})$. There is a vector $y \in N$ such that

$$\|y - w''\| \leq n^{-3} + O(n^{-9.5}).$$

It follows that there is a vector $y' \in N'$ such that

$$\|y' - w'\| \leq n^{-3} + O(n^{-9.5}) \leq 2n^{-3}.$$

For any matrix M of norm at most n,

$$\|Mw'\| \geq \|My'\| - 2n^{-3}n = \|My'\| - 2n^{-2}.$$

It follows that if $\|Mw'\| \leq 3n^{-B}$ for some $B \geq 2$, then $\|My'\| \leq 5n^{-2}$. Now take $M = M_n^\mu$. For each fixed y', the probability that $\|My'\| \leq 5n^{-2}$ is at most $\exp(-\Omega(n))$, by Lemma 7.6. Furthermore, the number of y' is subexponential (at most $O(n^3)^l = O(n)^{3n^{.2}} = \exp(o(n))$). Thus the claim follows directly by the union bound. $\square$

8. Discretization of progressions

The purpose of this section is to prove Theorem 3.6. The arguments here are elementary (based mostly on the pigeonhole principle and linear algebra, in particular Cramer's rule) and can be read independently of the rest of the paper.

We shall follow the informal strategy outlined in Section 3.1. We begin with a preliminary observation, which asserts the intuitive fact that progressions do not contain large lacunary subsets.

LEMMA 8.1. *Let $P \subset \mathbf{Z}$ be a symmetric generalized arithmetic progression of rank d and volume V, and let $x_1, \ldots, x_{d+1}$ be nonzero elements of P. Then there exist $1 \leq i < j \leq d+1$ such that*

$$C_d^{-1} V^{-1} |x_i| \leq |x_j| \leq C_d V |x_i|$$

for some constant $C_d > 0$ depending only on d.

Proof. We may order $|x_{d+1}| \geq |x_d| \geq \cdots \geq |x_1|$. If we write

$$P = \{m_1 v_1 + \cdots + m_d v_d : |m_i| \leq M_i \text{ for all } 1 \leq i \leq d\}$$

(so that $V = \Theta_d(M_1 \ldots M_d)$), then each of the $x_1, \ldots, x_{d+1}$ can be written as a linear combination of the $v_1, \ldots, v_d$. Applying Cramer's rule, we conclude that there exists a nontrivial relation

$$a_1 x_1 + \cdots + a_{d+1} x_{d+1} = 0$$

where $a_1, \ldots, a_{d+1} = O_d(V)$ are integers, not all zero. If we let j be the largest index such that a_j is nonzero, then $j > 1$ (since x_1 is nonzero) and in particular, we conclude that

$$|x_j| = O(|a_j x_j|) = O_d(V |x_{j-1}|)$$

from which the claim follows. $\square$

Proof of Theorem 3.6. We can assume that R_0 is very large compared to $(SV)^{O_d(1)}$ since otherwise the claim is trivial (take $P_{\text{sparse}} := P$ and $P_{\text{small}} := \{0\}$). We can also take $V \geq 2$.

Let $B = B_d$ be a large integer depending only on d to be chosen later. The first step is to subdivide the interval $[(SV)^{-B^{B+2}}R_0, (SV)^{B^{B+2}}R_0]$ into $\Theta(B)$ overlapping subintervals of the form $[(SV)^{-B^{B+1}}R, (SV)^{B^{B+1}}R]$, with every integer being contained in at most $O(1)$ of the subintervals. From Lemma 8.1 and the pigeonhole principle we see that at most $O_d(1)$ of the intervals can contain an element of $(SV)^{B^B}P$ (which has volume $O((SV)^{O_d(B^B)})$. If we let B be sufficiently large, we can thus find an interval $[(SV)^{-B^{B+1}}R, (SV)^{B^{B+1}}R]$ which is disjoint from $(SV)^{B^B}P$. Since P is symmetric, this means that every $x \in (SV)^{B^B}P$ is either larger than $(SV)^{B^{B+1}}R$ in magnitude, or smaller than $(SV)^{-B^{B+1}}R$ in magnitude.

Having located a good scale R to discretize, we now split P into small ($\ll R$) and sparse ($\gg R$-separated) components. We write P explicitly as

$$P = \{m_1v_1 + \cdots + m_dv_d : |m_i| \leq M_i \text{ for all } 1 \leq i \leq d\}$$

so that $V = \Theta_d(M_1 \dots M_d)$ and more generally

$$kP = \{m_1v_1 + \cdots + m_dv_d : |m_i| \leq kM_i \text{ for all } 1 \leq i \leq d\}$$

for any $k \geq 1$. For any $1 \leq s \leq B$, let $A_s \subset \mathbf{Z}^d$ denote the set

$$A_s := \{(m_1, \dots, m_d) : |m_i| \leq V^{B^s}M_i \text{ for all } 1 \leq i \leq d; \\ |m_1v_1 + \cdots + m_dv_d| \leq (SV)^{-B^{B+1}}R\}.$$

Roughly speaking, this space corresponds to the kernel of Φ as discussed in Section 3.1; the additional parameter s is a technicality needed to compensate for the fact that boxes, unlike vector spaces, are not quite closed under dilations. We now view A_s as a subset of the Euclidean space $\mathbf{R}^d$. As such it spans a vector space $X_s \subset \mathbf{R}^d$. Clearly

$$X_1 \subseteq X_2 \subseteq \cdots \subseteq X_B.$$

Therefore if B is large enough, by the pigeonhole principle (applied to the dimensions of these vector spaces) we can find $1 \leq s < B$ such that we have the stabilization property $X_s = X_{s+1}$. Let the dimension of this space be r; thus $0 \leq r \leq d$.

There are two cases, depending on whether $r = d$ or $r < d$. Suppose first that $r = d$ (so the kernel has maximal dimension). Then by definition of A_s we have d "equations" in d unknowns,

$$m_1^{(j)}v_1 + \cdots + m_d^{(j)}v_d = O((SV)^{-B^{B+1}}R) \text{ for all } 1 \leq j \leq d,$$

where $m_i^{(j)} = O(M_iV^{B^s})$ and the vectors $(m_1^{(j)}, \dots, m_d^{(j)}) \in A_s$ are linearly independent as j varies. Using Cramer's rule we conclude that

$$v_i = O_d((SV)^{O_d(B^s)}(SV)^{-B^{B+1}}R) \text{ for all } 1 \leq j \leq d$$

since all the determinants and minors which arise from Cramer's rule are integers that vary from 1 to $O_d(V^{O_d(B)})$ in magnitude. Since $M_i = O(V)$ for all i, we conclude that $x = O_d(V^{O_d(B^s)}(SV)^{-B^{B+1}}R)$ for all $x \in P$, which by construction of R (and the fact that $s < B$) shows that

$$P \subset [-(SV)^{-B^{B+1}}R, (SV)^{-B^{B+1}}R]$$

(if B is sufficiently large). Thus in this case we can take $P_{\text{small}} = P$ and $P_{\text{sparse}} = \{0\}$.

Now we consider the case when $r < d$ (so the kernel is proper). In this case we can write X_s as a graph of some linear transformation $T : \mathbf{R}^r \to \mathbf{R}^{d-r}$: after permutation of the coordinates, we have

$$X_s = \{(x, Tx) \in \mathbf{R}^r \times \mathbf{R}^{d-r} : x \in \mathbf{R}^r\}.$$

The coefficients of T form an $r \times d - r$ matrix, which can be computed by Cramer's rule to be rational numbers with numerator and denominator $O_d((SV)^{O_d(B^s)})$; this follows from X_s being spanned by A_s, and on the integrality and size bounds on the coefficients of elements of A_s.

Let $m \in A_s$ be arbitrary. Since A_s is also contained in X_s, we can write $m = (m_{[1,r]}, Tm_{[1,r]})$ for some $m_{[1,r]} \in \mathbf{Z}^r$ with magnitude $O_d((SV)^{O_d(B^s)})$. By definition of A_s, we conclude that

$$\langle m_r, v_{[1,r]}\rangle_{\mathbf{R}^r} + \langle Tm_r, v_{[r+1,d]}\rangle_{\mathbf{R}^{d-r}} = O((SV)^{-B^{B+1}}R)$$

where $v_{[1,r]} := (v_1, \ldots, v_r)$, $v_{[r+1,d]} := (v_{r+1}, \ldots, v_d)$, and the inner products on $\mathbf{R}^r$ and $\mathbf{R}^{d-r}$ are the standard ones. Thus

$$\langle m_r, v_{[1,r]} + T^* v_{[r+1,d]}\rangle_{\mathbf{R}^r} = O((SV)^{-B^{B+1}}R)$$

where $T^* : \mathbf{R}^{d-r} \to \mathbf{R}^r$ is the adjoint linear transformation to T. Now since A spans X, the $m_{[1,r]}$ will linearly span $\mathbf{R}^r$ as we vary over all elements m of A. Thus by Cramer's rule we conclude that

$$v_{[1,r]} + T^* v_{[r+1,d]} = O_d(V^{O_d(B^s)}(SV)^{-B^{B+1}}R). \tag{18}$$

Write $(w_1, \ldots, w_r) := T^* v_{[r+1,d]}$; thus $w_1, \ldots, w_r$ are rational numbers. Then construct the symmetric generalized arithmetic progressions P_{small} and P_{sparse} explicitly as

$$\begin{aligned} P_{\text{sparse}} := \{&m_1 w_1 + \cdots + m_r w_r + m_{r+1} v_{r+1} \\ &+ \cdots + m_d v_d : |m_i| \le M_i \text{ for all } 1 \le i \le d\} \end{aligned}$$

and

$$P_{\text{small}} := \{m_1(v_1 + w_1) + \cdots + m_r(v_r + w_r) : |m_i| \le M_i \text{ for all } 1 \le i \le d\}.$$

It is clear from construction that $P \subseteq P_{\text{sparse}} + P_{\text{small}}$, and that P_{sparse} and P_{small} have rank at most d and volume at most V. Now from (18),

$$v_i + w_i = O_d((SV)^{O_d(B^s)}(SV)^{-B^{B+1}}R)$$

and hence for any $x \in P_{\text{small}}$,

$$x = O_d((SV)^{O_d(B^s)}(SV)^{-B^{B+1}}R).$$

By choosing B large enough we conclude that

$$|x| \le R/S$$

which gives the desired smallness bound on P_{small}.

The only remaining task is to show that SP_{sparse} is sparse. It suffices to show that $SP_{\text{sparse}} - SP_{\text{sparse}}$ has no nonzero intersection with $[-RS, RS]$. Suppose for contradiction that this failed. Then we can find $m_1, \dots, m_d$ with $|m_i| \le 2SM_i$ for all i and

$$0 < m_1w_1 + \cdots + m_rw_r + m_{r+1}v_{r+1} + \cdots + m_dv_d < RS.$$

Let Q be the least common denominator of all the coefficients of T^*, then $Q = O_d((SV)^{O_d(B^s)})$. Multiplying the above equation by Q, we obtain

$$\begin{aligned} 0 &< m_1Qw_1 + \cdots + m_rQw_r + m_{r+1}Qv_{r+1} + \cdots + m_dQv_d \\ &< O(RSV^{O_d(B^s)}) < (SV)^{B^{B+1}}R. \end{aligned}$$

Since $(w_1, \dots, w_r) = T^*v_{[r+1,r+d]}$, the expression between the inequality signs is an integer linear combination of $v_{r+1}, \dots, v_d$, with all coefficients of size $O_d((SV)^{O_d(B^s)})$, for example

$$m_1Qw_1 + \cdots + m_rQw_r + m_{r+1}Qv_{r+1} + \cdots + m_dQv_d = a_{r+1}v_{r+1} + \cdots + a_dv_d.$$

In particular, this expression lies in $(SV)^{B^B}P$ (again taking B to be sufficiently large). Thus by construction of R, we can improve the upper bound of $(SV)^{B^{B+1}}R$ to $(SV)^{-B^{B+1}}R$:

$$0 < a_{r+1}v_{r+1} + \cdots + a_dv_d < (SV)^{-B^{B+1}}R. \tag{19}$$

Taking B to be large, this implies that $(0, \dots, 0, a_{r+1}, \dots, a_d)$ lies in X_{s+1}, which equals X_s. But X_s was a graph from $\mathbf{R}^r$ to $\mathbf{R}^{d-r}$, and thus $a_{r+1} = \cdots = a_d = 0$, which contradicts (19). This establishes the sparseness. □

9. Proof of Theorem 3.10

Let $Y = \{y_1, \dots, y_l\}$ be a set of l independent vectors in $\mathbf{R}^n$. Recall that $M_n^{\mu,Y}$ denote the random matrix with row vectors $X_1^\mu, \dots, X_{n-l}^\mu$, $y_1, \dots, y_l$, where X_i^μ are i.i.d. copies of $X^\mu = (\eta_1^\mu \dots, \eta_n^\mu)$.

Define $\delta(\mu) := \max\{1-\mu, \mu/2\}$. It is easy to show that for any subspace V of dimension d,

$$\mathbf{P}(X^\mu \in V) \le \delta(\mu)^{d-n}. \tag{20}$$

In the following, we use N to denote the quantity $(1/\delta(\mu))^n$. As $0 < \mu \le 1$, $\delta(\mu) > 0$ and thus N is exponentially large in n. Thus it will suffice to show that

$$\mathbf{P}(M_n^{\mu,Y} \text{ singular }) \le N^{-\varepsilon+o(1)}$$

for some $\varepsilon = \varepsilon(\mu, l) > 0$, where the $o(1)$ term is allowed to depend on μ, l, and ε. We may assume that n is large depending on μ and l since the claim is trivial otherwise.

Note that if $M_n^{\mu,Y}$ is singular, then the row vectors span a proper subspace V. To prove the theorem, it suffices to show that for any sufficiently small positive constant ε

$$\sum_{V, V \text{ proper subspace}} \mathbf{P}(X_1^\mu, \dots, X_{n-l}^\mu, y_1, \dots, y_l \text{ span } V) \le N^{-\varepsilon+o(1)}.$$

Arguing as in [25, Lemma 5.1], we can restrict ourselves to hyperplanes. Thus, it is enough to prove

$$\sum_{V, V \text{ hyperplane}} \mathbf{P}(X_1^\mu, \dots, X_{n-l}^\mu, y_1, \dots, y_l \text{ span } V) \le N^{-\varepsilon+o(1)}.$$

We may restrict our attention to those hyperplanes V which are spanned by their intersection with $\{-1, 0, 1\}^n$, together with $y_1, \dots, y_l$. Let us call such hyperplanes *nontrivial*. Furthermore, we call a hyperplane H *degenerate* if there is a vector v orthogonal to H and at most $\log\log n$ coordinates of v are nonzero. Following [25, Lemma 5.3], it is easy to see that the number of degenerate nontrivial hyperplanes is at most $N^{o(1)}$. Thus, their contribution in the sum is at most

$$N^{o(1)}\delta(\mu)^{n-l} = N^{-1+o(1)}$$

which is acceptable. Therefore, from now on we can assume that V is nondegenerate.

For each nontrivial hyperplane V, define the *discrete codimension* $d(V)$ of V to be the unique integer multiple of $1/n$ such that

$$N^{-\frac{d(V)}{n} - \frac{1}{n^2}} < \mathbf{P}(X^\mu \in V) \le N^{-\frac{d(V)}{n}}. \tag{21}$$

Thus $d(V)$ is large when V contains few elements from $\{-1, 0, 1\}^n$, and conversely.

Let B_V denote the event that $X_1^\mu, \dots, X_{n-l}^\mu, y_1, \dots, y_l$ span V. We denote by Ω_d the set of all nondegenerate, nontrivial hyperplanes with discrete codimension d. It is simple to see that $1 \le d(V) \le n^2$ for all nontrivial V. In particular, there are $n^2 = N^{o(1)}$ possible values of d, so to prove our theorem it suffices to show that

$$\sum_{V \in \Omega_d} \mathbf{P}(B_V) \le N^{-\varepsilon+o(1)} \tag{22}$$

for all $1 \le d \le n^2$.

We first handle the (simpler) case when d is large. Note that if

$$X_1^\mu, \dots, X_{n-l}^\mu, y_1, \dots, y_l \text{ span } V,$$

then some subset of $n-l-1$ vectors X_i together with the y_j's already span V (since the y_j's are independent). By symmetry, we have

$$\begin{aligned}
&\sum_{V\in\Omega_d} \mathbf{P}(B_V) \\
&\qquad \le (n-l) \sum_{V\in\Omega_d} \mathbf{P}(X_1^\mu, \dots, X_{n-l-1}^\mu, y_1, \dots, y_l \text{ span } V)\mathbf{P}(X_{n-l}^\mu \in V) \\
&\qquad \le nN^{-\frac{d}{n}} \sum_{V\in\Omega_d} \mathbf{P}(X_1^\mu, \dots, X_{n-l-1}^\mu, y_1, \dots, y_l \text{ span } V) \\
&\qquad \le nN^{-\frac{d}{n}} = N^{-\frac{d}{n}+o(1)}.
\end{aligned}$$

This disposes of the case when $d \ge \varepsilon n$. It remains to verify the following lemma.

LEMMA 9.1. *For all sufficiently small positive constant ε, the following holds. If d is any integer multiple of $1/n$ such that*

$$1 \le d \le (\varepsilon - o(1))n, \tag{23}$$

then

$$\sum_{V\in\Omega_d} \mathbf{P}(B_V) \le N^{-\varepsilon+o(1)}.$$

Proof. For $0 < \mu \le 1$ we define the quantity $0 < \mu^* \le 1/8$ as follows. If $\mu = 1$ then $\mu^* := 1/16$. If $1/2 \le \mu < 1$, then $\mu^* := (1-\mu)/4$. If $0 < \mu < 1/2$, then $\mu^* := \mu/4$. We will need the following inequality, which is a generalization of [25, Lemma 6.2].

LEMMA 9.2. *Let V be a nondegenerate nontrivial hyperplane. Then*

$$\mathbf{P}(X^\mu \in V) \le \left(\frac{1}{2} + o(1)\right) \mathbf{P}(X^{\mu^*} \in V).$$

The proof of Lemma 9.2 relies on some Fourier-analytic ideas of Halász [9] (see also [10], [25], [26]) and is deferred until the end of the section. Assuming it for now, we continue the proof of Lemma 9.1.

Let us set $\gamma := \frac{1}{2}$; this is not the optimal value of this parameter, but will suffice for this argument.

Let A_V be the event that $X_1^{\mu^*}, \dots, X_{(1-\gamma)n}^{\mu^*}, \overline{X}_1^\mu, \dots, \overline{X}_{(\gamma-\varepsilon)n}^\mu$ are linearly independent in V, where $X_i^{\mu^*}$'s are i.i.d. copies of X^{μ^*} and $\overline{X}_j^\mu$'s are i.i.d. copies of X^μ.

LEMMA 9.3.
$$\mathbf{P}(A_V) \geq N^{(1-\gamma)-(1-\varepsilon)d+o(1)}.$$

Proof. Note that the right-hand side on the bound in Lemma 9.3 is the probability of the event A'_V that $X_1^{\mu^*}, \dots, X_{(1-\gamma)n}^{\mu^*}, \overline{X}_1^{\mu}, \dots, \overline{X}_{(\gamma-\varepsilon)n}^{\mu}$ belong to V. Thus, by Bayes' identity it is sufficient to show that
$$\mathbf{P}(A_V|A'_V) = N^{o(1)}.$$

From (21),
$$\mathbf{P}(X^{\mu} \in V) = (1 + O(1/n))\delta(\mu)^d \tag{24}$$
and hence by Lemma 9.2
$$\mathbf{P}(X^{\mu^*} \in V) \geq (2 + O(1/n))\delta(\mu)^d. \tag{25}$$
On the other hand, by (20)
$$\mathbf{P}(X^{\mu^*} \in W) \leq (1 - \mu^*)^{n-\dim(W)}$$
for any subspace W. Thus by Bayes' identity, we have the conditional probability bound
$$\begin{aligned}&\mathbf{P}(X^{\mu^*} \in W|X^{(\mu^*)} \in V)\\ &\quad\leq (2 + O(1/n))^{-1}\delta(\mu)^{-d}(1 - \mu^*)^{n-\dim(W)} \leq \delta(\mu)^{-d}(1 - \mu^*)^{n-\dim(W)}.\end{aligned}$$
When $\dim(W) \leq (1 - \gamma)n$, the bound is less than one when ε is sufficiently small, thanks to the bound on d and the choice $\gamma = \frac{1}{2}$.

Let E_k be the event that $X_1^{\mu^*}, \dots, X_k^{\mu^*}$ are linearly independent. The above estimates imply that
$$\mathbf{P}(E_{k+1}|E_k \wedge A'_V) \geq 1 - \delta(\mu)^{-d}(1 - \mu^*)^{n-k}$$
for all $0 \leq k \leq (1 - \gamma)n$. Thus applying Bayes' identity repeatedly, we obtain
$$\mathbf{P}(E_{(1-\gamma)n}|A'_V) \geq N^{-o(1)}.$$

To complete the proof, observe that since
$$\mathbf{P}(X^{\mu} \in W) \leq \delta(\mu)^{n-\dim(W)}$$
for any subspace W, it follows that by (24),
$$\mathbf{P}(X^{\mu} \in W|X^{\mu} \in V) \leq (1 + O(1/n))\delta(\mu)^{-d}\delta(\mu)^{n-\dim(W)}.$$

Let us assume $E_{(1-\gamma)n}$ and denote by W the $(1-\gamma)n$-dimensional subspace spanned by $X_1^{\mu^*}, \dots, X_{(1-\gamma)n}^{\mu^*}$. Let U_k denote the event that $\overline{X}_1^{\mu}, \dots, \overline{X}_k^{\mu}, W$ are liearly independent. We have
$$\begin{aligned}p_k &= \mathbf{P}(U_{k+1}|U_k \wedge A'_V)\\ &\geq 1 - (1 + O(1/n))\delta(\mu)^{-d}\delta(\mu)^{n-k-(1-\gamma)n} \geq 1 - \frac{1}{100}\delta(\mu)^{-(\gamma-\varepsilon)n+k}\end{aligned}$$

for all $0 \le k < (\gamma - \varepsilon)n$, thanks to (23). Thus by Bayes' identity we obtain

$$\mathbf{P}(A_V | A'_V) \ge N^{o(1)} \prod_{0 \le k < (\gamma-\varepsilon)n} p_k = N^{o(1)}$$

as desired. □

Now we continue the proof of the theorem. Fix $V \in \Omega_d$. Since A_V and B_V are independent, by Lemma 9.3,

$$\mathbf{P}(B_V) = \frac{\mathbf{P}(A_V \wedge B_V)}{\mathbf{P}(A_V)} \le N^{-(1-\gamma)+(1-\varepsilon)d+o(1)} \mathbf{P}(A_V \wedge B_V).$$

Consider a set

$$X_1^{\mu^*}, \dots, X_{(1-\gamma)n}^{\mu^*}, \overline{X}_1^{\mu}, \dots, \overline{X}_{(\gamma-\varepsilon)n}^{\mu}, X_1^{\mu}, \dots, X_{n-l}^{\mu}$$

of vectors satisfying $A_V \wedge B_V$. Then there exists $\varepsilon n - l - 1$ vectors $X_{j_1}^{\mu}, \dots, X_{j_{\varepsilon n-l-1}}^{\mu}$ inside $X_1^{\mu}, \dots, X_{n-l}^{\mu}$, which together with

$$X_1^{\mu^*}, \dots, X_{(1-\gamma)n}^{\mu^*}, \overline{X}_1^{\mu}, \dots, \overline{X}_{(\gamma-\varepsilon)n}^{\mu}, y_1, \dots, y_l,$$

span V. Since the number of possible indices $j_1, \dots, j_{\varepsilon n-l-1}$ is $\binom{n-l}{\varepsilon n-l-1}$ $= 2^{(h(\varepsilon)+o(1))n}$ (with h being the entropy function), by conceding a factor of

$$2^{(h(\varepsilon)+o(1))n} = N^{ah(\varepsilon)+o(1)},$$

where $a = \log_{1/\delta(\mu)} 2$, we can assume that $j_i = i$ for all relevant i. Let C_V be the event that

$$X_1^{\mu^*}, \dots, X_{(1-\gamma)n}^{\mu^*}, \overline{X}_1^{\mu}, \dots, \overline{X}_{(\gamma-\varepsilon)n}^{\mu}, X_1^{\mu}, \dots, X_{\varepsilon n-l-1}^{\mu}, y_1, \dots, y_l \text{ span } V.$$

Then we have

$$\mathbf{P}(B_V) \le N^{-(1-\gamma)+(1-\varepsilon)d+ah(\varepsilon)+o(1)} \mathbf{P}\Big(C_V \wedge (X_{\varepsilon n}^{\mu}, \dots, X_{n-l}^{\mu} \text{ in } V)\Big).$$

On the other hand, C_V and the event $(X_{\varepsilon n}, \dots, X_n \text{ in } V)$ are independent, so

$$\mathbf{P}\Big(C_V \wedge (X_{\varepsilon n}^{\mu}, \dots, X_{n-l}^{\mu} \text{ in } V)\Big) = \mathbf{P}(C_V)\mathbf{P}(X^{\mu} \in V)^{(1-\varepsilon)n+1-l}.$$

Putting the last two estimates together we obtain

$$\begin{aligned}\mathbf{P}(B_V) &\le N^{-(1-\gamma)+(1-\varepsilon)d+ah(\varepsilon)+o(1)} N^{-((1-\varepsilon)n+1-l)d/n} \mathbf{P}(C_V) \\ &= N^{-(1-\gamma)+ah(\varepsilon)+(l-1)\varepsilon+o(1)} \mathbf{P}(C_V).\end{aligned}$$

Since any set of vectors can only span a single space V, we have $\sum_{V \in \Omega_d} \mathbf{P}(C_V)$ ≤ 1. Thus, by summing over Ω_d,

$$\sum_{V \in \Omega_d} \mathbf{P}(B_V) \le N^{-(1-\gamma)+ah(\varepsilon)+(l-1)\varepsilon+o(1)}.$$

With the choice $\gamma = \frac{1}{2}$, we obtain a bound of $N^{-\varepsilon+o(1)}$ as desired, by choosing ε sufficiently small. This provides the desired bound in Lemma 9.1. □

9.1. *Proof of Lemma* 9.2. To conclude, we prove Lemma 9.2. Let $v = (a_1, \dots, a_n)$ be the normal vector of V and define

$$F_\mu(\xi) := \prod_{i=1}^{n}((1-\mu) + \mu \cos 2\pi a_i \xi).$$

From Fourier analysis (cf. [25])

$$\mathbf{P}(X^\mu \in V) = \mathbf{P}(X^\mu \cdot v = 0) = \int_0^1 F_\mu(\xi) d\xi.$$

The proof of Lemma 9.2 is based on the following technical lemma.

LEMMA 9.4. *Let μ_1 and μ_2 be a positive numbers at most $1/2$ such that the following two properties hold for for any $\xi, \xi' \in [0,1]$:*

$$F_{\mu_1}(\xi) \leq F_{\mu_2}(\xi)^4 \tag{26}$$

and

$$F_{\mu_1}(\xi) F_{\mu_1}(\xi') \leq F_{\mu_2}(\xi + \xi')^2. \tag{27}$$

Furthermore,

$$\int_0^1 F_{\mu_1}(\xi)\ d\xi = o(1). \tag{28}$$

Then

$$\int_0^1 F_{\mu_1}(\xi)\ d\xi \leq (1/2 + o(1)) \int_0^1 F_{\mu_2}(\xi)\ d\xi. \tag{29}$$

Proof. Since $\mu_1, \mu_2 \leq 1/2$, $F_{\mu_1}(\xi)$ and $F_{\mu_2}(\xi)$ are positive for any ξ. From (27) we have the sumset inclusion

$$\{\xi \in [0,1] : F_{\mu_1}(\xi) > \alpha\} + \{\xi \in [0,1] : F_{\mu_1}(\xi)\alpha\} \subseteq \{\xi \in [0,1] : F_{\mu_2}(\xi) > \alpha\}$$

for any $\alpha > 0$. Taking measures of both sides and applying the Mann-Kneser-Macbeath "$\alpha + \beta$ inequality" $|A + B| \geq \min(|A| + |B|, 1)$ (see [17]), we obtain

$$\min(2|\{\xi \in [0,1] : F_{\mu_1}(\xi) > \alpha\}|, 1) \leq |\{\xi \in [0,1] : F_{\mu_2}(\xi) > \alpha\}|.$$

But from (28) we see that $|\{\xi \in [0,1] : F_{\mu_2}(\xi) > \alpha\}|$ is strictly less than 1 if $\alpha > o(1)$. Thus we conclude that

$$|\{\xi \in [0,1] : F_{\mu_1}(\xi) > \alpha\}| \leq \frac{1}{2}|\{\xi \in [0,1] : F_{\mu_2}(\xi) > \alpha\}|$$

when $\alpha > o(1)$. Integrating this in α, we obtain

$$\int_{[0,1]:F_{\mu_1}(\xi) > o(1)} F_{\mu_1}(\xi)\ d\xi \leq \frac{1}{2} \int_0^1 F_{\mu_2}(\xi)\ d\xi.$$

On the other hand, from (26) we see that when $F_{\mu_1}(\xi) \leq o(1)$, then $F_{\mu_1}(\xi) = o(F_{\mu_1}(\xi)^{1/4}) = o(F_{\mu_2}(\xi))$, and thus

$$\int_{[0,1]:F_{\mu_1}(\xi)\leq o(1)} F_{\mu_1}(\xi)\ d\xi \leq o(1) \int_0^1 F_{\mu_2}\ d\xi.$$

Adding these two inequalities we obtain (29) as desired. □

By Lemma 5.1

$$\mathbf{P}(X^\mu \cdot v = 0) \leq \mathbb{P}_\mu(\mathbf{v}) \leq \mathbb{P}_{\mu/4}(\mathbf{v}) = \int_0^1 F_{\mu/4}(\xi) d\xi.$$

It suffices to show that the conditions of Lemma 9.4 hold with $\mu_1 = \mu/4$ and $\mu_2 = \mu^* = \mu/16$. The last estimate $\int_0^1 F_{\mu_1}(\xi)\ d\xi \leq o(1)$ is a simple corollary of the fact that at least $\log\log n$ among the a_i are nonzero (instead of $\log\log n$, one can use any function tending to infinity with n), so we only need to verify the other two. Inequality (26) follows from the fact that $\mu_2 = \mu_1/4$ and the proof of the fourth property of Lemma 5.1.

To verify (27), it suffices to show that for any $\mu' \leq 1/2$ and any θ, θ'

$$((1-\mu') + \mu' \cos\theta)((1-\mu') + \mu' \cos\theta') \leq ((1-\mu'/4) + \frac{\mu'}{4}\cos(\theta+\theta')^2.$$

The left-hand side is bounded from above by $((1-\mu') + \mu' \cos\frac{\theta+\theta'}{2})^2$, due to convexity. Thus, it remains to show that

$$(1-\mu') + \mu' \cos\frac{\theta+\theta'}{2} \leq \left(1 - \frac{\mu'}{4}\right) + \frac{\mu'}{4}\cos(\theta+\theta')$$

since both expressions are positive for $\mu' < 1/2$. By defining $x := \cos\frac{\theta+\theta'}{2}$, the last inequality becomes

$$(1-\mu') + \mu' x \leq \left(1 - \frac{\mu'}{4}\right) + \frac{\mu'}{4}(2x^2 - 1)$$

which trivially holds. This completes the proof of Lemma 9.2. □

University of California, Los Angeles, CA
E-mail address: tao@math.ucla.edu

Rutgers University, Piscataway, NJ
E-mail address: vanvu@math.rutgers.edu

References

[1] D. Bau III and L. N. Trefethen, Numerical linear algebra, *Society for Industrial and Applied Math.* (SIAM), Philadelphia, PA, 1997.

[2] B. Bollobás, *Random Graphs*, Second edition, *Cambridge Studies in Adv. Math.* **73**, Cambridge Univ. Press, Cambridge, 2001.

[3] A. Edelman, Eigenvalues and condition numbers of random matrices, *SIAM J. Matrix Anal. Appl.* **9** (1988), 543–560.

[4] A. Edelman and B. Sutton, Tails of condition number distributions, *SIMAX* **27** (2005), 547–560.

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

[6] ———, Extremal problems in number theory, *Proc. Sympos. Pure Math.* **VIII** (1965), 181–189, A. M. S., Providence, R.I.

[7] P. Frankl and Z. Füredi, Solution of the Littlewood-Offord problem in high dimensions, *Ann. of Math.* **128** (1988), 259–270.

[8] J. Griggs, J. Lagarias, A. Odlyzko, and J. Shearer, On the tightest packing of sums of vectors, *European J. Combin.* **4** (1983), 231–236.

[9] G. Halász, Estimates for the concentration function of combinatorial number theory and probability, *Period. Math. Hungar.* **8** (1977), 197–211.

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

[11] G. Katona, On a conjecture of Erdős and a stronger form of Sperner's theorem, *Studia Sci. Math. Hungar.* **1** (1966), 59–63.

[12] D. Kleitman, On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors, *Advances in Math.* **5** (1970), 155–157.

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

[14] N. Alon, M. Krivelevich, and V. H. Vu, On the concentration of eigenvalues of random symmetric matrices, *Israel J. Math.* **131** (2002), 259–267.

[15] A. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, *Adv. in Math.* **195** (2005), 491–523.

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

[17] A. M. Macbeath, On measure of sum sets II. The sum-theorem for the torus, *Proc. Cambridge Philos. Soc.* **49** (1953), 40–43.

[18] L. P. Postnikova and A. A. Judin, An analytic method for estimates of the concentration function (Russian), *Analytic Number Theory, Mathematical Analysis and their Applications* (dedicated to I. M. Vinogradov on his 85th birthday), *Trudy Mat. Inst. Steklov*. **143** (1977), 143–151, 210.

[19] B. A. Rogozin, The concentration functions of sums of independent random variables, *Proc. of the Second Japan-USSR Symposium on Probability Theory* (Kyoto, 1972), pp. 370–376. *Lecture Notes in Math*. **330**, 370–376, Springer-Verlag, New York, 1973.

[20] M. Rudelson, Invertibility of random matrices: Norm of the inverse, *Annals of Math*. **168** (2008), 575–600.

[21] N. P. Salikhov, An estimate for the concentration function by the Esseen method. (Russian) *Teor. Veroyatnost. i Primenen.* **41** (1996), 561–577; translation in *Theory Probab. Appl*. **41** (1997), 504–518.

[22] A. Sárközy and E. Szemerédi, Über ein Problem von Erdős und Moser, *Acta Arithmetica* **11** (1965), 205–208.

[23] S. Smale, On the efficiency of algorithms of analysis, *Bull. Amer. Math. Soc.* **13** (1985), 87–121.

[24] R. Stanley, Weyl groups, the hard Lefschetz theorem, and the Sperner property, *SIAM J. Algebraic Discrete Methods* **1** (1980), 168–184.

[25] T. Tao and V. H. Vu, On random ± 1 matrices: Singularity and determinant, *Random Structures Algorithms* **28** (2006), 1–23.

[26] ———, On the singularity probability of random Bernoulli matrices, *J. Amer. Math. Soc.* **20** (2007), 603–628.

[27] ———, On the condition number of a randomly perturbed matrix, *Proc. Thirty-Ninth Annual ACM Sympos. on Theory of Computing* 2007, 248–255.

[28] ———, *Additive Combinatorics*, Cambridge Univ. Press, Cambridge, 2006.

(Received November 8, 2005)