Electronic Journal of Probability

Electron. J. Probab. **0** (2012), no. 0, 1–23.
ISSN: 1083-6489 DOI: 10.1214/EJP.vVOL-PID

On the least singular value of random symmetric matrices

Hoi H. Nguyen*

Abstract

Let F_n be an n by n symmetric matrix whose entries are bounded by n^γ for some $\gamma > 0$. Consider a randomly perturbed matrix $M_n = F_n + X_n$, where X_n is a *random symmetric matrix* whose upper diagonal entries $x_{ij}, 1 \le i \le j$, are iid copies of a random variable ξ. Under a very general assumption on ξ, we show that for any $B > 0$ there exists $A > 0$ such that $\mathbf{P}(\sigma_n(M_n) \le n^{-A}) \le n^{-B}$.

Keywords: Random symmetric matrices, least singular values.
AMS 2010 Subject Classification: 15A52; 15A63; 11B25.

Submitted to EJP on January 1, 2012, final version accepted on January 1, 2012.

1 Introduction

Let F_n be an n by n matrix whose entries are bounded by $n^{O(1)}$. Consider a randomly perturbed matrix $M_n = F_n + X_n$, where X_n is a *random matrix* whose entries are iid copies of a random variable. It has been shown, under a very general assumption on ξ, that the singular value of M_n cannot be too small.

Theorem 1.1. *[21, Theorem 2.1] Assume that $M_n = F_n + X_n$, where the entries of F_n are bounded by n^γ, and the entries of X_n are iid copies of a random variable of zero mean and unit variance. Then for any $B > 0$, there exists $A > 0$ such that*

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

Here $\sigma_n(M_n)$ is the smallest singular value of M_n, defined as

$$\sigma_n(M_n) := \inf_{\|x\|=1} \|M_n x\|.$$

The dependence among the parameters in Theorem 1.1 was made explicitly in [24]. Under the stronger assumption that ξ has sub-Gaussian distribution, Rudelson and Vershynin [16] obtained an

*Department of Mathematics, University of Pennsylvania, PA 19104, USA hoing@math.upenn.edu

almost best possible estimate on the tail bound of $\sigma_n(M_n)$. For more results regarding this random matrix ensemble we refer the reader to [16, 21, 24].

One important application of Theorem 1.1 is a polynomial bound for the condition number of M_n.

Corollary 1.2. *[21, Corollary 2.10] With the same assumption as in Theorem 1.1, for any $B > 0$, there exists $A > 0$ such that*

$$\mathbf{P}(\sigma_1(M_n)/\sigma_n(M_n) \le n^A) \ge 1 - n^{-B}.$$

The condition number $\kappa(M) = \sigma_1(M)/\sigma_n(M)$ of a matrix M plays a crucial role in numerical linear algebra. The above corollary implies that if one perturbs a fixed matrix F of small spectral norm by a (very general) random matrix X_n, the condition number of the resulting matrix will be relatively small with high probability. This fact has some nice applications in theoretical computer science. (See for instance [17, 18] for further discussions on these applications).

Another popular model of random matrices is that of *random symmetric matrices*; this is one of the simplest models that has non-trivial correlations between the matrix entries. A significant new difficulty in the study of the singularity of X_n (or of M_n in general) is that the symmetry ensures that $\det(X_n)$ is a quadratic function of each row, as opposed to the regular random ensembles in which $\det(X_n)$ is a linear function of each row.

A recent result of Costello, Tao and Vu [2] shows that if the upper diagonal entries x_{ij} of X_n are iid Bernoulli random variables, then X_n is non-singular with probability $1 - n^{-1/8+o(1)}$. In [12], the current author improved the bound to any polynomial order.

The goal of this note is to study the smallest singular value of randomly perturbed matrices M_n, under a general assumption on ξ.

Condition 1 (Anti-concentration)**.** *Assume that ξ has zero mean, unit variance, and there exist positive constants c_1 and c_s such that*

$$\mathbf{P}(c_1 \le |\xi - \xi'|) \ge c_2,$$

where ξ' is an independent copy of ξ.

Theorem 1.3 (Main theorem)**.** *Assume that the upper diagonal entries x_{ij} of X_n are iid copies of a random variable ξ satisfying Condition 1. Assume also that the entries f_{ij} of the symmetric matrix F_n satisfy $|f_{ij}| \le n^\gamma$ for some $\gamma > 0$. Then for any $B > 0$, there exists $A > 0$ depending on c_1, c_2, γ and B such that for all sufficiently large n,*

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

Our result immediately implies a polynomial bound for the condition number of M_n as follows.

Corollary 1.4. *With the same assumptions as of Theorem 1.3, for any $B > 0$, there exists $A > 0$ depending in c_1, c_2, γ and B such that for all sufficiently large n,*

$$\mathbf{P}(\kappa(M_n) \ge n^A) \le n^{-B}.$$

As another application, we provide a relatively fine lower bound for the determinant of random symmetric matrices. This result refines an important case of [25, Theorem 34] obtained by Tao and Vu.

Corollary 1.5. *Assume that the upper diagonal entries x_{ij} of X_n are iid copies of a random variable ξ of zero mean, unit variance, and there is a constant $C > 0$ such that $\mathbf{P}(|\xi| \le C) = 1$. Assume furthermore that the entries f_{ij} of the symmetric matrix F_n also satisfy $|f_{ij}| \le C$. Then for any positive constant B there exists a positive constant D depending on B and C such that the following holds with probability $1 - O(n^{-B})$,*

$$|\det(M_n)| \ge \exp(-Dn^{1/3}\log n)\mathbf{E}(|\det(M_n)|),$$

and

$$\det(M_n)^2 \ge \exp(-Dn^{1/3}\log n)\mathbf{E}(\det(M_n)^2).$$

This corollary complements previously known results on the concentration of the determinant of non-symmetric random matrices (cf. [1, 3, 7, 19]).

Remark. When a preliminary version of this paper was submitted to the arxiv, Vershynin also published a similar result with stronger bounds (see [27]). However, our result is different from Vershynin's in three ways. Firstly, our Condition 1 on ξ is weaker, as we do not require it to have bounded fourth-moment. Secondly, our bound for the least singular value works for perturbed matrices of the form $M_n = F_n + X_n$ with $\|F_n\| = n^{O(1)}$. Lastly, the techniques we use are very different. Our proof relies on an almost complete inverse-type result concerning the concentration of quadratic forms, which is of interest of its own.

Notation. For a matrix M we use the notations $\mathbf{r}_i(M)$ and $\mathbf{c}_j(M)$ to denote its i-th row vector and its j-th column vector respectively; we use the notation $(M)_{ij}$ to denote its ij entry.

We use η to denote random Bernoulli variables (thus η takes values ± 1 with probability 1/2).

Here and later, asymptotic notations such as $O, \Omega, \Theta, \omega$, and so for, are used under the assumption that $n \to \infty$. A notation such as $O_C(.)$ emphasizes that the hidden constant in O depends on C. If $a = \Omega(b)$, we write $b \ll a$ or $a \gg b$. If $a = \Omega(b)$ and $b = \Omega(a)$, we write $a \asymp b$.

2 The approach to prove Theorem 1.3

For the sake of simplicity, we will prove our result under the following condition.

Condition 2. *With probability one,*

$$|x_{ij}| \le n^{B+1},$$

for all i, j.

In fact, because ξ has unit variance, we have

$$\mathbf{P}(|x_{ij}| \ge n^{B+1}) = O(n^{-2B-2}).$$

Thus, we can assume that $|x_{ij}| \leq n^{B+1}$ at the cost of an additional negligible term $o(n^{-B})$ in probability.

We next assume that $\sigma_n(M_n) \leq n^{-A}$. Thus

$$M_n \mathbf{x} = \mathbf{y},$$

for some $\|\mathbf{x}\| = 1$ and $\|\mathbf{y}\| \leq n^{-A}$. There are two cases to consider.

Case 1. $\det(M_n) = 0$. This is the case to consider when ξ has discrete distribution.

We first show that it is enough to consider the case of M_n having rank $n-1$, thanks to the following result.

Lemma 2.1. *For any $1 \leq k \leq n-2$, we have*

$$\mathbf{P}(\text{rank}(M_n) = k \leq n-2) \leq O_{c_1}(1)\mathbf{P}(\text{rank}(M_{2n-k-1}) = 2n-k-2).$$

We deduce Lemma 2.1 from a useful observation by Odlyzko, whose simple proof is presented in Appendix A.

Lemma 2.2 (Odlyzko's lemma,[15]). *Let H be a linear subspace in $\mathbf{R}^n$ of dimension at most $k \leq n$. Then*

$$\mathbf{P}(\mathbf{u} \in H) \leq (\sqrt{1-c_3})^{n-k},$$

where $\mathbf{u} = (f_1 + x_1, \dots, f_n + x_n)$, f_i are fixed and x_i are iid copies of ξ.

Proof. (of Lemma 2.1) View M_{n+1} as the matrix obtained by adding the first row and first column to M_n. Let H be the vector space of dimension k spanned by the row vectors of M_n. Then the probability that the subvector formed by the last n components of the first row of M_{n+1} does not belong to H, by Lemma 2.2, is at least $1 - (\sqrt{1-c_3})^{n-k}$.

Observe that if this is the case then the last n columns of M_{n+1} span a vector space of dimension $k+1$. Additionally, by symmetry, as the subvector formed by the last n components of the first column of M_{n+1} does not belong to H, adding the first column will increase the rank of M_{n+1} to $k+2$.

Hence,

$$\mathbf{P}(\text{rank}(M_{n+1}) = k+2 | \text{rank}(M_n) = k) \geq 1 - (\sqrt{1-c_3})^{n-k}.$$

In general, for $1 \leq t \leq n-k$ we have

$$\mathbf{P}(\text{rank}(M_{n+t}) = k+2t | \text{rank}(M_{n+t-1}) = k+2(t-1)) \geq 1 - (\sqrt{1-c_3})^{n-t-k+1}.$$

Because the rows (and columns) added to M_{n+t-1} at each step (to create M_{n+t}) are independent, we have

$$
\begin{aligned}
&\mathbf{P}(\mathrm{rank}(M_{2n-k-1}) = 2n-k-2 | \mathrm{rank}(M_n) = k) \geq \\
&\geq \prod_{t=1}^{n-k-1} \mathbf{P}\big(\mathrm{rank}(M_{n+t}) = k+2t | \mathrm{rank}(M_{n+t-1}) = k+2(t-1)\big) \\
&\geq (1-(\sqrt{1-c_3})^{n-k})(1-(\sqrt{1-c_3})^{n-k-1}) \cdots (1-(\sqrt{1-c_3})) = \Omega_{c_3}(1).
\end{aligned}
$$

□

Next we show that in the case of M_n having rank $n-1$, it suffices to assume that $\mathrm{rank}(M_{n-1}) \geq n-2$, thanks to the following simple observation.

Lemma 2.3. *Assume that M_n has rank $n-1$. Then there exists $1 \leq i \leq n$ such that the removal of the i-th row and the i-column of M_n results in a matrix M_{n-1} of rank at least $n-2$.*

Proof. (of Lemma 2.3)Without loss of generality, assume that the last $n-1$ rows of M_n span a subspace of dimension $n-1$. Then the matrix obtained from M_n by removing the first row and the first column has rank at least $n-2$. □

Without loss of generality, we assume that the matrix M_{n-1} obtained from M_n by removing its first row and first column has rank at least $n-2$. We next express $\det(M_n)$ as a quadratic function of its first row $(m_{11}, \dots, m_{1n})$ as follows.

$$\det(M_n) = c_{11}(M_n)m_{11} + \sum_{2 \leq i,j \leq n} c_{ij}(M_{n-1})m_{1i}m_{1j}$$

where $c_{11}(M_n)$ is the first cofactor of M_n, while $c_{ij}(M_{n-1})$ are the corresponding cofactors of the matrix M_{n-1}.

It is crucial to note that, since M_{n-1} has rank at least $n-2$, at least one of the cofactors $c_{ij}(M_{n-1})$ is nonzero. Set $c := (\sum_{2 \leq i,j \leq n} c_{ij}(M_{n-1})^2)^{1/2}$ and $a_{ij} := c_{ij}(M_{n-1})/c$.

Roughly speaking, our approach consists of two main steps.

- *Step 1.* Assume that

$$\mathbf{P}_{x_{11},\dots,x_{1n}}((c_{11}(M_n)/c)m_{11} + \sum_{2 \leq i,j \leq n} a_{ij}m_{1i}m_{1j} = 0 | M_{n-1}) \geq n^{-B},$$

  Then there is a strong additive structure among the cofactors $c_{ij}(M_{n-1})$ of M_{n-1}.

- *Step 2.* The probability, with respect to M_{n-1}, that there is a strong additive structure among the $c_{ij}(M_{n-1})$ is negligible.

Here we use the subscript $\mathbf{P}_{x_{11},\dots,x_{1n}}$ to emphasize that the probability under consideration is taken with respect to the random variables $x_{11}, \dots, x_{1n}$.

We will execute Step 1 by proving Theorem 2.5 below (as a special case). Step 2 will be carried out by proving Theorem 2.6.

Case 2. $\det(M_n) \neq 0$. Let $C(M_n) = (c_{ij}(M_n))$, $1 \le i, j \le n$, be the matrix of the cofactors of M_n. We have

$$C(M_n)\mathbf{y} = \det(M_n) \cdot \mathbf{x}.$$

Thus

$$\|C(M_n)\mathbf{y}\| = |\det(M_n)|.$$

By paying a factor of n in probability, without loss of generality we can assume that

$$|c_{11}(M_n)y_1 + \dots c_{1n}(M_n)y_n| \ge |\det(M_n)|/n^{1/2}.$$

Note that $\|\mathbf{y}\| \le n^{-A}$, thus

$$\sum_{j=1}^{n} |c_{1j}(M_n)|^2 \ge n^{2A-1} \det(M_n)^2. \tag{2.1}$$

For $j \ge 2$, we write

$$c_{1j}(M_n) = \sum_{i=2}^{n} m_{i1} c_{ij}(M_{n-1}),$$

where M_{n-1} is the matrix obtained from M_n by removing its first row and first column, and $c_{ij}(M_{n-1})$ are the corresponding cofactors of M_{n-1}.

Hence, by the Cauchy-Schwarz inequality, by Condition 2, and by the bounds $f_{ij} \le n^{\gamma}$ for the entries of F_n, we have

$$\begin{aligned} c_{1j}(M_n)^2 &\le \textstyle\sum_{i=2}^{n} m_{i1}^2 \sum_{i=2}^{n} c_{ij}^2(M_{n-1}) \\ &\le n^{2B+2\gamma+3} \textstyle\sum_{i=2}^{n} c_{ij}^2(M_{n-1}). \end{aligned} \tag{2.2}$$

Similarly, for $j = 1$ we write

$$c_{11}(M_n) = \sum_{i=2}^{n} m_{i2} c_{i2}(M_{n-1}).$$

Thus,

$$c_{11}(M_n)^2 \le n^{2B+2\gamma+3}\sum_{i=2}^{n} c_{i2}^2(M_{n-1}). \tag{2.3}$$

It follows from (2.1),(2.2) and (2.3) that

$$\sum_{2\le i,j\le n} c_{ij}(M_{n-1})^2 \ge n^{2A-2B-2\gamma-4}\det(M_n)^2.$$

Hence, for proving Theorem 1.3, it suffices to justify the following result.

Theorem 2.4. *For any $B>0$, there exists $A>0$ such that*

$$\mathbf{P}\big((\sum_{2\le i,j\le n} c_{ij}(M_{n-1})^2)^{1/2} \ge n^A|\det(M_n)|\big) \le n^{-B}.$$

To prove Theorem 2.4, we again express $\det(M_n)$ as a quadratic form of its first row.

$$\det(M_n) = c_{11}(M_n)m_{11} + \sum_{2\le i,j\le n} c_{ij}(M_{n-1})m_{1i}m_{j1}.$$

In other words,

$$\det(M_n)/c = m_{11}c_{11}/c + \sum_{2\le i,j\le n} a_{ij}m_{1i}m_{1j},$$

where $c := (\sum_{2\le i,j\le n} c_{ij}(M_{n-1})^2)^{1/2}$ and $a_{ij} := c_{ij}(M_{n-1})/c$.

Roughly speaking, our approach in this case also consists of two main steps.

- *Step 1.* Assume that

  $$\mathbf{P}_{x_{11},\dots,x_{1n}}(|(c_{11}(M_n)/c)m_{11} + \sum_{2\le i,j\le n} a_{ij}m_{1i}m_{1j}| \le n^{-A}|M_{n-1}) \ge n^{-B}.$$

  Then there is a strong additive structure among the cofactors c_{ij}.

- *Step 2.* The probability, with respect to M_{n-1}, that there is a strong additive structure among the c_{ij} is negligible.

We now state our main supporting lemmas.

Theorem 2.5 (Step 1)**.** *Let $0<\epsilon<1$ be given constant. Assume that*

$$\sup_a \mathbf{P}_{x_2,\dots,x_n}(|\sum_{2\le i,j\le n} a_{ij}(x_i+f_i)(x_j+f_j) - a| \le n^{-A}) \ge n^{-B}$$

for some sufficiently large integer A, where M_{n-1} is the matrix obtained from M_n by removing its first row and first column, $a_{ij} = c_{ij}(M_{n-1})/c$, x_i are iid copies of ξ, and f_i are arbitrary fixed numbers. Then, there exists a vector $\mathbf{u} = (u_1, \dots, u_{n-1})$ satisfying the following properties.

- $\|\mathbf{u}\| \asymp 1$ *and* $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le n^{-A/2+O_{B,\epsilon}(1)}$ *for* $n - O_{B,\epsilon}(1)$ *rows of* M_{n-1}*.*
- *There exists a generalized arithmetic progression Q of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$ that contains at least $n - 2n^\epsilon$ components u_i.*
- *All the components u_i, and all the generators of the generalized arithmetic progression are rational numbers of the form p/q, where $|p|, |q| \le n^{A/2+O_{B,\epsilon}(1)}$.*

We refer the reader to Section 3 for a definition of generalized arithmetic progression.

In the second step of the approach, we show that the probability for M_{n-1} having the above properties is negligible.

Theorem 2.6 (Step 2)**.** *With respect to M_{n-1}, the probability that there exists a vector $\mathbf{u}$ as in Theorem 2.5 is* $\exp(-\Omega(n))$.

The rest of the paper is organized as follows. After a short discussion of the main lemmas, we prove Theorem 2.5 in Section 4 and conclude Theorem 2.6 in Section 5. The proof of Corollary 1.5 will be presented in Section 6.

3 The Lemmas

A classical result of Erdős [6] and Littlewood-Offord [11] asserts that if a_i are real numbers of magnitude $|a_i| \ge 1$, then the probability that the random sum $\sum_{i=1}^n a_i x_i$ concentrates on an interval of length one is of order $O(n^{-1/2})$, where x_i are iid copies of a Bernoulli random variable. This remarkable inequality has generated an impressive way of research, particularly from the early 1960s to the late 1980s. We refer the reader to [9, 10] and the references therein.

Motivated by inverse theorems from additive combinatorics (see [26, Chapter 5]), Tao and Vu brought a new view to the problem: find the underlying reason as to why the concentration probability of $\sum_{i=1}^n a_i x_i$ on a short interval is large.

Typical examples of a_i that have large concentration probability are *generalized arithmetic progressions* (GAPs).

A set Q is a *GAP of rank r* if it can be expressed as in the form

$$Q = \{g_0 + k_1 g_1 + \dots + k_r g_r | k_i \in \mathbf{Z}, K_i \le k_i \le K_i' \text{ for all } 1 \le i \le r\}$$

for some $\{g_0, \dots, g_r\}, \{K_1, \dots, K_r\}, \{K_1', \dots, K_r'\}$.

It is convenient to think of Q as the image of an integer box $B := \{(k_1, \dots, k_r) \in \mathbf{Z}^r | K_i \le k_i \le K_i'\}$ under the linear map

$$\Phi : (k_1, \dots, k_r) \mapsto g_0 + k_1 g_1 + \dots + k_r g_r.$$

The numbers g_i are the *generators* of P, the numbers K_i' and K_i are the *dimensions* of P, and $\mathrm{Vol}(Q) := |B|$ is the *size* of B. We say that Q is *proper* if this map is one to one, or equivalently if $|Q| = \mathrm{Vol}(Q)$. For non-proper GAPs, we of course have $|Q| < \mathrm{Vol}(Q)$. If $-K_i = K_i'$ for all $i \geq 1$ and $g_0 = 0$, we say that Q is *symmetric*.

A closer look at the definition of GAPs reveals that if a_i are very *close* to the elements of a *GAP* of rank $O(1)$ and size $n^{O(1)}$, then the probability that $\sum_{i=1}^n a_i x_i$ concentrates on a short interval is of order $n^{-O(1)}$, where x_i are iid copies of a Bernoulli random variable.

It was shown by Tao and Vu [22, 21, 24], in an implicit way, that these are essentially the only examples that have high concentration probability. An explicit and optimal version has been given in a recent paper by the current author and Vu.

We say that a is δ-close to a set Q if there exists $q \in Q$ such that $|a - q| \leq \delta$.

Theorem 3.1 (Inverse Littlewood-Offord theorem for linear forms, [14]). *Let $0 < \epsilon < 1$ and $B > 0$. Let $\beta > 0$ be an arbitrary real number that may depend on n. Suppose that $\sum_{i=1}^n a_i^2 = 1$, and*

$$\sup_a \mathbf{P}_{\mathbf{x}}(|\sum_{i=1}^n a_i(x_i + f_i) - a| \leq \beta) = \rho \geq n^{-B},$$

where $\mathbf{x} = (x_1, \dots, x_n)$, and x_i are iid copies of a random variable ξ satisfying Condition 1. Then, for any number n' between n^ϵ and n, there exists a proper symmetric GAP $Q = \{\sum_{i=1}^r k_i g_i : k_i \in \mathbf{Z}, |k_i| \leq L_i\}$ such that

- *At least $n - n'$ elements of a_i are β-close to Q.*
- *Q has small rank, $r = O_{B,\epsilon}(1)$, and small cardinality*

$$|Q| \leq \max\left(O_{B,\epsilon}(\frac{\rho^{-1}}{\sqrt{n'}}), 1\right).$$

- *There is a non-zero integer $p = O_{B,\epsilon}(\sqrt{n'})$ such that all steps g_i of Q have the form $g_i = \beta \frac{p_i}{p}$, with $p_i \in \mathbf{Z}$ and $p_i = O_{B,\epsilon}(\beta^{-1}\sqrt{n'})$.*

In this and all subsequent theorems, the hidden constants could also depend on c_1, c_2, c_3 of Condition 1. We could have written $O_{c_1,c_2,c_3}(.)$ everywhere, but these notations are somewhat cumbersome, and this dependence is not our focus, so we omit them.

Theorem 3.1 was proven in [14] with $c_1 = 1, c_2 = 2$ and $c_3 = 1/2$, but the proof there automatically extends to any constants $0 < c_1 < c_2$ and $0 < c_3$.

To prove Theorem 2.5, we need a similar inverse-type result for the quadratic form $\sum_i a_{ij}(x_i + f_i)(x_j + f_j)$. We will invoke the following theorem from [13].

Theorem 3.2 (Inverse Littlewood-Offord theorem for quadratic forms, [13]). *Let $0 < \epsilon < 1$ and $B > 0$. Let $\beta > 0$ be an arbitrary real number that may depend on n. Assume that $a_{ij} = a_{ji}$, where $\sum_{i,j} a_{ij}^2 = 1$, and*

$$\sup_a \mathbf{P}_{\mathbf{x}}(|\sum_{i,j \leq n} a_{ij}(x_i + f_i)(x_j + f_j) - a| \leq \beta) = \rho \geq n^{-B}.$$

Then, there exist an integer $k \neq 0, |k| = n^{O_{B,\epsilon}(1)}$, a set of $r = O(1)$ rows $\mathbf{r}_{i1}, \dots, \mathbf{r}_{ir}$ of $A_n = (a_{ij})$, and set I of size at least $n - 2n^{\epsilon}$ such that for each $i \in I$, there exist integers $k_{ii_1}, \dots, k_{ii_r}$, all bounded by $n^{O_{B,\epsilon}(1)}$, such that the following holds.

$$\mathbf{P}_{\mathbf{z}}(|\langle \mathbf{z}, k\mathbf{r}_i(A_n) + \sum_{j=1}^{r} k_{ii_j}\mathbf{r}_{i_j}(A_n)\rangle| \leq \beta n^{O_{B,\epsilon}(1)}) \geq n^{-O_{B,\epsilon}(1)}, \tag{3.1}$$

where $\mathbf{z} = (z_1, \dots, z_n)$ and z_i are iid copies of $\eta^{(1/2)}(\xi - \xi')$, where $\eta^{(1/2)}$ is a Bernoulli random variable of parameter $1/2$ independent of ξ and ξ'.

4 proof of Theorem 2.5

We first apply Theorem 3.2 to a_{ij} to obtain

$$\mathbf{P}_{\mathbf{z}}(|\langle \mathbf{z}, k\mathbf{r}_i(A_n) + \sum_{j} k_{ii_j}\mathbf{r}_{i_j}(A_n)\rangle| \leq n^{-A+O_{B,\epsilon}(1)}) \geq n^{-O_{B,\epsilon}(1)}.$$

For short, we denote by $\mathbf{r}'_i$ the vector $k\mathbf{r}_i(A_n) + \sum_j k_{ii_j}\mathbf{r}_{i_j}(A_n)$. Thus, for any $i \in I$,

$$\mathbf{P}_{\mathbf{z}}(|\langle \mathbf{z}, \mathbf{r}'_i\rangle| \leq n^{-A+O_{B,\epsilon}(1)}) \geq n^{-O_{B,\epsilon}(1)}. \tag{4.1}$$

Ideally, our next step is to apply Theorem 3.1 to the $\mathbf{r}'_i$. However, the application is meaningful only when $\|\mathbf{r}'_i\|$ is relatively large. Investigating the degenerate case is our next goal.

Set

$$K = n^{-A/2}.$$

We consider two cases.

Case 1.(*degenerate case*) $\|\mathbf{r}'_i\| \leq K$ for all $i \in I$. Hence, with $I_0 := \{i_1, \dots, i_r\}$

$$\|k\mathbf{r}_i(A_n) + \sum_{j\in I_0} k_{ij}\mathbf{r}_j(A_n)\| = \|\mathbf{r}'_i\| \leq K. \tag{4.2}$$

Next, because $\sum_j \|\mathbf{c}_j(A_n)\|^2 = 1$, there exists an index j_0 such that $\|\mathbf{c}_{j_0}(A_n)\| \geq n^{-1/2}$. Consider this column vector.

It follows from (4.2) that for any $i \in I$,

$$|k\mathbf{c}_{j_0}(i) + \sum_{j\in I_0} k_{ij}\mathbf{c}_{j_0}(j)| \leq K.$$

The above inequality means that the components $\mathbf{c}_{j_0}(i)$ of $\mathbf{c}_{j_0}(A_n)$ belong to a GAP generated by $\mathbf{c}_{j_0}(j)/k, j \in I_0$, up to an error K. This suggests us the following approximation.

For each $j \notin I$, we approximate $\mathbf{c}_{j_0}(j)$ by a number v_j of the form $(1/\lfloor 2K^{-1} \rfloor) \cdot \mathbf{Z}$ such that $|v_j - \mathbf{c}_{j_0}(j)| \le K$. We next set

$$v_i := \sum_{j \in I_0} k_{ij} v_j / k$$

for any $i \in I$.

Thus, v_i belongs to a GAP of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$ for all $i \in I$.

With $\mathbf{v} = (v_1, \dots, v_{n-1})$, we have

$$\|\mathbf{v} - \mathbf{c}_{j_0}(A_n)\| \le K n^{O_{B,\epsilon}(1)}.$$

Furthermore, by Condition 2, and because $\langle \mathbf{c}_{j_0}(A_n), \mathbf{r}_i(M_{n-1}) \rangle = 0$ for $i \ne j_0$, we infer that

$$|\langle \mathbf{v}, \mathbf{r}_i(M_{n-1}) \rangle| \le K n^{O_{B,\epsilon}(1)}.$$

Note that $\|\mathbf{v}\| \gg n^{-1/2}$. Set $\mathbf{u} := \lfloor 1/\|\mathbf{v}\| \rfloor \cdot \mathbf{v}$, we then obtain

- $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1}) \rangle| \le n^{-A/2 + O_{B,\epsilon}(1)}$ for $n-2$ rows of M_{n-1}.
- There exists a GAP of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$ that contains at least $n - 2n^\epsilon$ components u_i.
- All the components u_i, and all the generators of the GAP are rational numbers of the form p/q, where $|p|, |q| \le n^{A/2 + O_{B,\epsilon}(1)}$.

Case 2.(*non-degenerate case*). There exists $i_0 \in I$ such that $\|\mathbf{r}'_{i_0}\| \ge K$. Because $\mathbf{r}'_{i_0} = k\mathbf{r}_{i_0}(A_n) + \sum_{j \in I_0} k_{i_0 j} \mathbf{r}_j(A_n)$, $\mathbf{r}'_{i_0}$ is orthogonal to $n - |I_0| - 1 = n - O_{B,\epsilon}(1)$ column vectors of M_{n-1}. Consequently, because M_{n-1} is symmetric, $\mathbf{r}'_{i_0}$ is orthogonal to $n - O_{B,\epsilon}(1)$ row vectors of M_{n-1}.

Set

$$\mathbf{v} := \mathbf{r}'_{i_0} / \|\mathbf{r}'_{i_0}\|.$$

Hence, $\langle \mathbf{v}, \mathbf{r}_i(M_{n-1}) \rangle = 0$ for at least $n - O_{B,\epsilon}(1)$ row vectors of M_{n-1}.

Also, it follows from (4.1) that

$$\mathbf{P}_{\mathbf{z}}(|\langle \mathbf{z}, \mathbf{v} \rangle| \le n^{-A/2 + O_{B,\epsilon}(1)}) \ge n^{-O_{B,\epsilon}(1)}. \tag{4.3}$$

Next, because the z_i satisfy Condition 1, Theorem 3.1 applying to (4.3) implies that $\mathbf{v}$ can be approximated by a vector $\mathbf{u}$ as follows.

- $|u_i - v_i| \le n^{-A/2+O_{B,\epsilon}(1)}$ for all i.
- There exists a GAP of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$ that contains at least $n - n^{\epsilon}$ components u_i.
- All the components u_i, and all the generators of the GAP are rational numbers of the form p/q, where $|p|, |q| \le n^{A/2+O_{B,\epsilon}(1)}$.

Note that, by the approximation above, we have $\|\mathbf{u}\| \asymp 1$ and $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le n^{-A/2+O_{B,\epsilon}(1)}$ for at least $n - O_{B,\epsilon}(1)$ row vectors of M_{n-1}.

5 Proof of Theorem 2.6

We first bound the number N of vectors $\mathbf{u}$ satisfying the conclusion of Theorem 2.6.

Because each GAP is determined by its generators and dimensions, the number of Qs is bounded by $(n^{A+O_{B,\epsilon}(1)})^{O_{B,\epsilon}(1)}(n^{O_{B,\epsilon}(1)})^{O_{B,\epsilon}(1)} = n^{O_{A,B,\epsilon}(1)}$.

Next, for a given Q of rank $O_{B,\epsilon}(1)$ and size $n^{O_{B,\epsilon}(1)}$ obtained from Theorem 2.5, there are at most $n^{n-2n^{\epsilon}}|Q|^{n-2n^{\epsilon}} = n^{O_{B,\epsilon}(n)}$ ways to choose the $n - 2n^{\epsilon}$ components u_i that Q contains.

The remaining components belong to the set $\{p/q, |p|, |q| \le n^{A/2+O_{B,\epsilon}(1)}\}$, so there are at most $(n^{A+O_{B,\epsilon}(1)})^{2n^{\epsilon}} = n^{O_{A,B,\epsilon}(n^{\epsilon})}$ ways to choose them.

Hence, we obtain the key bound

$$N \le n^{O_{A,B,\epsilon}(1)} n^{O_{B,\epsilon}(n)} n^{O_{A,B,\epsilon}(n^{\epsilon})} = n^{O_{B,\epsilon}(n)}. \tag{5.1}$$

Set $\beta_0 := n^{-A/2+O_{B,\epsilon}(1)}$, the bound obtained from the conclusion of Theorem 2.5. For a vector $\mathbf{u}$, we define $\mathbf{P}_{\beta_0}(\mathbf{u})$ as follows

$$\mathbf{P}_{\beta_0}(\mathbf{u}) := \mathbf{P}(|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le \beta_0 \text{ for } n - O_{B,\epsilon}(1) \text{ rows of } M_{n-1}).$$

From (5.1), for our task of proving Theorem 2.6, it would be ideal if we can show that the probability $\mathbf{P}_{\beta_0}(\mathbf{u})$ is smaller than $\exp(-\Omega(n))/N$ for each $\mathbf{u}$.

Roughly speaking, our strategy is to classify $\mathbf{u}$ into two classes: one contains of $\mathbf{u}$ of very small $\mathbf{P}_{\beta_0}(\mathbf{u})$, and thus their contribution is negligible; the other contains of $\mathbf{u}$ of relatively large $\mathbf{P}_{\beta_0}(\mathbf{u})$. To deal with those $\mathbf{u}$ of the second type, we will not control $\sum \mathbf{P}_{\beta_0}(\mathbf{u})$ directly but pass to a class of new vectors $\mathbf{u}'$ that are also almost orthogonal to many rows of M_{n-1}, while the probability $\sum \mathbf{P}_{\beta_0}(\mathbf{u}')$ is relatively smaller than $\sum \mathbf{P}_{\beta_0}(\mathbf{u})$. More details follow.

5.1 Technical reductions and key observations

By paying a factor of $n^{O_{B,\epsilon}(1)}$ in probability and without loss of generality we may assume that $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le \beta_0$ for the first $n - O_{B,\epsilon}(1)$ rows of M_{n-1}. Also, by paying another factor of n^{n^ϵ} in probability, we may assume that the first n_0 components u_i of $\mathbf{u}$ belong to a GAP Q, and $u_{n_0} \ge 1/2\sqrt{n-1}$, where $n_0 := n - 2n^\epsilon$. We refer to remaining u_i as exceptional components. Note that these extra factors do not affect our final bound $\exp(-\Omega(n))$.

For given $\beta > 0$ and $i \le n_0$, we define

$$\rho_\beta^{(i)}(\mathbf{u}) := \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}(|x_i u_i + \dots + x_{n_0} u_{n_0} - a| \le \beta),$$

where $x_i, \dots, x_{n_0}$ are iid copies of ξ.

A crucial observation is that, by exposing the rows of M_{n-1} one by one, and due to symmetry, the probability $\mathbf{P}_\beta(\mathbf{u})$ that $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le \beta$ for all $i \le n - O_{B,\epsilon}(1)$ can be bounded by

$$\begin{aligned} \mathbf{P}_\beta(\mathbf{u}) &\le \prod_{1\le i\le n-O_{B,\epsilon}(1)} \sup_a \mathbf{P}_{x_i,\dots,x_{n-1}}(|x_i u_i + \dots + x_{n-1}u_{n-1} - a| \le \beta) \\ &\le \prod_{1\le i\le n_0} \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}(|x_i u_i + \dots + x_{n_0}u_{n_0} - a| \le \beta) \\ &= \prod_{1\le i\le n_0} \rho_\beta^{(i)}(\mathbf{u}). \end{aligned} \tag{5.2}$$

Also, because of Condition 1 and $u_{n_0} \ge 1/2\sqrt{n-1}$, for any $\beta < c_1/2\sqrt{n-1}$ we have

$$\begin{aligned} \rho_\beta^{(k)}(\mathbf{u}) &\le \sup_a \mathbf{P}_{x_{n_0}}(|x_{n_0}u_{n_0} - a| \le \beta) \\ &\le 1 - c_3, \end{aligned} \tag{5.3}$$

and thus,

$$\mathbf{P}_\beta(\mathbf{u}) \le (1-c_3)^{n_0} = (1-c_3)^{(1-o(1))n}.$$

Next, let C be a sufficiently large constant depending on B and ϵ. We classify $\mathbf{u}$ into two classes $\mathcal{B}$ and $\mathcal{B}'$, depending on whether $\mathbf{P}_{\beta_0}(\mathbf{u}) \ge n^{-Cn}$ or not.

Because of (5.1), and as C is large enough,

$$\sum_{\mathbf{u}\in\mathcal{B}'} \mathbf{P}_{\beta_0}(\mathbf{u}) \le n^{O_{B,\epsilon}(n)}/n^{Cn} \le n^{-n/2}. \tag{5.4}$$

For the rest of the section, we focus on $\mathbf{u} \in \mathcal{B}$.

5.2 Approximation for degenerate vectors

Let $\mathcal{B}_1$ be the collection of $\mathbf{u} \in \mathcal{B}$ satisfying the following property: for any $n' = n^{1-\epsilon}$ components $u_{i_1}, \dots, u_{i_{n'}}$ among the $u_1, \dots, u_{n_0}$, we have

$$\sup_a \mathbf{P}_{x_{i_1},\dots,x_{i_{n'}}}(|u_{i_1}x_{i_1} + \dots + u_{i_{n'}}x_{i_{n'}} - a| \le n^{-B-4}) \ge (n')^{-1/2+o(1)}. \tag{5.5}$$

For consision we set $\beta = n^{-B-4}$. It follows from Theorem 3.1 that, among any $u_{i_1}, \dots, u_{i_{n'}}$, there are, say, at least $n'/2+1$ components that belong to an interval of length 2β. This is because our GAP Q now has only one element as in the size estimate the upper bound $O(\rho^{-1}/\sqrt{n'/2})$ is now $o(1)$. (One may also deduce this fact from the original Littlewood-Offord theorem.)

A simple argument then implies that there is an interval of length 2β that contains all but $n'-1$ components u_i. (To prove this, arrange the components in increasing order, then all but perhaps the first $n'/2$ and the last $n'/2$ components will belong to an interval of length 2β).

Thus there exists a vector $\mathbf{u}' \in (2\beta) \cdot \mathbf{Z}$ satisfying the following conditions.

- $|u_i - u'_i| \le 2\beta$ for all i.
- $u'_i = u$ for at least $n_0 - n'$ indices i.

Because of the approximation and of Condition 2 that $|x_{ij}| \le n^{B+1}$, whenever $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le \beta_0$, we have

$$|\langle \mathbf{u}', \mathbf{r}_i(M_{n-1})\rangle| \le n^{B+2}(2\beta) + \beta_0 := \beta'.$$

It is clear, from the bound on β and β_0, that $\beta' \le c_1/2\sqrt{n-1}$, and thus by (5.3),

$$\mathbf{P}_{\beta'}(\mathbf{u}') \le (1-c_3)^{(1-o(1))n}.$$

Now we bound the number of $\mathbf{u}'$ obtained from the approximation. First, there are $O(n^{n-n_0+n'}) = O(n^{2n^{1-\epsilon}})$ ways to choose those u'_i that take the same value u, and there are just $O(\beta^{-1})$ ways to choose u. The remaining components belong to the set $(2\beta)^{-1} \cdot \mathbf{Z}$, and thus there are at most $O((\beta^{-1})^{n-n_0+n'}) = O(n^{O_{A,B,\epsilon}(n^{1-\epsilon})})$ ways to choose them.

Hence we obtain the total bound

$$\begin{aligned}\sum_{\mathbf{u}\in\mathcal{B}_1} \mathbf{P}_{\beta_0}(\mathbf{u}) \le \sum_{\mathbf{u}'} \mathbf{P}_{\beta'}(\mathbf{u}') &\le O(n^{2n^{1-\epsilon}})O(n^{O_{A,B,\epsilon}(n^{1-\epsilon})})(1-c_3)^{(1-o(1))n} \\ &\le (1-c_3)^{(1-o(1))n}.\end{aligned}$$

5.3 Approximation for non-degenerate vectors

Assume that $\mathbf{u} \in \mathcal{B}_2 := \mathcal{B} \backslash \mathcal{B}_1$. By exposing the rows of M_{n-1} accordingly, and by paying an extra factor $\binom{n_0}{n'} = O(n^{n^{1-\epsilon}})$ in probability, we may assume that the components $u_{n_0-n'+1}, \dots, u_{n_0}$ satisfy the property

$$\sup_a \mathbf{P}_{x_{n_0-n'+1},\dots,x_{n_0}}(|u_{n_0-n'+1}x_{n_0-n'+1} + \cdots + u_{n_0}x_{n_0} - a| \le n^{-B-4}) \le (n')^{-1/2+o(1)}$$
$$\le n^{-1/2+\epsilon/2+o(1)}. \quad (5.6)$$

Next, define the following sequence $\beta_k, k \ge 0$. $\beta_0 = n^{-A/2+O_{B,\epsilon}(1)}$ is the bound obtained from the conclusion of Theorem 2.5, and

$$\beta_{k+1} := (2n^{B+2}+1)\beta_k.$$

Recall from (5.2) that

$$\mathbf{P}_{\beta_k}(\mathbf{u}) \le \prod_{1\le i\le n_0-n'} \rho^{(i)}_{\beta_k}(\mathbf{u}) =: \pi_{\beta_k}(\mathbf{u}).$$

Roughly speaking, the reason we truncated the product here is that whenever $i \le n_0 - n^{1-\epsilon}$, and β_k is small enough, the terms $\rho^{(i)}_{\beta_k}(\mathbf{u})$ are smaller than $(n')^{-1/2+o(1)}$, owing to (5.6). This fact will allow us to gain some significant factors when applying Theorem 3.1.

Note that $\pi_{\beta_k}(\mathbf{u})$ increases with k, and recall that $\pi_{\beta_0}(\mathbf{u}) \ge n^{-Cn}$. Thus, by the pigeonhole principle, there exists $k_0 := k_0(\mathbf{u}) \le C\epsilon^{-1}$ such that

$$\pi_{\beta_{k_0+1}}(\mathbf{u}) \le n^{\epsilon n}\pi_{\beta_{k_0}}(\mathbf{u}). \quad (5.7)$$

It is crucial to note that, since A was chosen to be sufficiently large compared to $O_{B,\epsilon}(1)$ and C, we have

$$\beta_{k_0+1} \le n^{-B-4}.$$

Having mentioned the upper bound of $\rho^{(i)}_{\beta_i}(\mathbf{u})$, we now turn to its lower bound. Because of Condition 2 and $u_i \le 1$ for all i, the following trivial bound holds for any $\beta \ge \beta_0$ and $i \le n_0 - n'$ by pigeonhole principle,

$$\rho^{(i)}_{\beta}(\mathbf{u}) \ge \beta n^{-B-2} \ge \beta_0 n^{-B-2} = n^{-A/2+O_{B,\epsilon}(1)}.$$

We next divide the interval $I = [n^{-A/2+O_{B,\epsilon}(1)}, n^{-1/2+\epsilon/2+o(1)}]$ into $K = (A/2 + O_{B,\epsilon}(1))\epsilon^{-1}$ subintervals $I_k = [n^{-A/2+O_{B,\epsilon}(1)+k\epsilon}, n^{-A/2+O_{B,\epsilon}(1)+(k+1)\epsilon}]$. For short, we denote by ρ_k the left endpoint of each I_k. Thus $\rho_k = n^{-A/2+O_{B,\epsilon}(1)+k\epsilon}$.

With all the necessary settings above, we now classify $\mathbf{u}$ basing on the distributions of the $\rho^{(i)}_{\beta_{k_0}}(\mathbf{u}), 1 \le i \le n_0 - n^{1-\epsilon}$.

For each $0 \le k_0 \le C\epsilon^{-1}$ and each tuple $(m_0, \dots, m_K)$ satisfying $m_0 + \dots + m_K = n_0 - n^{1-\epsilon}$, we let $\mathcal{B}^{(m_0,\dots,m_K)}_{k_0}$ denote the collection of those $\mathbf{u}$ from $\mathcal{B}_2$ that satisfy the following conditions.

- $k_0(\mathbf{u}) = k_0$.
- There are exactly m_k terms of the sequence $(\rho^{(i)}_{\beta_{k_0}}(\mathbf{u}))$ belonging to the interval I_k. In other words, if $m_0 + \dots + m_{k-1} + 1 \le i \le m_0 + \dots + m_k$ then $\rho^{(i)}_{\beta_{k_0}}(\mathbf{u}) \in I_k$.

Now we will use Theorem 3.1 to approximate $\mathbf{u} \in \mathcal{B}^{(m_0,\dots,m_K)}_{k_0}$ as follows.

- *First step.* Consider each index i in the range $1 \le i \le m_0$. Because $\rho^{(1)}_{\beta_{k_0}} \in I_0$, we apply Theorem 3.1 to approximate u_i by u_i' such that $|u_i - u_i'| \le \beta_{k_0}$ and the u_i' belong to a GAP Q_0 of rank $O_{B,\epsilon}(1)$ and size $O(\rho_0^{-1}/n^{1/2-\epsilon})$ for all but $n^{1-2\epsilon}$ indices i. Furthermore, all u_i' have the form $\beta_{k_0} \cdot p/q$, where $|p|, |q| = O(n\beta_{k_0}^{-1}) = O(n^{A/2+O_{B,\epsilon}(1)})$.
- *k-th step,* $1 \le k \le K$. We focus on i from the range $n_0 + \dots + n_{k-1} + 1 \le i \le n_0 + \dots + n_k$. Because $\rho^{(n_0+\dots+n_{k-1}+1)}_{\beta_{k_0}} \in I_k$, we apply Theorem 3.1 to approximate u_i by u_i' such that $|u_i - u_i'| \le \beta_{k_0}$ and u_i belongs to a GAP Q_k of rank $O_{B,\epsilon}(1)$ and size $O(\rho_k^{-1}/n^{1/2-\epsilon})$ for all but $n^{1-2\epsilon}$ indices i. Furthermore, all u_i' have the form $\beta_{k_0} \cdot p/q$, where $|p|, |q| = O(n\beta_{k_0}^{-1}) = O(n^{A/2+O_{B,\epsilon}(1)})$.
- For the remaining components u_i, we just simply approximate them by the closest point in $\beta_{i_0} \cdot \mathbf{Z}$.

We have thus provided an approximation of $\mathbf{u}$ by $\mathbf{u}'$ satisfying the following properties.

1. $|u_i - u_i'| \le \beta_{k_0}$ for all i.
2. $u_i' \in Q_k$ for all but $n^{1-2\epsilon}$ indices i in the range $m_0 + \dots + m_{k-1} + 1 \le i \le m_0 + \dots + m_k$.
3. All the u_i', including the generators of Q_k, belong to the set $\beta_{k_0} \cdot \{p/q, |p|, |q| \le n^{A/2+O_{B,\epsilon}(1)}\}$.
4. Q_k has rank $O_{B,\epsilon}(1)$ and size $|Q_k| = O(\rho_k^{-1}/n^{1/2-\epsilon})$.

Let $\mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}$ be the collection of all $\mathbf{u}'$ obtained from $\mathbf{u} \in \mathcal{B}^{(m_1,\dots,m_K)}_{k_0}$ as above. Observe that, as $|\langle \mathbf{u}, \mathbf{r}_i(M_{n-1})\rangle| \le \beta_{k_0}$ for all $i \le n - O_{B,\epsilon}(1)$, we have

$$|\langle \mathbf{u}', \mathbf{r}_i(M_{n-1})\rangle| \le (n^{B+2} + 1)\beta_{k_0}. \tag{5.8}$$

Hence, in order to justify Theorem 2.6 in the case $\mathbf{u} \in \mathcal{B}_2$, it suffices to show that the probability that (5.8) holds for all $i \le n - O_{B,\epsilon}(1)$, for some $\mathbf{u}' \in \mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}$, is small.

Consider a $\mathbf{u}' \in \mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}$ and the probability $\mathbf{P}_{(n^{B+2}+1)\beta_{k_0}}(\mathbf{u}')$ that (5.8) holds for all $i \le n - O_{B,\epsilon}(1)$. We have

$$\begin{aligned}\mathbf{P}_{(n^{B+2}+1)\beta_{k_0}}(\mathbf{u}') &\le \prod_{1\le i\le n_0-n^{1-\epsilon}} \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}(|u'_i x_i + \dots + u'_{n-1}x_{n_0} - a| \le (n^{B+2}+1)\beta_{k_0}) \\ &\le \prod_{1\le i\le n_0-n^{1-\epsilon}} \sup_a \mathbf{P}_{x_i,\dots,x_{n_0}}(|u_i x_i + \dots + u_{n-1}x_{n_0} - a| \le (2n^{B+2}+1)\beta_{k_0}) \\ &= \pi_{\beta_{k_0+1}}(\mathbf{u}) \le n^{\epsilon n}\pi_{\beta_{k_0}}(\mathbf{u}),\end{aligned}$$

where in the last inequality we used (5.7).

We recall from the definition of $\mathcal{B}^{(m_1,\dots,m_K)}_{k_0}$ that

$$\begin{aligned}\pi_{\beta_{k_0}}(\mathbf{u}) \le \prod_{k=1}^{K} \rho_{k+1}^{m_k} = n^{\epsilon(m_1+\dots+m_k)} \prod_{k=1}^{K} \rho_k^{m_k} \\ \le n^{\epsilon n} \prod_{k=1}^{K} \rho_k^{m_k}.\end{aligned}$$

Hence,

$$\mathbf{P}_{(n^{B+2}+1)\beta_{k_0}}(\mathbf{u}') \le n^{2\epsilon n} \prod_{k=1}^{K} \rho_k^{m_k}. \tag{5.9}$$

In the next step we bound the size of $\mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}$.

Because each Q_k is determined by its $O_{B,\epsilon}(1)$ generators from the set $\beta_{k_0}\cdot\{p/q, |p|, |q| \le n^{A/2+O_{B,\epsilon}(1)}\}$, and its dimensions from the integers bounded by $n^{O_{B,\epsilon}(1)}$, there are $n^{O_{A,B,\epsilon}(1)}$ ways to choose each Q_k. So the total number of ways to choose $Q_1,\dots,Q_K$ is bounded by

$$(n^{O_{A,B,\epsilon}(1)})^K = n^{O_{A,B,\epsilon}(1)}.$$

Next, after locating Q_k, the number N_1 of ways to choose u'_i from each Q_k is

$$\begin{aligned} N_1 &\le \prod_{k=1}^{K} \binom{m_k}{n^{1-2\epsilon}} |Q_k|^{m_k - n^{1-2\epsilon}} \\ &\le 2^{m_1+\cdots+m_K} \prod_{k=1}^{K} |Q_k|^{m_k} \\ &\le (O(1))^n \prod_{k=1}^{K} \rho_k^{-m_k} / n^{(1/2-\epsilon)(m_1+\cdots+m_k)} \\ &\le \prod_{k=1}^{K} \rho_k^{-m_k} / n^{(1/2-\epsilon-o(1))n}, \end{aligned}$$

where we used the bound $|Q_k| = O(\rho_k^{-1}/n^{1/2-\epsilon})$.

The remaining components u_i' can take any value from the set $\beta_{k_0} \cdot \{p/q, |p|, |q| \le n^{A/2+O_{B,\epsilon}(1)}\}$, so the number N_2 of ways to choose them is bounded by

$$N_2 \le (n^{A+O_{B,\epsilon}(1)})^{2n^{\epsilon}+Kn^{1-2\epsilon}} = n^{O_{A,B,\epsilon}(n^{1-2\epsilon})}.$$

Putting the bound for N_1 and N_2 together, we obtain a bound N for $|\mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}|$,

$$N \le \prod_{k=1}^{K} \rho_k^{-m_k} / n^{(1/2-\epsilon-o(1))n}. \tag{5.10}$$

It follows from (5.9) and (5.10) that

$$\sum_{\mathbf{u}' \in \mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}} \mathbf{P}_{(n^{B+2}+1)\beta_{k_0}}(\mathbf{u}') \le n^{2\epsilon n} \prod_{k=1}^{K} \rho_k^{m_k} \prod_{k=1}^{K} \rho_k^{-m_k} / n^{(1/2-\epsilon-o(1))n} \le n^{-(1/2-3\epsilon-o(1))n}. \tag{5.11}$$

Summing over the choices of k_0 and $(m_1,\dots,m_K)$ we obtain the bound

$$\sum_{k_0,m_1,\dots,m_K} \sum_{\mathbf{u}' \in \mathcal{B}'^{(m_1,\dots,m_K)}_{k_0}} \mathbf{P}_{(n^{B+2}+1)\beta_{k_0}}(\mathbf{u}') \le n^{-(1/2-3\epsilon-o(1))n},$$

completing the proof of Theorem 2.6.

6 Proof of Corollary 1.5

Assume that the upper diagonal entries of M_n satisfy the conditions of Corollary 1.5. We denote by $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ the real eigenvalues of M_n.

Our first ingredient is the following special form of the spectral concentration result of Guionnet and Zeitouni.

Lemma 6.1. *[8, Theorem 1.1] Assume that f is a convex Lipschitz function. Then for any $\delta \geq \delta_0 := 16C\sqrt{\pi}|f|_L/n$,*

$$\mathbf{P}\left(|\sum_{i=1}^{n} f(\lambda_i) - \mathbf{E}(\sum_{i=1}^{n} f(\lambda_i))| \geq \delta n\right) \leq 4\exp(-\frac{n^2(\delta-\delta_0)^2}{16C^2|f|_L^2}).$$

Following [3] and [7], we will apply the above theorem to the cut-off functions $f_\epsilon^+(x) := \log(\max(\epsilon, x))$ and $f_\epsilon^-(x) = \log(\max(\epsilon, -x))$, for some $\epsilon > 0$ to be determined. The main reason we have to truncate the log function is because it is not Lipschitz. Note that f^+ and f^{-1} both have Lipschitz constant ϵ^{-1}. Although they are not convex, it is easy to write them as difference of convex functions of Lipschitz constant $O(\epsilon^{-1})$, and so Lemma 6.1 applies. Thus the following estimates hold for $\delta \gg (\epsilon n)^{-1}$

$$\mathbf{P}\left(|\sum_{\lambda_i \in S_\epsilon^+} \log \lambda_i - \mathbf{E}(\sum_{\lambda_i \in S_\epsilon^+} \log \lambda_i)| \geq \delta n\right) \leq \exp(-\Theta(n^2\delta^2\epsilon^2))$$

and

$$\mathbf{P}\left(|\sum_{\lambda_i \in S_\epsilon^-} \log|\lambda_i| - \mathbf{E}(\sum_{\lambda_i \in S_\epsilon^-} \log|\lambda_i|)| \geq \delta n\right) \leq \exp(-\Theta(n^2\delta^2\epsilon^2)),$$

where $S_\epsilon^+ := \{\lambda_i, \lambda_i \geq \epsilon\}$ and $S_\epsilon^- := \{\lambda_i, \lambda_i \leq -\epsilon\}$.

Hence,

$$\mathbf{P}\left(|\sum_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} \log|\lambda_i| - \mathbf{E}(\sum_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} \log|\lambda_i|)| \geq 2\delta n\right) \leq \exp(-\Theta(n^2\delta^2\epsilon^2)). \quad (6.1)$$

Roughly speaking, (6.1) implies that $\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} |\lambda_i|$ is well concentrated around its mean. It thus remains to control the factor $R := \prod_{|\lambda_i| \leq \epsilon} |\lambda_i|$. We will bound R away from zero, relying on Theorem 1.3 and Lemma 6.2 below.

Lemma 6.2. *[25, Proposition 66], [5, Theorem 5.1] Assume that M_n is a random symmetric matrix of entries satisfying the conditions of Corollary 1.5. Then for all $I \subset \mathbf{R}$ with $|I| \geq K^2 \log^2 n/n^{1/2}$, one has*

$$N_I \ll n^{1/2}|I|$$

with probability $1 - \exp(-\omega(\log n))$, where N_I is the number of λ_i belonging to I.

We refer the readers to [4] for a survey of recent results on the distribution of the eigenvalues of M_n.

By Lemma 6.2, we have $|\{i, |\lambda_i| \leq \epsilon\}| \ll n^{1/2}\epsilon$. Also, Theorem 1.3 implies that $\min_i\{|\lambda_i|\} \geq n^{-A}$ with probability $1 - O(n^{-B})$. Thus

$$R = \prod_{|\lambda_i| \leq \epsilon} |\lambda_i| \geq (\min_i\{|\lambda_i|\})^{n^{1/2}\epsilon} = n^{-O(n^{1/2}\epsilon)}. \tag{6.2}$$

Our next goal is the following result.

Proposition 6.3. *With probability* $1 - n^{-\omega(1)}$ *we have*

$$\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} |\lambda_i| = \exp(-O(\epsilon^{-1}\log n + \epsilon^{-2}))\mathbf{E}(\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} |\lambda_i|) - \exp(\frac{2\log n}{\epsilon}) \tag{6.3}$$

and

$$\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} \lambda_i^2 = \exp(-O(\epsilon^{-1}\log n + \epsilon^{-2}))\mathbf{E}(\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} \lambda_i^2) - \exp(\frac{2\log n}{\epsilon}). \tag{6.4}$$

Let us complete the proof of the first half of Corollary 1.5 assuming Proposition 6.3. The second half follows by the same reasoning.

Firstly, because $\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} |\lambda_i| \geq \prod_{i=1}^n |\lambda_i|/\epsilon^{n-|S_\epsilon^- \cup S_\epsilon^+|} \geq \prod_{i=1}^n |\lambda_i| = |\det(M_n)|$, it follows from Proposition 6.3 that with probability $1 - n^{-\omega(1)}$,

$$\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} |\lambda_i| = \exp(-O(\epsilon^{-1}\log n + \epsilon^{-2}))\mathbf{E}(|\det(M_n)|) - \exp(\frac{2\log n}{\epsilon}). \tag{6.5}$$

Secondly, by (6.2), the following holds with probability $1 - O(n^{-B})$

$$|\det(M_n)| = \prod_{\lambda_i \notin S_\epsilon^- \cup S_\epsilon^+} |\lambda_i| \prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} |\lambda_i| \geq n^{-O(n^{1/2}\epsilon)} \prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} |\lambda_i|.$$

Combining with (6.5), we have

$$|\det(M_n)| = \exp(-O(\epsilon^{-1}\log n + \epsilon^{-2} + \epsilon n^{1/2}\log n))\mathbf{E}(|\det(M_n|) - n^{-O(n^{1/2}\epsilon)}\exp(\frac{2\log n}{\epsilon}).$$

By choosing $\epsilon = n^{-1/6}$, we obtain the conclusion of Corollary 1.5, noting that $\mathbf{E}(|\det(M_n)|) \gg \exp(n)$.

It remains to prove Proposition 6.3.

Proof. (of Proposition 6.3) Set

$$U := \sum_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} \log|\lambda_i| - \mathbf{E}(\sum_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} \log|\lambda_i|).$$

By (6.1) we have

$$\mathbf{P}(|U| \geq 2\delta n) \leq \exp(-\Theta(n^2\delta^2\epsilon^2)), \tag{6.6}$$

for $\delta \gg (n\epsilon)^{-1}$.

Also, note that $\mathbf{E}(U) = 0$. Thus, by Jensen inequality and by (6.6),

$$\begin{aligned} 1 \leq \mathbf{E}(\exp(U)) &\leq \mathbf{E}(\exp(|U|)) \\ &\leq 1 + \int_0^\infty \exp(t)\mathbf{P}(|U| \geq t)dt \\ &\leq 1 + \int_0^{\log n/\epsilon} \exp(t)dt + \int_{\log n/\epsilon}^\infty \exp(t)\exp(-\Theta(t^2\epsilon^2))dt \\ &= \exp(O(\epsilon^{-1}\log n + \epsilon^{-2})). \end{aligned} \tag{6.7}$$

Observe that

$$\mathbf{E}(\exp(U)) = \mathbf{E}(\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon +} |\lambda_i|)/\exp(\mathbf{E}(\sum_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} \log|\lambda_i|)).$$

It thus follows from (6.7) that

$$\exp(\mathbf{E}(\sum_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} \log|\lambda_i|)) = \exp(-O(\epsilon^{-1}\log n + \epsilon^{-2}))\mathbf{E}(\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon +} |\lambda_i|).$$

This relation, together with (6.6), imply that with probability $1 - n^{-\omega(1)}$,

$$\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} |\lambda_i| = \exp(-O(\epsilon^{-1}\log n + \epsilon^{-2}))\mathbf{E}(\prod_{\lambda_i \in S_\epsilon^- \cup S_\epsilon^+} |\lambda_i|) - \exp(\frac{2\log n}{\epsilon}).$$

The second half of Proposition 6.3 follows from the identical calculation applied to $\exp(2U)$.

□

A Proof of Lemma 2.2

Assume that $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbf{R}^n$ are independent vectors that span H. Also, without loss of generality, we assume that the subvectors $(v_{11}, \dots, v_{1k}), \dots, (v_{k1}, \dots, v_{kk})$ generate a full space of dimension k.

Consider a random vector $\mathbf{u} = (f_1 + x_1, \dots, f_n + x_n)$, where $x_1, \dots, x_n$ are iid copies of ξ. If $\mathbf{u} \in H$, then there exist $\alpha_1, \dots, \alpha_k$ such that

$$\mathbf{u} = \sum_{i=1}^{k} \alpha_i \mathbf{v}_i.$$

Note that $\alpha_1, \dots, \alpha_k$ are uniquely determined once the first k components of $\mathbf{u}$ are exposed. Thus we have

$$\mathbf{P}(\mathbf{u} \in H) \le \prod_{k+1 \le j} \mathbf{P}_{x_j}(x_j + f_j = \sum_{i=1}^{k} \alpha_i v_{ij}) \le (\sqrt{1-c_3})^{n-k},$$

where in the last estimate we use the fact (which follows from Condition 1) that $\sup_a \mathbf{P}(\xi = a) \le \sqrt{1-c_3}$.

Acknowledgements. The author would like to thank R. Pemantle and V. Vu for valuable discussions and comments. He is grateful to the referee for careful reading this manuscript and providing very helpful remarks.

References

[1] A. Barvinok, *Polynomial time algorithms to approximate permanents and mixed discriminants within a simply exponential factor*, Random Structures Algorithms 14 (1999), 29-61.

[2] K. Costello, T. Tao and V. Vu, *Random symmetric matrices are almost surely non-singular*, Duke Math. J. 135 (2006), 395-413.

[3] K. P. Costello and V. Vu, *Concentration of random determinants and permanent estimators*, Siam J. Discrete Math 23 (2009), 1356-1371.

[4] L. Erdős, *Universality of Wigner random matrices: a survey of recent results*, `arxiv.org/abs/1004.0861`.

[5] L. Erdős, B. Schlein and H-T. Yau, *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Notices, 3 (2010), 436-479.

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

[7] S. Friedland, B. Rider, and O. Zeitouni, *Concentration of permanent estimators for certain large matrices*, Ann. Appl. Probab., 14 (2004) 1559-1576.

[8] A. Guionnet and O. Zeitouni, *Concentration of the spectral measure for large matrices*, Electron. Comm. Probab.5 (2000) 119-136.

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

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

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

[12] H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, Duke Mathematics Journal Vol. 161, 4 (2012), 545-586.

[13] H. Nguyen, *A continuous variant of the inverse Littlewood-Offord problem for quadratic forms*, to appear in Contribution to Discrete Mathematics.

[14] H. Nguyen and V. Vu, *Optimal Littlewood-Offord theorems*, Advances in Math., Vol. 226 6 (2011), 5298-5319.

[15] A. Odlyzko, *On subspaces spanned by random selections of* ± 1 *vectors*, J. Combin. Theory Ser. A 47 (1988), no. 1, 124-133.

[16] M. Rudelson and R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances in Mathematics 218 (2008), 600-633.

[17] S. Smale, *On the efficiency of algorithms of analysis*, Bullentin of the AMS (13) (1985), 87-121.

[18] D. A. Spielman and S. H. Teng, *Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time*, J.ACM 51 (2004), no. 3, 385-463.

[19] T. Tao and V. Vu, *On* ± 1 *matrices: singularity and determinant*, Random Structures Algorithms 28 (2006), 1-23.

[20] T. Tao and V. Vu, *On the singularity probability of random Bernoulli matrices*, Journal of the A. M. S 20 (2007), 603-673.

[21] T. Tao and V. Vu, *Random matrices: The Circular Law*, Communication in Contemporary Mathematics 10 (2008), 261-307.

[22] T. Tao and V. Vu, *From the Littlewood-Offord problem to the circular law: universality of the spectral distribution of random matrices,* Bull. Amer. Math. Soc. (N.S.) 46 (2009), no. 3, 377–396.

[23] T. Tao and V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random matrices*, Annals of Mathematics (2) 169 (2009), no 2, 595-632.

[24] T. Tao and V. Vu, *Smooth analysis of the condition number and the least singular value*, Mathematics of Computation, 79 (2010), 2333-2352.

[25] T. Tao and V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math. 206 (2011), no. 1, 127-204.

[26] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge Univ. Press, 2006.

[27] R. Vershynin, *Invertibility of symmetric random matrices*, to appear in Random Structure and Algorithms.