Combinatorics, Probability and Computing (2022), **31**, pp. 662–683
doi:10.1017/S0963548321000511

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

On the smallest singular value of symmetric random matrices

Vishesh Jain[1*], Ashwin Sah[2] and Mehtaab Sawhney[2]

[1]Department of Statistics, Stanford University, Stanford, CA 94305, USA and [2]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*Corresponding author. Email: visheshj@stanford.edu

(Received 25 June 2021; revised 17 August 2021; accepted 28 September 2021; first published online 5 November 2021)

Abstract

We show that for an $n \times n$ random symmetric matrix A_n, whose entries on and above the diagonal are independent copies of a sub-Gaussian random variable ξ with mean 0 and variance 1,

$$\mathbb{P}[s_n(A_n) \le \epsilon/\sqrt{n}] \le O_\xi(\epsilon^{1/8} + \exp(-\Omega_\xi(n^{1/2}))) \quad \text{for all } \epsilon \ge 0.$$

This improves a result of Vershynin, who obtained such a bound with $n^{1/2}$ replaced by n^c for a small constant c, and $1/8$ replaced by $(1/8) - \eta$ (with implicit constants also depending on $\eta > 0$). Furthermore, when ξ is a Rademacher random variable, we prove that

$$\mathbb{P}[s_n(A_n) \le \epsilon/\sqrt{n}] \le O(\epsilon^{1/8} + \exp(-\Omega((\log n)^{1/4} n^{1/2}))) \quad \text{for all } \epsilon \ge 0.$$

The special case $\epsilon = 0$ improves a recent result of Campos, Mattos, Morris, and Morrison, which showed that $\mathbb{P}[s_n(A_n) = 0] \le O(\exp(-\Omega(n^{1/2})))$. Notably, in a departure from the previous two best bounds on the probability of singularity of symmetric matrices, which had relied on somewhat specialized and involved combinatorial techniques, our methods fall squarely within the broad geometric framework pioneered by Rudelson and Vershynin, and suggest the possibility of a principled geometric approach to the study of the singular spectrum of symmetric random matrices. The main innovations in our work are new notions of arithmetic structure – the Median Regularized Least Common Denominator (MRLCD) and the Median Threshold, which are natural refinements of the Regularized Least Common Denominator (RLCD)introduced by Vershynin, and should be more generally useful in contexts where one needs to combine anticoncentration information of different parts of a vector.

Keywords: random matrix theory; least singular value; random symmetric matrices

2020 MSC Codes: Primary: 60B20

1. Introduction

Let M_n denote an $n \times n$ random matrix, each of whose entries is an independent copy of a sub-Gaussian random variable ξ with mean 0 and variance 1. Prominent well-studied examples include the Ginibre ensemble (corresponding to $\xi = \mathcal{N}(0, 1)$) and i.i.d. Rademacher matrices (corresponding to the Rademacher random variable $\xi = \pm 1$ with probability $1/2$ each).

A landmark result of Rudelson and Vershynin [26] shows that there are absolute constants $C, c > 0$, depending only on the sub-Gaussian norm of ξ, for which

$$\mathbb{P}[s_n(M_n) \le \epsilon/\sqrt{n}] \le C\epsilon + 2e^{-cn} \quad \text{for all } \epsilon \ge 0, \tag{1}$$

© The Author(s), 2021. Published by Cambridge University Press.

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

where $s_n(M_n) = \inf_{v \in \mathbb{S}^{n-1}} \|Mv_2\|$ denotes the smallest singular value of M_n. Up to the constants $C, c > 0$, the above result is optimal, as can be seen by considering the two examples mentioned above. In particular, this result shows that the probability that an i.i.d. Rademacher matrix is singular is at most $2\exp(-cn)$ (for some $c > 0$), thereby recovering (and substantially generalising) a well-known result of Kahn, Komlós, and Szemerédi [13]. We remark that after a series of intermediate works [2, 27, 28], a breakthrough result of Tikhomirov [29] established that the probability of singularity of an i.i.d. Rademacher matrix is at most $(1/2 + o_n(1))^n$, which is optimal up to the $o_n(1)$ term.

In this paper, we will be concerned with $n \times n$ *symmetric* random matrices A_n i.e. $(A_n)_{ij} = (A_n)_{ji}$, each of whose entries on and above the diagonal is an independent copy of a sub-Gaussian random variable ξ with mean 0 and variance 1. We note that the identical distribution assumption may be significantly relaxed (in particular, allowing for the diagonal entries to have a different distribution), although for the sake of simplicity, we do not deal with this modification here; the interested reader is referred to [16, 30].

While symmetric matrices are especially convenient to work with linear algebraically, the lack of independence between the entries of A_n makes the non-asymptotic study of its smallest singular value considerably more challenging than that of M_n. In the early 1990s, it was conjectured by Weiss that $A_n(\text{Rad})$ (i.e. A_n where ξ is a Rademacher random variable) is invertible with probability $1 - o_n(1)$. This was only resolved in 2005 by Costello, Tao, and Vu [6], despite the corresponding statement for M_n (due to Komlós [14]) having been established almost 40 years prior (see also [8] for a recent simple proof).

Vershynin [30] showed that for any sub-Gaussian random variable ξ with mean 0 and variance 1, there are constants c, C_η depending only on the sub-Gaussian norm of ξ such that

$$\mathbb{P}[s_n(A_n) \le \epsilon/\sqrt{n}] \le C_\eta \epsilon^{1/8-\eta} + 2e^{-n^c}. \tag{2}$$

This improves (and generalizes) the nearly concurrent estimate of $O_C(n^{-C})$ on the singularity probability of $A_n(\text{Rad})$ obtained by Nguyen [21] using a novel quadratic variant of the inverse Littlewood–Offord theory. We note that in a subsequent work [22], Nguyen obtained estimates on the lower tail of $s_n(A_n)$ for a large class of random variables ξ, including those not covered by [30], although the quantitative bounds in this work are much weaker than (2).

Recently, the upper bound on the singularity probability of $A_n(\text{Rad})$ has been improved in a couple of works. Building on novel combinatorial techniques in [10], it was shown by Ferber and Jain [9] that this probability is at most $\exp(-\Omega(n^{1/4}\sqrt{\log n}))$. Subsequently, using a different combinatorial method inspired by the method of hypergraph containers [1], Campos, Mattos, Morris, and Morrison improved the bound to $\exp(-\Omega(\sqrt{n}))$. We note that both of these works deal only with $A_n(\text{Rad})$, and only with the singularity probability as opposed to quantitative estimates on $s_n(A_n(\text{Rad}))$.

The first main result of this paper is a strengthening of (2); the quantitative bounds are sufficiently powerful to generalize the aforementioned result of Campos et al. to all sub-Gaussian random variables.

Theorem 1.1 *Let A_n denote an $n \times n$ random symmetric matrix, each of whose entries on and above the diagonal is an independent copy of a sub-Gaussian random variable ξ with mean 0 and variance 1. Then, there are constants $C_{1.1}, c_{1.1}$ depending only on the sub-Gaussian norm of ξ such that, for all $\epsilon \ge 0$,*

$$\mathbb{P}[s_n(A_n) \le \epsilon/\sqrt{n}] \le C_{1.1}\epsilon^{1/8} + 2e^{-c_{1.1}n^{1/2}}.$$

Next, we consider the particularly well-studied case $\xi = \text{Rad}$; setting $\epsilon = 0$ in the theorem below improves the result of Campos et al. (see (3) in the Remark below).

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

Theorem 1.2 *Let A_n denote an $n \times n$ random symmetric matrix, each of whose entries on and above the diagonal is an independent Rademacher random variable. Then, there are absolute constants $C_{1.2}, c_{1.2}$ such that, for all $\epsilon \geq 0$,*

$$\mathbb{P}[s_n(A_n) \leq \epsilon/\sqrt{n}] \leq C_{1.2}\epsilon^{1/8} + 2e^{-c_{1.2}n^{1/2}(\log n)^{1/4}}.$$

Remark

1. We note that Theorem 1.2 can be extended to the setting of discrete random variables covered in recent work of the authors [11]. We leave the details to an interested reader.
2. The $\epsilon^{1/8}$ term on the right-hand side in Theorem 1.1 improves on the $\epsilon^{1/8+\eta}$ in [30]. It is believed that the correct dependence on ϵ is $O(\epsilon)$, which would be optimal in light of the Gaussian example.
3. The term $\exp(-\Omega(n^{1/2}))$ on the right-hand side in Theorem 1.1 extends the result of Campos, Mattos, Morris, and Morrison to general sub-Gaussian random variables, whereas Theorem 1.2 improves this result by a factor of $(\log n)^{1/4}$ in the exponent, in the special case when $\xi = \text{Rad}$. A well-known conjecture is that one should be able to replace this with $\exp(-\Omega(n))$, although this will likely require significant new ideas.

Our proof follows the geometric framework for studying the smallest singular value of random matrices, pioneered by Rudelson and Vershynin [26] for random matrices with i.i.d. entries (see the notes [25] for a highly readable introduction) and adapted for random symmetric matrices by Vershynin [30]. For the purpose of studying the singularity probability, the key ingredient in this framework is to identify an appropriate quantitative notion of arithmetic structure of vectors on the unit sphere, which on the one hand controls the anti-concentration properties of the vector and on the other hand permits the construction of sufficiently small epsilon-nets for its sublevel sets. In the seminal work of Rudelson and Vershynin [26], the (essential) Least Common Denominator (LCD) plays this role, crucially relying on the independence of the rows of the matrix in order to 'tensorize' the corresponding anti-concentration estimates. For the study of random symmetric matrices, where the rows are highly dependent, the LCD proves to be insufficient. Consequently, Vershynin [30] introduced the notion of the RLCD, which allows a limited amount of tensorization of the corresponding anti-concentration estimates, but at the cost of being able to take advantage of only a very small amount of randomness in the matrix.

The main innovation in our work are new notions of arithmetic structure of vectors, which we call the MRLCD (see Section 3) and the Median Threshold (see Section 4). Compared to the RLCD and its natural threshold analogue, the MRLCD and median threshold are able to exploit the information that many different coordinate projections (i.e. subvectors) of a vector are arithmetically unstructured in a simple and transparent manner, thereby allowing us to take advantage of substantially more randomness in the matrix compared to [30]. Moreover, we are able to show that level sets of the MRLCD and median threshold admit sufficiently small nets at the appropriate scale – for the MRLCD, this follows by suitably adapting by-now standard bounds due to Rudelson and Vershynin [26], whereas for the median threshold, we adapt work of Tikhomirov [29] on the singularity of i.i.d Bernoulli random matrices. As the details are anyway short, we defer further discussion to Sections 3 and 4.

We note that since its first appearance in [30], the RLCD has been used in many works (see, e.g., [17, 18, 20, 23, 31]); the MRLCD (and median threshold, for discrete distributions) can replace these applications in a black-box manner, and likely lead to improved quantitative estimates. We also note that a related use of combinatorially incorporating arithmetic unstructure of different projections of a vector appeared in recent work of the authors [12] on the probability of singularity

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

of the adjacency matrix of random regular digraphs; however, the interaction with both the net and anticoncentration estimates is more delicate here.

1.1 Notation

We will drop the dimension in the subscript, henceforth denoting A_n by A, and denoting its rows by $A_1, \ldots, A_n$. For an integer N, $\mathbb{S}^{N-1}$ denotes the set of unit vectors in $\mathbb{R}^N$, and $\mathbb{B}_2^N$ denotes the unit ball in $\mathbb{R}^N$ (i.e., the set of vectors of Euclidean norm at most 1). $\|\cdot_2\|$ denotes the standard Euclidean norm of a vector, and for a matrix $A=(a_{ij})$, $\|A\|$ is its spectral norm (i.e., $\ell^2 \to \ell^2$ operator norm), and $\|A_{\mathrm{HS}}\|$ is its Hilbert-Schmidt norm, defined by $\|A^2_{\mathrm{HS}}\| = \sum_{i,j} a_{ij}^2$.

We will let $[N]$ denote the interval $\{1, \ldots, N\}$, $\mathfrak{S}_{[N]}$ denote the set of permutations of $[N]$, and $\binom{[N]}{k}$ denote the set of subsets of $[N]$ of size exactly k. We will denote multisets by $\{\{\}\}$, so that $\{\{a_1, \ldots, a_n\}\}$, with the a_i's possibly repeated, is a multi-set of size n. For a vector $v \in \mathbb{R}^N$ and $T \subseteq [N]$, $v|_T$ denotes the $|T|$-dimensional vector obtained by only retaining the coordinates of v in T. We write $u \parallel v$ for $u, v \in \mathbb{R}^N$ if there is $t \in \mathbb{R}$ such that $u = tv$ or $tu = v$.

We will also make use of asymptotic notation. For functions f,g, $f = O_\alpha(g)$ (or $f \lesssim_\alpha g$ means that $f \le C_\alpha g$, where C_α is some constant depending on α; $f = \Omega_\alpha(g)$ (or $f \gtrsim_\alpha g$) means that $f \ge c_\alpha g$, where $c_\alpha > 0$ is some constant depending on α, and $f = \Theta_\alpha(g)$ means that both $f = O_\alpha(g)$ and $f = \Omega_\alpha(g)$ hold.

All logarithms are natural, unless indicated otherwise, and floors and ceilings are omitted when they make no essential difference.

1.2 Concurrent work

After uploading the first version of this manuscript, we learned of concurrent and independent work of Campos et al. [3] which, building on the techniques of [5], proved a singularity bound of $\exp(-\Omega(\sqrt{n \log n}))$ for symmetric Bernoulli matrices, a slightly improved bound over the $\epsilon = 0$ case of Theorem 1.2. However, the combinatorial methods used there do not provide quantitative information on the least singular value, and their results do not apply in as general a setting.

1.3 Subsequent work

Several months after the appearance of this article on the arXiv, a remarkable breakthrough work of Campos et al. [4] obtained an upper bound of the form $O(\exp(-cn))$ on the probability of singularity of $n \times n$ symmetric Rademacher matrices, also using geometric techniques.

2. Preliminaries

We will need the decomposition of the unit sphere into compressible and incompressible vectors, as formalized by Rudelson and Vershynin [26].

Definition 2.1 (Compressible and incompressible vectors) For $c_0, c_1 \in (0,1)$, $\mathrm{Comp}(c_0, c_1)$ consists of all vectors $v \in \mathbb{S}^{n-1}$ which are within Euclidean distance c_1 of some vector $w \in \mathbb{R}^n$ satisfying $|\operatorname{Supp}(w)| \le c_0 n$. Moreover, $\mathrm{Incomp}(c_0, c_1) := \mathbb{S}^{n-1} \setminus \mathrm{Comp}(c_0, c_1)$.

In order to prove Theorem 1.1, it suffices to analyze $\inf_{x \in \mathrm{Incomp}(c_0, c_1)} \|Ax_2\|$ due to the following.

Lemma 2.2 *There exist $c_0, c_1, c \in (0,1)$ depending only on the sub-Gaussian moment of ξ so that for any vector $u \in \mathbb{R}^n$, we have*

$$\mathbb{P}\left[\inf_{v \in \mathrm{Comp}(c_0, c_1)} \|Av - u_2\| < c\sqrt{n}\right] \le 2\exp(-cn).$$

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

Proof. This follows immediately by combining [30, Proposition 4.2] with the concentration of the operator norm of random matrices with independent, uniformly sub-Gaussian centred entries (cf. [30, Lemma 2.3]). □

Lemma 2.3 (Incompressible vectors are spread, cf. [30, Lemma 3.8]) *For every $c_0, c_1 \in (0,1)$, we can choose $c_{2.3} := c_{2.3}(c_0, c_1) \in (0, 1/5)$ depending only on c_0, c_1 such that the following holds. For every $v \in \mathrm{Incomp}\,(c_0, c_1)$, there are at least $2\lceil c_{2.3} n\rceil$ indices $k \in [n]$ such that*

$$\frac{c_1}{\sqrt{2n}} \le |v_k| \le \frac{1}{\sqrt{c_0 n}}.$$

Definition 2.4 (Spread set) For every $c_0, c_1 \in (0,1)$, and for every $v \in \mathrm{Incomp}\,(c_0, c_1)$, we assign a subset $\mathrm{Spread}\,(v) \subseteq [n]$ such that

$$|\,\mathrm{Spread}\,(v)| = \lceil c_{2.3} n\rceil, \textit{ and}$$
$$\frac{c_1}{\sqrt{2n}} \le |v_k| \le \frac{1}{\sqrt{c_0 n}} \text{ for all } k \in \mathrm{Spread}\,(v).$$

Definition 2.5 For every $c_0, c_1 \in (0,1)$ and for $\lambda \in (0, c_{2.3}/2)$, let $c_{2.5}(\lambda)n$ be the largest multiple of $\lceil \lambda n\rceil$ less than or equal to $\lceil c_{2.3} n\rceil$. Note that $c_{2.5}(\lambda) \ge \frac{c_{2.3}}{2}$.

To every $v \in \mathrm{Incomp}\,(c_0, c_1)$, we assign $\mathrm{Spread}_\lambda\,(v) \subseteq \mathrm{Spread}\,(v)$ such that

$$|\,\mathrm{Spread}_\lambda\,(v)| = c_{2.5}(\lambda)n,$$

and choose a partition

$$\mathrm{Spread}_\lambda\,(v) = \bigsqcup_{j=1}^{k} \mathrm{Spread}_\lambda^j\,(v)$$

into $k = c_{2.3}(\lambda)n/\lfloor \lambda n\rfloor$ disjoint subsets of size $\lfloor \lambda n\rfloor$. We further assume that the choice of $\mathrm{Spread}_\lambda\,(v)$ and $\mathrm{Spread}_\lambda^j\,(v)$ is uniform for a given choice of λ and $\mathrm{Spread}\,(v)$ (in particular, these choices do *not* depend directly on v).

In the above definition, we will ultimately choose $\lambda = \Theta(1/\sqrt{n})$ for Theorem 1.1 and $\lambda = \Theta((\log n)^{1/4}/\sqrt{n})$ for Theorem 1.2. We recall the definition of the Lévy concentration function.

Definition 2.6 For a random variable X and $\epsilon \ge 0$, the Lévy concentration of X of width ϵ is

$$\mathcal{L}(X, \epsilon) = \sup_{x\in\mathbb{R}} \mathbb{P}[|X - x| \le \epsilon].$$

We will also need a slight variant of the standard tensorization lemma, whose proof follows from the usual argument (cf. [26, Lemma 2.2]). We include the details for completeness.

Lemma 2.7 (Tensorization) *Let $X = (X_1, \ldots, X_N)$ be a random vector in $\mathbb{R}^N$ with independent coordinates. Suppose that for all $k \in [N]$, there exist $a_k, b_k \ge 0$ such that*

$$\sup_{X_1,\ldots,X_{k-1}} \mathcal{L}(X_k | X_1, \ldots, X_{k-1}, \epsilon) \le a_k\epsilon + b_k \quad \textit{for all } \epsilon \ge 0.$$

Then

$$\mathcal{L}(X, \epsilon\sqrt{N}) \le e^N \prod_{k=1}^{N} (a_k\epsilon + b_k).$$

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

Proof. We have

$$\begin{aligned}\mathbb{P}[|X| \le \epsilon\sqrt{N}] &= \mathbb{P}\left[\sum_{j=1}^{N} X_j^2 \le \epsilon^2 N\right] = \mathbb{P}\left[N - \frac{1}{\epsilon^2}\sum_{j=1}^{N} X_j^2 \ge 0\right] \\ &\le \mathbb{E}\exp\left(N - \frac{1}{\epsilon^2}\sum_{j=1}^{N} X_j^2\right) \\ &\le e^N \prod_{j=1}^{N} \sup_{X_1,\ldots,X_{j-1}} \mathbb{E}\exp\left(-X_j^2/\epsilon^2 | X_1,\ldots,X_{j-1}\right).\end{aligned}$$

We finish by noting that for any realization of $X_1, \ldots, X_{j-1}$,

$$\begin{aligned}\mathbb{E}\exp\left(-X_j^2/\epsilon^2 | X_1,\ldots,X_{j-1}\right) &= \int_0^\infty 2ue^{-u^2}\mathbb{P}[|X_k| < \epsilon u | X_1,\ldots,X_{j-1}]\, du \\ &\le \int_0^\infty 2ue^{-u^2}(a_j\epsilon u + b_j)\, du \le a_j\epsilon + b_j.\end{aligned}$$ □

We also recall the definition of essential least common denominator (LCD). We use a log-normalized version due to Rudelson (unpublished), which also appears in [30].

Definition 2.8 (LCD). For $L \ge 1$ and $v \in \mathbb{S}^{N-1}$, the least common denominator (LCD) $D_L(v)$ is defined as

$$D_L(v) = \inf\left\{\theta > 0 : \operatorname{dist}(\theta v, \mathbb{Z}^N) < L\sqrt{\log_+(\theta/L)}\right\}.$$

Finally we will require the following anticoncentration inequality of Miroshnikov and Rogozin [19]; this generalizes a well-known inequality of Lévy-Kolmogorov-Rogozin [24], which is itself an extension of Erdös's solution [7] to the Littlewood-Offord problem.

Lemma 2.9 ([19, Corollary 1]). *Let $\xi_1, \ldots, \xi_N$ be independent random variables. Then, for any real numbers $r_1, \ldots, r_N > 0$ and any real $r \ge \max_{i\in[N]} r_N$, we have*

$$\mathcal{L}\left(\sum_{i=1}^{N} \xi_i, r\right) \le C_{2.9}\, r\left(\sum_{i=1}^{N} \frac{r_i^2(1-\mathcal{L}(\xi_i, r_i))}{\mathcal{L}(\xi_i, r_i)^2}\right)^{-1/2},$$

where $C_{2.9} > 0$ is an absolute constant.

3. Median regularized LCD (MRLCD)

In this section, we introduce the median regularized LCD (MRLCD), which is the notion of arithmetic structure that we will use in the proof of Theorem 1.1. As opposed to the regularized LCD (RLCD) (introduced in [30]) which guarantees only one arithmetically unstructured projection of the vector, a large MRLCD guarantees many arithmetically unstructured projections of the vector. This simple change allows the MRLCD to piece together various unstructured parts of the vector to obtain significantly better small-ball probability estimates (Proposition 3.5), while at the same time not significantly impacting the size of nets of level sets (Proposition 3.3).

Definition 3.1 (Median Regularized LCD). For $v \in \operatorname{Incomp}(c_0, c_1)$, $\lambda \in (0, c_{2.3})$, and $L \ge 1$, the median regularized LCD, denoted $\widehat{MD}_L(v, \lambda)$, is defined as

$$\widehat{MD}_L(v,\lambda) = \operatorname{median}\{\{D_L(v_I/\|v_I\|_2) : I = \operatorname{Spread}_\lambda^j(v) \text{ for some } j\}\}.$$

Here, the median of an even number of elements is not an average, but instead the value of the upper half. We denote by $I_M(v)$ the set $\mathrm{Spread}_\lambda^j(v)$ achieving the median (arbitrarily chosen from among all such sets), and $\mathcal{I}_M(v)$ the collection of sets attaining values at least that of the median. We let $\mathcal{J}_M(v)$ be the collection of sets attaining values at most that of the median.

We will consider level sets obtained by dyadically chopping the range of the MRLCD.

Definition 3.2 (Level sets of MRLCD) For $\lambda \in (0, c_{2.3})$, $L \geq 1$, and $D \geq 1$, we define the set

$$S_D = \{v \in \mathrm{Incomp}\,(c_0, c_1) : \widehat{MD}_L(v, \lambda) \in [D, 2D]\}.$$

3.1 Nets for level sets of MRLCD

The main result of this subsection is the following.

Proposition 3.3 *Let $c_0, c_1 \in (0, 1)$. There exists $C_{3.3} = C_{3.3}(c_0, c_1) > 0$ for which the following holds. Let $\lambda \in (C_{3.3}/n, c_{2.3}/2)$ and $L \geq 1$. For every $D \geq 1$, S_D has a β-net $\mathcal{N}$ such that*

$$\beta = \frac{L\sqrt{\log(2D)}}{D}, \quad |\mathcal{N}| \leq D^{1/\lambda}\left(\frac{C_{3.3}D}{\sqrt{\log(2D)}}\right)^n \cdot \left(\frac{\sqrt{\log(2D)}}{\sqrt{\lambda n}}\right)^{c_{2.3}n/8}.$$

Remark By changing $C_{3.3}$ by a constant factor, we can further assume that $\mathcal{N} \subseteq S_D$. Also, the L dependence here is not optimal – one can save a factor of $L^{n(1-c_{2.3}/8)}$ by being more careful, although this does not affect the overall bounds if L is of constant order as in our application.

The proof of Proposition 3.3 relies on a bound on the size of nets for level sets of the LCD.

Lemma 3.4 (Corollary of Lemma 7.8 in [30]) *Let $m \in \mathbb{N}$, $D \geq 1$, and $c \in (0, 1)$ be such that $D > c\sqrt{m} \geq 2$. There exists a constant C depending only on c for which the following holds. Let $\chi > 1$, $L \geq 1$, and $\lambda > 0$. Then the set*

$$\{x \in \sqrt{\chi\lambda}\mathbb{B}_2^m : c\sqrt{m} < D_L(x/\|x_2\|) \leq D\}$$

has a $\beta\sqrt{\chi\lambda}$-net $\mathcal{N}$ such that

$$\beta = \frac{4L\sqrt{\log(2D)}}{D}, \quad |\mathcal{N}| \leq \left(\frac{CD}{\sqrt{m}}\right)^m D^2.$$

Now we conclude the result.

Proof of Proposition 3.3. Let $r = \lceil \lambda n \rceil$ and $k = c_{2.5}(\lambda)n$, so that $r|k$ by definition.

Let $v \in S_D$. By paying a factor of at most 2^n in the size of our net, we may fix a realization of $\mathrm{Spread}\,(v)$, which determines $\mathrm{Spread}_\lambda(v)$ and $\mathrm{Spread}_\lambda^j(v)$ for $1 \leq j \leq k/r$. We pay an additional factor of at most 2^n to reveal which sets $\mathrm{Spread}_\lambda^j(v)$ are in $\mathcal{J}_M(v)$. Let $J \subseteq [k/r]$ be the collection of corresponding indices j. We see that $t := |J| \geq k/(2r)$ by definition of median. Write $J = \{j_1, \ldots, j_t\}$ and let $I_i = \mathrm{Spread}_\lambda^{j_i}(v)$.

Note that, given $I_1, \ldots, I_t$, we know that $D_L(v_{I_i}/\|v_{I_i 2}\|) \leq 2D$ for all $1 \leq i \leq t$. Moreover, since $I_i \subseteq \mathrm{Spread}_\lambda(v)$, it follows that $\|v_{I_i 2}\| \leq \sqrt{\chi\lambda}$ for some χ depending only on c_0. Further, by [30, Lemma 6.2], it follows (again, since $I_i \subseteq \mathrm{Spread}_\lambda(v)$) that $D_L(v_{I_i}/\|v_{I_i}\|_2) \geq c\sqrt{\lceil \lambda n \rceil}$ for some c depending only on c_0, c_1. Hence, by Lemma 3.4, we have a $\beta\sqrt{\chi\lambda}$-net for $\{v_{I_i}\}_{v \in S_D}$ where

$$\beta = \frac{2L\sqrt{\log(4D)}}{D}$$

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

of size at most

$$\left(\frac{CD}{\sqrt{\lceil \lambda n \rceil}}\right)^{\lceil \lambda n \rceil} D^2.$$

Finally, we take a product of these nets over $1 \leq i \leq t$, along with a standard β-net of $\mathbb{B}_2^{I_0}$ (this net has size at most $(1+3/\beta)^{|I_0|}$), where we let $I_0 = [n] \setminus (I_1 \cup \cdots \cup I_t)$, to obtain the desired conclusion upon adjusting the value of β by standard arguments. □

3.2 Anticoncentration via MRLCD

We derive anticoncentration for a fixed vector with respect to MRLCD; the key idea is to patch together anticoncentration estimates on different segments of the vector through the use of Lemma 2.9.

Proposition 3.5 (Anticoncentration via the MRLCD). *Let $\xi_1, \ldots, \xi_n$ be i.i.d. random variables. Suppose that there exist $\epsilon_0, p_0, M_0 > 0$ such that $\mathcal{L}(\xi_k, \epsilon_0) \leq 1 - p_0$ and $\mathbb{E}[|\xi_k|] \leq M_0$ for all k. Finally, let $c_0, c_1 \in (0, 1)$. Then, there exist $C_{3.5}$, depending only on ϵ_0, p_0, M_0 and $C'_{3.5}$ depending on $\epsilon_0, p_0, M_0, c_0, c_1$ such that the following holds.*

Let $L \geq p_0^{-1/2}$, $\lambda \in (C'_{3.5}L^2/n, c_{2.3})$, $v \in \mathrm{Incomp}\,(c_0, c_1)$, $J \subseteq \mathrm{Spread}_\lambda\,(v)$, and $S_J = \sum_{k \in J} v_k \xi_k$. Suppose that J is a union of sets in $\mathcal{I}_M(v)$. Then for every $\epsilon \geq 0$, we have for sufficiently large n (depending on $\epsilon_0, p_0, M_0, c_0, c_1$) that

$$\mathcal{L}(S_J, \epsilon) \leq C_{3.5} L \left(\frac{\epsilon}{\sqrt{|J|/n}} + \frac{\sqrt{\lambda n/|J|}}{\widehat{MD}_L(v, \lambda)}\right).$$

Remark The above proposition should be compared with [30, Proposition 6.9], which bounds the Lévy concentration function in terms of the regularized LCD. The key difference is that the term $\sqrt{|J|/n}$ in the denominator of our bound is replaced by $\sqrt{\lambda}$, which is always smaller. In fact, in the application considered here, λ must be chosen to be $O(1/\sqrt{n})$, which makes the above proposition significantly more efficient than the corresponding proposition in [30] for most of the matrix row-vector products (which satisfy $|J|/n = \Theta(1)$).

Proof. Since $v \in \mathrm{Incomp}\,(c_0, c_1)$, we have $\widehat{MD}_L(v, \lambda) \geq c\sqrt{\lceil \lambda n \rceil}$ for some c depending only on c_0, c_1 ([30, Lemma 6.2]).

First, assume that $\epsilon \leq 1/(c\sqrt{n})$. Let $r = \lceil \lambda n \rceil$ and $k = c_{2.5}(\lambda)n$, so that $r|k$ by definition. For $i \in [k/r]$, let

$$S_i = \sum_{k \in \mathrm{Spread}_\lambda^i\,(v)} v_k \xi_k.$$

Let I be such that $J = \cup_{i \in I} \mathrm{Spread}_\lambda^i\,(v)$. Since J is a union of sets in $\mathcal{I}_M(v)$, and $D_L(v_I/\|v_I\|_2) \geq \widehat{MD}_L(v, \lambda)$ for each $I \in \mathcal{I}_M(v)$, it follows by standard anticoncentration estimates based on the LCD (see [30, Proposition 6.9] for the logarithmic version), that there exists an absolute constant $C > 0$ such that

$$\mathcal{L}(S_i, \epsilon) \leq CL\left(\frac{\epsilon}{\sqrt{\lambda}} + \frac{1}{\widehat{MD}_L(v, \lambda)}\right) < \frac{1}{2}$$

for all $i \in I$, where the latter inequality follows from the assumption that $\epsilon \leq 1/(c\sqrt{n})$, along with the lower bound on λ (by taking $C'_{3.5}$ sufficiently large depending on various parameters).

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

Now note that

$$S_J = \sum_{i \in I} S_i$$

and that the S_i are independent. Also, note that $|I| = |J|/\lceil \lambda n \rceil$. Therefore, by Lemma 2.9, we have

$$\begin{aligned}
\mathcal{L}(S_J, \epsilon) &\le C_{2.9}\epsilon \left(\sum_{i \in I} \frac{\epsilon^2(1 - \mathcal{L}(S_i, \epsilon))}{\mathcal{L}(S_i, \epsilon)^2} \right)^{-1/2} \\
&\le \frac{C_{2.9}\sqrt{2}}{\sqrt{|I|}} \max_{i \in I} \mathcal{L}(S_i, \epsilon) \le \frac{2C_{2.9}}{\sqrt{|J|/n}} \cdot CL \left(\epsilon + \frac{\sqrt{\lambda}}{\widehat{MD}_L(v, \lambda)} \right),
\end{aligned}$$

which proves the desired conclusion for $\epsilon \le 1/c\sqrt{n}$.

Finally, for $\epsilon > 1/(c\sqrt{n})$, we note that any interval of length 2ϵ can be tiled by at most $2\epsilon/\epsilon_0$ intervals of length $2\epsilon_0$, where $\epsilon_0 = 1/(2c\sqrt{n})$. Moreover, for such ϵ_0, we have

$$\epsilon_0 + \frac{\sqrt{\lambda}}{\widehat{MD}_L(v, \lambda)} \le 4\epsilon_0.$$

Hence, we have that for all $\epsilon > 1/(c\sqrt{n})$,

$$\begin{aligned}
\mathcal{L}(S_J, \epsilon) &\le \frac{2\epsilon}{\epsilon_0} \cdot \mathcal{L}(S_J, \epsilon_0) \\
&\le \frac{2\epsilon}{\epsilon_0} \cdot \frac{2C_{2.9}CL}{\sqrt{|J|/n}} \cdot 4\epsilon_0 \le 16C_{2.9}CL \cdot \frac{\epsilon}{\sqrt{\lambda}},
\end{aligned}$$

as desired. □

Next, we derive a small-ball result for (symmetric) matrix-vector products.

Lemma 3.6 *Fix $K \ge 1$, $c_0, c_1 \in (0, 1)$ and $v \in \text{Incomp}\,(c_0, c_1)$. There exists L depending only on the sub-Gaussian norm of ξ, and $c_{3.6}$, $C_{3.6}$ depending on the sub-Gaussian norm of ξ and on c_0, c_1 such that the following holds.*

Let $\lambda \in (C_{3.6}/n, c_{3.6})$ and suppose that $v \in S_D$ (with MRLCD defined with respect to λ, L). Then, for any $u \in \mathbb{R}^n$, we have

$$\mathbb{P}[\|Av - u\|_2 \le K\beta\sqrt{n}] \le \left(\frac{C_{3.6}L^2\sqrt{\log{(2D)}}}{D} \right)^{n - \lceil \lambda n \rceil},$$

where

$$\beta = \frac{L\sqrt{\log{(2D)}}}{D}.$$

Proof. Fix $u \in \mathbb{R}^n$ and $v \in S_D$. Note that for any permutation matrix P, $\|Av - u\|_2 \le K\beta\sqrt{n}$ occurs if and only if

$$\|(PAP^{-1})Pv - Pu\| \le K\beta\sqrt{n}.$$

Furthermore, $PAP^{-1} = PAP^{\intercal}$ has the same distribution as A. Therefore, we will be able to permute the indices of $[n]$ at our convenience (depending on v).

In particular, we may assume that $\text{Spread}_{\lambda}\,(v) = [c_{2.5}(\lambda)n]$. Let $A_t = \{k \in [n] : t\lceil \lambda n \rceil < k \le (t + 1)\lceil \lambda n \rceil\}$ (defined for $0 \le t \le T - 1$), where

$$T = \frac{c_{2.5}(\lambda)n}{\lceil \lambda n \rceil}.$$

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

We may also assume that $A_t \in \mathcal{I}_M(v)$ for all $t \le \lceil T/2 \rceil - 1$. Then, for all $\lceil \lambda n \rceil \le j \le c_{2.5}(\lambda)n$, the set $J = [j]$ has a subset of at least half the size which satisfies the assumptions of Proposition 3.5 (namely, the union of the first $\lfloor j/\lceil \lambda n \rceil \rfloor$ sets A_t).

Therefore, Proposition 3.5 implies that for all L sufficiently large depending on the sub-Gaussian norm of ξ, if $j \ge \lceil \lambda n \rceil$ and $\epsilon \ge 0$, then

$$\mathcal{L}((Av-u)_j | (Av-u)_{j+1,\ldots,n}, \epsilon) \le CC_{3.5}L\left(\frac{\epsilon}{\sqrt{j/n}} + \frac{\sqrt{\lambda n/j}}{D}\right),$$

where C depends only on c_0, c_1. Here, we have used that the first j elements of the jth row are independent of rows $j+1, \ldots, n$, and that the Lévy concentration is monotone under removing independent random variables from a sum.

Let $j' = \min(j, c_{2.5}(\lambda)n)$. Then, by Lemma 2.7, we deduce

$$\begin{aligned}
\mathbb{P}[\|Av-u\|_2 \le K\beta\sqrt{n}] &\le \mathbb{P}\left[\sum_{j=\lceil \lambda n \rceil}^{n} (Av-u)_j^2 \le K^2\beta^2 n\right] \\
&\le (CL)^n \prod_{j=\lceil \lambda n \rceil}^{n} \left(\frac{K\beta\sqrt{n/(n-\lceil \lambda n \rceil)}}{\sqrt{j'/n}} + \frac{\sqrt{\lambda n/j'}}{D}\right) \\
&\le (CL)^n \prod_{j=\lceil \lambda n \rceil}^{n} \left(\frac{KL\sqrt{n\log(2D)}}{D\sqrt{j'}}\right) \\
&\le \left(\frac{CL^2\sqrt{\log(2D)}}{D}\right)^{n-\lceil \lambda n \rceil},
\end{aligned}$$

where the last inequality uses $\prod_{j=1}^{n}(n/j) \le e^n$. □

3.3 Structure theorem

We will need the following structure theorem, which shows that, except with exponentially small probability, the preimage under A of any fixed vector is highly unstructured. This is our replacement for the key [30, Theorem 7.1]. As usual, A denotes a random $n \times n$ symmetric matrix with independent ξ entries on and above the diagonal.

Theorem 3.7 *Fix $K \ge 1$. Depending on the sub-Gaussian norm of ξ, we can choose L, c, C so that the following holds. For all $\lambda \in (C/n, 1/\sqrt{n})$, we have for any $u \in \mathbb{R}^n$ that*

$$\mathbb{P}[\exists v \in \mathbb{S}^{n-1} : (Av \parallel u) \wedge (v \in \mathrm{Comp}(c_0, c_1) \vee \widehat{MD}_L(v, \lambda) \le 2^{\lambda n/C}) \wedge (\|A\| \le K\sqrt{n})] \le 2e^{-cn}.$$

Proof. This is an immediate consequence of Proposition 3.3, Lemmas 3.2 and 3.6. Note that if $Av = tu$ for $t \in \mathbb{R}$, then $\|A\| \le K\sqrt{n}$ implies $\|tu\|_2 \le K\sqrt{n}$. For compressible vectors v, we use Lemma 2.2 on a constant amount of target vectors parallel to u so as to cover the full range. For the rest, if the MRLCD is between D and $2D$ (for some $D \le 2^{\lambda n/C}$), we take a net constructed in Lemma 3.4 along with a $1/D$-net for $\{tu : \|tu\|_2 \le K\sqrt{n}\}$, which adds an additional (unimportant) factor of $KD\sqrt{n}$ to the size of our nets. Since

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

$$KD\sqrt{n}\cdot D^{1/\lambda}\left(\frac{C_{3.3}D}{\sqrt{\log(2D)}}\right)^{n}\cdot\left(\frac{\sqrt{\log(2D)}}{\sqrt{\lambda n}}\right)^{c_{2.3}n/8}\times\left(\frac{C_{3.6}L^2\sqrt{\log(2D)}}{D}\right)^{n-\lceil\lambda n\rceil}$$
$$=KD\sqrt{n}\cdot D^{1/\lambda}\left(\frac{C'D}{\sqrt{\log(2D)}}\right)^{\lceil\lambda n\rceil}\left(\frac{\sqrt{\log(2D)}}{\sqrt{\lambda n}}\right)^{c_{2.3}n/8}$$
$$\le C'^{n}\cdot 2^{\lambda^2 n^2/C}\cdot(1/C)^{c_{2.3}n/16}$$
$$\le C'^{n}2^{n/C}(1/C)^{c_{2.3}n/16},$$

the result follows by a union bound upon taking C sufficiently large. We omit the standard details, referring the reader to the proof of [30, Theorem 7.1] for a more detailed calculation. □

4. Median threshold

We begin by defining an alternate notion of structure, based on the so-called threshold function (Definition 4.1), which will allow us to use results of Tikhomirov [29] to obtain a stronger bound for the probability of singularity Rademacher random symmetric matrices. We note that, although we have chosen to focus on the Rademacher case, our analysis can be extended to general real discrete distributions using recent results of the authors [11].

For a technical reason that will become clear later, we fix a sufficiently small absolute constant $p\in(0,1/2]$ throughout this section; for the case of Rademacher random variables, which is our focus, one can take $p=1/10$.

Definition 4.1 For $p\in(0,1)$, $L\ge 1$, and $v\in\mathbb{S}^{N-1}$, the threshold $\mathcal{T}_{p,L}(v)$ is defined as

$$\mathcal{T}_{p,L}(v)=\sup\left\{t\in(0,1):\mathcal{L}\left(\sum_{i=1}^{N}b_i{}'v_i,t\right)>Lt\right\},$$

where the $b_i{}'$ are i.i.d. random variables distributed as $\mathrm{Ber}\,(p)-\mathrm{Ber}\,'(p)$.

Definition 4.2 For $p\in(0,1)$, $v\in\mathrm{Incomp}\,(c_0,c_1)$, $\lambda\in(0,c_{2.3})$, and $L\ge 1$, the median threshold, denoted $\widehat{\mathcal{T}}_{p,L}(v,\lambda)$, is defined as

$$\widehat{\mathcal{T}}_{p,L}(v,\lambda)=\mathrm{median}\{\{\mathcal{T}_{p,L}(v_I/\|v_I\|_2):I=\mathrm{Spread}_{\lambda}^{j}\,(v)\text{ for some }j\}\}.$$

4.1. Threshold of random lattice points

We next recall the key technical result of Tikhomirov [29], which upper bounds the number of vectors with “large” threshold within a lattice of appropriate size. The important fact is that the number of such vectors is *superexponentially* small compared to the size of the lattice, which is the key difference with the results coming from the MRLCD. First, we must establish some notation.

Definition 4.3 Choose $N,n\ge 1$ and $\delta\in(0,1]$, as well as $K\ge 1$. We say that $\mathcal{A}\subseteq\mathbb{Z}^n$ is (N,n,K,δ)-*admissible* if the following hold:

- $\mathcal{A}=A_1\times\cdots\times A_n$, where each A_i is an origin-symmetric subset of $\mathbb{Z}\cap(-nN,nN)$,
- A_i is an integer interval of size at least $2N+1$ for all $i>\delta n$,
- A_i is a union of two integer intervals of total size at least $2N$ and $A_i\cap[-N,N]=\emptyset$ for all $i\le\delta n$, and
- $|\mathcal{A}|\le(KN)^n$.

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

Theorem 4.4 (From [29, Corollary 4.3]). *Let $\delta, \epsilon \in (0, 1]$, $p \in (0, 1/2]$, and $K, M \geq 1$. There exist $n_{4.4} = n_{4.4}(\delta, \epsilon, K, M) \geq 1$ and $L_{4.4} = L_{4.4}(\delta, \epsilon, K) > 0$ such that the following holds. If $n \geq n_{4.4}$, $1 \leq N \leq (1 - p + \epsilon)^{-n}$, and $\mathcal{A}$ is (N, n, K, δ)-admissible, then*

$$\left|\left\{x \in \mathcal{A} : \mathcal{L}\left(\sum_{i=1}^{n} b_i x_i, \sqrt{n}\right) \geq L_{4.4} N^{-1}\right\}\right| \leq \exp(-Mn)|\mathcal{A}|,$$

where the b_i are i.i.d. Ber (p) *random variables. Furthermore, $n_{4.4} = \exp(C_{4.4}(\delta, \epsilon, K)M^2)$ is allowable.*

Remark. This is the same as [29, Corollary 4.3], except that we have claimed an explicit dependence between n and M, namely that one can take M growing as $(\log n)^{1/2}$ (all other parameters fixed). This is an immediate consequence of unraveling the parameter dependencies in [29, Theorem 4.2]. We give a brief sketch, using the notation of [29, Theorem 4.2]. In the proof of [29, Theorem 4.2], one sets $L = L_{4.5}(2M, p, \delta, \epsilon/2)$, which can be checked to grow exponentially in M by examining the last line of the proof of [29, Proposition 4.5]. This shows that the parameter q in the proof of [29, Theorem 4.2] is chosen to be linear in M, and hence, the parameter $\widetilde{\epsilon}$ grows as M^{-1}. Next, it is required that $n \geq n_{4.10}(p, \widetilde{\epsilon}, \max(16\widetilde{R}, L), \widetilde{R}, 2M)$ and $n \geq n_{4.5}(2M, p, \delta, \epsilon/2)$. The more restrictive condition comes from [29, Proposition 4.10], and indeed, an examination of the first few lines of the proof of this proposition reveals that it suffices to have n growing as $\exp(\Theta(M^2))$. One also sees that $\eta_{4.2} = \eta_{4.10}(p, \widetilde{\epsilon}, \max(16\widetilde{R}, L), \widetilde{R}, 2M)$ decays as $\exp(-\Theta(M^2))$. Finally, the deduction of [29, Corollary 4.3] from [29, Theorem 4.2] requires $n^{-1/2} \leq \eta$, for which n growing as $\exp(\Theta(M^2))$ is sufficient in light of the decay of η discussed above.

4.2. Replacement

In order to relate the anticoncentration of a vector with respect to Rademacher random variables to Definition 4.1 and 4.2, we will require the following inequality. This is closely related to the replacement trick employed by Kahn et al. [13] and later by Tao and Vu [28] (although the application here is substantially simpler).

Lemma 4.5 *There exists an absolute constant $C_{4.5}$ for which the following holds. Let $v \in \mathbb{R}^n$ and $r > 0$. Then, for any $0 < p \leq (2 - \sqrt{2})/4$,*

$$\mathcal{L}\left(\sum_{i=1}^{n} b_i v_i, r\right) \leq C_{4.5} \mathcal{L}\left(\sum_{i=1}^{n} b_i' v_i, r\right),$$

where b_i are independent Rademacher random variables and b_i' are distributed as Ber $(p) -$ Ber $'(p)$.

Proof. Note that by scaling v, we may assume without loss of generality that $r = 1$. Let $X = \sum_{i=1}^{n} b_i' v_i$. By Esseen's inequality and $|\cos t| \leq (3 + \cos(2t))/4$, we find

$$\mathcal{L}\left(\sum_{i=1}^{n} b_i v_i, 1\right) \leq C \int_{-2}^{2} \prod_{i=1}^{n} |\cos(v_i \theta)| d\theta \leq C \int_{-2}^{2} \prod_{i=1}^{n} \left(\frac{3}{4} + \frac{1}{4}\cos(2v_i\theta)\right) d\theta$$

$$\leq C \int_{-2}^{2} \prod_{i=1}^{n} \mathbb{E}\exp\left(i\theta \cdot 2b_i' v_i\right) d\theta \leq 2C \int_{\mathbb{R}} \mathbb{1}_{[-2,2]} * \mathbb{1}_{[-2,2]}(\theta)\mathbb{E}\exp(i\theta(2X))d\theta$$

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

$$
\begin{aligned}
&= 4C\mathbb{E}\left(\frac{\sin(4X)}{2X}\right)^2 \\
&\le 4C\mathbb{E}\left(\frac{\sin(4X)}{2X}\cdot \mathbb{1}_{X\in[-1,1]}\right)^2 + \sum_{k=1}^{\infty} 4C\mathbb{E}\left(\frac{\sin(4X)}{2X}\cdot \mathbb{1}_{\pm X\in[2k-1,2k+1]}\right)^2 \\
&\le 16C\mathcal{L}(X,1) + C\sum_{k=1}^{\infty}\frac{\mathcal{L}(X,1)}{(2k-1)^2} \le C'\mathcal{L}(X,1).
\end{aligned}
$$

The third inequality uses $p \le (2-\sqrt{2})/4$, and the penultimate inequality uses $\sin(4x)/(2x) \le 2$ for $x \in [-1, 1]$. □

4.3 Randomized rounding

We will make use of a slight modification of [29, Lemma 5.3], proved using randomized rounding (cf. [15]). As the proof is identical we omit the details.

Lemma 4.6 *Let $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ be a vector, Δ be a fixed distribution supported in $[-1, 1]^n$, and let $\mu > 0$, $\psi \in \mathbb{R}$ be fixed. There exist absolute constants $c_{4.6}$ and $C_{4.6}$ for which the following holds.*

Suppose that for all $t \ge \sqrt{n}$,

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} b_i y_i - \psi\right| \le t\right] \le \mu t,$$

where $(b_1, \ldots, b_n)$ are independent and distributed as Δ. Then, there exists a vector $y' \in \mathbb{Z}^n$ satisfying

(R1) $\|y - y'\|_\infty \le 1$,
(R2) $\mathbb{P}[|\sum_{i=1}^n b_i y_i' - \psi| \le t] \le C_{4.6}\mu t$ for all $t \ge \sqrt{n}$, and
(R3) $\mathcal{L}(\sum_{i=1}^n b_i y_i', \sqrt{n}) \ge c_{4.6}\mathcal{L}(\sum_{i=1}^n b_i y_i, \sqrt{n})$.

Next, we prove a version of the above proposition for the case when $\Delta = \mathrm{Ber}(p) - \mathrm{Ber}'(p)$ with p sufficiently small. The main difference is that the left hand side in (R2) above can be replaced by the Lévy concentration at width t; this can be done since for a distribution with non-negative characteristic function, the maximum concentration of given width is essentially obtained around $\psi = 0$.

Lemma 4.7 *Let $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ be a vector, $p \in (0, 1)$, and let $\mu > 0$, $\psi \in \mathbb{R}$ be fixed. There exist absolute constants $c_{4.7}$ and $C_{4.7}$ for which the following holds.*

Suppose that for all $t \ge \sqrt{n}$,

$$\mathcal{L}\left(\sum_{i=1}^{n} b_i' y_i, t\right) \le \mu t,$$

where the b_i' are independent and distributed as $\mathrm{Ber}(p) - \mathrm{Ber}'(p)$. Then, there exists a vector $y' = (y_1', \ldots, y_n') \in \mathbb{Z}^n$ satisfying

(R1) $\|y - y'\|_\infty \le 1$,
(R2) $\mathcal{L}(\sum_{i=1}^n b_i' y_i', t) \le C_{4.7}\mu t$ for all $t \ge \sqrt{n}$, and
(R3) $\mathcal{L}(\sum_{i=1}^n b_i' y_i', \sqrt{n}) \ge c_{4.7}\mathcal{L}(\sum_{i=1}^n b_i' y_i, \sqrt{n})$.

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

Proof. We apply Lemma 4.6 to the distribution $\Delta = \mathrm{Ber}\,(p) - \mathrm{Ber}\,'(p)$ and $\psi = 0$. From (R2),

$$\mathbb{P}\left[\left|\sum_{i=1}^{n} b_i' y_i\right| \le t\right] \le C_{4.6}\mu t$$

for all $t \ge \sqrt{n}$. Now let $t \ge \sqrt{n}$ and $X = (\sum_{i=1}^{n} b_i' y_i')/t$, and note that X has nonnegative characteristic function since b_i' does. Thus, for all $\psi \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}[|X-\psi| \le 1] &= \mathbb{E}[\mathbb{1}_{[-1,1]}(X-\psi)] \le \mathbb{E}[\mathbb{1}_{[-1,1]} * \mathbb{1}_{[-1,1]}(X-\psi)] \\ &= \int_{\mathbb{R}} \left(\frac{2\sin\theta}{\theta}\right)^2 \mathbb{E}\exp(i\theta(X-\psi))d\theta \\ &\le \int_{\mathbb{R}} \left(\frac{2\sin\theta}{\theta}\right)^2 |\mathbb{E}\exp(i\theta X)|d\theta \\ &= \int_{\mathbb{R}} \left(\frac{2\sin\theta}{\theta}\right)^2 \mathbb{E}\exp(i\theta X)d\theta \\ &= \mathbb{E}[\mathbb{1}_{[-1,1]} * \mathbb{1}_{[-1,1]}(X)] \le 2\mathbb{P}[|X| \le 2] \le 4C_{4.6}\mu t.\end{aligned}$$

□

4.4 Threshold structure theorem

We now prove the following improved version of Theorem 3.7.

Theorem 4.8 *Fix $K \ge 1$ and $0 < p \le (2-\sqrt{2})/4$. We can choose $L, c > 0$ and $c' = c'(p)$ so that the following holds for sufficiently large n. For all $\lambda \in (n^{-2/3}, c(\log n)^{1/4} n^{-1/2})$, we have for any $u \in \mathbb{R}^n$ that*

$$\mathbb{P}[\exists v \in \mathbb{S}^{n-1} : (Av \parallel u) \wedge (v \in \mathrm{Comp}\,(c_0, c_1) \vee \widehat{\mathcal{T}}_{p,L}(v,\lambda) \ge 2^{-c'\lambda n}) \wedge (\|A\| \le K\sqrt{n})] \le 2e^{-cn},$$

where A is a symmetric matrix with entries on and above the diagonal i.i.d. and distributed as the sum of a Rademacher random variable, and a Gaussian random variable of mean 0 and variance n^{-2n}.

Remark The Gaussian perturbation of the entries of A is not important here and will only be used later, where it will be convenient to assume that various sub-matrices of A are invertible almost surely. Moreover, the variance of the Gaussian is chosen sufficiently small so that all anticoncentration claims that we need are essentially unaffected by this perturbation.

Proof. As in the proof of Theorem 3.7, we can deal with compressible vectors using Lemma 2.2. Therefore, it remains to deal with incompressible vectors with ‘large’ median threshold.

By standard small-ball estimates for incompressible vectors using a quantitative central limit theorem (see [29, Lemma 5.1]), for $v \in \mathrm{Incomp}\,(c_0, c_1)$, there is $C_0 = C_0(p, c_0, c_1)$ such that $\widehat{\mathcal{T}}_{p,L}(v,\lambda) \le C_0(\lambda n)^{-1/2}$. We let $r = \lceil \lambda n \rceil$ and $k = c_{2.5}(\lambda)n$, so that $r|k$ by definition, and let $m = \lfloor k/(2r) \rfloor$.

Step 1: Randomized rounding. We consider the case $\widehat{\mathcal{T}}_{p,L}(v,\lambda) \in [1/T, 2/T]$, where $T \in [C_0^{-1}\sqrt{\lambda n}, 2^{c'\lambda n}]$. Then, by definition, there exist intervals $I_1, \ldots, I_m$ of the form $\mathrm{Spread}^j_\lambda(v)$ with

$$\mathcal{T}_{p,L}(v_{I_i}/\|v_{I_i}\|_2) \le 2/T$$

for all $i \in [m]$.

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

Let $D = C_1\sqrt{n}T$, where $C_1 = C_1(p, c_0, c_1) \geq 1$ will be an integer chosen later. Let $y = Dv$. By the definition of the threshold, for all $t \geq \sqrt{\lceil \lambda n \rceil}$ we have

$$\mathcal{L}\left(\sum_{j\in I_i} b_j{}'y_j, t\right) = \mathcal{L}\left(\sum_{j\in I_i} b_j{}'v_j, \frac{t}{D}\right) = \mathcal{L}\left(\sum_{j\in I_i} b_j{}'\frac{v_j}{\|v_{I_i}\|_2}, \frac{t}{D\|v_{I_i}\|_2}\right)$$
$$\leq \mathcal{L}\left(\sum_{j\in I_i} b_j{}'\frac{v_j}{\|v_{I_i}\|_2}, \frac{\sqrt{2}t}{c_1 C_1\sqrt{\lceil \lambda n\rceil}T}\right) \leq \frac{L}{T}\cdot\frac{2t}{\sqrt{\lceil \lambda n\rceil}},$$

as long as we chose $C_1 > \sqrt{2}/c_1$, where the $b_i{}'$ are independent random variables distributed as $\mathrm{Ber}\,(p) - \mathrm{Ber}\,'(p)$. Applying Lemma 4.7 to the $\lceil \lambda n\rceil$-dimension vector y_{I_i}, we see that there is $y'_{I_i} \in \mathbb{Z}^{\lceil \lambda n\rceil}$ satisfying the conclusions of Lemma 4.7 (with n replaced by $\lceil \lambda n\rceil$). In particular, by (R3), we see that

$$\mathcal{L}\left(\sum_{j\in I_i} b'_j y'_j, \sqrt{\lceil \lambda n\rceil}\right) \geq c_{4.7}\mathcal{L}\left(\sum_{j\in I_i} b'_j v_j, \sqrt{\lambda}/(C_1 T)\right)$$
$$\geq c_{4.7}\mathcal{L}\left(\sum_{j\in I_i} b'_j\frac{v_j}{\|v_{I_i}\|_2}, \frac{2\sqrt{\lambda}}{C_1 T\cdot\sqrt{\lambda/c_0}}\right)$$
$$\geq C_1^{-1}\sqrt{c_0}\cdot c_{4.7}\mathcal{L}\left(\sum_{j\in I_i} b'_j\frac{v_j}{\|v_{I_i}\|_2}, 2/T\right)$$
$$\geq C_1^{-1}\sqrt{c_0}\cdot c_{4.7}\cdot 2LT^{-1}. \tag{3}$$

Let $I_0 = [n]\setminus(I_1\cup\cdots\cup I_m)$. Then, by approximating each coordinate of y_{I_0} by the nearest integer, and combining with the above integer approximations of $y_{I_1},\ldots,y_{I_m}$, we obtain an integer vector $y' \in \mathbb{Z}^n$ such that for all $1\leq i\leq m$, the $\lceil \lambda n\rceil$-dimensional integer vector y'_{I_i} satisfies $\|y_{I_i} - y'_{I_i}\|_\infty \leq 1$, (3), and

$$\mathcal{L}\left(\sum_{j\in I_i} b'_j y'_j, t\right) \leq C_{4.7}\frac{2L}{T\sqrt{\lceil \lambda n\rceil}}t, \quad \text{for all } t\geq\sqrt{\lceil \lambda n\rceil}.$$

Step 2: Size of nets of level sets. We now estimate the number of possible realizations y'. This is the analogue of Proposition 3.3 in the present context. By paying an overall factor of at most 6^n, we may fix $\mathrm{Spread}\,(v)$ (hence all the $\mathrm{Spread}^j_\lambda\,(v)$), as well as which $\mathrm{Spread}^j_\lambda\,(v)$ are in $\mathcal{I}_M(v)$ and $\mathcal{J}_M(v)$. As above, let us denote the intervals $\mathrm{Spread}^j_\lambda\,(v)$ in $\mathcal{I}_M(v)$ by $I_1,\ldots,I_m$, and let $I_0 = [n]\setminus(I_1\cup\cdots\cup I_m)$.

First, note that the number of choices for y'_{I_0} is at most $(CD/\sqrt{n})^{|I_0|}$ for an absolute constant C – this follows since y'_{I_0} is an integer point in a ball of radius $D\geq\sqrt{n}\geq\sqrt{|I_0|}$ (provided that C_1 is chosen sufficiently large), at which point, we can use a standard volumetric estimate for the number of integer points in $\mathbb{R}^{I_0}$ in a ball of radius $R\geq\sqrt{|I_0|}$, together with the bound $|I_0|\geq n/2$.

Next, we fix $i\in[m]$, and bound the number of choices for y'_{I_i}. Note that for any $r\geq 0$,

$$\mathcal{L}\left(\sum_{j\in I_i} b_j y'_j, r\right) \geq \mathcal{L}\left(\sum_{j\in I_i} (b_j - \tilde{b}_j)y'_j, r\right),$$

where $b_i, \widetilde{b}_i$ are independent copies of Ber (p). Since b'_j is distributed as $b_j - \widetilde{b}_j$, it follows from (3) that

$$\mathcal{L}\left(\sum_{j\in I_i} b_j y'_j, \sqrt{\lceil \lambda n \rceil}\right) \geq c_2 L N^{-1},$$

where b_j are i.i.d. Ber (p) random variables. From the definition of Spread$_\lambda$ (v) and (R1), we see that y'_{I_i} lies within a $(D/(C_2\sqrt{n}), \lceil \lambda n \rceil, K', 1)$-admissible set $\mathcal{A}$ for C_2 and K' sufficiently large depending on c_0, c_1. Then, for L sufficiently large depending on c_0, c_1, p, by Theorem 4.4 (noting that D is bounded by $n2^{c'\lambda n}$ for all sufficiently large n, and that we can take c' to be sufficiently small depending on p), we deduce that the number of potential $y'_{I_i} \in \mathbb{Z}^{I_i}$ is bounded by

$$\exp\,(-M|I_i|)(CD/\sqrt{n})^{|I_i|},$$

where C depends on c_0, c_1, and M grows as $\sqrt{\log\lceil \lambda n \rceil}$, hence as $\sqrt{\log n}$. Explicitly, we can pick $M \geq c_3\sqrt{\log n}$ for some small $c_3 > 0$ depending only on c_0, c_1, p. Multiplying the total number of possibilities for $y'_0, y'_1, \ldots, y'_m$, we see that the total number of possibilities for $y' \in \mathbb{Z}^n$ is at most

$$\exp\,(-c_{2.3}Mn/8)(CD/\sqrt{n})^n,$$

for C depending on c_0, c_1 and $M \geq c_3\sqrt{\log n}$ with c_3 depending on c_0, c_1, p.

Step 3: Small-ball probability for net points. Fix $y' \in \mathbb{Z}^n$ resulting from the randomized rounding process and $u \in \mathbb{R}^n$. Our goal is to bound $\mathbb{P}[\|Ay' - u\|_2 \leq Kn]$. As in the proof of Lemma 3.6, we can without loss of generality permute the coordinates of y' so that $I_1, \ldots, I_m$ are the first m blocks of size $\lceil \lambda n \rceil$ within $[n]$. Then, for all $\lceil \lambda n \rceil \leq j \leq c_{2.5}(\lambda)n$, we have for all $\epsilon \geq 0$ that

$$\mathcal{L}((Ay' - u)_j | (Ay' - u)_{j+1,\ldots,n}, D\epsilon) \leq CL\left(\frac{\epsilon}{\sqrt{j/n}} + \frac{\sqrt{\lambda n/j}}{T}\right),$$

where C is an absolute constant. To deduce this, we use that the first j elements of row j are independent of rows $j+1, \ldots, n$, then use Lemma 4.5 to replace the Rademacher entries of A (plus the small Gaussian perturbation, which has variance so small that it can be disregarded) by Ber (p) − Ber $'(p)$, and finally use Lemma 2.9 (as in the proof of Proposition 3.5) to stitch together the Lévy concentration properties of each y'_{I_i} (guaranteed by (R3) of Lemma 4.7). Combining this with Lemma 2.7, we see that for y', u as above,

$$\mathbb{P}[\|Ay' - u\|_2 \leq Kn] \leq \left(\frac{C''L\sqrt{n}}{D}\right)^{n-\lceil \lambda n \rceil},$$

where C'' depends only on c_0, c_1, p.

Step 4: Union bound. On the event $\|A\| \leq K\sqrt{n}$, $Av = tu$ with $\|tu\|_2 \leq K\sqrt{n}$. By splitting the range $\{tu\}$ into $(4D/\sqrt{n})^2$ intervals, and rounding v as in Step 2, we see that the probability of the event in question is bounded above by

$$\exp\,(-c_{2.3}Mn/8)\left(\frac{CD}{\sqrt{n}}\right)^{n+2} \sup_{y',u} \mathbb{P}[\|Ay' - u\|_2 \leq Kn],$$

where the supremum is over $u \in \mathbb{R}^n$ and $y' \in \mathbb{Z}^n$ such that each y'_{I_i} for $i \in [m]$ satisfies the conclusions of Lemma 4.7 (with n replaced by $\lceil \lambda n \rceil$). Controlling the final factor by Step 3, we see that the probability is bounded above by

$$\exp\,(-c_{2.3}Mn/8)C^n D^{\lceil \lambda n \rceil + 2},$$

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

where C depends on c_0, c_1, p. Finally, since $D \le 2^{2c'\lambda n}$ for all n sufficiently large (depending on c_0, c_1, p), we obtain an overall upper bound of

$$\exp\left(-c_{2.3}Mn/8 + n\log C + 2c'\lambda^2 n^2 + 6c'\lambda n\right).$$

Since $M \ge c_3\sqrt{\log n}$, for c_3 depending on c_0, c_1, p, the desired result follows by choosing $\lambda < c(\log n)^{1/4}n^{-1/2}$ for c sufficiently small depending on c_0, c_1, p, so that the quantity above is bounded by $\exp\left(-\Omega(n(\log n)^{1/2})\right)$. □

5. Proof of Theorem 1.1

In this section (along with Appendix A), we complete the proof of Theorem 1.1 by closely following [30] with appropriate modifications. Since the smallest singular value is a continuous function of the entries of the matrix, by perturbing each entry of the random matrix by a Gaussian variable with arbitrarily small variance, one may assume that ξ is absolutely continuous with respect to the Lebesgue measure; in particular, one may freely assume that various square matrices whose entries are independent copies of ξ are invertible.

5.1 Quadratic small-ball probabilities

To prove Theorem 1.1 and 1.2, we need the following small-ball inequalities for quadratic forms. The derivation is almost identical to the approach in [30, Theorem 8.1], with improvements coming from Theorem 3.7 and Proposition 3.5 (respectively Theorem 4.8). We include details in the appendix for the reader's convenience.

Theorem 5.1 *Let A be an $n \times n$ symmetric random matrix whose independent entries are identical copies of a sub-Gaussian random variable ξ with variance 1. Suppose X is a random vector (independent of A) whose entries are independent copies of ξ. Then, for every $\epsilon \ge 0$ and $u \in \mathbb{R}$, we have*

$$\mathbb{P}\left[\frac{|\langle A^{-1}X, X\rangle - u|}{\sqrt{1 + \|A^{-1}X\|_2^2}} \le \epsilon \wedge \|A\| \le K\sqrt{n}\right] \le C_{5.1}\epsilon^{1/8} + 2\exp\left(-c_{5.1}n^{1/2}\right).$$

We similarly derive the following strengthening for Rademacher entries.

Theorem 5.2 *Let A be an $n \times n$ symmetric random matrix whose independent entries are distributed as the sum of a Rademacher random variable and a centred Gaussian with variance n^{-2n}. Suppose X is a random vector (independent of A) whose entries are independent Rademachers. Then, for all sufficiently large n, and for every $\epsilon \ge 0$ and $u \in \mathbb{R}$, we have*

$$\mathbb{P}\left[\frac{|\langle A^{-1}X, X\rangle - u|}{\sqrt{1 + \|A^{-1}X\|_2^2}} \le \epsilon \wedge \|A\| \le K\sqrt{n}\right] \le C_{5.1}\epsilon^{1/8} + 2\exp\left(-c_{5.1}n^{1/2}(\log n)^{1/4}\right).$$

5.2 Putting it together

Given the above, the proofs of Theorems 1.1 and 1.2 follows from a modification (due to Vershynin) of the invertibility-via-distance paradigm due to Rudelson and Vershynin. We reproduce the details from [30] for the reader's convenience

Proof of Theorems 1.1 and 1.2. Fix $c_0, c_1, c \in (0, 1)$, as guaranteed by Lemma 2.2. We can clearly assume that $\epsilon \le c$. Then, by the union bound and Lemma 2.2, we have

$$\begin{aligned}\mathbb{P}[s_n(A) \le \epsilon/\sqrt{n}] &\le \mathbb{P}[\exists v \in \mathrm{Comp}\,(c_0, c_1) : \|Av\|_2 \le c\sqrt{n}] + \mathbb{P}[\exists v \in \mathrm{Incomp}\,(c_0, c_1) : \|Av\|_2 \le \epsilon/\sqrt{n}] \\ &\le 2\exp(-cn) + \mathbb{P}[\exists v \in \mathrm{Incomp}\,(c_0, c_1) : \|Av\|_2 \le \epsilon/\sqrt{n}].\end{aligned}$$

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

Let $A_1, \ldots, A_n$ denote the rows of A, and note that, by symmetry,

$$Av = \sum_{i=1}^{n} v_i A_i^T.$$

In particular,

$$\|Av\|_2 \geq |v_i| \operatorname{dist}(A_i, H_i),$$

where H_i is the span of the rows A_j for $j \neq i$. Since $|v_i| \geq c_1/2\sqrt{n}$ for all $i \in \operatorname{Spread}(v)$, it follows that if $\|Av\|_2 \leq \epsilon/\sqrt{n}$ for some $v \in \operatorname{Incomp}(c_0, c_1)$, then we must necessarily have

$$\operatorname{dist}(A_i, H_i) \leq \frac{\epsilon\sqrt{2}}{c_1}$$

for at least $c_{2.3}n$ indices $i \in [n]$. Thus, we see that the probability that $s_n(A) \leq \epsilon/\sqrt{n}$ is at most

$$2\exp(-cn) + \frac{1}{c_{2.3}n} \sum_{i=1}^{n} \mathbb{P}\left[\operatorname{dist}(A_i, H_i) \leq \frac{\epsilon\sqrt{2}}{c_1}\right].$$

Therefore, for Theorem 1.1 it suffices to show that

$$\mathbb{P}[\operatorname{dist}(A_1, H_1) \leq \epsilon] \leq C\epsilon^{1/8} + 2\exp(-cn^{1/2}).$$

A direct computation ([30, Proposition 5.1]) shows that

$$\operatorname{dist}(A_1, H_1) = \frac{|\langle (A')^{-1}X, X\rangle - a_{11}|}{\sqrt{1 + \|(A')^{-1}X\|_2^2}},$$

where A' is the bottom right $(n-1) \times (n-1)$ block of A, and X is the first column of A with the top element removed. At this point, we can apply Theorem 5.1 to conclude. If A has Rademacher entries, by continuity we can transfer the singular value estimate to the model where the distribution is perturbed by a centred Gaussian with sufficiently small variance, at which point, an application of Theorem 5.2 allows us to conclude. □

Acknowledgements

We thank Asaf Ferber and Roman Vershynin for comments on the manuscript. We also thank Dhruv Rohatgi for pointing out some typographical errors in a previous version. We are grateful to an anonymous referee for their careful reading of the manuscript and valuable comments. The last two authors were supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. This work was done while the first author was participating in a program at the Simons Institute for the Theory of Computing.

References

[1] Balogh, J., Morris, R. and Samotij, W. (2018) The method of hypergraph containers. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018*, Vol. IV of *Invited Lectures, World Scientific Publishing*, Hackensack, NJ, pp. 3059–3092.

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

[3] Campos, M., Jenssen, M., Michelen, M. and Sahasrabudhe, J. Singularity of random symmetric matrices revisited. arXiv:2011.03013.

[4] Campos, M., Jenssen, M., Michelen, M. and Sahasrabudhe, J. The singularity probability of a random symmetric matrix is exponentially small. arXiv:2105.11384.

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

[5] Campos, M., Mattos, L., Morris, R. and Morrison, N. (2021) On the singularity of random symmetric matrices. *Duke Math. J.* **170**(5) 881–907.
[6] Costello, K. P., Tao, T. and Vu, V. (2006) Random symmetric matrices are almost surely nonsingular. *Duke Math. J.* **135** 395–413.
[7] Erdös, P. (1945) On a lemma of Littlewood and Offord. *Bull. Am. Math. Soc.* **51** 898–902.
[8] Ferber, A. Singularity of random symmetric matrices–simple proof. arXiv:2006.07439.
[9] Ferber, A. and Jain, V. (2019) Singularity of random symmetric matrices–a combinatorial approach to improved bounds. *In Forum of Mathematics, Sigma*, Vol. **7**, Cambridge University Press.
[10] Ferber, A., Jain, V., Luh, K. and Samotij, W. On the counting problem in inverse Littlewood–Offord theory. arXiv:1904.10425.
[11] Jain, V., Sah, A. and Sawhney, M. Singularity of discrete random matrices II. arXiv:2010.06554.
[12] Jain, V., Sah, A. and Sawhney, M. The smallest singular value of dense random regular digraphs. arXiv:2008.04755.
[13] Kahn, J., Komlós, J. and Szemerédi, E. (1995) On the probability that a random ± 1-matrix is singular. *J. Am. Math. Soc.* **8** 223–240.
[14] Komlós, J. (1967) On determinant of (0, 1) matrices. *Studia Science Mathematics Hungarica* 27–21.
[15] Livshyts, G. V. The smallest singular value of heavy-tailed not necessarily iid random matrices via random rounding. arXiv:1811.07038.
[16] Livshyts, G. V., Tikhomirov, K. and Vershynin, R. The smallest singular value of inhomogeneous square random matrices. arXiv:1909.04219.
[17] Lopatto, P. and Luh, K. Tail bounds for gaps between eigenvalues of sparse random matrices. arXiv:1901.05948.
[18] Luh, K. and Vu, V. Sparse random matrices have simple spectrum. arXiv:1802.03662.
[19] Mirošnikov, A. L. and Rogozin, B. A. (1980) Inequalities for concentration functions. *Teor. Veroyatnost. i Primenen.* **25** 178–183.
[20] Nguyen, H., Tao, T. and Vu, V. (2017) Random matrices: tail bounds for gaps between eigenvalues. *Prob. Theory Related Fields* **167** 777–816.
[21] Nguyen, H. H. (2012) Inverse Littlewood-Offord problems and the singularity of random symmetric matrices. *Duke Math. J.* **161** 545–586.
[22] Nguyen, H. H. (2012) On the least singular value of random symmetric matrices. *Electron. J. Probab.* **17**(53) 19.
[23] O'Rourke, S. and Touri, B. (2016) On a conjecture of Godsil concerning controllable random graphs. *SIAM J. Control Opt.* **54** 3347–3378.
[24] Rogozin, B. A. (1961) On the increase of dispersion of sums of independent random variables. *Teor. Verojatnost. i Primenen* **6** 106–108.
[25] Rudelson, M. Recent developments in non-asymptotic theory of random matrices. In *Modern Aspects of Random Matrix Theory, Proceedings of Symposia in Applied Mathematics*, Vol. 72 of *American Mathematical Society*, Providence, RI, 2014, pp. 83–120.
[26] Rudelson, M. and Vershynin, R. (2008) The Littlewood–Offord problem and invertibility of random matrices. *Adv. Math.* **218** 600–633.
[27] Tao, T. and Vu, V. (2006) On random± 1 matrices: singularity and determinant. *Random Struct. Algorithms* **28** 1–23.
[28] Tao, T. and Vu, V. H. (2007) On the singularity probability of random Bernoulli matrices. *J. Am. Math. Soc.* **20** 603–628.
[29] Tikhomirov, K. (2020) Singularity of random Bernoulli matrices. *Ann. Math. (2)* **191** 593–634.
[30] Vershynin, R. (2014) Invertibility of symmetric random matrices, *Random Struct. Algorithms* **44** 135–182.
[31] Wei, F. Investigate invertibility of sparse symmetric matrix. arXiv:1712.04341.

Appendix A: Quadratic small-ball probabilities

The purpose of this appendix is to prove Theorem 5.1 for completeness. We will also briefly note the necessary modifications to deduce Theorem 5.2.

We essentially replicate the argument in [30, Section 8] with the obvious modifications. In brief, our improved anticoncentration estimate Proposition 3.5 will allow us to replace the $\epsilon^{1/8+\eta}$ dependence in [30] with $\epsilon^{1/8}$, and the improved range of arithmetic structure derived in Theorem 3.7 will allow us to achieve an error term of $\exp(-\Omega(\sqrt{n}))$. For the sake of simplicity, we define the event

$$\mathcal{E}_K := \{\|A\| \le K\sqrt{n}\}.$$

Proposition A.1 (Analogue of [30, Proposition 8.2]) *Let A be a symmetric random matrix whose independent entries are identical copies of a sub-Gaussian random variable with mean 0 and variance 1. Let X be a random vector (independent of A) whose coordinates are i.i.d. copies of ξ. There*

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

exist constants $C, c > 0$ depending only on the sub-Gaussian moment of ξ for which the following holds.

For $\lambda \in (C/n, 1/\sqrt{n})$, A satisfies the following with probability at least $1 - 2e^{-cn}$: if $\mathcal{E}_K$ holds, then for every $\epsilon > 0$:

- $\|A^{-1}X\|_2 \geq c$ with probability at least $1 - e^{-cn}$ in the randomness of X.
- $\|A^{-1}X\|_2 \leq \epsilon^{-1/2}\|A^{-1}\|_{\mathrm{HS}}$ with probability at least $1 - \epsilon$ in the randomness of X.
- $\|A^{-1}X\|_2 \geq \epsilon\|A^{-1}\|_{\mathrm{HS}}$ with probability at least $1 - C\epsilon - 2e^{-c\lambda n}$ in the randomness of X.

Proof. The first two parts have the same proof as in [30, Proposition 8.2]. The last part also has essentially the same proof, except that we use Proposition 3.5 in place of [30, Proposition 6.9] and use Theorem 3.7 in place of [30, Theorem 7.1]. □

Remark For Theorem 5.2, we note that if A has entries that are Rademacher plus a centred Gaussian with sufficiently small variance, we can prove the same statement with $n^{-2/3} < \lambda < c(\log n)^{1/4}n^{-1/2}$. We use Theorem 4.8 instead of Theorem 3.7 and the analogue of Proposition 3.5 for the threshold. The remaining part of the proof is exactly the same.

Next, we require the following decoupling lemma from [30]; this use of decoupling to establish singularity for symmetric random matrices originates in work of Costello et al. [6] and has been used in essentially all follow-up works.

Lemma A.2 ([30, Lemma 8.4]) *Let G be an arbitrary symmetric $n \times n$ matrix, and let X, X' be independent samples of a random vector in $\mathbb{R}^n$ with independent coordinates. Let $J \subseteq [n]$. Then, for every $\epsilon \geq 0$ we have*

$$\mathcal{L}(\langle GX, X\rangle, \epsilon)^2 \leq \mathbb{P}_{X,X'}[|\langle GP_{J^c}(X - X'), P_J X\rangle - v| \leq \epsilon]$$

for some random variable v determined by $G|_{J^c \times J^c}$ and $P_{J^c}X, P_{J^c}X'$.

We can now prove Theorem 5.1; we refer the reader to [30] for a more detailed exposition.

Proof of Theorem 5.1. We randomly choose $J \subseteq [n]$ by sampling elements independently with probability $1 - c_{2.3}/2$. We trivially see by the Chernoff bound that if $\mathcal{E}_J = \{|J^c| \leq c_{2.3}n\}$, then

$$\mathbb{P}[\mathcal{E}_J] \geq 1 - 2e^{-cn}.$$

For J satisfying $\mathcal{E}_J$, let us assign the set Spread (v) for $v \in$ Incomp (c_0, c_1) in a way such that Spread $(x) \subseteq |J|$. We can do this since, in Lemma 2.3, we chose $c_{2.3}$ so as to have at least $2c_{2.3}n$ spread coordinates. We will then use this assignment to obtain the median regularized LCDs that are used.

Next, consider the event $\mathcal{E}_D$ given by

$$\epsilon_0^{1/2}\sqrt{1 + \|A^{-1}X\|_2^2} \leq \|A^{-1}\|_{\mathrm{HS}} \leq \frac{1}{\epsilon_0}\|A^{-1}P_{J^c}(X - X')\|_2.$$

Applying Proposition A.1 to X and $Y_i = \delta_i(X_i - X_i')$, where δ_i is the indicator of $i \in J^c$, and adjusting constants appropriately, we find that

$$\mathbb{P}_{J,A,X,X'}[\mathcal{E}_D \vee \mathcal{E}_K^c] \geq 1 - C\epsilon_0 - 2e^{-c\lambda n} - 2e^{-cn},$$

where the constants depend only on the sub-Gaussian norm of ξ. Now define

$$x_0 = \frac{A^{-1}P_{J^c}(X - X')}{\|A^{-1}P_{J^c}(X - X')\|_2},$$

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

which is a random vector. If the denominator is 0, we can use an arbitrary fixed vector. Let $\mathcal{E}_U$ be the event (analogue of [30, Equation 8.11]) that

$$x_0 \in \text{Incomp}\,(c_0, c_1), \qquad \widehat{MD}_L(x_0, \lambda) \geq 2^{\lambda n/C},$$

where we choose L as in Theorem 3.7.

Now condition on some J satisfying $\mathcal{E}_J$ and some X,X'. By Theorem 3.7, we deduce that

$$\mathbb{P}_A[\mathcal{E}_U \vee \mathcal{E}_K^c | X, X', J] \geq 1 - 2e^{-cn}.$$

Thus (analogue of [30, Equation 8.12])

$$\mathbb{P}_{J,A,X,X'}[(\mathcal{E}_J \wedge \mathcal{E}_D \wedge \mathcal{E}_U) \vee \mathcal{E}_K^c] \geq 1 - p_0,$$

where $p_0 = C \max(\epsilon_0, 2^{-\lambda n/C})$. Hence, there is a realization of J such that $\mathcal{E}_J$ holds and

$$\mathbb{P}_{A,X,X'}[(\mathcal{E}_D \wedge \mathcal{E}_U) \vee \mathcal{E}_K^c] \geq 1 - p_0.$$

We fix this choice of J for the remainder of the proof. Now let $\mathcal{E}_A$ be the event, dependent only on A, that simultaneously $\mathcal{E}_K$ and

$$\mathbb{P}_{X,X'}[\mathcal{E}_D \wedge \mathcal{E}_U | A] \geq 1 - p_0^{1/2}.$$

By Fubini's theorem, Markov's inequality, and the fact that $\mathcal{E}_K$ depends only on A, we see from the above that (analogue of [30, Equation 8.13])

$$\mathbb{P}_A[\mathcal{E}_A \vee \mathcal{E}_K^c] \geq 1 - p_0^{1/2}.$$

Now, if $\mathcal{E}$ is the desired event

$$\frac{|\langle A^{-1}X, X\rangle - u|}{\sqrt{1 + \|A^{-1}X\|_2^2}} \leq \epsilon,$$

then

$$\mathbb{P}_{A,X}[\mathcal{E}] \leq \mathbb{P}[\mathcal{E}_K^c] + p_0^{1/2} + \sup_{A \in \mathcal{E}_A} \mathbb{P}_X[\mathcal{E}|A].$$

Fix some $A \in \mathcal{E}_A$ for the remainder of the proof. We need to bound

$$\mathbb{P}_X[\mathcal{E}|A] \leq \mathbb{P}_{X,X'}[\mathcal{E} \wedge \mathcal{E}_D | A] + p_0^{1/2}.$$

Using $\mathcal{E}_D$ along with $\mathcal{E}$, we see that

$$\mathbb{P}_{X,X'}[\mathcal{E} \wedge \mathcal{E}_D | A] \leq \mathbb{P}_{X,X'}[|\langle A^{-1}X, X\rangle - u| \leq \epsilon\epsilon_0^{-1/2}\|A^{-1}\|_{\mathrm{HS}} | A] =: p_1.$$

Then by Lemma A.2, we find that this satisfies

$$p_1^2 \leq \mathbb{P}_{X,X'}[|\langle A^{-1}P_{J^c}(X - X'), P_J X\rangle - v| \leq \epsilon\epsilon_0^{-1/2}\|A^{-1}\|_{\mathrm{HS}} | A],$$

where $v = v(A^{-1}, P_{J^c}X, P_{J^c}X')$ is some random variable depending on these parameters only. This last probability is at most

$$p_0^{1/2} + \mathbb{P}_{X,X'}[|\langle A^{-1}P_{J^c}(X - X'), P_J X\rangle - v| \leq \epsilon\epsilon_0^{-1/2}\|A^{-1}\|_{\mathrm{HS}} \wedge \mathcal{E}_D \wedge \mathcal{E}_U | A].$$

Now by using $\mathcal{E}_D$ again, and dividing the inequality in question by $\|A^{-1}P_{J^c}(X - X')\|_2$, we see that (analogue of [30, Equation 8.15])

$$p_1^2 \leq p_0^{1/2} + \mathbb{P}_{X,X'}[|\langle x_0, P_J X\rangle - w| \leq \epsilon_0^{-3/2}\epsilon \wedge \mathcal{E}_U | A].$$

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press

Finally, we can apply Proposition 3.5 to this random variable. Note that x_0, w do not depend on $P_J X$. Also, we know from $\mathcal{E}_U$ that $\widehat{MD}_L(x_0, \lambda) \geq 2^{\lambda n/C}$. It suffices to check that $\mathrm{Spread}_\lambda(x_0) \subseteq J$ by the definitions chosen at the beginning; hence, we can drop the randomness in $P_J X$ of all coordinates except for those in the spread set and apply the result. We again technically need to check that $\mathrm{Spread}_\lambda(x_0)$ satisfies the conditions on Proposition 3.5, which we have already implicitly verified before. Overall, we deduce

$$\mathbb{P}_{X,X'}[|\langle x_0, P_J X\rangle - w| \leq \epsilon_0^{-3/2}\epsilon \wedge \mathcal{E}_U | A] \leq C\epsilon_0^{-3/2}\epsilon + 2^{-\lambda n/C}.$$

Finally, tracing it all back, we have

$$\mathbb{P}[\mathcal{E}] \leq \mathbb{P}[\mathcal{E}_K^c] + 2p_0^{1/2} + p_1 \leq 2e^{-cn} + C\epsilon_0^{1/2} + C\epsilon_0^{1/4} + C\epsilon_0^{-3/4}\epsilon^{1/2} + 2^{-\lambda n/C}$$

by sub-Gaussian concentration of the operator norm of A ([30, Lemma 2.3]). Now choosing $\epsilon_0 = \epsilon^{1/2}$ and $\lambda = 1/\sqrt{n}$, which is the biggest permitted by Theorem 3.7, we are done. □

Cite this article: Jain V, Sah A and Sawhney M (2022). On the smallest singular value of symmetric random matrices. *Combinatorics, Probability and Computing* **31**, 662–683. https://doi.org/10.1017/S0963548321000511

https://doi.org/10.1017/S0963548321000511 Published online by Cambridge University Press