Forum of Mathematics, Pi (2024), **Vol. 12**:e3 1–69
doi:10.1017/fmp.2023.29

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

The least singular value of a random symmetric matrix

Marcelo Campos[1], Matthew Jenssen[2], Marcus Michelen[3] and Julian Sahasrabudhe[4]

[1]Department of Pure Mathematics and Mathematical Statistics (DPMMS), University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, United Kingdon; E-mail: mc2482@cam.ac.uk.
[2]Department of Mathematics, King's College London, Strand, London, WC2R 2LS, United Kingdom; E-mail: matthew.jenssen@kcl.ac.uk.
[3]Department of Mathematics, Statistics and Computer Science, University of Illinois Chicago, 851 South Morgan Street, Chicago, IL 60607, USA; E-mail: michelen@uic.edu.
[4]Department of Pure Mathematics and Mathematical Statistics (DPMMS), University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, United Kingdon; E-mail: jdrs2@cam.ac.uk.

Received: 4 November 2022; **Revised:** 10 September 2023; **Accepted:** 29 September 2023

2020 Mathematics Subject Classification: *Primary* – 60B20; *Secondary* – 15B52

Abstract
Let A be an $n\times n$ symmetric matrix with $(A_{i,j})_{i\leqslant j}$ independent and identically distributed according to a subgaussian distribution. We show that

$$\mathbb{P}(\sigma_{\min}(A)\leqslant \varepsilon n^{-1/2})\leqslant C\varepsilon+e^{-cn},$$

where $\sigma_{\min}(A)$ denotes the least singular value of A and the constants $C,c>0$ depend only on the distribution of the entries of A. This result confirms the folklore conjecture on the lower tail of the least singular value of such matrices and is best possible up to the dependence of the constants on the distribution of $A_{i,j}$. Along the way, we prove that the probability that A has a repeated eigenvalue is $e^{-\Omega(n)}$, thus confirming a conjecture of Nguyen, Tao and Vu [*Probab. Theory Relat. Fields* 167 (2017), 777–816].

Contents

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

1. Introduction

Let A be an $n \times n$ random symmetric matrix whose entries on and above the diagonal $(A_{i,j})_{i\leqslant j}$ are independent and identically distributed (i.i.d.) with mean 0 and variance 1. This matrix model, sometimes called the Wigner matrix ensemble, was introduced in the 1950s in the seminal work of Wigner [50], who established the famous "semicircular law" for the eigenvalues of such matrices.

In this paper, we study the extreme behavior of the *least singular value* of A, which we denote by $\sigma_{\min}(A)$. Heuristically, we expect that $\sigma_{\min}(A) = \Theta(n^{-1/2})$, and thus it is natural to consider

$$\mathbb{P}(\sigma_{\min}(A) \leqslant \varepsilon n^{-1/2}), \tag{1.1}$$

for all $\varepsilon \geqslant 0$ (see Section 1.2). In this paper, we prove a bound on this quantity which is optimal up to constants, for all random symmetric matrices with i.i.d. *subgaussian* entries. This confirms the folklore conjecture, explicitly stated by Vershynin in [46].

Theorem 1.1. *Let ζ be a subgaussian random variable with mean 0 and variance 1, and let A be an $n\times n$ random symmetric matrix whose entries above the diagonal $(A_{i,j})_{i\leqslant j}$ are independent and distributed according to ζ. Then for every $\varepsilon \geqslant 0$,*

$$\mathbb{P}_A(\sigma_{\min}(A) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon + e^{-cn}, \tag{1.2}$$

where $C, c > 0$ depend only on ζ.

This conjecture is sharp up to the value of the constants $C, c > 0$ and resolves the "up-to-constants" analogue of the Spielman–Teng [38] conjecture for random symmetric matrices (see Section 1.2). Also note that the special case $\varepsilon = 0$ tells us that the singularity probability of any random symmetric A with subgaussian entry distribution is exponentially small, generalizing our previous work [4] on the $\{-1, 1\}$ case.

1.1. Repeated eigenvalues

Before we discuss the history of the least singular value problem, we highlight one further contribution of this paper: a proof that a random symmetric matrix has no repeated eigenvalues with probability $1 - e^{-\Omega(n)}$.

In the 1980s, Babai [43] conjectured that the adjacency matrix of the binomial random graph $G(n, 1/2)$ has no repeated eigenvalues with probability $1 - o(1)$ (see [43]). Tao and Vu [43] proved this conjecture in 2014 and, in subsequent work on the topic with Nguyen [24], went on to conjecture the probability that a random symmetric matrix with i.i.d. subgaussian entries has no repeated eigenvalues is $1 - e^{-\Omega(n)}$. In this paper, we prove this conjecture en route to proving Theorem 1.1, our main theorem.

Theorem 1.2. *Let ζ be a subgaussian random variable with mean 0 and variance 1, and let A be an $n \times n$ random symmetric matrix, where $(A_{i,j})_{i\leqslant j}$ are independent and distributed according to ζ. Then A has no repeated eigenvalues with probability at least $1 - e^{-cn}$, where $c > 0$ is a constant depending only on ζ.*

Theorem 1.2 is easily seen to be sharp whenever $A_{i,j}$ is discrete: consider the event that three rows of A are identical; this event has probability $e^{-\Theta(n)}$ and results in two 0 eigenvalues. Also note that the constant in Theorem 1.2 can be made arbitrary small; consider the entry distribution ζ, which takes value 0 with probability $1 - p$ and each of $\{-p^{-1/2}, p^{-1/2}\}$ with probability $p/2$. Here, the probability of 0 being a repeated root is $\geqslant e^{-(3+o(1))pn}$.

We in fact prove a more refined version of Theorem 1.2, which gives an upper bound on the probability that two eigenvalues of A fall into an interval of length ε. This is the main result of Section 7. For this, we let $\lambda_1(A) \geqslant \ldots \geqslant \lambda_n(A)$ denote the eigenvalues of the $n \times n$ real symmetric matrix A.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Theorem 1.3. *Let ζ be a subgaussian random variable with mean* 0 *and variance* 1*, and let A be an $n \times n$ random symmetric matrix, where $(A_{i,j})_{i \leqslant j}$ are independent and distributed according to ζ. Then for each $\ell < cn$ and all $\varepsilon \geqslant 0$, we have*

$$\max_{k \leqslant n-\ell} \mathbb{P}\big(|\lambda_{k+\ell}(A) - \lambda_k(A)| \leqslant \varepsilon n^{-1/2}\big) \leqslant (C\varepsilon)^{\ell} + 2e^{-cn}\,,$$

where $C, c > 0$ are constants, depending only on ζ.

In the following subsection, we describe the history of the least singular value problem. In Section 1.3, we discuss a technical theme which is developed in this paper, and then, in Section 2, we go on to give a sketch of Theorem 1.1.

1.2. History of the least singular value problem

The behavior of the least singular value was first studied for random matrices B_n with *i.i.d.* coefficients, rather than for *symmetric* random matrices. For this model, the history goes back to von Neumann [48], who suggested that one typically has

$$\sigma_{\min}(B_n) \approx n^{-1/2},$$

while studying approximate solutions to linear systems. This was then more rigorously conjectured by Smale [36] and proved by Szarek [39] and Edelman [8] in the case that $B_n = G_n$ is a random matrix with i.i.d. *standard gaussian* entries. Edelman found an exact expression for the density of the least singular value in this case. By analyzing this expression, one can deduce that

$$\mathbb{P}(\sigma_{\min}(G_n) \leqslant \varepsilon n^{-1/2}) \leqslant \varepsilon, \tag{1.3}$$

for all $\varepsilon \geqslant 0$ (see, e.g. [38]). While this gives a very satisfying understanding of the gaussian case, one encounters serious difficulties when trying to extend this result to other distributions. Indeed, Edelman's proof relies crucially on an exact description of the joint distribution of eigenvalues that is available in the gaussian setting. In the last 20 or so years, intense study of the least singular value of i.i.d. random matrices has been undertaken with the overall goal of proving an appropriate version of (1.3) for different entry distributions and models of random matrices.

An important and challenging feature of the more general problem arises in the case of *discrete* distributions, where the matrix B_n can become singular with nonzero probability. This singularity event will affect the quantity (1.1) for very small ε and thus estimating the probability that $\sigma_{\min}(B_n) = 0$ is a crucial aspect of generalizing (1.3). This is reflected in the famous and influential Spielman–Teng conjecture [37] which proposes the bound

$$\mathbb{P}(\sigma_{\min}(B_n) \leqslant \varepsilon n^{-1/2}) \leqslant \varepsilon + 2e^{-cn}, \tag{1.4}$$

where B_n is a Bernoulli random matrix. Here, this added exponential term "comes from" the singularity probability of B_n. In this direction, a key breakthrough was made by Rudelson [30], who proved that if B_n has i.i.d. *subgaussian* entries, then

$$\mathbb{P}(\sigma_{\min}(B_n) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon n + n^{-1/2}\,.$$

This result was extended in a series of works [32, 40, 44, 49], culminating in the influential work of Rudelson and Vershynin [31], who showed the "up-to-constants" version of Spielman-Teng:

$$\mathbb{P}(\sigma_{\min}(B_n) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon + e^{-cn}, \tag{1.5}$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

where B_n is a matrix with i.i.d. entries that follow any subgaussian distribution and $C, c > 0$ depend only on ζ. A key ingredient in the proof of (1.5) is a novel approach to the "inverse Littlewood-Offord problem," a perspective pioneered by Tao and Vu [44] (see Section 1.3 for more discussion).

Another very different approach was taken by Tao and Vu [41], who showed that the distribution of the least singular value of B_n is identical to the least singular value of the Gaussian matrix G_n, up to scales of size n^{-c}. In particular, they prove that

$$\left|\mathbb{P}(\sigma_{\min}(B_n) \leqslant \varepsilon n^{-1/2}) - \mathbb{P}(\sigma_{\min}(G_n) \leqslant \varepsilon n^{-1/2})\right| = O(n^{-c_0}), \tag{1.6}$$

thus resolving the Spielman-Teng conjecture for $\varepsilon \geqslant n^{-c_0}$, in a rather strong form. While falling just short of the Spielman-Teng conjecture, the work of Tao and Vu [41], Rudelson and Vershynin [31], and subsequent refinements by Tikhomirov [45] and Livshyts et al. [22] (see also [21, 29]) leave us with a very strong understanding of the least singular value for *i.i.d.* matrix models. However, progress on the analogous problem for random symmetric matrices, or *Wigner random matrices*, has come somewhat more slowly and more recently: In the symmetric case, even proving that A_n is nonsingular with probability $1 - o(1)$ was not resolved until the important 2006 paper of Costello et al. [7].

Progress on the symmetric version of Spielman–Teng continued with Nguyen [25, 26] and, independently, Vershynin [46]. Nguyen proved that for any $B > 0$, there exists an $A > 0$ for which[1]

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant n^{-A}) \leqslant n^{-B}.$$

Vershynin [46] proved that if A_n is a matrix with subgaussian entries then, for all $\varepsilon > 0$, we have

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \leqslant C_\eta \varepsilon^{1/8-\eta} + 2e^{-n^c}, \tag{1.7}$$

for all $\eta > 0$, where the constants $C_\eta, c > 0$ may depend on the underlying subgaussian random variable. He went on to conjecture that ε should replace $\varepsilon^{1/8-\varepsilon}$ as the correct order of magnitude, and that e^{-cn} should replace e^{-n^c}.

After Vershynin, a series of works [3, 5, 16, 17, 19] made progress on singularity probability (i.e., the $\varepsilon = 0$ case of Vershynin's conjecture), and we, in [4], ultimately showed that the singularity probability is exponentially small, when $A_{i,j}$ is uniform in $\{-1, 1\}$:

$$\mathbb{P}(\det(A_n) = 0) \leqslant e^{-cn},$$

which is sharp up to the value of $c > 0$.

However, for general ε, the state of the art is due to Jain et al. [19], who improved on Vershynin's bound (1.7) by showing

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \leqslant C\varepsilon^{1/8} + e^{-\Omega(n^{1/2})},$$

under the subgaussian hypothesis on A_n.

For large ε, for example, $\varepsilon \geqslant n^{-c}$, another very different and powerful set of techniques have been developed, which in fact apply more generally to the distribution of other "bulk" eigenvalues and additionally give distributional information on the eigenvalues. The works of Tao and Vu [40, 42], Erdős, Schlein and Yau [10, 11, 13], Erdős et al. [9], and specifically, Bourgade et al. [2] tell us that

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \leqslant \varepsilon + o(1), \tag{1.8}$$

[1]Nguyen in [26] actually proves the same result for random matrices of the form $A_n + F$, where F is a fixed symmetric $n \times n$ matrix satisfying $\|F\|_{op} \leqslant n^{O(1)}$.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

thus obtaining the correct dependence[2] on ε when n is sufficiently large compared to ε. These results are similar in flavor to (1.6) in that they show the distribution of various eigenvalue statistics is closely approximated by the corresponding statistics in the gaussian case. We note, however, that it appears these techniques are limited to these large ε and different ideas are required for $\varepsilon < n^{-C}$, and certainly for ε as small as $e^{-\Theta(n)}$.

Our main theorem, Theorem 1.1, proves Vershynin's conjecture and thus proves the optimal dependence on ε for all $\varepsilon > e^{-cn}$, up to constants.

1.3. Approximate negative correlation

Before we sketch the proof of Theorem 1.1, we highlight a technical theme of this paper: the approximate negative correlation of certain "linear events." While this is only one of several new ingredients in this paper, we isolate these ideas here, as they seem to be particularly amenable to wider application. We refer the reader to Section 2 for a more complete overview of the new ideas in this paper.

We say that two events A, B in a probability space are *negatively correlated* if

$$\mathbb{P}(A \cap B) \leqslant \mathbb{P}(A)\mathbb{P}(B).$$

Here, we state and discuss two *approximate* negative correlation results: one of which is from our paper [4], but is used in an entirely different context, and one of which is new.

We start by describing the latter result, which says that a "small ball" event is approximately negatively correlated with a large deviation event. This complements our result from [4], which says that two "small ball events," of different types, are negatively correlated. In particular, we prove something in the spirit of the following inequality, though in a slightly more technical form.

$$\mathbb{P}_X\big(|\langle X, v\rangle| \leqslant \varepsilon \text{ and } \langle X, u\rangle > t\big) \lesssim \mathbb{P}_X(|\langle X, v\rangle| \leqslant \varepsilon)\mathbb{P}_X(\langle X, u\rangle > t), \tag{1.9}$$

where u, v are unit vectors and $t, \varepsilon > 0$ and $X = (X_1, \ldots, X_n)$ with i.i.d. subgaussian random variables with mean 0 and variance 1.

To state and understand our result, it makes sense to first consider, in isolation, the two events present in (1.9). The easier of the two events is $\langle X, u\rangle > t$, which is a large deviation event for which we may apply the essentially sharp and classical inequality (see Chapter 3.4 in [47])

$$\mathbb{P}_X(\langle X, u\rangle > t) \leqslant e^{-ct^2},$$

where $c > 0$ is a constant depending only on the distribution of X.

We now turn to understand the more complicated small-ball event $|\langle X, v\rangle| \leqslant \varepsilon$ appearing in (1.9). Here, we have a more subtle interaction between v and the distribution of X, and thus we first consider the simplest possible case: when X has i.i.d. standard *gaussian* entries. Here, one may calculate

$$\mathbb{P}_X(|\langle X, v\rangle| \leqslant \varepsilon) \leqslant C\varepsilon, \tag{1.10}$$

for all $\varepsilon > 0$, where $C > 0$ is an absolute constant. However, as we depart from the case when X is gaussian, a much richer behavior emerges when the vector v admits some "arithmetic structure." For example, if $v = n^{-1/2}(1, \ldots, 1)$ and the X_i are uniform in $\{-1, 1\}$, then

$$\mathbb{P}_X(|\langle X, v\rangle| \leqslant \varepsilon) = \Theta(n^{-1/2}),$$

[2] Tao and Vu, with Corollary 24 in [42], prove that the distribution of $\sigma_{\min}$ remains asymptotically invariant if the distribution of the entries $A_{i,j}$ is replaced by a distribution that matches four moments with the original distribution. A follow-up work [9] joint with Erdős et al. describes an approach to combine ideas from the works [10, 11, 13] to remove the moment matching assumptions of [42], but does not explicitly address the problem of the least singular value. The work [2] builds on these works to prove the sharp, nonquantitative statement at (1.8). See the discussion below Theorem 2.2 of [2] for more detail.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

for any $0 < \varepsilon < n^{-1/2}$. This, of course, stands in contrast to (1.10) for all $\varepsilon \ll n^{-1/2}$ and suggests that we employ an appropriate measure of the arithmetic structure of v.

For this, we use the notion of the "least common denominator" of a vector, introduced by Rudelson and Vershynin [31]. For parameters $\alpha, \gamma \in (0,1)$ define the *least common denominator* (LCD) of $v \in \mathbb{R}^n$ to be

$$D_{\alpha,\gamma}(v) := \inf\left\{\phi > 0 : \ \|\phi v\|_{\mathbb{T}} \leqslant \min\left\{\gamma\phi\|v\|_2, \sqrt{\alpha n}\right\}\right\}, \tag{1.11}$$

where $\|v\|_{\mathbb{T}} := \mathrm{dist}(v, \mathbb{Z}^n)$, for all $v \in \mathbb{R}^n$. What makes this definition useful is the important "inverse Littlewood-Offord theorem" of Rudelson and Vershynin [31], which tells us (roughly speaking) that one has (1.10) whenever $D_{\alpha,\gamma}(v) = \Omega(\varepsilon^{-1})$. This notion of least common denominator is inspired by Tao and Vu's introduction and development of "inverse Littlewood-Offord theory," which is a collection of results guided by the meta-hypothesis: "If $\mathbb{P}_X(\langle X, v\rangle = 0)$ is large *then* v must have structure." We refer the reader to the paper of Tao and Vu [44] and the survey of Nguyen and Vu [28] for more background and history on inverse Littlewood-Offord theory and its role in random matrix theory. We may now state our version of (1.9), which uses $D_{\alpha,\gamma}(v)^{-1}$ as a proxy for $\mathbb{P}(|\langle X, v\rangle| \leqslant \varepsilon)$.

Theorem 1.4. *For $n \in \mathbb{N}$, $\varepsilon, t > 0$ and $\alpha, \gamma \in (0,1)$, let $v \in \mathbb{S}^{n-1}$ satisfy $D_{\alpha,\gamma}(v) > C/\varepsilon$ and let $u \in \mathbb{S}^{n-1}$. Let ζ be a subgaussian random variable, and let $X \in \mathbb{R}^n$ be a random vector whose coordinates are i.i.d. copies of ζ. Then*

$$\mathbb{P}_X\left(|\langle X, v\rangle| \leqslant \varepsilon \text{ and } \langle X, u\rangle > t\right) \leqslant C\varepsilon e^{-ct^2} + e^{-c(\alpha n + t^2)},$$

where $C, c > 0$ depend only on γ and the distribution of ζ.

In fact, we need a significantly more complicated version of this result (Lemma 5.2), where the small-ball event $|\langle X, v\rangle| \leqslant \varepsilon$ is replaced with a small-ball event of the form

$$|f(X_1, \ldots, X_n)| \leqslant \varepsilon,$$

where f is a quadratic polynomial in variables $X_1, \ldots, X_n$. The proof of this result is carried out in Section 5 and is an important aspect of this paper. Theorem 1.4 is stated here to illustrate the general flavor of this result, and is not actually used in this paper. We do provide a proof in Appendix 9 for completeness and to suggest further inquiry into inequalities of the form (1.9).

We now turn to discuss our second approximate negative dependence result, which deals with the intersection of two different small ball events. This was originally proved in our paper [4], but is put to a different use here. This result tells us that the events

$$|\langle X, v\rangle| \leqslant \varepsilon \qquad \text{and} \qquad |\langle X, w_1\rangle| \ll 1, \ldots, |\langle X, w_k\rangle| \ll 1 \tag{1.12}$$

are approximately negatively correlated, where $X = (X_1, \ldots, X_n)$ is a vector with i.i.d. subgaussian entries and $w_1, \ldots, w_k$ are orthonormal. That is, we prove something in the spirit of

$$\mathbb{P}_X\left(\{|\langle X, v\rangle| \leqslant \varepsilon\} \cap \bigcap_{i=1}^{k}\{|\langle X, w_i\rangle| \ll 1\}\right) \lesssim \mathbb{P}_X\left(|\langle X, v\rangle| \leqslant \varepsilon\right)\mathbb{P}_X\left(\bigcap_{i=1}^{k}\{|\langle X, w_i\rangle| \ll 1\}\right),$$

though in a more technical form.

To understand our result, again, it makes sense to consider the two events in (1.12) in isolation. Since we have already discussed the subtle event $|\langle X, v\rangle| \leqslant \varepsilon$, we consider the event on the right of (1.12). Returning to the gaussian case, we note that if X has independent standard gaussian entries, then one may compute directly that

$$\mathbb{P}_X\left(|\langle X, w_1\rangle| \ll 1, \ldots, |\langle X, w_k\rangle| \ll 1\right) = \mathbb{P}(|X_1| \ll 1, \ldots |X_k| \ll 1) \leqslant e^{-\Omega(k)}, \tag{1.13}$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

by rotational invariance of the gaussian. Here, the generalization to other random variables is not as subtle, and the well-known Hanson-Wright [18] inequality tells us that (1.13) holds more generally when X has general i.i.d. subgaussian entries.

Our innovation in this line is our second "approximate negative correlation theorem," which allows us to control these two events *simultaneously*. Again, we use $D_{\alpha,\gamma}(v)^{-1}$ as a proxy for $\mathbb{P}(|\langle X, v\rangle| \leqslant \varepsilon)$.

Here, for ease of exposition, we state a less general version for $X = (X_1, \ldots, X_n) \in \{-1, 0, 1\}$ with i.i.d. c-lazy coordinates, meaning that $\mathbb{P}(X_i = 0) \geqslant 1 - c$. Our theorem is stated in full generality in Section 9 (see Theorem 9.2).

Theorem 1.5. *Let $\gamma \in (0, 1)$, $d \in \mathbb{N}$, $\alpha \in (0, 1)$, $0 \leqslant k \leqslant c_1\alpha d$, and $\varepsilon \geqslant \exp(-c_1\alpha d)$. Let $v \in \mathbb{S}^{d-1}$, let $w_1, \ldots, w_k \in \mathbb{S}^{d-1}$ be orthogonal, and let W be the matrix with rows $w_1, \ldots, w_k$.*

If $X \in \{-1, 0, 1\}^d$ is a $1/4$-lazy random vector and $D_{\alpha,\gamma}(v) > 16/\varepsilon$, then

$$\mathbb{P}_X\left(|\langle X, v\rangle| \leqslant \varepsilon \text{ and } \|WX\|_2 \leqslant c_2\sqrt{k}\right) \leqslant C\varepsilon e^{-c_1 k},$$

where $C, c_1, c_2 > 0$ are constants, depending only on γ.

In this paper, we will put Theorem 1.5 to a very different use than to that in [4], where we used it to prove a version of the following statement.

Let $v \in \mathbb{S}^{d-1}$ be a vector on the sphere, and let H be an $n \times d$ random $\{-1, 0, 1\}$-matrix *conditioned on the event* $\|Hv\|_2 \leqslant \varepsilon n^{1/2}$, for some $\varepsilon > e^{-cn}$. Here, $d = cn$ and $c > 0$ is a sufficiently small constant. Then the probability that the rank of H is $n - k$ is $\leqslant e^{-ckn}$.

In this paper, we use (the generalization of) Theorem 1.5 to obtain good bounds on quantities of the form

$$\mathbb{P}_X(\|BX\|_2 \leqslant \varepsilon n^{1/2}),$$

where B is a fixed matrix with an exceptionally large eigenvalue (possibly as large as e^{cn}), but is otherwise pseudo-random, meaning (among other things) that the rest of the spectrum does not deviate too much from that of a random matrix. We use Theorem 1.5 to decouple the interaction of X with the largest eigenvector of B, from the interaction of X with the rest of B. We refer the reader to (2.10) in the sketch in Section 2 and to Section 9 for more details.

The proof of Theorem 9.2 follows closely along the lines of the proof of Theorem 1.5 from [4], requiring only technical modifications and adjustments. So as not to distract from the new ideas of this paper, we have sidelined this proof to the Appendix.

Finally, we note that it may be interesting to investigate these approximate negative correlation results in their own right, and investigate to what extent they can be sharpened.

2. Proof sketch

Here, we sketch the proof of Theorem 1.1. We begin by giving the rough "shape" of the proof, while making a few simplifying assumptions, (2.2) and (2.3). We shall then come to discuss the substantial new ideas of this paper in Section 2.2, where we describe the considerable lengths we must go to in order to remove our simplifying assumptions. Indeed, if one were to only tackle these assumptions using standard tools, one cannot hope for a bound much better than $\varepsilon^{1/3}$ in Theorem 1.1 (see Section 2.2.2).

2.1. The shape of the proof

Recall that A_{n+1} is a $(n+1) \times (n+1)$ random symmetric matrix with subgaussian entries. Let $X := X_1, \ldots, X_{n+1}$ be the columns of A_{n+1}, let

$$V = \mathrm{Span}\{X_2, \ldots, X_{n+1}\},$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

and let A_n be the matrix A_{n+1} with the first row and column removed. We now use an important observation from Rudelson and Vershynin [31] that allows for a geometric perspective on the least singular value problem[3]

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \lesssim \mathbb{P}(\operatorname{dist}(X, V) \leqslant \varepsilon).$$

Here, our first significant challenge presents itself: X and V are not independent, and thus the event $\operatorname{dist}(X, V) \leqslant \varepsilon$ is hard to understand directly. However, one can establish a formula for $\operatorname{dist}(X, V)$ that is a rational function in the vector X with coefficients that depend only on V. This brings us to the useful inequality[4] due to Vershynin [46],

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \lesssim \sup_{r\in\mathbb{R}} \mathbb{P}_{A_n,X}\left(|\langle A_n^{-1}X, X\rangle - r| \leqslant \varepsilon\|A_n^{-1}X\|_2\right), \tag{2.1}$$

where we are ignoring the possibility of A_n being singular for now. We thus arrive at our main technical focus of this paper, bounding the quantity on the right-hand side of (2.1).

We now make our two simplifying assumptions that shall allow us to give the overall shape of our proof without any added complexity. We shall then layer-on further complexities as we discuss how to remove these assumptions.

As a first simplifying assumption, let us assume that the collection of X that dominates the probability at (2.1) satisfies

$$\|A_n^{-1}X\|_2 \approx \|A_n^{-1}\|_{\mathrm{HS}}. \tag{2.2}$$

This is not, at first blush, an unreasonable assumption to make as $\mathbb{E}_X\, \|A_n^{-1}X\|_2^2 = \|A_n^{-1}\|_{\mathrm{HS}}^2$. Indeed, the Hanson-Wright inequality tells us that $\|A_n^{-1}X\|_2$ is concentrated about its mean, for all reasonable A_n^{-1}. However, as we will see, this concentration is not strong enough for us here.

As a second assumption, we assume that the relevant matrices A_n in the right-hand side of (2.1) satisfy

$$\|A_n^{-1}\|_{\mathrm{HS}} \approx cn^{1/2}. \tag{2.3}$$

This turns out to be a *very* delicate assumption, as we will soon see, but is not entirely unreasonable to make for the moment: for example, we have $\|A_n^{-1}\|_{\mathrm{HS}} = \Theta_\delta(n^{1/2})$ with probability $1-\delta$. This, for example, follows from Vershynin's theorem [46] along with Corollary 8.4, which is based on the work of [13].

With these assumptions, we return to (2.1) and obverse our task has reduced to proving

$$\min_r \mathbb{P}_X\left(|\langle A^{-1}X, X\rangle - r| \leqslant \varepsilon n^{1/2}\right) \lesssim \varepsilon, \tag{2.4}$$

for all $\varepsilon > e^{-cn}$, where we have written $A^{-1} = A_n^{-1}$ and think of A^{-1} as a fixed (pseudo-random) matrix.

We observe, for a general fixed matrix A^{-1}, there is no hope in proving such an inequality: Indeed, if $A^{-1} = n^{-1/2}J$, where J is the all-ones matrix, then the left-hand side of (2.4) is $\geqslant cn^{-1/2}$ for *all* $\varepsilon > 0$, falling vastly short of our desired (2.4).

Thus, we need to introduce a collection of fairly strong "quasi-randomness properties" of A that hold with, probably $1-e^{-cn}$. These will ensure that A^{-1} is sufficiently "non-structured" to make our goal (2.4) possible. The most important and difficult of these quasi-randomness conditions is to show that the eigenvectors v of A satisfy

$$D_{\alpha,\gamma}(v) > e^{cn},$$

[3]Here and throughout, we understand $A \lesssim B$ to mean that there exists an absolute constant $C > 0$ for which $A \leqslant CB$.

[4]In this sketch, we will be ignoring a few exponentially rare events, and so the inequalities listed here should be understood as "up to an additive error of e^{-cn}."

for some appropriate α, γ, where $D_{\alpha,\gamma}(v)$ is the *least common denominator* of v defined at (1.11). Roughly, this means that none of the eigenvectors of A "correlate" with a rescaled copy of the integer lattice $t\mathbb{Z}^n$, for any $e^{-cn} \leqslant t \leqslant 1$.

To prove that these quasi-randomness properties hold with probability $1 - e^{-cn}$ is a difficult task and depends fundamentally on the ideas in our previous paper [4]. Since we don't want these ideas to distract from the new ideas in this paper, we have opted to carry out the details in the Appendix. With these quasi-randomness conditions in tow, we can return to (2.4) and apply Esseen's inequality to bound the left-hand side of (2.4) in terms of the characteristic function $\varphi(\theta)$ of the random variable $\langle A^{-1}X, X\rangle$,

$$\min_r \mathbb{P}_X\left(|\langle A^{-1}X, X\rangle - r| \leqslant \varepsilon n^{1/2}\right) \lesssim \varepsilon \int_{-1/\varepsilon}^{1/\varepsilon} |\varphi(\theta)|\, d\theta.$$

While this maneuver has been quite successful in work on characteristic functions for (linear) sums of independent random variables, the characteristic function of such quadratic functions has proved to be a more elusive object. For example, even the analogue of the Littlewood-Offord theorem is not fully understood in the quadratic case [6, 23]. Here, we appeal to our quasi-random conditions to avoid some of the traditional difficulties: we use an application of Jensen's inequality to *decouple* the quadratic form and bound $\varphi(\theta)$ pointwise in terms of an average over a related collection of characteristic functions of *linear* sums of independent random variables

$$|\varphi(\theta)|^2 \leqslant \mathbb{E}_Y |\varphi(A^{-1}Y; \theta)|,$$

where Y is a random vector with i.i.d. entries and $\varphi(v;\theta)$ denotes the characteristic function of the sum $\sum_i v_i X_i$, where X_i are i.i.d. distributed according to the original distribution ζ. We can then use our pseudo-random conditions on A to bound

$$|\varphi(A^{-1}Y; \theta)| \lesssim \exp\left(-c\theta^2\right),$$

for all but exponentially few Y, allowing us to show

$$\int_{-1/\varepsilon}^{1/\varepsilon} |\varphi(\theta)|\, d\theta \leqslant \int_{-1/\varepsilon}^{1/\varepsilon} \left[\mathbb{E}_Y |\varphi(A^{-1}Y; \theta)|\right]^{1/2} \leqslant \int_{-1/\varepsilon}^{1/\varepsilon} \left(\exp\left(-c\theta^2\right) + e^{-cn}\right) d\theta = O(1)$$

and thus completing the proof, up to our simplifying assumptions.

2.2. *Removing the simplifying assumptions*

While this is a good story to work with, the challenge starts when we turn to remove our simplifying assumptions (2.2), (2.3). We also note that if one only applies standard methods to remove these assumptions, then one would get stuck at the "base case" outlined below. We start by discussing how to remove the simplifying assumption (2.3), whose resolution governs the overall structure of the paper.

2.2.1. Removing the assumption (2.3)

What is most concerning about making the assumption $\|A_n^{-1}\|_{\mathrm{HS}} \approx n^{-1/2}$ is that it is, in a sense, *circular*: If we assume the modest-looking hypothesis $\mathbb{E}\,\|A^{-1}\|_{\mathrm{HS}} \lesssim n^{1/2}$, we would be able to deduce

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) = \mathbb{P}(\sigma_{\max}(A_n^{-1}) \geqslant n^{1/2}/\varepsilon) \leqslant \mathbb{P}(\|A_n^{-1}\|_{\mathrm{HS}} \geqslant n^{1/2}/\varepsilon) \lesssim \varepsilon,$$

by Markov. In other words, showing that $\|A^{-1}\|_{\mathrm{HS}}$ is concentrated about $n^{-1/2}$ (in the above sense) actually *implies* Theorem 1.1. However, this is not as worrisome as it appears at first. Indeed, if we are trying to prove Theorem 1.1 for $(n+1) \times (n+1)$ matrices using the above outline, we only need to control the Hilbert-Schmidt norm of the inverse of the *minor* A_n^{-1}. This suggests an inductive or (as we use) an *iterative* "bootstrapping argument" to successively improve the bound. Thus, in effect, we look to prove

$$\mathbb{E}\,\|A_n^{-1}\|_{\mathrm{HS}}^{\alpha}\mathbf{1}(\sigma_{\min}(A_n)\geqslant e^{-cn})\lesssim n^{\alpha/2},$$

for successively larger $\alpha\in(0,1]$. Note, we have to cut out the event of A being singular from our expectation, as this event has nonzero probability.

2.2.2. Base case

In the first step of our iteration, we prove a "base case" of

$$\mathbb{P}(\sigma_{\min}(A_n)\leqslant\varepsilon n^{-1/2})\lesssim\varepsilon^{1/4}+e^{-cn} \tag{2.5}$$

without the assumption (2.3) which is equivalent to

$$\mathbb{E}\,\|A_n^{-1}\|_{\mathrm{HS}}^{1/4}\mathbf{1}(\sigma_{\min}(A_n)\geqslant e^{-cn})\lesssim n^{1/8}.$$

To prove this "base case," we upgrade (2.1) to

$$\mathbb{P}\left(\sigma_{\min}(A_{n+1})\leqslant\varepsilon n^{-1/2}\right)\lesssim\varepsilon+\sup_{r\in\mathbb{R}}\mathbb{P}\left(\frac{|\langle A_n^{-1}X,X\rangle-r|}{\|A_n^{-1}X\|_2}\leqslant C\varepsilon,\|A_n^{-1}\|_{\mathrm{HS}}\leqslant\frac{n^{1/2}}{\varepsilon}\right). \tag{2.6}$$

In other words, we can intersect with the event

$$\|A_n^{-1}\|_{\mathrm{HS}}\leqslant n^{1/2}/\varepsilon \tag{2.7}$$

at a loss of only $C\varepsilon$ in probability.

We then push through the proof outlined in Section 2.1 to obtain our initial weak bound of (2.5). For this, we first use the Hanson-Wright inequality to give a weak version of (2.2), and then use (2.7) as a weak version of our assumption (2.3). We note that this base step (2.5) already improves the best known bounds on the least singular value problem for random symmetric matrices.

2.2.3. Bootstrapping

To improve on this bound, we use a "bootstrapping" lemma which, after applying it three times, allows us to improve (2.5) to the near-optimal result

$$\mathbb{P}(\sigma_{\min}(A_n)\leqslant\varepsilon n^{-1/2})\lesssim\varepsilon\sqrt{\log 1/\varepsilon}+e^{-cn}\,. \tag{2.8}$$

Proving this bootstrapping lemma essentially reduces to the problem of getting good estimates on

$$\mathbb{P}_X\left(\|A^{-1}X\|_2\leqslant s\right)\qquad\text{for}\qquad s\in(\varepsilon,n^{-1/2}), \tag{2.9}$$

where A is a matrix with $\|A^{-1}\|_{op}=\delta^{-1}$ and $\delta\in(\varepsilon,cn^{-1/2})$ but is "otherwise pseudo-random." Here, we require two additional ingredients.

To start unpacking (2.9), we use that $\|A^{-1}\|_{op}=\delta^{-1}$ to see that if v is a unit eigenvector corresponding to the largest eigenvalue of A^{-1}, then

$$\|A^{-1}X\|_2\leqslant s\qquad\text{implies that}\qquad|\langle X,v\rangle|<\delta s.$$

While this leads to a decent first bound of $O(\delta s)$ on the probability (2.9) (after using the quasi-randomness properties of A), however, this is *not* enough for our purposes, and in fact, we have to use

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

the additional information that X must *also* have small inner product with many other eigenvectors of A (assuming s is sufficiently small). Working along these lines, we show that (2.9) is bounded above by

$$\mathbb{P}_X\Big(|\langle X, v_1\rangle| \leqslant s\delta \text{ and } |\langle X, v_i\rangle| \leqslant \sigma_i s \text{ for all } i = 2, \ldots, n-1\Big), \tag{2.10}$$

where w_i is a unit eigenvector of A corresponding to the singular value $\sigma_i = \sigma_i(A)$. Now, appealing to the quasi-random properties of the eigenvectors of A^{-1}, we may apply our approximate negative correlation theorem (Theorem 1.5) to see that (2.10) is at most

$$O(\delta s)\exp(-cN_A(-c/s, c/s)), \tag{2.11}$$

where $c > 0$ is a constant and $N_A(a,b)$ denotes the number of eigenvalues of the matrix A in the interval (a,b). The first $O(\delta s)$ factor comes from the event $|\langle X, v_1\rangle| \leqslant s\delta$, and the second factor comes from approximating

$$\mathbb{P}_X\Big(|\langle X, w_i\rangle| < c \text{ for all } i \text{ s.t. } s\sigma_i < c\Big) = \exp\big(-\Theta(N_A(-c/s, c/s))\big)\,. \tag{2.12}$$

This bound is now sufficiently strong for our purposes, *provided* the spectrum of A adheres sufficiently closely to the typical spectrum of A_n. This now leads us to understand the rest of the spectrum of A_n and, in particular, the next smallest singular values $\sigma_{n-1}, \sigma_{n-2}, \ldots$.

Now, this might seem like a step in the wrong direction, as we are now led to understand the behavior of *many* singular values and not just the smallest. However, this "loss" is outweighed by the fact that we need only to understand these eigenvalues on scales of size $\Omega(n^{-1/2})$, which is now well understood due to the important work of Erdős et al. [13].

These results ultimately allow us to derive sufficiently strong results on quantities of the form (2.9), which, in turn, allow us to prove our "bootstrapping lemma." We then use this lemma to prove the near-optimal result

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \lesssim \varepsilon\sqrt{\log 1/\varepsilon} + e^{-cn}\,. \tag{2.13}$$

2.2.4. Removing the assumption (2.2) and the last jump to Theorem 1.1

We now turn to discuss how to remove our simplifying assumption (2.2), made above, which will allow us to close the gap between (2.13) and Theorem 1.1.

To achieve this, we need to consider how $\|A^{-1}X\|_2$ varies about $\|A^{-1}\|_{\mathrm{HS}}$, where we are, again, thinking of $A^{-1} = A_n^{-1}$ as a sufficiently quasi-random matrix. Now, the Hanson-Wright inequality tells us that, indeed, $\|A^{-1}X\|_2$ is concentrated about $\|A^{-1}\|_{\mathrm{HS}}$, on a scale $\lesssim \|A^{-1}\|_{op}$. While this is certainly useful for us, it is far from enough to prove Theorem 1.1. For this, we need to rule out any "macroscopic" correlation between the events

$$\{|\langle A^{-1}X, X\rangle - r| < K\varepsilon\|A^{-1}\|_{\mathrm{HS}}\} \text{ and } \{\|A^{-1}X\|_2 > K\|A^{-1}\|_{\mathrm{HS}}\} \tag{2.14}$$

for all $K > 0$. Our first step toward understanding (2.14) is to replace the *quadratic* large deviation event $\|A^{-1}X\|_2 > K\|A^{-1}\|_{\mathrm{HS}}$ with a collection of *linear* large deviation events:

$$\langle X, w_i\rangle > K\log(i+1),$$

where $w_n, w_{n-1}, \ldots, w_1$ are the eigenvectors of A corresponding to singular values $\sigma_n \leqslant \sigma_{n-1} \leqslant \ldots \leqslant \sigma_1$, respectively, and the $\log(i+1)$ factor should be seen as a weight function that assigns more weight to the smaller singular values.

Interestingly, we run into a similar obstacle as before: If the "bulk" of the spectrum of A^{-1} is sufficiently erratic, this replacement step will be too lossy for our purposes. Thus, we are led to prove

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

another result, showing that we may assume that the spectrum of A^{-1} adheres sufficiently to the typical spectrum of A_n. This reduces to proving

$$\mathbb{E}_{A_n}\left[\frac{\sum_{i=1}^{n}\sigma_{n-i-1}^{-2}(\log i)^2}{\sum_{i=1}^{n}\sigma_{n-i-1}^{-2}}\right]=O(1),$$

where the left-hand side is a statistic which measures the degree of distortion of the smallest singular values of A_n. To prove this, we again lean on the work of Erdős et al. [13].

Thus, we have reduced the task of proving the approximate independence of the events at (2.14) to proving the approximate independence of the collection of events

$$\{|\langle A^{-1}X,X\rangle - r| < K\varepsilon\|A^{-1}\|_{\mathrm{HS}}\} \text{ and } \{\langle v_i,X\rangle > K\log(i+1)\}.$$

This is something, it turns out, that we can handle on the Fourier side by using a quadratic analogue of our negative correlation inequality, Theorem 1.4. The idea, here, is to prove an Esseen-type bound of the form

$$\mathbb{P}(|\langle A^{-1}X,X\rangle - t| < \delta, \langle X,u\rangle \geqslant s) \lesssim \delta e^{-s}\int_{-1/\delta}^{1/\delta}\left|\mathbb{E}e^{2\pi i\theta\langle A^{-1}X,X\rangle+\langle X,u\rangle}\right| d\theta\,. \tag{2.15}$$

Which introduces this extra “exponential tilt” to the characteristic function. From here, one can carry out the plan sketched in Section 2.1 with this more complicated version of Esseen, then integrate over s to upgrade (2.13) to Theorem 1.1.

2.3. *Outline of the rest of the paper*

In the next short section, we introduce some key definitions, notation, and preliminaries that we use throughout the paper. In Section 4, we establish a collection of crucial quasi-randomness properties that hold for the random symmetric matrix A_n with probability $1-e^{-\Omega(n)}$. We shall condition on these events for most of the paper. In Section 5, we detail our Fourier decoupling argument and establish an inequality of the form (2.15). This allows us to prove our new approximate negative correlation result Lemma 5.2. In Section 6, we prepare the ground for our iterative argument by establishing (2.6), thereby switching our focus to the study of the quadratic form $\langle A_n^{-1}X,X\rangle$. In Section 7, we prove Theorem 1.2 and Theorem 1.3, which tell us that the eigenvalues of A cannot “crowd” small intervals. In Section 8, we establish regularity properties for the bulk of the spectrum of A^{-1}. In Section 9, we deploy the approximate negative correlation result (Theorem 1.5) in order to carry out the portion of the proof sketched between (2.9) and (2.12). In Section 10, we establish our base step (2.5) and bootstrap this to prove the near optimal bound (2.13). In the final section, Section 11, we complete the proof of our main Theorem 1.1.

3. Key definitions and preliminaries

We first need a few notions out of the way, which are related to our paper [4] on the singularity of random symmetric matrices.

3.1. *Subgaussian and matrix definitions*

Throughout, ζ will be a mean 0, variance 1 random variable. We define the *subgaussian moment* of ζ to be

$$\|\zeta\|_{\psi_2} := \sup_{p\geqslant 1} p^{-1/2}(\mathbb{E}\,|\zeta|^p)^{1/p}\,.$$

A mean 0, variance 1 random variable is said to be *subgaussian* if $\|\zeta\|_{\psi_2}$ is finite. We define Γ to be the set of subgaussian random variables and, for $B > 0$, we define $\Gamma_B \subseteq \Gamma$ to be a subset of ζ with $\|\zeta\|_{\psi_2} \leqslant B$.

For $\zeta \in \Gamma$, define $\mathrm{Sym}_n(\zeta)$ to be the probability space on $n \times n$ symmetric matrices A for which $(A_{i,j})_{i \geqslant j}$ are independent and distributed according to ζ. Similarly, we write $X \sim \mathrm{Col}_n(\zeta)$ if $X \in \mathbb{R}^n$ is a random vector whose coordinates are i.i.d. copies of ζ.

We shall think of the spaces $\{\mathrm{Sym}_n(\zeta)\}_n$ as coupled in the natural way: The matrix $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$ can be sampled by first sampling $A_n \sim \mathrm{Sym}_n(\zeta)$, which we think of as the principal minor $(A_{n+1})_{[2,n+1]\times[2,n+1]}$, and then generating the first row and column of A_{n+1} by generating a random column $X \sim \mathrm{Col}_n(\zeta)$. In fact, it will make sense to work with a random $(n+1)\times(n+1)$ matrix, which we call A_{n+1} throughout. This is justified, as much of the work is done with the principal minor A_n of A_{n+1}, due to the bound (2.1) as well as Lemma 6.1.

3.2. Compressible vectors

We shall require the now-standard notions of *compressible vectors*, as defined by Rudelson and Vershynin [31].

For parameters $\rho, \delta \in (0,1)$, we define the set of compressible vectors $\mathrm{Comp}(\delta,\rho)$ to be the set of vectors in $\mathbb{S}^{n-1}$ that are distance at most ρ from a vector supported on at most δn coordinates. We then define the set of *incompressible* vectors to be all other unit vectors, that is $\mathrm{Incomp}(\delta,\rho) := \mathbb{S}^{n-1} \setminus \mathrm{Comp}(\delta,\rho)$. The following basic fact about incompressible vectors from [31] will be useful throughout:

Fact 3.1. For each $\delta, \rho \in (0,1)$, there is a constant $c_{\rho,\delta} \in (0,1)$, so that for all $v \in \mathrm{Incomp}(\delta,\rho)$, we have that $|v_j| n^{1/2} \in [c_{\rho,\delta}, c_{\rho,\delta}^{-1}]$ for at least $c_{\rho,\delta} n$ values of j.

Fact 3.1 assures us that for each incompressible vector, we can find a large subvector that is "flat." Using the work of Vershynin [46], we will safely be able to ignore compressible vectors. In particular, [46, Proposition 4.2] implies the following lemma. We refer the reader to Appendix XII for details.

Lemma 3.2. *For $B > 0$ and $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$. Then there exist constants $\rho, \delta, c \in (0,1)$, depending only on B, so that*

$$\sup_{u \in \mathbb{R}^n} \mathbb{P}\big(\exists x \in \mathrm{Comp}(\delta,\rho), \exists t \in \mathbb{R} : Ax = tu\big) \leqslant 2e^{-cn}$$

and

$$\mathbb{P}\big(\exists u \in \mathrm{Comp}(\delta,\rho), \exists t \in \mathbb{R} : Au = tu\big) \leqslant 2e^{-cn}.$$

The first statement says, roughly, that $A^{-1}u$ is incompressible for each fixed u; the second states that all unit eigenvectors are incompressible.

Remark 3.3 (Choice of constants, $\rho, \delta, c_{\rho,\delta}$). Throughout, we let ρ, δ denote the constants guaranteed by Lemma 3.2 and $c_{\rho,\delta}$ the corresponding constant from Fact 3.1. These constants shall appear throughout the paper and shall always be considered as fixed.

Lemma 3.2 follows easily from [46, Proposition 4.2] with a simple net argument.

3.3. Notation

We quickly define some notation. For a random variable X, we use the notation $\mathbb{E}_X$ for the expectation with respect to X and we use the notation $\mathbb{P}_X$ analogously. For an event $\mathcal{E}$, we write $\mathbf{1}_{\mathcal{E}}$ or $\mathbf{1}\{\mathcal{E}\}$ for the indicator function of the event $\mathcal{E}$. We write $\mathbb{E}^{\mathcal{E}}$ to be the expectation defined by $\mathbb{E}^{\mathcal{E}}[\,\cdot\,] = \mathbb{E}[\,\cdot\,\mathbf{1}_{\mathcal{E}}]$. For a vector $v \in \mathbb{R}^n$ and $J \subset [n]$, we write v_J for the vector whose ith coordinate is v_i if $i \in J$ and 0 otherwise.

We shall use the notation $X \lesssim Y$ to indicate that there exists a constant $C > 0$ for which $X \leqslant CY$. In a slight departure from convention, *we will always allow this constant to depend on the subgaussian constant* B, if present. We shall also let our constants implicit in big-O notation to depend on B, if this constant is relevant in the context. We hope that we have been clear as to where the subgaussian constant is relevant, and so this convention is to just reduce added clutter.

4. Quasi-randomness properties

In this technical section, we define a list of "quasi-random" properties of A_n that hold with probability $1 - e^{-\Omega(n)}$. This probability is large enough that we can assume that these properties hold for all the principal minors of A_{n+1}. Showing that several of these quasi-random properties hold with probability $1 - e^{-\Omega(n)}$ will prove to be a challenging task, and our proof will depend deeply on ideas from our previous paper [4], on the singularity probability of a random symmetric matrix. So as not to distract from the new ideas in this paper, we do most of this work in the Appendix.

4.1. Defining the properties

It will be convenient to assume throughout that every minor of A_{n+1} is invertible, and so we will perturb the matrix slightly so that we may assume this. If we add to A_{n+1} an independent random symmetric matrix whose upper triangular entries are independent gaussian random variables with mean 0 and variance n^{-n}, then with probability $1 - e^{-\Omega(n)}$, the singular values of A_{n+1} move by at most, say, $n^{-n/3}$. Further, after adding this random gaussian matrix, every minor of the resulting matrix is invertible with probability 1. Thus, we will assume without loss of generality throughout that every minor of A_{n+1} is invertible.

In what follows, we let $A = A_n \sim \mathrm{Sym}_n(\zeta)$ and let $X \sim \mathrm{Col}_n(\zeta)$ be a random vector, independent of A. Our first quasi-random property is standard from the concentration of the operator norm of a random symmetric matrix. We define $\mathcal{E}_1$ by

$$\mathcal{E}_1 = \{\|A\|_{op} \leqslant 4\sqrt{n}\}. \tag{4.1}$$

For the next property, we need a definition. Let $X, X' \sim \mathrm{Col}_n(\zeta)$, and define the random vector in $\mathbb{R}^n$ as $\tilde{X} := X_J - X'_J$, where $J \subseteq [n]$ is a μ-random subset, that is, for each $j \in [n]$, we have $j \in J$ independently with probability μ. The reason behind this definition is slightly opaque at present, but will be clear in the context of Lemma 5.2 in Section 5. Until we get there, it is reasonable to think of $\tilde{X}$ as being essentially X; in particular, it is a random vector with i.i.d. subgaussian entries with mean 0 and variance μ. We now define $\mathcal{E}_2$ to be the event in A defined by

$$\mathcal{E}_2 = \left\{\mathbb{P}_{\widetilde{X}}\left(A^{-1}\widetilde{X}/\|A^{-1}\widetilde{X}\|_2 \in \mathrm{Comp}\,(\delta,\rho)\right) \leqslant e^{-c_2 n}\right\}. \tag{4.2}$$

We remind the reader that $\mathrm{Comp}\,(\delta,\rho)$ is defined in Section 3.2, and $\delta, \rho \in (0,1)$ are constants, fixed throughout the paper, and chosen according to Lemma 3.2. In the (rare) case that $\widetilde{X} = 0$, we interpret $\mathbb{P}_{\widetilde{X}}(A^{-1}\widetilde{X}/\|A^{-1}\widetilde{X}\|_2 \in \mathrm{Comp}\,(\delta,\rho)) = 1$.

Recalling the least common denominator defined at (1.11), we now define the event $\mathcal{E}_3$ by

$$\mathcal{E}_3 = \{D_{\alpha,\gamma}(u) \geqslant e^{c_3 n} \text{ for every unit eigenvector } u \text{ of } A\}. \tag{4.3}$$

The next condition tells us that the random vector $A^{-1}\widetilde{X}$ is typically unstructured. We will need a slightly stronger notion of structure than just looking at the LCD, in that, we will need all sufficiently large subvectors to be unstructured. For $\mu \in (0,1)$, define the *subvector least common denominator*, as

$$\hat{D}_{\alpha,\gamma,\mu}(v) := \min_{\substack{I \subset [n] \\ |I| \geqslant (1-2\mu)n}} D_{\alpha,\gamma}\left(v_I/\|v_I\|_2\right).$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

We note that this is closely related to the notion of "regularized least common denominator" introduced by Vershynin in [46].

Now, if we define the random vector $v = v(\widetilde{X}) := A^{-1}\widetilde{X}$, then we define $\mathcal{E}_4$ to be the event that A satisfies

$$\mathcal{E}_4 = \left\{\mathbb{P}_{\widetilde{X}}\left(\hat{D}_{\alpha,\gamma,\mu}(v) < e^{c_4 n}\right) \leqslant e^{-c_4 n}\right\}. \tag{4.4}$$

As is the case for $\mathcal{E}_2$, under the event that $\widetilde{X} = 0$, we interpret $\mathbb{P}_{\widetilde{X}}(\hat{D}_{\alpha,\gamma,\mu}(v) < e^{c_4 n}) = 1$.

We now define our main quasi-randomness event $\mathcal{E}$ to be the intersection of these events:

$$\mathcal{E} := \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4. \tag{4.5}$$

The following lemma essentially allows us to assume that $\mathcal{E}$ holds in what follows.

Lemma 4.1. *For $B > 0$, $\zeta \in \Gamma_B$, and all sufficiently small $\alpha, \gamma, \mu \in (0,1)$, there exist constants $c_2, c_3, c_4 \in (0,1)$ appearing in (4.2), (4.3), and (4.4) so that*

$$\mathbb{P}_A(\mathcal{E}^c) \leqslant 2e^{-\Omega(n)}. \tag{4.6}$$

Remark 4.2 (Choice of constants, α, γ, μ). We take $\alpha, \gamma \in (0,1)$ to be sufficiently small so that Lemma 4.1 holds. For μ, we will choose it to be sufficiently small so that (1) Lemma 4.1 holds; (2) we have $\mu \in (0, 2^{-15})$; and so that (3) $\mu > 0$ is small enough to guarantee that every set $I \subseteq [n]$ with $|I| \geqslant (1-2\mu)n$ satisfies

$$\|w\|_2 \leqslant c_{\rho,\delta}^{-2}\|w_I\|_2, \tag{4.7}$$

for every $w \in \mathrm{Incomp}(\delta, \rho)$. This is possible by Fact 3.1. These constants α, γ, μ will appear throughout the paper and will always be thought of as fixed according to this choice.

4.2. Statement of our master quasi-randomness theorem and the deduction of Lemma 4.1

We will deduce Lemma 4.1 from a "master quasi-randomness theorem" together with a handful of now-standard results in the area.

For the purposes of the following sections, we shall informally consider a vector as "structured" if

$$\hat{D}_{\alpha,\gamma,\mu}(v) \leqslant e^{c_\Sigma n},$$

where $c_\Sigma \in (0,1)$ is a small constant, to be chosen shortly. Thus, it makes sense to define the set of "structured directions" on the sphere

$$\Sigma = \Sigma_{\alpha,\gamma,\mu} := \{v \in \mathbb{S}^{n-1} : \hat{D}_{\alpha,\gamma,\mu}(v) \leqslant e^{c_\Sigma n}\}. \tag{4.8}$$

We now introduce our essential quasi-randomness measure of a random matrix. For $\zeta \in \Gamma$, $A \sim \mathrm{Sym}_n(\zeta)$, and a given vector $w \in \mathbb{R}^n$, define

$$q_n(w) = q_n(w; \alpha, \gamma, \mu) := \mathbb{P}_A\left(\exists v \in \Sigma \text{ and } \exists s, t \in [-4\sqrt{n}, 4\sqrt{n}] : Av = sv + tw\right) \tag{4.9}$$

and set

$$q_n = q_n(\alpha, \gamma, \mu) := \sup_{w \in \mathbb{S}^{n-1}} q_n(w). \tag{4.10}$$

We now state our "master quasi-randomness theorem," from which we deduce Lemma 4.1.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Theorem 4.3 (Master quasi-randomness theorem). *For $B > 0$ and $\zeta \in \Gamma_B$, there exist constants $\alpha, \gamma, \mu, c_\Sigma, c \in (0,1)$ depending only on B so that*

$$q_n(\alpha,\gamma,\mu) \leqslant 2e^{-cn}\,.$$

The proof of Theorem 4.3 is quite similar to the main theorem of [4], albeit with a few technical adaptations, and is proved in the Appendix. Note that $q_n(\alpha,\gamma,\mu)$ is monotone decreasing as α, γ, and μ decrease. As such, Theorem 4.3 implies that its conclusion holds for all sufficiently small α, γ, μ as well.

We now prove that our pseudo-random event $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4$ holds with probability $1 - e^{-\Omega(n)}$.

Proof of Lemma 4.1. The event $\mathcal{E}_1$: From [15], we may deduce[5] the following concentration bound

$$\mathbb{P}\big(\,\|A\|_{op} \geqslant (3+t)\sqrt{n}\big) \lesssim e^{-ct^{3/2}n}, \tag{4.11}$$

which holds[6] for all $t \geqslant 0$. Thus, by (4.11), the event $\mathcal{E}_1$ at (4.1) fails with probability $\lesssim e^{-\Omega(n)}$.

The event $\mathcal{E}_2$: By Lemma 3.2, there is a $c > 0$ so that for each $u \neq 0$, we have

$$\mathbb{P}_A(A^{-1}u/\|A^{-1}u\|_2 \in \mathrm{Comp}\,(\delta,\rho)) \leqslant e^{-cn}\,.$$

Applying Markov's inequality shows

$$\mathbb{P}_A\left(\mathbb{P}_{\widetilde{X}}\left(A^{-1}\widetilde{X}/\|A^{-1}\widetilde{X}\|_2 \in \mathrm{Comp}\,(\delta,\rho), \widetilde{X} \neq 0\right) > e^{-cn/2}\right) \leqslant e^{-cn/2}\,,$$

and so the event in (4.2) fails with probability at most $O\left(e^{-\Omega(n)}\right)$, under the event $\widetilde{X} \neq 0$. By Theorem 3.1.1, in [47], we have that

$$\mathbb{P}_{\widetilde{X}}(\widetilde{X} = 0) \leqslant e^{-\Omega(\mu n)}\,. \tag{4.12}$$

Choosing c_2 small enough shows an exponential bound on $\mathbb{P}(\mathcal{E}_2^c)$.

The event $\mathcal{E}_3$: If $D_{\alpha,\gamma}(u) \leqslant e^{c_3 n}$, for an u an eigenvector $Au = \lambda v$, we have that

$$\hat{D}_{\alpha,\gamma,\mu}(u) \leqslant D_{\alpha,\gamma}(u) \leqslant e^{c_3 n},$$

where the first inequality is immediate from the definition. Now, note that if $\mathcal{E}_1$ holds, then $\lambda \in [-4\sqrt{n}, 4\sqrt{n}]$, and so

$$\mathbb{P}(\mathcal{E}_3^c) \leqslant \mathbb{P}\big(\exists u \in \Sigma, \lambda \in [-4\sqrt{n}, 4\sqrt{n}] : Au = \lambda u\big) + \mathbb{P}(\mathcal{E}_1^c) \leqslant q_n(0) + e^{-\Omega(n)},$$

where the first inequality holds if we choose $c_3 \leqslant c_\Sigma$. We now apply Theorem 4.3 to see $q_n(0) \leqslant q_n \lesssim e^{-\Omega(n)}$, yielding the desired result.

The event $\mathcal{E}_4$: Note first that, by (4.12), we may assume $\widetilde{X} \neq 0$. For a fixed instance of $\widetilde{X} \neq 0$, we have

$$\mathbb{P}_A\left(\hat{D}_{\alpha,\gamma,\mu}\left(A^{-1}\tilde{X}/\|\tilde{X}\|_2\right) < e^{c_4 n}\right) \leqslant \mathbb{P}_A\big(\exists v \in \Sigma : Av = \tilde{X}/\|\tilde{X}\|_2\big) \leqslant q_n\left(\tilde{X}/\|\tilde{X}\|_2\right), \tag{4.13}$$

which is at most $e^{-\Omega(n)}$, by Theorem 4.3. Here, the first inequality holds when $c_4 \leqslant c_\Sigma$.

We now write $v = A^{-1}\tilde{X}/\|\tilde{X}\|_2$ and apply Markov's inequality

$$\mathbb{P}(\mathcal{E}_4^c) = \mathbb{P}_A\left(\mathbb{P}_{\tilde{X}}\left(\hat{D}_{\alpha,\gamma,\mu}(v) < e^{c_4 n}\right) \geqslant e^{-c_4 n}\right) \leqslant e^{c_4 n}\mathbb{E}_{\tilde{X}}\mathbb{P}_A(\hat{D}_{\alpha,\gamma,\mu}(v) < e^{c_4}n) = e^{-\Omega(n)},$$

[5]Technically, the result of [15] is sharper and for random matrices whose entries are symmetric random variables. However, (4.11) follows from [15] along with a "symmetrization trick."

[6]We use this bound rather than the more standard concentration bounds for the norm of subgaussian random matrices for aesthetic purposes: It allows us to use the absolute constant "4" in (4.1).

where the last line follows when c_4 is taken small relative to the implicit constant in the bound on the right-hand side of (4.13).

Since we have shown that each of $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_4$ holds with probability $1 - e^{-\Omega(n)}$, the intersection fails with exponentially small probability. □

5. Decoupling quadratic forms

In this section, we will prove our Esseen-type inequality that will allow us to deal with a small ball event and a large deviation event simultaneously.

Lemma 5.1. *For $B > 0$, let $\zeta \in \Gamma_B$ and $X \sim \mathrm{Col}_n(\zeta)$. Let M be an $n \times n$ symmetric matrix, $u \in \mathbb{R}^n$, $t \in \mathbb{R}$, and $s, \delta \geqslant 0$. Then*

$$\mathbb{P}(|\langle MX, X\rangle - t| < \delta, \langle X, u\rangle \geqslant s) \lesssim \delta e^{-s} \int_{-1/\delta}^{1/\delta} \left|\mathbb{E}\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right| d\theta\,. \tag{5.1}$$

We will then bound the integrand (our so-called "titled" characteristic function) with a *decoupling* maneuver, somewhat similar to a "van der Corput trick" in classical Fourier analysis. This amounts to a clever application of Cauchy-Schwarz, inspired by Kwan and Sauermann's work on Costello's conjecture [20] (a similar technique appears in [1] and [25]). We shall then be able to mix in our quasi-random conditions on our matrix A to ultimately obtain Lemma 5.2, which gives us a rather tractable bound on the left-hand side of (5.1). To state this lemma, let us recall that $\mathcal{E}$ (defined at (4.5)) is the set of symmetric matrices satisfying the quasi-randomness conditions in the previous section, Section 4. Also recall that the constant $\mu \in (0, 2^{-15})$ is defined in Section 4 so that Lemma 4.1 holds and is treated as a fixed constant throughout this paper.

Lemma 5.2. *For $B > 0$, let $\zeta \in \Gamma_B$, $X \sim \mathrm{Col}_n(\zeta)$ and let A be a real symmetric $n \times n$ matrix with $A \in \mathcal{E}$ and set $\mu_1 := \sigma_{\max}(A^{-1})$. Also let $s \geqslant 0, \delta > e^{-cn}$ and $u \in \mathbb{S}^{n-1}$. Then*

$$\mathbb{P}_X\left(\left|\langle A^{-1}X, X\rangle - t\right| \leqslant \delta\mu_1, \langle X, u\rangle \geqslant s\right) \lesssim \delta e^{-s} \int_{-1/\delta}^{1/\delta} I(\theta)^{1/2}\, d\theta + e^{-\Omega(n)}\,,$$

where

$$I(\theta) := \mathbb{E}_{J, X_J, X'_J}\, \exp\left(\langle (X + X')_J, u\rangle - c\theta^2\mu_1^{-2}\|A^{-1}(X - X')_J\|_2^2\right),$$

$X' \sim \mathrm{Col}_n(\zeta)$ is independent of X, and $J \subseteq [n]$ is a μ-random set. Here, $c > 0$ is a constant depending only on B.

While the definition of $I(\theta)$ (and therefore the conclusion of the lemma) is a bit mysterious at this point, we assure the reader that this is a step in the right direction.

All works bounding the singularity probability for random symmetric matrices contain a related decoupling step [3, 4, 5, 16, 25, 46], starting with Costello et al.'s breakthrough [7], building off of Costello's earlier work [6] on anticoncentration of bilinear and quadratic forms. A subtle difference in the decoupling approach from [20] used here is that the quadratic form is decoupled *after* bounding a small ball probability in terms of the integral of a characteristic function rather than on the probability itself; the effect of this approach is that we do not lose a power of δ, but only lose by a square root "under the integral" on the integrand $I(\theta)$.

5.1. Proofs

We now dive in and prove our Esseen-type inequality. For this, we shall appeal to the classical Esseen inequality [14]: If Z is a random variable taking values in $\mathbb{R}$ with characteristic function $\varphi_Z(\theta) :=$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

$\mathbb{E}_Z\, e^{2\pi i\theta Z}$, then for all $t \in \mathbb{R}$, we have

$$\mathbb{P}_X(|Z-t| \leqslant \delta) \lesssim \delta \int_{-1/\delta}^{1/\delta} |\varphi_Z(\theta)|\, d\theta.$$

We shall also use the following basic fact about subgaussian random vectors (see, for example, [47, Proposition 2.6.1]): If $\zeta \in \Gamma_B$ and $Y \sim \mathrm{Col}_n(\zeta)$, then for every vector $u \in \mathbb{R}^n$, we have

$$\mathbb{E}_Y\, e^{\langle Y,u\rangle} \leqslant \exp(2B^2\|u\|_2^2)\,. \tag{5.2}$$

Proof of Lemma 5.1. Since $\mathbf{1}\{x \geqslant s\} \leqslant e^{x-s}$, we may bound

$$\mathbb{P}_X(|\langle MX,X\rangle - t| < \delta, \langle X,u\rangle \geqslant s) \leqslant e^{-s}\mathbb{E}\left[\mathbf{1}\{|\langle MX,X\rangle - t| < \delta\}e^{\langle X,u\rangle}\right]. \tag{5.3}$$

Define the random variable $Y \in \mathbb{R}^n$ by

$$\mathbb{P}(Y \in U) = (\mathbb{E}\, e^{\langle X,u\rangle})^{-1}\mathbb{E}[\mathbf{1}_U e^{\langle X,u\rangle}], \tag{5.4}$$

for all open $U \subseteq \mathbb{R}^n$. Note that the expectation $\mathbb{E}_X e^{\langle X,u\rangle}$ is finite by (5.2). We now use this definition to rewrite the expectation on the right-hand side of (5.3),

$$\mathbb{E}_X\left[\mathbf{1}\{|\langle MX,X\rangle - t| < \delta\}e^{\langle X,u\rangle}\right] = \left(\mathbb{E}\, e^{\langle X,u\rangle}\right)\mathbb{P}_Y(|\langle MY,Y\rangle - t| \leqslant \delta)\,.$$

Thus, we may apply Esseen's lemma to the random variable Y to obtain

$$\mathbb{P}_Y(|\langle MY,Y\rangle - t| \leqslant \delta) \lesssim \delta \int_{-1/\delta}^{1/\delta} |\mathbb{E}_Y\, e^{2\pi i\theta\langle MY,Y\rangle}|\, d\theta\,.$$

By the definition of Y, we have

$$\mathbb{E}_Y\, e^{2\pi i\theta\langle MY,Y\rangle} = \left(\mathbb{E}_X\, e^{\langle X,u\rangle}\right)^{-1}\mathbb{E}\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle},$$

completing the lemma. □

To control the integral on the right-hand side of Lemma 5.1, we will appeal to the following decoupling lemma, which is adapted from Lemma 3.3 from [20].

Lemma 5.3 (Decoupling with an exponential tilt). *Let $\zeta \in \Gamma$, let $X, X' \sim \mathrm{Col}_n(\zeta)$ be independent, and let $J \cup I = [n]$ be a partition of $[n]$. Let M be an $n \times n$ symmetric matrix and let $u \in \mathbb{R}^n$. Then*

$$\left|\mathbb{E}_X\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2 \leqslant \mathbb{E}_{X_J,X_J'}\, e^{\langle (X+X')_J,u\rangle}\cdot\left|\mathbb{E}_{X_I} e^{4\pi i\theta\langle M(X-X')_J,X_I\rangle + 2\langle X_I,u\rangle}\right|.$$

Proof. After partitioning the coordinates of X according to J and writing $\mathbb{E}_X = \mathbb{E}_{X_I}\mathbb{E}_{X_J}$, we apply Jensen's inequality to obtain

$$E := \left|\mathbb{E}_X\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2 = \left|\mathbb{E}_{X_I}\mathbb{E}_{X_J}\, e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2 \leqslant \mathbb{E}_{X_I}\left|\mathbb{E}_{X_J} e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2.$$

We now expand the square $\left|\mathbb{E}_{X_J} e^{2\pi i\theta\langle MX,X\rangle + \langle X,u\rangle}\right|^2$ as

$$\begin{aligned}&\mathbb{E}_{X_J,X_J'} e^{2\pi i\theta\langle M(X_I+X_J),(X_I+X_J)\rangle + \langle (X_I+X_J),u\rangle - 2\pi i\theta\langle M(X_I+X_J'),(X_I+X_J')\rangle + \langle (X_I+X_J'),u\rangle}\\ &= \mathbb{E}_{X_J,X_J'} e^{4\pi i\theta\langle M(X_J-X_J'),X_I\rangle + \langle X_J+X_J',u\rangle + 2\langle X_I,u\rangle + 2\pi i\langle MX_J,X_J\rangle - 2\pi i\langle MX_J',X_J'\rangle},\end{aligned}$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

where we used the fact that M is symmetric. Thus, swapping expectations yields

$$\begin{aligned} E &\leqslant \mathbb{E}_{X_J,X'_J}\mathbb{E}_{X_I} e^{4\pi i\theta\langle M(X_J-X'_J),X_I\rangle+\langle X_J+X'_J,u\rangle+2\langle X_I,u\rangle+2\pi i\langle MX_J,X_J\rangle-2\pi i\langle MX'_J,X'_J\rangle} \\ &\leqslant \mathbb{E}_{X_J,X'_J}\left|\mathbb{E}_{X_I} e^{4\pi i\theta\langle M(X_J-X'_J),X_I\rangle+\langle X_J+X'_J,u\rangle+2\langle X_I,u\rangle+2\pi i\langle MX_J,X_J\rangle-2\pi i\langle MX'_J,X'_J\rangle}\right| \\ &= \mathbb{E}_{X_J,X'_J}\, e^{\langle X_J+X'_J,u\rangle}\left|\mathbb{E}_{X_I} e^{4\pi i\theta\langle M(X-X')_J,X_I\rangle+2\langle X_I,u\rangle}\right|, \end{aligned}$$

as desired. Here, we could swap expectations, since all expectations are finite, due to the subgaussian assumption on ζ. □

We need a basic bound that will be useful for bounding our tilted characteristic function. This bound appears in the proof of Theorem 6.3 in Vershynin's paper [46].

Fact 5.4. For $B > 0$, let $\zeta \in \Gamma_B$, let ζ' be an independent copy of ζ, and set $\xi = \zeta - \zeta'$. Then for all $a \in \mathbb{R}^n$, we have

$$\prod_j \mathbb{E}_\xi\,|\cos(2\pi\xi a_j)| \leqslant \exp\left(-c \min_{r\in[1,c^{-1}]} \|ra\|_{\mathbb{T}}^2\right),$$

where $c > 0$ depends only on B.

A simple symmetrization trick along with Cauchy-Schwarz will allow us to prove a similar bound for the tilted characteristic function.

Lemma 5.5. *For $B > 0$, let $\zeta \in \Gamma_B$, $X \sim \mathrm{Col}_n(\zeta)$ and let $u, v \in \mathbb{R}^n$. Then*

$$\left|\mathbb{E}_X e^{2\pi i\langle X,v\rangle+\langle X,u\rangle}\right| \leqslant \exp\left(-c \min_{r\in[1,c^{-1}]} \|rv\|_{\mathbb{T}}^2 + c^{-1}\|u\|_2^2\right), \tag{5.5}$$

where $c \in (0, 1)$ depends only on B.

Proof. Let ζ' be an independent copy of ζ, and note that

$$\left|\mathbb{E}_\zeta\, e^{2\pi i\zeta v_j+\zeta u_j}\right|^2 = \mathbb{E}_{\zeta,\zeta'}\, e^{2\pi i(\zeta-\zeta')v_j+(\zeta+\zeta')u_j} = \mathbb{E}_{\zeta,\zeta'}\left[e^{(\zeta+\zeta')u_j}\cos(2\pi(\zeta-\zeta')v_j)\right].$$

Let $\widetilde{X} = (\widetilde{X}_i)_{i=1}^n$, $\widetilde{Y} = (Y_i)_{i=1}^n$ denote vectors with i.i.d. coordinates distributed as $\xi := \zeta - \zeta'$ and $\zeta + \zeta'$, respectively. We have

$$\left|\mathbb{E}_X e^{2\pi i\langle X,v\rangle+\langle X,u\rangle}\right|^2 \leqslant \mathbb{E}\, e^{\langle\widetilde{Y},u\rangle}\prod_j \cos(2\pi\widetilde{X}_j v_j) \leqslant \left(\mathbb{E}_{\widetilde{Y}} e^{2\langle\widetilde{Y},u\rangle}\right)^{1/2}\left(\prod_j \mathbb{E}_\xi|\cos(2\pi\xi v_j)|\right)^{1/2}, \tag{5.6}$$

where we have applied the Cauchy-Schwarz inequality along with the bound $|\cos(x)|^2 \leqslant |\cos(x)|$ to obtain the last inequality. By (5.2), the first expectation on the right-hand side of (5.6) is at most $\exp(O(\|u\|_2^2))$. Applying Fact 5.4 completes the lemma. □

5.2. *Quasi-random properties for triples* (J, X_J, X'_J)

We now prepare for the proof of Lemma 5.2 by introducing a quasi-randomness notion on triples (J, X_J, X'_J). Here, $J \subseteq [n]$ and $X, X' \in \mathbb{R}^n$. For this, we fix an $n \times n$ real symmetric matrix $A \in \mathcal{E}$ and define the event $\mathcal{F} = \mathcal{F}(A)$ as the intersection of the events $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$, and $\mathcal{F}_4$, which are defined as follows. Given a triple (J, X_J, X'_J), we write $\tilde{X} := X_J - X'_J$.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Define events $\mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3(A)$ by

$$\mathcal{F}_1 := \{|J| \in [\mu n/2, 2\mu n]\} \tag{5.7}$$

$$\mathcal{F}_2 := \{\|\widetilde{X}\|_2 n^{-1/2} \in [c, c^{-1}]\} \tag{5.8}$$

$$\mathcal{F}_3(A) := \{A^{-1}\widetilde{X}/\|A^{-1}\widetilde{X}\|_2 \in \mathrm{Incomp}\,(\delta, \rho)\}\,. \tag{5.9}$$

Finally, we write $v = v(\tilde{X}) := A^{-1}\widetilde{X}$ and $I := [n] \setminus J$ and then define $\mathcal{F}_4(A)$ by

$$\mathcal{F}_4(A) := \left\{D_{\alpha,\gamma}\left(\frac{v_I}{\|v_I\|}\right) > e^{cn}\right\}\,. \tag{5.10}$$

We now define $\mathcal{F}(A) := \mathcal{F}_1 \cap \mathcal{F}_2 \cap \mathcal{F}_3(A) \cap \mathcal{F}_4(A)$ and prove the following basic lemma that will allow us to essentially assume that (5.7),(5.8),(5.9),(5.10) hold in all that follows. We recall that the constants $\delta, \rho, \mu, \alpha, \gamma$ were chosen in Lemmas 3.2 and 4.1 as a function of the subgaussian moment B. Thus, the only new parameter in $\mathcal{F}$ is the constant c in lines (5.8) and (5.10).

Lemma 5.6. *For $B > 0$, let $\zeta \in \Gamma_B$, let $X, X' \sim \mathrm{Col}_n(\zeta)$ be independent, and let $J \subseteq [n]$ be a μ-random subset. Let A be an $n \times n$ real symmetric matrix with $A \in \mathcal{E}$. We may choose the constant $c \in (0, 1)$ appearing in (5.8) and (5.10) as a function of B and μ so that*

$$\mathbb{P}_{J,X_J,X'_J}(\mathcal{F}^c) \lesssim e^{-cn}\,.$$

Proof. For $\mathcal{F}_1$, we use Hoeffding's inequality to see $\mathbb{P}(\mathcal{F}_1^c) \lesssim e^{-\Omega(n)}$. To bound $\mathbb{P}(\mathcal{F}_2^c)$, we note that the entries of $\widetilde{X}$ are independent, subgaussian, and have variance 2μ, and so $\widetilde{X}/(\sqrt{2\mu})$ has i.i.d. entries with mean zero, variance 1 and subgaussian moment bounded by $B/\sqrt{2\mu}$. Thus, from Theorem 3.1.1 in [47], we have

$$\mathbb{P}\big(\,\big|\|\widetilde{X}\|_2 - \sqrt{2n\mu}\big| > t\big) < \exp(-c\mu t^2/B^4).$$

For $\mathcal{F}_3(A), \mathcal{F}_4(A)$, recall that $A \in \mathcal{E}$ means that (4.2) and (4.4) hold, thus exponential bounds on $\mathbb{P}(\mathcal{F}_3^c)$ and $\mathbb{P}(\mathcal{F}_4^c)$ follow from Markov's inequality. □

5.3. *Proof of Lemma 5.2*

We now prove Lemma 5.2 by applying the previous three lemmas in sequence.

Proof of Lemma 5.2. Let $\delta \geqslant e^{-c_1 n}$, where we will choose $c_1 > 0$ to be sufficiently small later in the proof. Apply Lemma 5.1 to write

$$\mathbb{P}_X\left(\left|\langle A^{-1}X, X\rangle - t\right| \leqslant \delta\mu_1, \langle X, u\rangle \geqslant s\right) \lesssim \delta e^{-s}\int_{-1/\delta}^{1/\delta}\left|\mathbb{E}_X\, e^{2\pi i\theta\frac{\langle A^{-1}X,X\rangle}{\mu_1} + \langle X,u\rangle}\right| d\theta\,, \tag{5.11}$$

where we recall that $\mu_1 = \sigma_{\max}(A^{-1})$. We now look to apply our decoupling lemma, Lemma 5.3. Let J be a μ-random subset of $[n]$, define $I := [n] \setminus J$, and let X' be an independent copy of X. By Lemma 5.3, we have

$$\left|\mathbb{E}_X\, e^{2\pi i\theta\frac{\langle A^{-1}X,X\rangle}{\mu_1} + \langle X,u\rangle}\right|^2 \leqslant \mathbb{E}_J\mathbb{E}_{X_J,X'_J}\, e^{\langle (X+X')_J,u\rangle} \cdot \left|\mathbb{E}_{X_I}\, e^{4\pi i\theta\left\langle\frac{A^{-1}\widetilde{X}}{\mu_1}, X_I\right\rangle + 2\langle X_I,u\rangle}\right|\,, \tag{5.12}$$

where we recall that $\widetilde{X} = (X - X')_J$.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

We first consider the contribution to the expectation on the right-hand side of (5.12) from triples $(J, X_J, X'_J) \notin \mathcal{F}$. For this, let Y be a random vector, such that $Y_j = X_j + X'_j$, if $j \in J$, and $Y_j = 2X_j$, if $j \in I$. Applying the triangle inequality, we have

$$\mathbb{E}^{\mathcal{F}^c}_{J,X_J,X'_J}\, e^{\langle (X+X')_J, u\rangle} \cdot \left|\mathbb{E}_{X_I}\, e^{4\pi i\theta\langle \frac{A^{-1}\widetilde{X}}{\mu_1}, X_I\rangle + 2\langle X_I, u\rangle}\right| \leqslant \mathbb{E}^{\mathcal{F}^c}_{J,X_J,X'_J}\, e^{\langle (X+X')_J, u\rangle}\mathbb{E}_{X_I}\, e^{2\langle X_I, u\rangle}$$
$$= \mathbb{E}^{\mathcal{F}^c}_{J,X,X'} e^{\langle Y, u\rangle}.$$

By Cauchy-Schwarz, (5.2), and Lemma 5.6, we have

$$\mathbb{E}^{\mathcal{F}^c}_{J,X,X'}\, e^{\langle Y, u\rangle} \leqslant \mathbb{E}_{J,X,X'}\left[e^{\langle Y, 2u\rangle}\right]^{1/2} \mathbb{P}_{J,X_J,X'_J}(\mathcal{F}^c)^{1/2} \lesssim e^{-\Omega(n)}\,. \tag{5.13}$$

We now consider the contribution to the expectation on the right-hand side of (5.12) from triples $(J, X_J, X'_J) \in \mathcal{F}$. For this, let $w = w(X) := \frac{A^{-1}\widetilde{X}}{\mu_1}$ and assume $(J, X_J, X'_J) \in \mathcal{F}$. By Lemma 5.5, we have

$$\left|\mathbb{E}_{X_I} e^{4\pi i\theta\langle X_I, w\rangle + \langle X_I, 2u\rangle}\right| \lesssim \exp\left(-c \min_{r\in[1,c^{-1}]} \|2r\theta w_I\|_{\mathbb{T}}^2\right). \tag{5.14}$$

Note that $\|w_I\|_2 \leqslant \|\widetilde{X}\|_2 \leqslant c^{-1}\sqrt{n}$, by the definition of $\mu_1 = \sigma_{\max}(A^{-1})$ and line (5.8) in the definition of $\mathcal{F}(A)$.

Now, from property (5.10) in that definition and by the hypothesis $\delta > e^{-c_1 n}$, we may choose $c_1 > 0$ small enough so that

$$D_{\alpha,\gamma}(w_I/\|w_I\|_2) \geqslant 2c^{-2}n^{1/2}/\delta \geqslant 2c^{-1}\|w_I\|_2/\delta.$$

By the definition of the least common denominator, for $|\theta| \leqslant 1/\delta$, we have

$$\min_{r\in[1,c^{-1}]} \|2r\theta w_I\|_{\mathbb{T}} = \min_{r\in[1,c^{-1}]} \left\|2r\theta\|w_I\|_2 \cdot \frac{w_I}{\|w_I\|_2}\right\|_{\mathbb{T}} \geqslant \min\left\{\gamma\theta\|w_I\|_2, \sqrt{\alpha|I|}\right\}. \tag{5.15}$$

So, for $|\theta| \leqslant 1/\delta$, we use (5.15) in (5.14) to bound the right-hand side of (5.12) as

$$\mathbb{E}^{\mathcal{F}}_{J,X_J,X'_J} e^{\langle (X+X')_J, u\rangle} \cdot \left|\mathbb{E}_{X_I}\, e^{4\pi i\theta\langle w, X_I\rangle + 2\langle X_I, u\rangle}\right| \lesssim \mathbb{E}^{\mathcal{F}}_{J,X_J,X'_J}\, e^{\langle (X+X')_J, u\rangle} e^{-c\min\{\gamma^2\theta^2\|w_I\|_2^2, \alpha|I|\}}. \tag{5.16}$$

We now use that $(J, X_J, X'_J) \in \mathcal{F}$ to see that $w \in \text{Incomp}(\delta, \rho)$ and that we chose μ to be sufficiently small, compared to ρ, δ, to guarantee that

$$\|w\|_2 \leqslant C\|w_I\|_2,$$

for some $C > 0$ (see (4.7)). Thus, the right-hand side of (5.16) is

$$\lesssim \mathbb{E}^{\mathcal{F}}_{J,X_J,X'_J} e^{\langle (X+X')_J, u\rangle} e^{-c'\theta^2\|w\|_2^2} + e^{-\Omega(n)}\,.$$

Combining this with (5.16), (5.12) obtains the desired bound in the case $(J, X_J, X'_J) \in \mathcal{F}$. Combining this with (5.13) completes the proof of Lemma 5.2. □

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

6. Preparation for the "base step" of the iteration

As we mentioned at (2.1), Vershynin [46] gave a natural way of bounding the least singular value of a random symmetric matrix:

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \lesssim \sup_{r\in\mathbb{R}} \mathbb{P}_{A_n,X}\left(|\langle A_n^{-1}X, X\rangle - r| \leqslant \varepsilon \|A_n^{-1}X\|_2\right),$$

where we recall that A_n is obtained from A_{n+1} by deleting its first row and column. The main goal of this section is to prove the following lemma which tells us that we may intersect with the event $\sigma_{\min}(A_n) \geqslant \varepsilon n^{-1/2}$ in the probability on the right-hand side, at a loss of $C\varepsilon$. This will be crucial for the base step in our iteration, since the bound we obtain on $\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2})$ deteriorates as $\sigma_{\min}(A_n)$ decreases.

Lemma 6.1. *For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$ and let $X \sim \mathrm{Col}_n(\zeta)$. Then*

$$\mathbb{P}\left(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}\right) \lesssim \varepsilon + \sup_{r\in\mathbb{R}} \mathbb{P}\left(\frac{|\langle A_n^{-1}X, X\rangle - r|}{\|A_n^{-1}X\|_2} \leqslant C\varepsilon, \sigma_{\min}(A_n) \geqslant \varepsilon n^{-1/2}\right) + e^{-\Omega(n)},$$

for all $\varepsilon > 0$. Here, $C > 0$ depends only on B.

We deduce Lemma 6.1 from a geometric form of the lemma, which we state here. Let X_j denote the jth column of A_{n+1}, and let

$$H_j = \mathrm{Span}\{X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_{n+1}\} \text{ and } d_j(A_{n+1}) := \mathrm{dist}(X_j, H_j).$$

We shall prove the following "geometric" version of Lemma 6.1.

Lemma 6.2. *For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$. Then for all $\varepsilon > 0$,*

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \lesssim \varepsilon + \mathbb{P}\left(d_1(A_{n+1}) \leqslant C\varepsilon \text{ and } \sigma_{\min}(A_n) \geqslant \varepsilon n^{-1/2}\right) + e^{-\Omega(n)},$$

where $C > 0$ depends only on B.

The deduction of Lemma 6.1 from Lemma 6.2 is straightforward given the ideas from [46]; so we turn to discuss the proof of Lemma 6.2.

For this, we want to intersect the event $\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}$ with the event $\sigma_{\min}(A_n) \geqslant \varepsilon n^{-1/2}$, where we understand A_n to be the principal minor $A_{n+1}^{(n+1)}$ of A_{n+1}. To do this, we first consider the related "pathological" event

$$\mathcal{P} := \left\{\sigma_{\min}(A_{n+1}^{(i)}) \leqslant \varepsilon n^{-1/2} \text{ for at least } cn \text{ values of } i \in [n+1]\right\}$$

and then split our probability of interest into the sum

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \cap \mathcal{P}) + \mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \cap \mathcal{P}^c), \tag{6.1}$$

and work with each term separately. Here, $c = c_{\rho,\delta}/2$, where $c_{\rho,\delta}$ is the constant defined in Section 4.

We deal with the second term on the right-hand side by showing

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \cap \mathcal{P}^c) \leqslant \mathbb{P}(d_1(A_{n+1}) \lesssim \varepsilon \text{ and } \sigma_{\min}(A_n) \geqslant \varepsilon n^{-1/2}) + e^{-\Omega(n)}, \tag{6.2}$$

by a straightforward argument in a manner similar to Rudelson and Vershynin in [31]. We then deal with the first term on the right-hand side of (6.1) by showing that

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \cap \mathcal{P}) \lesssim \varepsilon + e^{-\Omega(n)}. \tag{6.3}$$

Putting these two inequalities together then implies Lemma 6.2.

6.1. Proof of the inequality at (6.2)

Here, we prove (6.2) in the following form.

Lemma 6.3. *For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$. Then, for all $\varepsilon > 0$, we have*

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \cap \mathcal{P}^c) \lesssim \mathbb{P}\big(d_1(A_{n+1}) \lesssim \varepsilon \text{ and } \sigma_{\min}(A_n) \geqslant \varepsilon n^{-1/2}\big) + e^{-\Omega(n)}.$$

For this, we use a basic but important fact which is at the heart of the geometric approach of Rudelson and Vershynin (see, e.g. [31, Lemma 3.5]).

Fact 6.4. Let M be an $n \times n$ matrix and v be a unit vector satisfying $\|Mv\|_2 = \sigma_{\min}(M)$. Then

$$\sigma_{\min}(M) \geqslant |v_j| \cdot d_j(M) \qquad \text{for each } j \in [n]\,.$$

We are now ready to prove the inequality mentioned at (6.2).

Proof of Lemma 6.3. We rule out another pathological event: Let v denote a unit eigenvector corresponding to the least singular value of A_{n+1}, and let $\mathcal{C}$ denote the event that v is (ρ, δ)-compressible.[7] By Lemma 3.2, $\mathbb{P}(\mathcal{C}) \leqslant e^{-\Omega(n)}$. Thus

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{P}^c) \leqslant \mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{C}^c \cap \mathcal{P}^c) + e^{-\Omega(n)}. \tag{6.4}$$

We now look to bound this event in terms of the distance of the column X_j to the subspace H_j, in the style of [31]. For this, we define

$$S := \{j : d_j(A_{n+1}) \leqslant \varepsilon/c_{\rho,\delta} \text{ and } \sigma_{\min}(A^{(j)}_{n+1}) \geqslant \varepsilon n^{-1/2}\}.$$

We now claim

$$\{\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}\} \cap \mathcal{C}^c \cap \mathcal{P}^c \Longrightarrow |S| \geqslant c_{\rho,\delta} n/2. \tag{6.5}$$

To see this, fix a matrix A satisfying the left-hand side of (6.5) and let v be an eigenvector corresponding to the least singular value. Now, since v is not compressible, there are $\geqslant c_{\rho,\delta} n$ values of $j \in [n+1]$ for which $|v_j| \geqslant c_{\rho,\delta} n^{-1/2}$. Thus, Fact 6.4 immediately tells us there are $\geqslant c_{\rho,\delta} n$ values of $j \in [n+1]$ for which $d_j(A) \leqslant \varepsilon/c_{\rho,\delta}$. Finally, by definition of $\mathcal{P}^c$, at most $c_{\rho,\delta} n/2$ of these values of j satisfy $\sigma_{n+1}(A^{(j)}) \leqslant \varepsilon n^{-1/2}$, and so (6.5) is proved.

We now use (6.5) along with Markov's inequality to bound

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{C}^c \cap \mathcal{P}^c) \leqslant \mathbb{P}(|S| \geqslant c_{\rho,\delta} n/2) \leqslant \frac{2}{c_{\rho,\delta} n}\mathbb{E}|S|. \tag{6.6}$$

Now, by definition of S and symmetry of the coordinates, we have

$$\begin{aligned}\mathbb{E}|S| &= \sum_j \mathbb{P}\big(d_j(A_{n+1}) \leqslant \varepsilon/c_{\rho,\delta},\ \sigma_{\min}(A^{(j)}_{n+1}) \geqslant \varepsilon n^{-1/2}\big)\\ &= n \cdot \mathbb{P}\big(d_1(A_{n+1}) \leqslant \varepsilon/c_{\rho,\delta},\ \sigma_{\min}(A^{(1)}_{n+1}) \geqslant \varepsilon n^{-1/2}\big)\,.\end{aligned}$$

Putting this together with (6.6) and (6.5) finishes the proof. □

6.2. Proof of the inequality at (6.3)

We now prove the inequality discussed at (6.3) in the following form.

[7]See Section 3 for definition and Section 4 for choice of δ, ρ.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Lemma 6.5. *For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$. Then, for all $\varepsilon > 0$, we have*

$$\mathbb{P}\left(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{P}\right) \lesssim \varepsilon + e^{-\Omega(n)}\,. \tag{6.7}$$

For the proof of this lemma, we will need a few results from the random matrix literature. The first such result is a more sophisticated version of Lemma 3.2, which tells us that the mass of the eigenvectors of A does not "localize" on a set of coordinates of size $o(n)$. The theorem we need, due to Rudelson and Vershynin (Theorem 1.5 in [35]), tells us that the mass of the eigenvectors of our random matrix does not "localize" on a set of coordinates of size $(1-c)n$, for any fixed $c > 0$. We state this result in a way to match our application.

Theorem 6.6. *For $B > 0$, $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$ and let v denote the unit eigenvector of A corresponding to the least singular value of A. Then there exists $c_2 > 0$, such that for all sufficiently small $c_1 > 0$, we have*

$$\mathbb{P}\big(\,|v_j| \geqslant (c_2c_1)^6 n^{-1/2} \text{ for at least } (1-c_1)n \text{ values of } j\big) \geqslant 1 - e^{-c_1 n}\,,$$

for n sufficiently large.

We also require an elementary, but extremely useful, fact from linear algebra. This fact is a key step in the work of Nguyen et al. on eigenvalue repulsion in random matrices (see [24, Section 4]); we state it here in a form best suited for our application.

Fact 6.7. Let M be an $n \times n$ real symmetric matrix, and let λ be an eigenvalue of M with corresponding unit eigenvector u. Let $j \in [n]$, and let λ' be an eigenvector of the minor $M^{(j)}$ with corresponding unit eigenvector v. Then

$$|\langle v, X^{(j)}\rangle| \leqslant |\lambda - \lambda'|/|u_j|,$$

where $X^{(j)}$ is the jth column of M with the jth entry removed.

Proof. Without loss of generality, take $j = n$ and express $u = (w, u_n)$, where $w \in \mathbb{R}^{n-1}$. Then we have $(M^{(n)} - \lambda I)w + X^{(n)}u_n = 0$. Multiplying on the left by v^T yields

$$|u_n\langle v, X^{(n)}\rangle| = |\lambda - \lambda'||\langle v, w\rangle| \leqslant |\lambda - \lambda'|\,. \qquad \square$$

We shall also need the inverse Littlewood-Offord theorem of Rudelson and Vershynin [31], which we have stated here in simplified form. Recall that $D_{\alpha,\gamma}(v)$ is the least common denominator of the vector v, as defined at (1.11).

Theorem 6.8. *For $n \in \mathbb{N}$, $B > 0$, $\gamma, \alpha \in (0,1)$, and $\varepsilon > 0$, let $v \in \mathbb{S}^{n-1}$ satisfy $D_{\alpha,\gamma}(v) > c\varepsilon^{-1}$ and let $X \sim \mathrm{Col}_n(\zeta)$, where $\zeta \in \Gamma_B$. Then*

$$\mathbb{P}(|\langle X, v\rangle| \leqslant \varepsilon) \lesssim \varepsilon + e^{-c\alpha n}\,.$$

Here, $c > 0$ depends only on B and γ.

We are now in a position to prove Lemma 6.5.

Proof of Lemma 6.5. Let A be an instance of our random matrix, and let v be the unit eigenvector corresponding to the least singular value of A. Let $w_j = w(A^{(j)})$ denote a unit eigenvector of $A^{(j)}$ corresponding to the least singular value of $A^{(j)}$.

We introduce two "quasi-randomness" events $\mathcal{Q}$ and $\mathcal{A}$ that will hold with probability $1 - e^{\Omega(n)}$. Indeed, define

$$\mathcal{Q}_j = \{D_{\alpha,\gamma}(w_j) \geqslant e^{c_3 n}\} \text{ for all } j \in [n+1] \text{ and set } \mathcal{Q} = \bigcap \mathcal{Q}_j\,.$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Here, α, γ, c_3 are chosen according to Lemma 4.1, which tells us that $\mathbb{P}(\mathcal{Q}^c) \leqslant e^{-\Omega(n)}$. Define

$$S_1 = \{j : \sigma_n(A_{n+1}^{(j)}) \leqslant \varepsilon n^{-1/2}\} \text{ and } S_2 = \{j : |v_j| \geqslant (cc_2/2)^6 n^{-1/2}\}.$$

Note that $\mathcal{P}$ holds exactly when $|S_1| \geqslant cn$. Let $\mathcal{A}$ be the "non-localization" event that $|S_2| \geqslant (1 - c/2)n$. By Theorem 6.6, we have $\mathbb{P}(\mathcal{A}^c) \leqslant e^{-\Omega(n)}$. Here, $c/2 = c_{\rho,\delta}/4$. Now, if we let $X^{(j)}$ denote the jth column of A with the jth entry removed, we define

$$T = \{j : |\langle w_j, X^{(j)}\rangle| \leqslant C\varepsilon\},$$

where $C = 2^7/(c_2c)^6$. We now claim

$$\{\sigma_{\min}(A) \leqslant \varepsilon n^{-1/2}\} \cap \mathcal{P} \cap \mathcal{A} \Longrightarrow |T| \geqslant cn/2. \tag{6.8}$$

To see this, first note that if $\mathcal{P} \cap \mathcal{A}$ holds, then $|S_1 \cap S_2| \geqslant cn/2$. Also, for each $j \in S_1 \cap S_2$, we may apply Fact 6.7 to see that $|\langle w_j, X^{(j)}\rangle| \leqslant C\varepsilon$ since j is such that $\sigma_{\min}(A^{(j)}) \leqslant \varepsilon n^{-1/2}$ and $\sigma_{\min}(A) \leqslant \varepsilon n^{-1/2}$. This proves (6.8).

To finish the proof of Lemma 6.5, we define the random variable

$$R = n^{-1} \sum_j \mathbf{1}\left(|\langle w_j, X^{(j)}\rangle| \leqslant C\varepsilon \text{ and } \mathcal{Q}_j\right),$$

and observe that $\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \cap \mathcal{P})$ is at most

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{A} \cap \mathcal{Q} \cap \mathcal{P}) + e^{-\Omega(n)} \leqslant \mathbb{P}(R \geqslant c/4) + e^{-\Omega(n)}.$$

We now apply Markov and expand the definition of R to bound

$$\mathbb{P}(R \geqslant c/4) \lesssim n^{-1} \sum_j \mathbb{E}_{A_{n+1}^{(j)}} \mathbb{P}_{X^{(j)}}\left(|\langle w_j, X^{(j)}\rangle| \leqslant C\varepsilon \cap \mathcal{Q}_j\right) \lesssim \varepsilon + e^{-\Omega(n)},$$

where the last inequality follows from the fact that $X^{(j)}$ is independent of the events Q_j and w_j, and therefore we may put the property $\mathcal{Q}_j$ to use by applying the inverse Littlewood-Offord theorem of Rudelson and Vershynin, Theorem 6.8. □

6.3. *Proofs of Lemmas 6.2 and 6.1*

All that remains is to put the pieces together and prove Lemmas 6.2 and 6.1.

Proof of Lemma 6.2. As we saw at (6.1), we simply express $\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2})$ as

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{P}) + \mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2} \text{ and } \mathcal{P}^c),$$

and then apply Lemma 6.5 to the first term and Lemma 6.3 to the second term. □

Proof of Lemma 6.1. If we set $a_{1,1}$ to be the first entry of $A = A_{n+1}$, then, by [46, Proposition 5.1], we have that

$$d_1(A_{n+1}) = \frac{|\langle A^{-1}X, X\rangle - a_{1,1}|}{\sqrt{1 + \|A^{-1}X\|_2^2}}.$$

Additionally, by [46, Proposition 8.2], we have $\|A^{-1}X\|_2 > 1/15$ with probability at least $1 - e^{-\Omega(n)}$. Replacing $a_{1,1}$ with r and taking a supremum completes the proof of Lemma 6.1. □

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

7. Eigenvalue crowding (and the proofs of Theorems 1.2 and 1.3)

The main purpose of this section is to prove the following theorem, which gives an upper bound on the probability that $k \geqslant 2$ eigenvalues of a random matrix fall in an interval of length ε. The case $\varepsilon = 0$ of this theorem tells us that the probability that a random symmetric matrix has *simple* spectrum (that is, has no repeated eigenvalue) is $1 - e^{-\Omega(n)}$, which is sharp and confirms a conjecture of Nguyen et al. [24].

Given an $n \times n$ real symmetric matrix M, we let $\lambda_1(M) \geqslant \ldots \geqslant \lambda_n(M)$ denote its eigenvalues.

Theorem 7.1. *For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$. Then for each $j \leqslant cn$ and all $\varepsilon \geqslant 0$, we have*

$$\max_{k \leqslant n-j} \mathbb{P}(|\lambda_{k+j}(A_n) - \lambda_k(A_n)| \leqslant \varepsilon n^{-1/2}) \leqslant (C\varepsilon)^j + 2e^{-cn},$$

where $C, c > 0$ are constants depending on B.

We suspect that the bound in Lemma 1.3 is actually *far* from the truth, for $\varepsilon > e^{-cn}$ and $j \geqslant 1$. In fact, one expects *quadratic* dependence on j in the exponent of ε. This type of dependence was recently confirmed by Nguyen [27] for $\varepsilon > e^{-n^c}$.

For the proof of Lemma 1.3, we remind the reader that if $u \in \mathbb{R}^n \cap \mathrm{Incomp}\,(\rho, \delta)$, then at least $c_{\rho,\delta} n$ coordinates of u have absolute value at least $c_{\rho,\delta} n^{-1/2}$.

In what follows, for an $n \times n$ symmetric matrix A, we use the notation $A^{(i_1,\ldots,i_r)}$ to refer to the minor of A for which the rows and columns indexed by $i_1, \ldots, i_r$ have been deleted. We also use the notation $A_{S \times T}$ to refer to the $|S| \times |T|$ submatrix of A defined by $(A_{i,j})_{i \in S, j \in T}$.

The following fact contains the key linear algebra required for the proof of Theorem 1.3.

Fact 7.2. For $1 \leqslant k + j < n$, let A be an $n \times n$ symmetric matrix for which

$$|\lambda_{k+j}(A) - \lambda_k(A)| \leqslant \varepsilon n^{-1/2}.$$

Let $(i_1, \ldots, i_j) \in [n]^j$ be such that $i_1, \ldots, i_j$ are distinct. Then there exist unit vectors $w^{(1)}, \ldots, w^{(k)}$ for which

$$\langle w^{(r)}, X_r \rangle \leqslant (\varepsilon n^{-1/2}) \cdot (1/|w^{(r-1)}_{i_r}|),$$

where $X_r \in \mathbb{R}^{n-r}$ is the i_rth column of A with coordinates indexed by $i_1, \ldots, i_r$ removed. That is, $X_r := A_{[n]\setminus\{i_1,\ldots,i_r\} \times \{i_r\}}$ and $w^{(r)}$ is a unit eigenvector corresponding to $\lambda_k(A^{(i_1,\ldots,i_r)})$.

Proof. For $(i_1, \ldots, i_j) \in [n]^j$, define the matrices $M_0, M_1, \ldots, M_j$ by setting $M_r = A^{(i_1,\ldots,i_r)}$ for $r = 1, \ldots, j$ and then $M_0 := A$. Now if

$$|\lambda_{k+j}(A) - \lambda_k(A)| \leqslant \varepsilon n^{-1/2},$$

then Cauchy's interlacing theorem implies

$$|\lambda_k(M_r) - \lambda_k(M_{r-1})| \leqslant \varepsilon n^{-1/2},$$

for all $r = 1, \ldots, j$. So let $w^{(r)}$ denote a unit eigenvector of M_r corresponding to eigenvalue $\lambda_k(M_r)$. Thus, by Fact 6.7, we see that

$$|\langle w^{(r)}, X_r \rangle| \leqslant (\varepsilon n^{-1/2}) \cdot (1/|w^{(r-1)}_{i_r}|),$$

for $r = 1, \ldots, j$, where $X_r \in \mathbb{R}^{n-r}$ is the i_rth column of M_{r-1}, with the diagonal entry removed. In other words, $X_r \in \mathbb{R}^{n-r}$ is the i_rth column of A with coordinates indexed by $i_1, \ldots, i_r$ removed. This completes the proof of Fact 7.2. □

Proof of Theorem 1.3. Note, we may assume that $\varepsilon > e^{-cn}$; the general case follows by taking c sufficiently small. Now, define $\mathcal{A}$ to be the event that *all* unit eigenvectors v of *all* $\binom{n}{j}$ of the minors

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

$A_n^{(i_1,\ldots,i_j)}$ lie in Incomp(ρ,δ) and satisfy $D_{\alpha,\gamma}(v) > e^{c_3 n}$, where α,γ,c_3 are chosen according to Lemma 4.1. Note that by Lemmas 4.1 and 3.2, we have

$$\mathbb{P}(\mathcal{A}^c) \leqslant \binom{n}{j+1} e^{-\Omega(n)} \leqslant n\left(\frac{en}{j}\right)^j e^{-\Omega(n)} \lesssim e^{-cn},$$

by taking c small enough, so that $j\log(en/j) < cn$ is smaller than the $\Omega(n)$ term.

With Fact 7.2 in mind, we define the event, $\mathcal{E}_{i_1,\ldots,i_j}$, for each $(i_1,\ldots,i_j) \in [n]^j$, i_r distinct, to be the event that

$$|\langle w^{(r)}, X_r\rangle| \leqslant \varepsilon/c_{\rho,\delta} \quad \text{for all } r \in [j]\,,$$

where $X_r \in \mathbb{R}^{n-r}$ is the i_rth column of A with coordinates indexed by $i_1,\ldots,i_r$ removed and $w^{(r)}$ is a unit eigenvector corresponding to $\lambda_k(A^{(i_1,\ldots,i_r)})$.

If $\mathcal{A}$ holds, then each $w^{(r)}$ has at least $c_{\rho,\delta}n$ coordinates with absolute value at least $c_{\rho,\delta}n^{-1/2}$. Thus, if additionally we have

$$|\lambda_{k+j}(A_n) - \lambda_k(A_n)| \leqslant \varepsilon n^{-1/2}.$$

Fact 7.2 tells us that $\mathcal{E}_{i_1,\ldots,i_j}$ occurs for at least $(c_{\rho,\delta}n/2)^j$ tuples $(i_1,\ldots,i_j)$.

Define N to be the number of indices $(i_1,\ldots,i_j)$ for which $\mathcal{E}_{i_1,\ldots,i_j}$ occurs, and note

$$\mathbb{P}(\,|\lambda_{k+j}(A_n) - \lambda_k(A_n)| \leqslant \varepsilon n^{-1/2}) \leqslant \mathbb{P}\big(N \geqslant (c_{\rho,\delta}n/2)^j \text{ and } \mathcal{A}\big) + O(e^{-cn}) \tag{7.1}$$

$$\leqslant \left(\frac{2}{c_{\rho,\delta}}\right)^j \mathbb{P}(\mathcal{E}_{1,\ldots,j} \cap \mathcal{A}) + O(e^{-cn}), \tag{7.2}$$

where, for the second inequality, we applied Markov's inequality and used the symmetry of the events $\mathcal{E}_{i_1,\ldots,i_j}$.

Thus, we need only show that there exists $C > 0$, such that $\mathbb{P}(\mathcal{E}_{1,\ldots,j} \cap \mathcal{A}) \leqslant (C\varepsilon)^j$. To use independence, we replace each of $w^{(r)}$ with the worst case vector, under $\mathcal{A}$

$$\mathbb{P}(\mathcal{E}_{1,\ldots,j} \cap \mathcal{A}) \leqslant \max_{w_1,\ldots,w_j : D_{\alpha,\gamma}(w_i) > e^{c_3 n}} \mathbb{P}_{X_1,\ldots,X_r}\big(\,|\langle w_r, X_r\rangle| \leqslant \varepsilon/c_{\rho,\delta} \text{ for all } r \in [j]\,\big) \tag{7.3}$$

$$\leqslant \max_{w_1,\ldots,w_j : D_{\alpha,\gamma}(w_i) > e^{c_3 n}} \prod_{r=1}^{j} \mathbb{P}_{X_r}\big(\,|\langle w_r, X_r\rangle| \leqslant \varepsilon/c_{\rho,\delta}\,\big), \leqslant (C\varepsilon)^j, \tag{7.4}$$

where the first inequality follows from the independence of the vectors $\{X_r\}_{r\leqslant j}$ and the last inequality follows from the fact that $D_{\alpha,\gamma}(w_r) > e^{c_3 n} \gtrsim 1/\varepsilon$ (by choosing $c > 0$ small enough relative to c_3), and the Littlewood-Offord theorem of Rudelson and Vershynin, Lemma 6.8. Putting (7.2) and (7.4) together completes the proof of Theorem 1.3. □

Of course, the proof of Theorem 1.2 follows immediately.

Proof of Theorem 1.2. Simply take $\varepsilon = 0$ in Theorem 1.3. □

8. Properties of the spectrum

In this section, we describe and deduce Lemma 8.1 and Corollary 8.2, which are the tools we will use to control the "bulk" of the eigenvalues of A^{-1}. Here, we understand "bulk" relative to the spectral measure of A^{-1}: our interest in an eigenvalue λ of A^{-1} is proportional to its contribution to $\|A^{-1}\|_{\mathrm{HS}}$. Thus, the behavior of *smallest* singular values of A are of the highest importance for us.

For this, we let $\sigma_n \leqslant \sigma_{n-1} \leqslant \cdots \leqslant \sigma_1$ be the singular values of A and let $\mu_1 \geqslant \ldots \geqslant \mu_n$ be the singular values of A^{-1}. Of course, we have $\mu_k = 1/\sigma_{n-k+1}$ for $1 \leqslant k \leqslant n$.

In short, these two lemmas, when taken together, tell us that

$$\sigma_{n-k+1} \approx kn^{-1/2}, \tag{8.1}$$

for all $n \geqslant k \gg 1$ in some appropriate sense.

Lemma 8.1. *For $p > 1$, $B > 0$ and $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$. There is a constant C_p depending on B, p so that*

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \leqslant C_p\,,$$

for all k.

We shall deduce Lemma 8.1 from the "local semicircular law" of Erdős et al. [13], which gives us good control of the bulk of the spectrum at "scales" of size $\gg n^{-1/2}$.

We also record a useful corollary of this lemma. For this, we define the function $\|\cdot\|_*$ for an $n \times n$ symmetric matrix M to be

$$\|M\|_*^2 = \sum_{k=1}^{n} \sigma_k(M)^2(\log(1+k))^2. \tag{8.2}$$

The point of this definition is to give some measure to how the spectrum of A^{-1} is "distorted" from what it "should be," according to the heuristic at (8.1). Indeed, if we have $\sigma_{n-k+1} = \Theta(k/\sqrt{n})$ for all k, say, then we have that

$$\|A^{-1}\|_* = \Theta(\mu_1).$$

Conversely, any deviation from this captures some macroscopic misbehavior on the part of the spectrum. In particular, the "weight function" $k \mapsto (\log(1+k))^2$ is designed to bias the smallest singular values, and thus we are primarily looking at this range for any poor behavior.

Corollary 8.2. *For $p > 1$, $B > 0$, and $\zeta \in \Gamma_B$, let $A \sim \mathrm{Sym}_n(\zeta)$. Then there exists constants $C_p, c_p > 0$ depending on B, p, such that*

$$\mathbb{E}\left[\left(\frac{\|A^{-1}\|_*}{\mu_1}\right)^p\right] \leqslant C_p\,.$$

In the remainder of this section, we describe the results of Erdős et al. [13] and deduce Lemma 8.1. We then deduce Corollary 8.2.

8.1. The local semicircular law and Lemma 8.1

For $a < b$, we define $N_A(a, b)$ to be the number of eigenvalues of A in the interval (a, b). One of the most fundamental results in the theory of random symmetric matrices is the *semicircular law*, which says that

$$\lim_{n\to\infty} \frac{N_A(a\sqrt{n}, b\sqrt{n})}{n} = \frac{1}{2\pi}\int_a^b (4-x^2)_+^{1/2}\,dx,$$

almost surely, where $A \sim \mathrm{Sym}_n(\zeta)$.

We use a powerful "local" version of the semicircle law developed by Erdős et al. in a series of important papers [10, 11, 13]. Their results show that the spectrum of a random symmetric matrix

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

actually adheres surprisingly closely to the semicircular law. In this paper, we need control on the number of eigenvalues in intervals of the form $[-t, t]$, where $1/n^{1/2} \ll t \ll n^{1/2}$. The semicircular law predicts that

$$N_A(-t,t) \approx \frac{n}{2\pi}\int_{-tn^{-1/2}}^{tn^{-1/2}} (4-x^2)_+^{1/2}\,dx = \frac{2tn^{1/2}}{\pi}(1+o(1)).$$

Theorem 1.11 of [12] makes this prediction rigorous.[8]

Theorem 8.3. *Let $B > 0$, $\zeta \in \Gamma_B$, and let $A \sim \mathrm{Sym}_n(\zeta)$. Then, for $t \in [Cn^{-1/2}, n^{1/2}]$,*

$$\mathbb{P}\left(\left|\frac{N_A(-t,t)}{n^{1/2}t} - 2\pi^{-1}\right| > \pi\right) \lesssim \exp\left(-c_1(t^2n)^{1/4}\right), \tag{8.3}$$

where $C, c_1 > 0$ are absolute constants.

Lemma 8.1 follows quickly from Theorem 8.3. In fact, we shall only use two corollaries.

Corollary 8.4. *Let $B > 0$, $\zeta \in \Gamma_B$, and let $A \sim \mathrm{Sym}_n(\zeta)$. Then for all $s \geqslant C$ and $k \in \mathbb{N}$ satisfying $sk \leqslant n$, we have*

$$\mathbb{P}\left(\frac{\sqrt{n}}{\mu_k k} \geqslant s\right) \lesssim \exp\left(-c(sk)^{1/2}\right),$$

where $C, c > 0$ are absolute constants.

Proof. Let C be the maximum of the constant C from Lemma 8.3 and π. If $\frac{\sqrt{n}}{\mu_k k} \geqslant s$, then $N_A(-skn^{-1/2}, skn^{-1/2}) \leqslant k$. We now apply Lemma 8.3 with $t = skn^{-1/2} \geqslant sn^{-1/2} \geqslant Cn^{-1/2}$ to see that this event occurs with probability $\lesssim \exp(-c\sqrt{sk})$. □

An identical argument provides a similar bound in the other direction.

Corollary 8.5. *Let $B > 0$, $\zeta \in \Gamma_B$, and let $A \sim \mathrm{Sym}_n(\zeta)$. Then for all $k \in \mathbb{N}$, we have*

$$\mathbb{P}\left(\mu_k \geqslant \frac{C\sqrt{n}}{k}\right) \lesssim \exp\left(-ck^{1/2}\right),$$

where $C, c > 0$ are absolute constants.

Proof of Lemma 8.1. Let C be the constant from Corollary 8.4. From the standard tail estimates on $\|A\|_{op}$, like (4.11) for example, we immediately see that for all $k \geqslant n/C$, we have

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \leqslant \mathbb{E}_A\left(\frac{\sigma_1(A)\sqrt{n}}{k}\right)^p = O_p((n/k)^p) = O_p(1).$$

Thus, we can restrict our attention to the case when $k \leqslant n/C$. Define the events

$$E_1 = \left\{\frac{\sqrt{n}}{\mu_k k} \leqslant C\right\}, \quad E_2 = \left\{\frac{\sqrt{n}}{\mu_k k} \in [C, n/k]\right\}, \quad E_3 = \left\{\frac{\sqrt{n}}{\mu_k k} \geqslant \frac{n}{k}\right\}.$$

We may bound

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \leqslant C^p + \mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbf{1}_{E_2} + \mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbf{1}_{E_3}. \tag{8.4}$$

[8] Theorem 1.11 of the survey [12] is based on Corollary 3.2 in [13]. In the paper [13], the results are technically stated for (complex) Hermitian random matrices. However, the same proof goes through for real symmetric matrices. This is why we cite the later survey [12], where this more general version is stated.

To deal with the second term in (8.4), we use Corollary 8.4 to see that

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbf{1}_{E_2} \lesssim \int_C^{n/k} ps^{p-1}e^{-c\sqrt{sk}}ds = O_p(1).$$

To deal with the third term in (8.4), we note that since $n/k \geqslant C$, we may apply Corollary 8.4, with $s = n/k$, to conclude that $\mathbb{P}(E_3) \lesssim e^{-c\sqrt{n}}$. Thus, by Cauchy-Schwarz, we have

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbf{1}_{E_3} \leqslant \left(\mathbb{E}\left(\frac{\sigma_1\sqrt{n}}{k}\right)^{2p}\right)^{1/2} \mathbb{P}(E_3)^{1/2} \leqslant O_p(1)\cdot n^p e^{-c\sqrt{n}} = O_p(1),$$

where we have used the upper tail estimate in σ_1 from (4.11) to see $\mathbb{E}\,\sigma_1^{2p} = O_p(n^p)$. □

8.2. *Deduction of Corollary 8.2*

We now conclude this section by deducing Corollary 8.2 from Lemma 8.1 and Corollary 8.5.

Proof of Corollary 8.2. Recall

$$\|A^{-1}\|_*^2 = \sum_{k=1}^n \mu_k^2(\log(1+k))^2.$$

By Hölder's inequality, we may assume without loss of generality that $p \geqslant 2$. Applying the triangle inequality for the $L^{p/2}$ norm gives

$$\left[\mathbb{E}\left(\sum_{k=1}^n \frac{\mu_k^2(\log(1+k))^2}{\mu_1^2}\right)^{p/2}\right]^{2/p} \leqslant \sum_{k=1}^n (\log(1+k))^2 \mathbb{E}\left[\frac{\mu_k^p}{\mu_1^p}\right]^{2/p}.$$

Taking C to be the constant from Corollary 8.5 bound

$$\mathbb{E}\left[\frac{\mu_k^p}{\mu_1^p}\right] \leqslant C^p k^{-p}\mathbb{E}\left[\left(\frac{\sqrt{n}}{\mu_1}\right)^p\right] + \mathbb{P}\left(\mu_k \geqslant C\frac{\sqrt{n}}{k}\right) \lesssim C^p k^{-p},$$

where we used Lemma 8.1 and Corollary 8.5 for the second inequality. Combining the previous two equations completes the proof. □

9. Controlling small balls and large deviations

The goal of this section is to prove the following lemma, which will be a main ingredient in our iteration in Section 10. We shall then use it again in the final step and proof of Theorem 1.1, in Section 11.

Lemma 9.1. *For $B > 0$ and $\zeta \in \Gamma_B$, let $A = A_n \sim \mathrm{Sym}_n(\zeta)$ and let $X \sim \mathrm{Col}_n(\zeta)$. Let $u \in \mathbb{R}^{n-1}$ be a random vector with $\|u\|_2 \leqslant 1$ that depends only on A. Then, for $\delta, \varepsilon > e^{-cn}$ and $s \geqslant 0$, we have*

$$\mathbb{E}_A \sup_r \mathbb{P}_X\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_*} \leqslant \delta,\ \langle X, u\rangle \geqslant s,\ \frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1}\right)$$
$$\lesssim \delta e^{-s}\left[\mathbb{E}_A\left(\frac{\mu_1}{\sqrt{n}}\right)^{7/9}\mathbf{1}\left\{\frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\}\right]^{6/7} + e^{-cn}, \tag{9.1}$$

where $c > 0$ depends only on $B > 0$.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Note that with this lemma, we have eliminated all "fine-grained" information about the spectrum of A^{-1} and all that remains is μ_1, which is the reciprocal of the least singular value of the matrix A. We also note that we will only need the full power of Lemma 9.1 in Section 11; until then, we will apply it with $s = 0, u = 0$.

We now turn our attention to proving Lemma 9.1. We start with an application of Theorem 1.5, our negative correlation theorem, which we restate here in its full-fledged form.

Theorem 9.2. *For $n \in \mathbb{N}$, $\alpha, \gamma \in (0,1), B > 0$, and $\mu \in (0, 2^{-15})$, there are constants $c, R > 0$ depending only on α, γ, μ, B so that the following holds. Let $0 \leqslant k \leqslant c\alpha n$ and $\varepsilon \geqslant \exp(-c\alpha n)$, let $v \in \mathbb{S}^{n-1}$, and let $w_1, \ldots, w_k \in \mathbb{S}^{n-1}$ be orthogonal. For $\zeta \in \Gamma_B$, let ζ' be an independent copy of ζ and Z_μ a Bernoulli variable with parameter μ; let $\widetilde{X} \in \mathbb{R}^n$ be a random vector whose coordinates are i.i.d. copies of the random variable $(\zeta - \zeta')Z_\mu$.*

If $D_{\alpha,\gamma}(v) > 1/\varepsilon$, then

$$\mathbb{P}_X\left(|\langle \widetilde{X}, v\rangle| \leqslant \varepsilon \text{ and } \sum_{j=1}^{k} \langle w_j, \widetilde{X}\rangle^2 \leqslant ck\right) \leqslant R\varepsilon \cdot e^{-ck}\,. \tag{9.2}$$

The proof of Theorem 9.2 is provided in the Appendix. We now prove Lemma 9.3.

Lemma 9.3. *Let A be an $n \times n$ real symmetric matrix with $A \in \mathcal{E}$, and set $\mu_i := \sigma_i(A^{-1})$, for all $i \in [n]$. For $B > 0$, $\zeta \in \Gamma_B$, let $X, X' \sim \mathrm{Col}_n(\zeta)$ be independent, let $J \subseteq [n]$ be a μ-random subset with $\mu \in (0, 2^{-15})$, and set $\widetilde{X} := (X - X')_J$. If $k \in [1, cn]$ is such that $s \in (e^{-cn}, \mu_k/\mu_1)$, then*

$$\mathbb{P}_{\widetilde{X}}\left(\|A^{-1}\widetilde{X}\|_2 \leqslant s\mu_1\right) \lesssim se^{-ck}\,, \tag{9.3}$$

where $c > 0$ depends only on B.

Proof. For each $j \in [n]$, we let v_j denote a unit eigenvector of A^{-1} corresponding to μ_j. Using the resulting singular value decomposition of A^{-1}, we may express

$$\|A^{-1}\widetilde{X}\|_2^2 = \langle A^{-1}\widetilde{X}, A^{-1}\widetilde{X}\rangle = \sum_{j=1}^{n} \mu_j^2 \langle \widetilde{X}, v_j\rangle^2,$$

and thus

$$\mathbb{P}_{\widetilde{X}}\left(\|A^{-1}\widetilde{X}\|_2\mu_1^{-1} \leqslant s\right) \leqslant \mathbb{P}_{\widetilde{X}}\left(|\langle v_1, \widetilde{X}\rangle| \leqslant s \text{ and } \sum_{j=2}^{k} \frac{\mu_j^2}{\mu_1^2}\langle v_j, \widetilde{X}\rangle^2 \leqslant s^2\right). \tag{9.4}$$

We now use that $s \leqslant 1$ and $\mu_k/\mu_1 \leqslant 1$ in (9.4) to obtain

$$\mathbb{P}_{\widetilde{X}}\left(\|A^{-1}\widetilde{X}\|_2\mu_1^{-1} \leqslant s\right) \leqslant \mathbb{P}_{\widetilde{X}}\left(|\langle v_1, \widetilde{X}\rangle| \leqslant s \text{ and } \sum_{j=2}^{k} \langle v_j, \widetilde{X}\rangle^2 \leqslant 1\right). \tag{9.5}$$

We now carefully observe that we are in a position to apply Theorem 1.5 to the right-hand side of (9.5). The coordinates of $\widetilde{X}$ are of the form $(\zeta - \zeta')Z_\mu$, where Z_μ is a Bernoulli random variable taking 1 with probability $\mu \in (0, 2^{-15})$ and 0 otherwise. Also, the $v_2, \ldots, v_k$ are orthogonal and, importantly, we use that $A \in \mathcal{E}$ to learn that[9] $D_{\alpha,\gamma}(v_1) > 1/s$ by property (4.3), provided we choose the constant

[9]Recall, here, that the constants $\alpha, \gamma > 0$ are implicit in the definition of $\mathcal{E}$ and are chosen so that Lemma 4.1 holds.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

$c > 0$ (in the statement of Lemma 9.3) to be sufficiently small, depending on μ, B. Thus, we may apply Theorem 1.5 and complete the proof of the Lemma 9.3. $\square$

With this lemma in hand, we establish the following corollary of Lemma 5.2.

Lemma 9.4. *For $B > 0$ and $\zeta \in \Gamma_B$, let $X \sim \mathrm{Col}_n(\zeta)$ and let A be an $n \times n$ real symmetric matrix with $A \in \mathcal{E}$. If $s > 0$, $\delta \in (e^{-cn}, 1)$ and $u \in \mathbb{S}^{n-1}$, then*

$$\sup_r \mathbb{P}_X\left(\left|\langle A^{-1}X, X\rangle - r\right| \leqslant \delta\mu_1, \langle X, u\rangle \geqslant s\right) \lesssim \delta e^{-s} \sum_{k=2}^{cn} e^{-ck} \left(\frac{\mu_1}{\mu_k}\right)^{2/3} + e^{-cn}, \tag{9.6}$$

where $c > 0$ is a constant depending only on B.

Proof. We apply Lemma 5.2 to the left-hand side of (9.6) to get

$$\sup_r \mathbb{P}_X\left(\left|\langle A^{-1}X, X\rangle - r\right| \leqslant \delta\mu_1, \langle X, u\rangle \geqslant s\right) \lesssim \delta e^{-s} \int_{-1/\delta}^{1/\delta} I(\theta)^{1/2}\, d\theta + e^{-\Omega(n)}, \tag{9.7}$$

where

$$I(\theta) := \mathbb{E}_{J, X_J, X'_J} \exp\left(\langle (X + X')_J, u\rangle - c'\theta^2\mu_1^{-2}\|A^{-1}(X - X')_J\|_2^2\right),$$

and $c' = c'(B) > 0$ is a constant depending only on B and $J \subseteq [n]$ is a μ-random subset. Set

$$\widetilde{X} = (X - X')_J \qquad \text{and} \qquad v = A^{-1}\widetilde{X},$$

and apply Hölder's inequality

$$I(\theta) = \mathbb{E}_{J, X_J, X'_J}\left[e^{\langle (X+X')_J, u\rangle} e^{-c'\theta^2\|v\|_2^2/\mu_1^2}\right] \lesssim \left(\mathbb{E}_{\widetilde{X}} e^{-c''\theta^2\|v\|_2^2/\mu_1^2}\right)^{8/9} \left(\mathbb{E}_{J, X_J, X'_J} e^{9\langle (X+X')_J, u\rangle}\right)^{1/9}. \tag{9.8}$$

Thus, we apply (5.2) to see that the second term on the right-hand side of (9.8) is $O(1)$. Thus, for each $\theta > 0$, we have

$$I(\theta)^{9/8} \lesssim_B \mathbb{E}_{\widetilde{X}} e^{-c''\theta^2\|v\|_2^2/\mu_1^2} \leqslant e^{-c''\theta^{1/5}} + \mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1\theta^{-9/10}).$$

As a result, we have

$$\int_{-1/\delta}^{1/\delta} I(\theta)^{1/2}\, d\theta \lesssim 1 + \int_1^{1/\delta} \mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1\theta^{-9/10})^{4/9}\, d\theta \lesssim 1 + \int_\delta^1 s^{-19/9}\mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1 s)^{4/9}\, ds.$$

To bound this integral, we partition $[\delta, 1] = [\delta, \mu_{cn}/\mu_1] \cup \bigcup_{k=2}^{cn} [\mu_k/\mu_1, \mu_{k-1}/\mu_1]$ and apply Lemma 9.3 to bound the integrand depending on which interval s lies in. Note, this lemma is applicable since $A \in \mathcal{E}$. We obtain

$$\int_{\mu_k/\mu_1}^{\mu_{k-1}/\mu_1} s^{-19/9}\mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1 s)^{4/9} \leqslant e^{-ck} \int_{\mu_k/\mu_1}^{\mu_{k-1}/\mu_1} s^{-15/9}\, ds \leqslant e^{-ck}(\mu_1/\mu_k)^{2/3},$$

while

$$\int_\delta^{\mu_{cn}/\mu_1} s^{-19/9}\mathbb{P}_{\widetilde{X}}(\|v\|_2 \leqslant \mu_1 s)^{4/9} \leqslant e^{-cn}\delta^{-3/2} \leqslant e^{-\Omega(n)}.$$

Summing over all k and plugging the result into (9.7) completes the lemma. $\square$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

We may now prove Lemma 9.1 by using the previous Lemma 9.4 along with the properties of the spectrum of A established in Section 8.

Proof of Lemma 9.1. Let $\mathcal{E}$ be our quasi-random event as defined in Section 4, and let

$$\mathcal{E}_0 = \mathcal{E} \cap \left\{ \frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1} \right\}.$$

For fixed $A \in \mathcal{E}_0$ and $u = u(A) \in \mathbb{R}^n$ with $\|u\|_2 \leqslant 1$, we may apply Lemma 9.4 with $\delta' = \delta \frac{\|A^{-1}\|_*}{\mu_1}$ to see that

$$\sup_{r \in \mathbb{R}} \mathbb{P}_X\left(\left|\langle A^{-1}X, X\rangle - r\right| \leqslant \delta\|A\|_*, \langle X, u\rangle \geqslant s \right) \lesssim \delta e^{-s} \left(\frac{\|A^{-1}\|_*}{\mu_1} \right) \sum_{k=2}^{cn} e^{-ck} \left(\frac{\mu_1}{\mu_k} \right)^{2/3} + e^{-cn}.$$

By Lemma 4.1, $\mathbb{P}_A(\mathcal{E}^c) \lesssim \exp(-\Omega(n))$. Therefore, it is enough to show that

$$\mathbb{E}_A^{\mathcal{E}_0} \left(\frac{\|A^{-1}\|_*}{\mu_1} \right) \left(\frac{\mu_1}{\mu_k} \right)^{2/3} \lesssim k \cdot \mathbb{E}_A^{\mathcal{E}_0} \left[\left(\frac{\mu_1}{\sqrt{n}} \right)^{7/9} \right]^{6/7}, \tag{9.9}$$

for each $k \in [2, cn]$. For this, apply Hölder's inequality to the left-hand side of (9.9) to get

$$\mathbb{E}_A^{\mathcal{E}_0} \left(\frac{\|A^{-1}\|_*}{\mu_1} \right) \left(\frac{\mu_1}{\mu_k} \right)^{2/3} \leqslant \mathbb{E}_A^{\mathcal{E}_0} \left[\left(\frac{\|A^{-1}\|_*}{\mu_1} \right)^{14} \right]^{1/14} \mathbb{E}_A^{\mathcal{E}_0} \left[\left(\frac{\sqrt{n}}{\mu_k} \right)^{28/3} \right]^{1/14} \mathbb{E}_A^{\mathcal{E}_0} \left[\left(\frac{\mu_1}{\sqrt{n}} \right)^{7/9} \right]^{6/7}.$$

We now apply Corollary 8.2 to see the first term is $O(1)$ and Lemma 8.1 to see that the second term is $O(k)$. This establishes (9.9) and thus Lemma 9.1. □

10. Intermediate bounds: Bootstrapping the lower tail

In this short section, we will use the tools developed so far to prove an "up-to-logarithms" version of Theorem 1.1. In the next section, Section 11, we will bootstrap this result (once again) to prove Theorem 1.1.

Lemma 10.1. *For $B > 0$, let $\zeta \in \Gamma_B$, and let $A_n \sim \mathrm{Sym}_n(\zeta)$. Then for all $\varepsilon > 0$*

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \lesssim \varepsilon \cdot (\log \varepsilon^{-1})^{1/2} + e^{-\Omega(n)}.$$

To prove Lemma 10.1, we first prove the following "base step" (Lemma 10.3), which we then improve upon in three increments, ultimately arriving at Lemma 10.1.

The "base step" is an easy consequence of Lemmas 6.2 and 9.1 and actually already improves upon the best known bounds on the least-singular value problem for random symmetric matrices. For this, we will need the well-known theorem due to Hanson and Wright [18, 51]. See [47, Theorem 6.2.1]) for a modern exposition.

Theorem 10.2 (Hanson-Wright). *For $B > 0$, let $\zeta \in \Gamma_B$, let $X \sim \mathrm{Col}_n(\zeta)$, and let M be an $m \times n$ matrix. Then for any $t \geqslant 0$, we have*

$$\mathbb{P}_X\left(\left|\|MX\|_2 - \|M\|_{\mathrm{HS}}\right| > t \right) \leqslant 2\exp\left(-\frac{ct^2}{B^4\|M\|^2} \right),$$

where $c > 0$ is absolute constant.

We now prove the base step of our iteration.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Lemma 10.3 (Base step). *For $B > 0$, let $\zeta \in \Gamma_B$ and let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$. Then*

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \lesssim \varepsilon^{1/4} + e^{-\Omega(n)}\,,$$

for all $\varepsilon > 0$.

Proof. As usual, we let $A := A_n$. By Lemma 6.1, it will be sufficient to show that for $r \in \mathbb{R}$,

$$\mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}X\|_2} \leqslant C\varepsilon,\ \sigma_n(A) \geqslant \varepsilon n^{-1/2}\right) \lesssim \varepsilon^{1/4} + e^{-\Omega(n)}\,. \tag{10.1}$$

By the Hanson-Wright inequality (Theorem 10.2), there exists $C' > 0$ so that

$$\mathbb{P}_X\big(\,\|A^{-1}X\|_2 \geqslant C'(\log \varepsilon^{-1})^{1/2} \cdot \|A^{-1}\|_{\mathrm{HS}}\,\big) \leqslant \varepsilon \tag{10.2}$$

and so the left-hand side of (10.1) is bounded above by

$$\varepsilon + \mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_{\mathrm{HS}}} \leqslant \delta,\ \sigma_n(A) \geqslant \varepsilon n^{-1/2}\right)\,,$$

where $\delta := C''\varepsilon \cdot (\log \varepsilon^{-1})^{1/2}$. Now, by Lemma 9.1 with the choice of $u = 0, s = 0$, we have

$$\mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_{\mathrm{HS}}} \leqslant \delta,\ \sigma_n(A) \geqslant \varepsilon n^{-1/2}\right) \lesssim \delta\varepsilon^{-2/3} + e^{-\Omega(n)} \lesssim \varepsilon^{1/4} + e^{-\Omega(n)}\,, \tag{10.3}$$

where we have used that $\|A^{-1}\|_* \geqslant \|A^{-1}\|_{\mathrm{HS}}$. We also note that Lemma 9.1 actually gives an upper bound on $\mathbb{E}_A \sup_r \mathbb{P}_X(\mathcal{A})$, where $\mathcal{A}$ is the event on the left-hand side of (10.7). Since $\sup_r \mathbb{P}_{A,X}(\mathcal{A}) \leqslant \mathbb{E}_A \sup_r \mathbb{P}_X(\mathcal{A})$, the bound (10.3), and thus Lemma 10.3, follows. □

The next lemma is our "bootstrapping step": Given bounds of the form

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \lesssim \varepsilon^{\kappa} + e^{-cn},$$

this lemma will produce better bounds for the same problem with A_{n+1} in place of A_n.

Lemma 10.4 (Bootstrapping step). *For $B > 0$, let $\zeta \in \Gamma_B$, let $A_{n+1} \sim \mathrm{Sym}_{n+1}(\zeta)$, and let $\kappa \in (0,1) \setminus \{7/10\}$. If for all $\varepsilon > 0$, and all n, we have*

$$\mathbb{P}\big(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}\big) \lesssim \varepsilon^{\kappa} + e^{-\Omega(n)}\,, \tag{10.4}$$

then for all $\varepsilon > 0$ and all n, we have

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leqslant \varepsilon n^{-1/2}) \lesssim (\log \varepsilon^{-1})^{1/2} \cdot \varepsilon^{\min\{1, 6\kappa/7 + 1/3\}} + e^{-\Omega(n)}\,.$$

Proof. Let $c > 0$ denote the implicit constant in the exponent on the right-hand side of (10.4). Note that if $0 < \varepsilon < e^{-cn}$, by the assumption of the lemma, then we have

$$\mathbb{P}(\sigma_{\min}(A_n) \leqslant \varepsilon n^{-1/2}) \lesssim e^{-\Omega(n)},$$

for all n, in which case, we are done. So we may assume $\varepsilon > e^{-cn}$.

As in the proof of the "base step," Lemma 10.3, we look to apply Lemmas 6.2 and 9.1 in sequence. For this, we write $A = A_n$ and bound (9.1) as in the conclusion of Lemma 9.1

$$\mathbb{E}_A \left(\frac{\mu_1}{\sqrt{n}}\right)^{7/9} \mathbf{1}\left\{\frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\} \leqslant \int_0^{\varepsilon^{-7/9}} \mathbb{P}\left(\sigma_{\min}(A) \leqslant x^{-9/7} n^{-1/2}\right) dx, \tag{10.5}$$

where we used that $\sigma_{\min}(A) = 1/\mu_1(A)$. Now use assumption (10.4) to see the right-hand side of (10.5) is

$$\lesssim 1 + \int_1^{\varepsilon^{-7/9}} (x^{-9\kappa/7} + e^{-cn})\, dx \lesssim \max\left\{1, \varepsilon^{\kappa-7/9}\right\}. \tag{10.6}$$

Now, we apply Lemma 9.1 with $\delta = C\varepsilon \cdot (\log \varepsilon^{-1})^{1/2}$, $s = 0$, and $u = 0$ to see that

$$\mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_{\mathrm{HS}}} \leqslant \delta,\ \frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1}\right) \lesssim \max\left\{\varepsilon, \varepsilon^{6\kappa/7+1/3}\right\} \cdot (\log \varepsilon^{-1})^{1/2} + e^{-\Omega(n)}, \tag{10.7}$$

for all r. Here, we used that $\|A^{-1}\|_{\mathrm{HS}} \leqslant \|A^{-1}\|_*$.

Now, by Hanson-Wright (Theorem 10.2), there exists $C' > 0$, such that

$$\mathbb{P}_X\left(\|A^{-1}X\|_2 \geqslant C'\|A^{-1}\|_{\mathrm{HS}} \cdot (\log \varepsilon^{-1})^{1/2}\right) \leqslant \varepsilon.$$

Thus, we choose C'' to be large enough, so that

$$\mathbb{P}_{A,X}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}X\|_2} \leqslant C''\varepsilon, \sigma_n(A) \geqslant \varepsilon n^{-1/2}\right) \lesssim \max\left\{\varepsilon, \varepsilon^{6\kappa/7+1/3}\right\} \cdot (\log \varepsilon^{-1})^{1/2} + e^{-\Omega(n)},$$

for all r. Lemma 6.1 now completes the proof of Lemma 10.4. □

Lemma 10.1 now follows by iterating Lemma 10.4 three times.

Proof of Lemma 10.1. By Lemmas 10.3 and 10.4, we have

$$\mathbb{P}(\sigma_{\min}(A) \leqslant \varepsilon n^{-1/2}) \lesssim \varepsilon^{13/21} \cdot (\log \varepsilon^{-1})^{1/2} + e^{-\Omega(n)} \lesssim \varepsilon^{13/21-\eta} + e^{-\Omega(n)},$$

for some small $\eta > 0$. Applying Lemma 10.4 twice more gives an exponent of $\frac{127}{147} - \frac{6}{7}\eta$ and then 1, for η small, thus completing the proof. □

11. Proof of Theorem 1.1

We are now ready to prove our main result, Theorem 1.1. We use Lemma 6.1 (as in the proof of Lemma 10.1) and the inequality at (4.5) to see that it is enough to prove

$$\mathbb{P}^{\mathcal{E}}\left(\frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}X\|_2} \leqslant C\varepsilon, \text{ and } \sigma_n(A) \geqslant \varepsilon n^{-1/2}\right) \lesssim \varepsilon + e^{-\Omega(n)}, \tag{11.1}$$

where C is as in Lemma 6.1 and the implied constants do not depend on r. Recall that $\mathcal{E}$ is the quasi-random event defined in Section 4.

To prepare ourselves for what follows, we put $\mathcal{E}_0 := \mathcal{E} \cap \{\sigma_{\min}(A) \geqslant \varepsilon n^{-1/2}\}$ and

$$Q(A, X) := \frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}X\|_2} \quad \text{and} \quad Q_*(A, X) := \frac{|\langle A^{-1}X, X\rangle - r|}{\|A^{-1}\|_*},$$

where

$$\|A^{-1}\|_*^2 = \sum_{k=1}^{n} \mu_k^2 (\log(1+k))^2,$$

as defined in Section 8. We now split the left-hand side of (11.1) as

$$\mathbb{P}^{\mathcal{E}_0}\left(Q(A,X) \leqslant C\varepsilon\right) \leqslant \mathbb{P}^{\mathcal{E}_0}\left(Q_*(A,X) \leqslant 2C\varepsilon\right) + \mathbb{P}^{\mathcal{E}_0}\left(Q(A,X) \leqslant C\varepsilon, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geqslant 2\right). \tag{11.2}$$

We can take care of the first term easily by combining Lemmas 9.1 and 10.1.

Lemma 11.1. *For $\varepsilon > 0$,*

$$\mathbb{P}^{\mathcal{E}_0}(Q_*(A,X) \leqslant 2C\varepsilon) \lesssim \varepsilon + e^{-\Omega(n)}.$$

Proof. Apply Lemma 9.1, with $\delta = 2C\varepsilon$, $u = 0$, and $s = 0$ to obtain

$$\mathbb{P}^{\mathcal{E}_0}(Q_*(A,X) \leqslant 2C\varepsilon) \lesssim \varepsilon\left(\mathbb{E}_A\left(\frac{\mu_1}{\sqrt{n}}\right)^{7/9}\mathbf{1}\left\{\frac{\mu_1}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\}\right)^{6/7} + e^{-\Omega(n)}.$$

By Lemma 10.1 and the calculation at (10.6), the expectation on the right is bounded by a constant. □

We now focus on the latter term on the right-hand side of (11.2). By considering the dyadic partition $2^j \leqslant \|A^{-1}X\|_2/\|A^{-1}\|_* \leqslant 2^{j+1}$, we see the second term on the right hand side (RHS) of (11.2) is

$$\lesssim \sum_{j=1}^{\log n}\mathbb{P}^{\mathcal{E}_0}\left(Q_*(A,X) \leqslant 2^{j+1}C\varepsilon\,, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geqslant 2^j\right) + e^{-\Omega(n)}. \tag{11.3}$$

Here, we have dealt with the terms for which $j \geqslant \log n$ by using the fact that

$$\mathbb{P}_X\left(\|A^{-1}X\|_2 \geqslant \sqrt{n}\|A^{-1}\|_*\right) \lesssim e^{-\Omega(n)},$$

which follows from Hanson-Wright and the inequality $\|A^{-1}\|_* \geqslant \|A^{-1}\|_{\mathrm{HS}}$.

We now show that the event $\|A^{-1}X\|_2 \geqslant t\|A^{-1}\|_*$ implies that X must correlate with one of the eigenvectors of A.

Lemma 11.2. *For $t > 0$, we have*

$$\mathbb{P}_X\left(Q_*(A,X) \leqslant 2Ct\varepsilon, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geqslant t\right) \leqslant 2\sum_{k=1}^{n}\mathbb{P}_X\left(Q_*(A,X) \leqslant 2Ct\varepsilon, \langle X, v_k\rangle \geqslant t\log(1+k)\right),$$

where $\{v_k\}$ is an orthonormal basis of eigenvectors of A.

Proof. Assume that $\|A^{-1}X\|_2 \geqslant t\|A^{-1}\|_*$, and use the singular value decomposition associated with $\{v_k\}_k$ to write

$$t^2\sum_k \mu_i^2(\log(k+1))^2 = t^2\|A\|_*^2 \leqslant \|A^{-1}X\|_2^2 = \sum_k \mu_k^2\langle v_k, X\rangle^2.$$

Thus

$$\{\|A^{-1}X\|_2 \geqslant t\|A^{-1}\|_*\} \subset \bigcup_k \left\{|\langle X, v_k\rangle| \geqslant t\log(k+1)\right\}.$$

To finish the proof of Lemma 11.2, we union bound and treat the case of $-X$ the same as X (by possibly changing the sign of v_k) at the cost of a factor of 2. □

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Proof of Theorem 1.1. Recall that it suffices to establish (11.1). Combining (11.2) with Lemma 11.2 and Lemma 11.1 tells us that

$$\mathbb{P}^{\mathcal{E}_0}\left(Q(A,X) \leqslant C\varepsilon\right) \lesssim \varepsilon + 2\sum_{j=1}^{\log n}\sum_{k=1}^{n} \mathbb{P}^{\mathcal{E}_0}\left(Q_*(A,X) \leqslant 2^{j+1}C\varepsilon, \langle X, v_k\rangle \geqslant 2^j \log(1+k)\right) + e^{-\Omega(n)}\,. \tag{11.4}$$

We now apply Lemma 9.1 for all $t > 0$, with $\delta = 2Ct\varepsilon$, $s = t\log(k+1)$ and $u = v_k$ to see that,

$$\mathbb{P}^{\mathcal{E}_0}\left(Q_*(A,X) \leqslant 2Ct\varepsilon, \langle X, v_k\rangle \geqslant t\log(1+k)\right) \lesssim \varepsilon t(k+1)^{-t} \cdot I^{6/7} + e^{-\Omega(n)}\,, \tag{11.5}$$

where

$$I := \mathbb{E}_A\left(\frac{\mu_1(A)}{\sqrt{n}}\right)^{7/9} \mathbf{1}\left\{\frac{\mu_1(A)}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\}.$$

Using (11.5) in (11.4) yields

$$\mathbb{P}^{\mathcal{E}_0}(Q(A,X) \leqslant C\varepsilon) \lesssim \varepsilon I^{6/7}\sum_{j=1}^{\log n}\sum_{k=1}^{n} 2^j(k+1)^{-2^j} + e^{-\Omega(n)} \lesssim \varepsilon \cdot I^{6/7} + e^{-\Omega(n)},$$

since $\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} 2^j(k+1)^{-2^j} = O(1)$. Now we write

$$I = \mathbb{E}_A\left(\frac{\mu_1(A)}{\sqrt{n}}\right)^{7/9} \mathbf{1}\left\{\frac{\mu_1(A)}{\sqrt{n}} \leqslant \varepsilon^{-1}\right\} \leqslant \int_0^{\varepsilon^{-7/9}} \mathbb{P}\left(\sigma_{\min}(A) \leqslant x^{-9/7}n^{-1/2}\right)\,dx$$

and apply Lemma 10.1 to see

$$\int_0^{\varepsilon^{-7/9}} \mathbb{P}\left(\sigma_{\min}(A) \leqslant x^{-9/7}n^{-1/2}\right)\,dx \lesssim \int_1^{\infty} s^{-9/7}\,ds + 1 \lesssim 1.$$

Thus, Lemma 6.1 completes the proof of Theorem 1.1. □

I. Introduction to the appendices

In these appendices, we lay out the proof of Theorem 4.3, the "master quasi-randomness theorem," which we left unproved in the main body of the paper, and the proof of Theorem 9.2. The proofs of these results are technical adaptations of the authors' previous work on the singularity of random symmetric matrices [4]. The last three appendices also tie up some other loose ends in the main body of the text.

In particular, the proof of Theorem 4.3 is similar to the proof of the main theorem in [4], with only a few tweaks and additions required to make the adaptation go through. In several places, we need only update the constants and will be satisfied in pointing the interested reader to [4] for more detail. Elsewhere, more significant adaptations are required, and we outline these changes in full detail. As such, parts of these appendices will bore the restless expert, but we hope it will provide a useful source for those who are taking up the subject or want to avoid writing out the (sometimes extensive) details for oneself.

I.1. Definitions

We collect a few definitions from the main body of the text that are most relevant for us here. Throughout, ζ will be a random variable with mean 0 and variance 1. Such a random variable is said to be *subgaussian* if the *subgaussian moment*

$$\|\zeta\|_{\psi_2} := \sup_{p \geqslant 1} p^{-1/2} (\mathbb{E}|\zeta|^p)^{1/p}$$

is finite. For $B > 0$, we let Γ_B denote the set of mean 0 variance 1 random variables with subgaussian moment $\leqslant B$, and we let $\Gamma = \bigcup_{B>0} \Gamma_B$.

For $\zeta \in \Gamma$, let $\mathrm{Sym}_n(\zeta)$ denote the probability space of $n \times n$ symmetric matrices with $(A_{i,j})_{i \leqslant j}$ i.i.d. distributed according to ζ. Let $\mathrm{Col}_n(\zeta)$ be the probability space on vectors of length n with independent coordinates distributed according to ζ.

For $v \in \mathbb{S}^{n-1}$ and $\mu, \alpha, \gamma \in (0,1)$, define the *least common denominator* (LCD) of the vector v via

$$D_{\alpha,\gamma}(v) := \inf\left\{t > 0 : \|tv\|_{\mathbb{T}} < \min\{\gamma\|tv\|_2, \sqrt{\alpha n}\}\right\}, \tag{I.1}$$

where $\|w\|_{\mathbb{T}} := \mathrm{dist}(w, \mathbb{Z}^n)$. We also define

$$\hat{D}_{\alpha,\gamma,\mu}(v) := \min_{\substack{I \subset [n] \\ |I| \geqslant (1-2\mu)n}} D_{\alpha,\gamma}(v_I)\,. \tag{I.2}$$

Remark I.1. We note that in the main body of the paper, we work with a slightly different notion of $\hat{D}$, where we define $\hat{D}_{\alpha,\gamma,\mu}(v) = \min_I D_{\alpha,\gamma}(v_I/\|v_I\|_2)$. This makes no difference for us, as Lemma II.6 below eliminates those v for which $\|v_I\|_2$ is less than a constant. Thus, we work with the slightly simpler definition (I.2) throughout.

We define the set of "structured direction on the sphere"

$$\Sigma = \Sigma_{\alpha,\gamma,\mu} := \left\{v \in \mathbb{S}^{n-1} : \hat{D}_{\alpha,\gamma,\mu}(v) \leqslant e^{c_\Sigma n}\right\}.$$

Now, for $\zeta \in \Gamma$, $A \sim \mathrm{Sym}_n(\zeta)$ and a given vector $w \in \mathbb{R}^n$, we define the quantity (as in Section 4)

$$q_n(w) = q_n(w; \alpha, \gamma, \mu) := \mathbb{P}_A\left(\exists v \in \Sigma \text{ and } \exists s,t \in [-4\sqrt{n}, 4\sqrt{n}] : \ Av = sv + tw\right).$$

We then recall (see (4.10))

$$q_n := \max_{w \in \mathbb{S}^{n-1}} q_n(w)\,.$$

I.2. Main theorems of the appendix

Let us now restate the two main objectives of this appendix. Our first goal is to prove the following.

Theorem I.2 (Master quasi-randomness theorem). *For $B > 0$ and $\zeta \in \Gamma_B$, there exist constants $\alpha, \gamma, \mu, c_\Sigma, c \in (0,1)$ depending only on B so that*

$$q_n(\alpha, \gamma, \mu) \leqslant 2e^{-cn}\,.$$

The second main goal of this appendix is to prove Theorem 9.2, which we will prove on our way to proving Theorem I.2.

Theorem I.3. *For $B > 0$, let $\zeta \in \Gamma_B$. For $d \in \mathbb{N}$, $\alpha, \gamma \in (0,1)$, and $\nu \in (0, 2^{-15})$, there are constants $c_0, R > 0$ depending only on α, γ, ν, B so that the following holds. Let $0 \leqslant k \leqslant c_0 \alpha d$ and $t \geqslant \exp(-c_0 \alpha d)$; let $v \in \mathbb{S}^{d-1}$, and let $w_1, \ldots, w_k \in \mathbb{S}^{d-1}$ be orthogonal.*

Let ζ' be an independent copy of ζ, let Z_ν be a Bernoulli random variable with parameter ν, and let $\tau \in \mathbb{R}^d$ be a random vector whose coordinates are i.i.d. copies of the random variable with distribution $(\zeta - \zeta')Z_\nu$.

If $D_{\alpha,\gamma}(v) > 1/t$, then

$$\mathbb{P}\left(|\langle \tau, v\rangle| \leqslant t \ \text{ and } \ \sum_{j=1}^{k} \langle w_j, \tau\rangle^2 \leqslant c_0 k\right) \leqslant Rt \cdot e^{-c_0 k}\,.$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

The proofs of Theorems I.2 and I.3 follow the same path as [4], where the authors proved analogous statements for the case where the entries of A are uniform in $\{-1, 1\}$. We refer the reader to the following Section I.3 for a discussion of how this appendix is structured relative to [4].

I.3. A Reader's guide for the appendices

Here, we describe the correspondence between sections in this appendix and sections in [4] and point out the key changes that come up.

In Section II, we set up many of the basic notions that we will need for the proof of Theorem I.2. The main novelty here is in the definitions of several auxiliary random variables, related to ζ, that will be used to study ζ in the course of the paper.

In Section III, we turn to prove Theorem I.2, while assuming several key results that we either import from [4] or prove in later sections. This section is the analogue of Section 9 in [4], and the main difference between these sections arises from the different definitions of q_n in these two papers (see (4.10)). Here, q_n is defined in terms of the least common denominator $D_{\alpha,\gamma}$, rather than the threshold $\mathcal{T}_L$ (see (II.7)). In the course of the proof, we *also* need to break things up according to $\mathcal{T}_L$, and define nets as we did in [4], but another net argument is required to exclude vectors with $\mathcal{T}_L$ small but $D_{\alpha,\gamma}$ large.

In Section IV, we define many of the key Fourier-related notions that we will need to prove the remaining results, including Theorem I.3. The main differences between the two papers in these sections comes from the different definition of the sublevel sets S_W (see (IV.1)). This new definition requires us to reprove a few of our basic lemmas from [4], however, the proofs go through easily.

In Section IV.2, we state our main inverse Littlewood-Offord Theorem for conditioned random walks and deduce Theorem I.3 from it. Lemma IV.3 in this section is also one of the main ingredients that goes into Theorem III.2. This section corresponds to Section 3 of [4].

Section V deals with Fourier replacement and is the analogue of Appendix B in [4]. Here, the only difference between the sections is that here we lack an explicit form for the Fourier transform. However, this difficulty is easily overcome.

In Section VI, we prove Lemma IV.3. This corresponds to Sections 4 and 5 of [4], from which several key geometric facts are imported wholesale, making our task significantly lighter here. The difference in the definitions from Section IV are salient here, but the majority of the proof is the same as in [4, Section 5], up to the constants involved.

The next three sections, Sections VII, VIII, and IX, correspond to Sections 6, 7, and 8 respectively of [4]. Here, the adaptation to this paper requires little more than updating constants. These three sections amount to converting Lemma IV.3 into the main net bound Theorem III.2.

Finally, in Section X, we deduce the Hanson-Wright inequality, Lemma VI.7, from Talagrand's inequality; this corresponds to Appendix E of [4] where the difference, again, is only up to constants.

II. Preparations

II.1. Symmetrizing and truncating the random variable

We will work with symmetrized, truncated, and lazy versions of the variable ζ. This is primarily because these altered versions will have better behaved Fourier properties. Here, we introduce these random variables and also note some properties of their characteristic functions. These properties are not so important until Section IV, but we have them here to help motivate some of the definitions.

Let ζ' be an independent copy of ζ and define

$$\tilde{\zeta} = \zeta - \zeta'.$$

We will want to truncate $\tilde{\zeta}$ to a bounded window, as this will be useful for our construction of a nondegenerate and not-too-large LCD in Section VI. In this direction, define $I_B = (1, 16B^2)$ and $p := \mathbb{P}(|\tilde{\zeta}| \in I_B)$. Our first step is to uniformly bound p in terms of B.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Lemma II.1. $p \geqslant \frac{1}{2^7 B^4}$.

Proof. By the Paley-Zygmund inequality

$$\mathbb{P}(|\tilde{\zeta}| > 1) = \mathbb{P}(|\tilde{\zeta}|^2 > \mathbb{E}|\tilde{\zeta}|^2/2) \geqslant \frac{(1-\frac{1}{2})^2(\mathbb{E}\tilde{\zeta}^2)^2}{(\mathbb{E}\tilde{\zeta}^4)} \geqslant \frac{1}{2^6 B^4},$$

where we have used $\mathbb{E}\tilde{\zeta}^4 = 2\mathbb{E}\zeta^4 + 6 \leqslant 2^5 B^4 + 6$ and $B \geqslant 1$. By Chebyshev's inequality, we have

$$\mathbb{P}(|\tilde{\zeta}| \geqslant 16B^2) \leqslant \frac{2}{2^8 B^4}.$$

Combining the bounds completes the proof. □

For a parameter $\nu \in (0,1)$, define ξ_ν by

$$\xi_\nu := \mathbf{1}\{|\tilde{\zeta}| \in I_B\}\tilde{\zeta}Z_\nu,$$

where Z_ν is an independent Bernoulli variable with mean ν. For $\nu \in (0,1)$ and $d \in \mathbb{N}$, we write $X \sim \Xi_\nu(d;\zeta)$ to indicate that X is a random vector in $\mathbb{R}^d$ whose entries are i.i.d. copies of the variable ξ_ν; similarly, we write $X \sim \Phi_\nu(d;\zeta)$ to denote a random vector whose entries are i.i.d. copies of the random variable $\tilde{\zeta}Z_\nu$.

We compute the characteristic function of ξ_ν to be

$$\phi_{\xi_\nu}(t) = \mathbb{E}e^{i2\pi t\xi_\nu} = 1 - \nu + \nu(1-p) + \nu p\mathbb{E}_{\tilde{\zeta}}[\cos(2\pi t\tilde{\zeta}) \mid |\tilde{\zeta}| \in (1, 16B^2)].$$

Define the variable $\bar{\zeta}$ as $\tilde{\zeta}$ conditioned on $|\tilde{\zeta}| \in I_B$, where we note that this conditioning makes sense since Lemma II.1 shows $p > 0$. In other words, for every Borel set S,

$$\mathbb{P}(\bar{\zeta} \in S) = p^{-1}\mathbb{P}(\tilde{\zeta} \in S \cap (I_B \cup -I_B)).$$

Therefore we can write the characteristic function of ξ_ν as

$$\phi_{\xi_\nu}(t) = 1 - \nu p + \nu p\mathbb{E}_{\bar{\zeta}}\cos(2\pi t\bar{\zeta}). \tag{II.1}$$

For $x \in \mathbb{R}$, define $\|x\|_{\mathbb{T}} := \operatorname{dist}(x, \mathbb{Z})$, and note the elementary inequalities

$$1 - 20\|a\|_{\mathbb{T}}^2 \leqslant \cos(2\pi a) \leqslant 1 - \|a\|_{\mathbb{T}}^2,$$

for $a \in \mathbb{R}$. These imply that

$$\exp\left(-32\nu p \cdot \mathbb{E}_{\bar{\zeta}}\|t\bar{\zeta}\|_{\mathbb{T}}^2\right) \leqslant \phi_{\xi_\nu}(t) \leqslant \exp\left(-\nu p \cdot \mathbb{E}_{\bar{\zeta}}\|t\bar{\zeta}\|_{\mathbb{T}}^2\right). \tag{II.2}$$

Also note that since $\phi_{\tilde{\zeta}Z_\nu}(t) = 1 - \nu + \nu\mathbb{E}_{\tilde{\zeta}}[\cos(2\pi t\tilde{\zeta})]$, we have

$$\phi_{\tilde{\zeta}Z_\nu}(t) \leqslant 1 - \nu + \nu(1-p) + \nu p\mathbb{E}_{\tilde{\zeta}}[\cos(2\pi t\tilde{\zeta}) \mid |\tilde{\zeta}| \in I_B] = \phi_{\xi_\nu}(t). \tag{II.3}$$

II.2. Properties of subgaussian random variables and matrices

We will use a basic fact about exponential moments of one-dimensional projections of subgaussian random variables (see, e.g. [47, Proposition 2.6.1]).

Fact II.2. For $B > 0$, let $Y = (Y_1, \ldots, Y_d)$ be a random vector with $Y_1, \ldots, Y_d \in \Gamma_B$. Then for all $u \in \mathbb{S}^{d-1}$, we have $\mathbb{E}\, e^{\langle Y,u\rangle} = O_B(1)$.

We will also use a large deviation bound for the operator norm of A (see (4.11)).

Fact II.3. For $B > 0$, let $\zeta \in \Gamma$ and $A \sim \mathrm{Sym}_n(\zeta)$. Then

$$\mathbb{P}(\|A\|_{op} \geqslant 4\sqrt{n}) \leqslant 2e^{-\Omega(n)}\,.$$

We also define the event $\mathcal{K} = \{\|A\|_{op} \geqslant 4\sqrt{n}\}$, and define the measure $\mathbb{P}^{\mathcal{K}}$ by

$$\mathbb{P}^{\mathcal{K}}(\mathcal{E}) = \mathbb{P}(\mathcal{K} \cap \mathcal{E}), \tag{II.4}$$

for every event $\mathcal{E}$.

II.3. Compressibility and eliminating nonflat vectors

As in [4], we may limit our attention to vectors that are "flat" on a constant proportion of their coordinates. This reduction is a consequence of the now-classical work of Rudelson and Vershynin on *compressible* and *incompressible* vectors [31].

Following [31], we say that a vector in $\mathbb{S}^{n-1}$ is (δ, ρ)-compressible if it has distance at most ρ from a vector with support of size at most δn. For $\delta, \rho \in (0, 1)$, let $\mathrm{Comp}\,(\delta, \rho)$ denote the set of all such compressible vectors in $\mathbb{S}^{n-1}$. Proposition 4.2 from Vershynin's paper [46] takes care of all compressible vectors.

Lemma II.4. *For $B > 0$, let $\zeta \in \Gamma_B$, let $A_n \sim \mathrm{Sym}_n(\zeta)$, and let $K \geqslant 1$. Then there exist $\rho, \delta, c > 0$ depending only on K, B, so that for every $\lambda \in \mathbb{R}$ and $w \in \mathbb{R}^n$, we have*

$$\mathbb{P}\Big(\inf_{x \in \mathrm{Comp}\,(\delta,\rho)} \|(A_n + \lambda I)x - w\|_2 \leqslant c\sqrt{n} \text{ and } \|A_n + \lambda I\|_{op} \leqslant K\sqrt{n}\Big) \leqslant 2e^{-cn}\,.$$

For the remainder of the paper, we let δ, ρ be the constants given in Lemma II.4. Define

$$\mathrm{Incomp}\,(\delta, \rho) := \mathbb{S}^{n-1} \setminus \mathrm{Comp}\,(\delta, \rho)$$

to be the set of (δ, ρ)-*incompressible* vectors. The key property of incompressible vectors is that they are "flat" for a constant proportion of coordinates. This is made quantitative in the following lemma of Rudelson and Vershynin [31].

Lemma II.5. *Let $v \in \mathrm{Incomp}\,(\delta, \rho)$. Then*

$$(\rho/2)n^{-1/2} \leqslant |v_i| \leqslant \delta^{-1/2}n^{-1/2}$$

for at least $\rho^2\delta n/2$ values of $i \in [n]$.

We now fix a few more constants to be held fixed throughout the paper. Let $\kappa_0 = \rho/3$ and $\kappa_1 = \delta^{-1/2} + \rho/6$, where δ, ρ are as in Lemma II.4. For $D \subseteq [n]$, define the set of directions in $\mathbb{S}^{n-1}$ that are "flat on D":

$$\mathcal{I}(D) = \left\{v \in \mathbb{S}^{n-1} : (\kappa_0 + \kappa_0/2)n^{-1/2} \leqslant |v_i| \leqslant (\kappa_1 - \kappa_0/2)n^{-1/2} \text{ for all } i \in D\right\},$$

and let

$$\mathcal{I} = \mathcal{I}_d := \bigcup_{D \subseteq [n], |D| = d} \mathcal{I}(D).$$

Applying Lemmas II.4 and II.5 in tandem will allow us to eliminate vectors outside of $\mathcal{I}$.

Lemma II.6. *Let $\delta, \rho, c > 0$ be the constants defined in Lemma II.4, and let $d < \rho^2\delta n/2$. Then*

$$\max_{w \in \mathbb{S}^{n-1}} \mathbb{P}_A\left(\exists v \in \mathbb{S}^{n-1} \setminus \mathcal{I} \text{ and } \exists s, t \in [-4\sqrt{n}, +4\sqrt{n}] : \|Av - sv - tw\|_2 \leqslant c\sqrt{n}/2\right) \leqslant 2e^{-\Omega(n)}\,. \tag{II.5}$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Proof. Lemma II.5, along with the definitions of κ_0, κ_1, and $\mathcal{I}$, implies that

$$\mathbb{S}^{n-1} \setminus \mathcal{I} \subseteq \mathrm{Comp}\,(\delta, \rho).$$

Now, fix a $w \in \mathbb{R}^n$ and take a $c\sqrt{n}/8$-net $\mathcal{N}$ for $[-4\sqrt{n}, 4\sqrt{n}]^2$ of size $O(c^{-2})$ to see that $\|Av - sv - tw\|_2 \leqslant c\sqrt{n}/2$ implies that there exists $(s', t') \in \mathcal{N}$ for which

$$\|(A - s'I)v - t'w\|_2 \leqslant c\sqrt{n}.$$

Thus, the left-hand side of (II.5) is

$$\leqslant \sum_{(s',t')\in\mathcal{N}} \mathbb{P}_A\left(\exists v \in \mathrm{Comp}\,(\delta, \rho) : \|(A - s'I)v - t'w\|_2 \leqslant c\sqrt{n}\right) \leqslant |\mathcal{N}| \cdot 2e^{-\Omega(n)},$$

where the final inequality follows by first intersecting each term in the sum with the event $\mathcal{E} := \{\|A - s'I\|_{op} \leqslant 16n^{1/2}\}$ (noting that $\mathbb{P}(\mathcal{E}^c) \leqslant 2e^{-\Omega(n)}$, by Fact II.3) and applying Lemma II.4 to each term in the sum with $\lambda = -s'$ and $K = 16$. □

II.4. Zeroed out matrices

To study our original matrix A, it will be useful to work with random symmetric matrices that have large blocks that are "zeroed out" and entries that are distributed like $\tilde{\zeta} Z_\nu$ elsewhere (see [4] for more discussion on this). For this, we set $d := c_0^2 n$ (where $c_0 > 0$ is a small constant to be determined later) and write $M \sim \mathcal{M}_n(\nu)$ for the $n \times n$ random matrix

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[d+1,n]\times[d+1,n]} \end{bmatrix}, \tag{II.6}$$

where H_1 is a $(n - d) \times d$ random matrix whose entries are i.i.d. copies of $\tilde{\zeta} Z_\nu$.

In particular, the matrix M will be useful for analyzing events of the form $\|Av\|_2 \leqslant \varepsilon n^{1/2}$, when $v \in \mathcal{I}([d])$.

We now use the definition of $\mathcal{M}_n(\nu)$ to define another notion of "structure" for vectors $v \in \mathbb{S}^{n-1}$. This is a very different measure of "structure" from that provided by the LCD, which we saw above. For $L > 0$ and $v \in \mathbb{R}^n$, define the *threshold* of v as

$$\mathcal{T}_L(v) := \sup\left\{t \in [0, 1] : \mathbb{P}(\|Mv\|_2 \leqslant t\sqrt{n}) \geqslant (4Lt)^n\right\}. \tag{II.7}$$

One can think of this $\mathcal{T}_L(v)$ as the "scale" at which the structure of v (relative to M) starts to emerge. So "large threshold" means "more structured."

III. Proof of Theorem I.2

Here, we recall some key notions from [4], state analogous lemmas, and prove Theorem I.2 assuming these lemmas.

III.1. Efficient nets

Our goal is to obtain an exponential bound on the quantity

$$q_n = \max_{w\in\mathbb{S}} \mathbb{P}_A\left(\exists v \in \Sigma \text{ and } \exists s, t \in [-4\sqrt{n}, 4\sqrt{n}] : \ Av = sv + tw\right),$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

defined at (4.10), where

$$\Sigma = \Sigma_{\alpha,\gamma,\mu} := \left\{ v \in \mathbb{S}^{n-1} : \hat{D}_{\alpha,\gamma,\mu}(v) \leqslant e^{c_\Sigma n} \right\}.$$

In the course of the proof, we will choose α, γ, μ to be sufficiently small.

We cover $\Sigma \subseteq \mathbb{S}^{n-1}$ with two regions which will be dealt with in very different ways. First, we define

$$S := \left\{ v \in \mathbb{S}^{n-1} :\ \mathcal{T}_L(v) \geqslant \exp(-2c_\Sigma n) \right\}.$$

This will be the trickier region and will depend on the net construction from [4]. We also need to take care of the region

$$S' := \{ v \in \mathbb{S}^{n-1} : \hat{D}_{\alpha,\gamma,\mu}(v) \leqslant \exp(c_\Sigma n),\ \mathcal{T}_L(v) \leqslant \exp(-2c_\Sigma n) \},$$

which we take care of using the nets constructed by Rudelson and Vershynin in [31]. We recall that $\mathcal{T}_L$ is defined at (II.7).

We also note that since the event $\mathcal{K} := \{\|A\|_{\text{op}} \geqslant 4n^{1/2}\}$ fails with probability $2e^{-cn}$ (Fact II.3) and we only need to deal with incompressible vectors $v \in \mathcal{I}$ (by Lemma II.6), it is enough to show

$$\sup_{w \in \mathbb{S}^{n-1}} \mathbb{P}_A^{\mathcal{K}} \left(\exists v \in \mathcal{I} \cap S,\ s,t \in [-4\sqrt{n}, +4\sqrt{n}] :\ Av = sv + tw \right) \leqslant e^{-\Omega(n)}, \tag{III.1}$$

and the same with S' replacing S. We recall that we define $\mathbb{P}^{\mathcal{K}}(\mathcal{E}) := \mathbb{P}(\mathcal{K} \cap \mathcal{E})$ for every event $\mathcal{E}$. To deal with the above probability, we will construct nets to approximate vectors in $\mathcal{I} \cap S$ and $\mathcal{I} \cap S'$. To define the nets used, we recall a few definitions from [4]. For a random variable $Y \in \mathbb{R}^d$ and $\varepsilon > 0$, we define the Lévy concentration of Y by

$$\mathcal{L}(Y, \varepsilon) = \sup_{w \in \mathbb{R}^d} \mathbb{P}(\|Y - w\|_2 \leqslant \varepsilon). \tag{III.2}$$

Now, for $v \in \mathbb{R}^n$, $\varepsilon > 0$, define

$$\mathcal{L}_{A,op}(v, \varepsilon\sqrt{n}) := \sup_{w \in \mathbb{R}^n} \mathbb{P}^{\mathcal{K}}(\|Av - w\|_2 \leqslant \varepsilon\sqrt{n}). \tag{III.3}$$

Slightly relaxing the requirements of $\mathcal{I}$, we define

$$\mathcal{I}'([d]) := \left\{ v \in \mathbb{R}^n : \kappa_0 n^{-1/2} \leqslant |v_i| \leqslant \kappa_1 n^{-1/2} \text{ for all } i \in [d] \right\}.$$

Define the (trivial) net

$$\Lambda_\varepsilon := B_n(0,2) \cap \left(4\varepsilon n^{-1/2} \cdot \mathbb{Z}^n \right) \cap \mathcal{I}'([d]).$$

III.1.1. Definition of net for $v \in S$

To deal with vectors in S, for $\varepsilon \geqslant \exp(-2c_\Sigma n)$, define

$$\Sigma_\varepsilon := \left\{ v \in \mathcal{I}([d]) :\ \mathcal{T}_L(v) \in [\varepsilon, 2\varepsilon] \right\}. \tag{III.4}$$

If $v \in \Sigma_\varepsilon$, for some $\varepsilon \geqslant \exp(-2c_\Sigma n)$, then the proof will be basically the same as in [4]. As such, we approximate Σ_ε by $\mathcal{N}_\varepsilon$, where we define

$$\mathcal{N}_\varepsilon := \left\{ v \in \Lambda_\varepsilon : (L\varepsilon)^n \leqslant \mathbb{P}(\|Mv\|_2 \leqslant 4\varepsilon\sqrt{n}) \text{ and } \mathcal{L}_{A,op}(v, \varepsilon\sqrt{n}) \leqslant (2^{10}L\varepsilon)^n \right\},$$

and show that $\mathcal{N}_\varepsilon$ is appropriately small.

First, the following lemma allows us to approximate Σ_ε by $\mathcal{N}_\varepsilon$.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Lemma III.1. *Let $\varepsilon \in (\exp(-2c_\Sigma n), \kappa_0/8)$. For each $v \in \Sigma_\varepsilon$, then there is $u \in \mathcal{N}_\varepsilon$, such that $\|u-v\|_\infty \leqslant 4\varepsilon n^{-1/2}$.*

This lemma is analogous to Lemma 8.2 in [4], and we postpone its proof to Section IX. The main difficulty faced in [4] is to prove an appropriate bound on $|\mathcal{N}_\varepsilon|$. In our case, we have an analogous bound.

Theorem III.2. *For $L \geqslant 2$ and $0 < c_0 \leqslant 2^{-50}B^{-4}$, let $n \geqslant L^{64/c_0^2}$, $d \in [c_0^2 n/4, c_0^2 n]$, and $\varepsilon > 0$ be so that $\log \varepsilon^{-1} \leqslant nL^{-32/c_0^2}$. Then*

$$|\mathcal{N}_\varepsilon| \leqslant \left(\frac{C}{c_0^6 L^2 \varepsilon}\right)^n,$$

where $C > 0$ is an absolute constant.

The proof of Theorem III.2 will follow mostly from Lemma IV.3, with the rest of the deduction following exactly the same path as in [4], which we present in Sections VII and VIII.

III.1.2. Definition of net for $v \in S'$

We now need to tackle the vectors in S'; that is, those with

$$\mathcal{T}_L(v) \leqslant \exp(-2c_\Sigma n) \text{ and } \hat{D}_{\alpha,\gamma,\mu}(v) \leqslant \exp(c_\Sigma n).$$

Here, we construct the nets using only the second condition using a construction of Rudelson and Vershynin [31]. Then the condition $\mathcal{T}_L(v) \leqslant \exp(-2c_\Sigma n)$ will come in when we union bound over nets. With this in mind, let

$$\Sigma'_\varepsilon := \left\{v \in \mathcal{I}([d]) \cap S' : \hat{D}_{\alpha,\gamma,\mu}(v) \in [(4\varepsilon)^{-1}, (2\varepsilon)^{-1}]\right\}.$$

We will approximate $v \in \Sigma'_\varepsilon$ by the net G_ε, where we define

$$G_\varepsilon := \bigcup_{|I| \geqslant (1-2\mu)n} \left\{\frac{p}{\|p\|_2} : p \in \left(\mathbb{Z}^I \oplus \sqrt{\alpha}\mathbb{Z}^{I^c}\right) \cap B_n(0, \varepsilon^{-1}) \setminus \{0\}\right\}. \tag{III.5}$$

The following two lemmas tell us that G_ε is a good $\varepsilon\sqrt{\alpha n}$-net for Σ'_ε. Here, this $\sqrt{\alpha}$ is the "win" over trivial nets.

Lemma III.3. *Let $\varepsilon > 0$ satisfy $\varepsilon \leqslant \gamma(\alpha n)^{-1/2}/4$. If $v \in \Sigma'_\varepsilon$, then there exists $u \in G_\varepsilon$, such that $\|u - v\|_2 \leqslant 16\varepsilon\sqrt{\alpha n}$.*

Proof. Set $D = \min_{|I| \geqslant (1-2\mu)n} D_{\alpha,\gamma}(v_I)$, and let I be a set attaining the minimum. By definition of $D_{\alpha,\gamma}$, there is $p_I \in \mathbb{Z}^I \cap B_n(0, \varepsilon^{-1})$ so that

$$\|Dv_I - p_I\|_2 < \min\{\gamma D\|v_I\|_2, \sqrt{\alpha n}\} \leqslant \sqrt{\alpha n},$$

and thus $p_I \neq 0$. We now may greedily choose $p_{I^c} \in \sqrt{\alpha}\mathbb{Z}^{I^c} \cap B_n(0, \varepsilon^{-1})$ so that

$$\|Dv_{I^c} - p_{I^c}\|_2 \leqslant \sqrt{\alpha n}.$$

Thus, if we set $p = p_I \oplus p_{I^c}$, by the triangle inequality, we have

$$\left\|v - \frac{p}{\|p\|_2}\right\|_2 \leqslant \frac{1}{D}(\|Dv - p\|_2 + |D - \|p\|_2|) \leqslant 4D^{-1}\sqrt{\alpha n} \leqslant 16\varepsilon\sqrt{\alpha n},$$

as desired. □

We also note that this net is sufficiently small for our purposes (see [31]).

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Fact III.4. For $\alpha, \mu \in (0,1)$, $K \geqslant 1$ and $\varepsilon \leqslant Kn^{-1/2}$, we have

$$|G_\varepsilon| \leqslant \left(\frac{32K}{\alpha^{2\mu}\varepsilon\sqrt{n}} \right)^n ,$$

where G_ε is as defined at (III.5).

The following simple corollary tells us that we can modify G_ε to build a net $G'_\varepsilon \subseteq \Sigma_\varepsilon$, at the cost of a factor of 2 in the accuracy of the next. That is, it is a $32\varepsilon\sqrt{\alpha n}$-net rather than a $16\varepsilon\sqrt{\alpha n}$ net.

Corollary III.5. *For $\alpha, \mu \in (0,1)$, $K \geqslant 1$ and $\varepsilon \leqslant Kn^{-1/2}$ there is a $32\varepsilon\sqrt{\alpha n}$-net G'_ε for Σ'_ε with $G'_\varepsilon \subset \Sigma'_\varepsilon$ and*

$$|G'_\varepsilon| \leqslant \left(\frac{32K}{\alpha^{2\mu}\varepsilon\sqrt{n}} \right)^n .$$

This follows from a standard argument.

III.2. Proof of Theorem I.2

We need the following easy observation to make sure we can use Corollary III.5.

Fact III.6. Let $v \in \mathcal{I}$, $\mu < d/4n$, and $\gamma < \kappa_0\sqrt{d/2n}$, then $\hat{D}_{\alpha,\gamma,\mu}(v) \geqslant (2\kappa_1)^{-1}\sqrt{n}$.

Proof. Since $v \in \mathcal{I}$, there is $D \subset [n]$, such that $|D| = d$ and $\kappa_0 n^{-1/2} \leqslant |v_i| \leqslant \kappa_1 n^{-1/2}$ for all $i \in D$. Now, write $\hat{D}(v) = \min_{|I| \geqslant (1-2\mu)n} D_{\alpha,\gamma}(v_I)$, and let I be a set attaining the minimum. Since $|I| \geqslant (1-2\mu)n \geqslant n - d/2$, we have $|I \cap D| \geqslant d/2$. So put $D' := I \cap D$, and note that for all $t \leqslant (2\kappa_1)^{-1}\sqrt{n}$, we have

$$\min_I d(tv_I, \mathbb{Z}^n) \geqslant d(tv_{D'}, \mathbb{Z}^{D'}) = t\|v_{D'}\|_2 \geqslant t\kappa_0\sqrt{d/2n} > \gamma t.$$

Therefore, $D_{\alpha,\gamma}(v_I) \geqslant (2\kappa_1)^{-1}\sqrt{n}$, by definition. □

When union bounding over the elements of our net, we will also want to use the following lemma to make sure $\mathcal{L}(Av, \varepsilon)$ is small whenever $\mathcal{T}_L(v) \leqslant \varepsilon$.

Lemma III.7. *Let $\nu \leqslant 2^{-8}$. For $v \in \mathbb{R}^n$ and $t \geqslant \mathcal{T}_L(v)$, we have*

$$\mathcal{L}(Av, t\sqrt{n}) \leqslant (50Lt)^n .$$

We prove this lemma in Section V using a fairly straightforward argument on the Fourier side. We now prove our main theorem, Theorem I.2.

Proof of Theorem I.2. We pick up from (III.1) and look to show that

$$q_{n,S} := \sup_{w \in \mathbb{S}^{n-1}} \mathbb{P}_A^{\mathcal{K}}\left(\exists v \in \mathcal{I} \cap S,\ s,t \in [-4\sqrt{n}, +4\sqrt{n}] :\ Av = sv + tw \right) \leqslant e^{-\Omega(n)}, \tag{III.6}$$

and the same with S' in place of S. We do this in three steps.

We first pause to describe how we choose the constants. We let $c_0 > 0$ to be sufficiently small so that Theorem III.2 holds, and we let $d := c_0^2 n$. The parameters μ, γ will be chosen small compared to d/n and κ_0 so that Fact III.6 holds. L will be chosen to be large enough so that $L > 1/\kappa_0$ and so that it is larger than some absolute constants that appear in the proof. We will choose $\alpha > 0$ to be small compared to $1/L$ and $1/\kappa_0$, and we will choose c_Σ small compared to $1/L$.

Step 1: Reduction to Σ_ε and Σ'_ε. Using that $\mathcal{I} = \bigcup_D \mathcal{I}(D)$, we union bound over all choices of D. By symmetry of the coordinates, we have

$$q_{n,S} \leqslant 2^n \sup_{w\in\mathbb{S}^{n-1}} \mathbb{P}_A^{\mathcal{K}}\left(\exists v \in \mathcal{I}([d]) \cap S,\ s,t \in [-4\sqrt{n}, +4\sqrt{n}] :\ Av = sv + tw\right). \tag{III.7}$$

Thus, it is enough to show that the supremum at (III.7) is at most 4^{-n}, and the same with S replaced by S'.

Now, let $\mathcal{W} = (2^{-n}\mathbb{Z}) \cap [-4\sqrt{n}, +4\sqrt{n}]$ and notice that for all $s,t \in [-4\sqrt{n}, +4\sqrt{n}]$, there is $s', t' \in \mathcal{W}$ with $|s-s'| \leqslant 2^{-n}$ and $|t-t'| \leqslant 2^{-n}$. So, union bounding over all (s',t'), the supremum term in (III.7) is at most

$$\leqslant 8^n \sup_{w\in\mathbb{R}^n,\ |s|\leqslant 4\sqrt{n}} \mathbb{P}_A^{\mathcal{K}}\left(\exists v \in \mathcal{I}([d]) \cap (S \cup S') :\ \|Av - sv - w\|_2 \leqslant 2^{-n+1}\right)$$

and the same with S replaced with S'.

We now need to treat S and S' a little differently. Starting with S, we let $\eta := \exp(-2c_\Sigma n)$, and note that for $v \in S$, we have, by definition, that

$$\eta \leqslant \mathcal{T}_L(v) \leqslant 1/L \leqslant \kappa_0/8, \tag{III.8}$$

where we will guarantee the last inequality holds by our choice of L later.

Now, recalling the definition of $\Sigma_\varepsilon := \Sigma_\varepsilon([d])$ at (III.4), we may write

$$\mathcal{I}([d]) \cap S \subseteq \bigcup_{j=0}^{n} \left\{v \in \mathcal{I} : \mathcal{T}_L(v) \in [2^j\eta, 2^{j+1}\eta]\right\} = \bigcup_{j=0}^{j_0} \Sigma_{2^j\eta},$$

where j_0 is the largest integer, such that $2^{j_0}\eta \leqslant \kappa_0/2$. Thus, by the union bound, it is enough to show

$$Q_\varepsilon := \max_{w\in\mathbb{R}^n,\ |s|\leqslant 4\sqrt{n}} \mathbb{P}_A^{\mathcal{K}}\left(\exists v \in \Sigma_\varepsilon :\ \|Av - sv - w\|_2 \leqslant 2^{-n+1}\right) \leqslant 2^{-4n}, \tag{III.9}$$

for all $\varepsilon \in [\eta, \kappa_0/4]$.

We now organize S' in a similar way, relative to the sets Σ'_ε. For this, notice that for $v \in \mathcal{I}([d]) \cap S'$, we have

$$(2\kappa_1)^{-1}\sqrt{n} \leqslant \hat{D}_{\alpha,\gamma,\mu}(v) \leqslant \exp(c_\Sigma n) = \eta^{-1/2},$$

by Fact III.6. So, if we recall the definition

$$\Sigma'_\varepsilon := \{v \in \mathcal{I}([d]) \cap S' : \hat{D}_{\alpha,\gamma,\mu}(v) \in [(4\varepsilon)^{-1}, (2\varepsilon)^{-1}]\},$$

then

$$\mathcal{I}([d]) \cap S' \subseteq \bigcup_{j=-1}^{j_1} \Sigma'_{2^j\sqrt{\eta}},$$

where j_1 is the least integer, such that $2^{j_1}\sqrt{\eta} \geqslant \kappa_1/(2\sqrt{n})$. Union bounding over j shows that it is sufficient to show

$$Q'_\varepsilon := \max_{w\in\mathbb{R}^n,\ |s|\leqslant 4\sqrt{n}} \mathbb{P}_A^{\mathcal{K}}\left(\exists v \in \Sigma'_\varepsilon :\ \|Av - sv - w\|_2 \leqslant 2^{-n+1}\right) \leqslant 2^{-6n}, \tag{III.10}$$

for all $\varepsilon \in [\sqrt{\eta}, \kappa_1/\sqrt{n}]$.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Step 2: A Bound on Q_ε: Take $w \in \mathbb{R}^n$ and $|s| \leqslant 4\sqrt{n}$; we will bound the probability uniformly over w and s. Since $\exp(-2c_\Sigma n) < \varepsilon < \kappa_0/8$, for $v \in \Sigma_\varepsilon$, we apply Lemma III.1, to find a $u \in \mathcal{N}_\varepsilon = \mathcal{N}_\varepsilon([d])$ so that $\|v-u\|_2 \leqslant 4\varepsilon$. So if $\|A\|_{op} \leqslant 4\sqrt{n}$, we see that

$$\begin{aligned}\|Au - su - w\|_2 &\leqslant \|Av - sv - w\|_2 + \|A(v-u)\|_2 + |s|\|v-u\|_2\\ &\leqslant \|Av - sv - w\|_2 + 8\sqrt{n}\|(v-u)\|_2\\ &\leqslant 33\varepsilon\sqrt{n},\end{aligned}$$

and thus

$$\{\exists v \in \Sigma_\varepsilon :\ \|Av - sv - w\|_2 \leqslant 2^{-n+1}\} \cap \{\|A\| \leqslant 4\sqrt{n}\} \subseteq \{\exists u \in \mathcal{N}_\varepsilon : \|Au - su - w\| \leqslant 33\varepsilon\sqrt{n}\}.$$

So, by union bounding over our net $\mathcal{N}_\varepsilon$, we see that

$$\begin{aligned}Q_\varepsilon \leqslant \mathbb{P}_A^{\mathcal{K}}\left(\exists v \in \mathcal{N}_\varepsilon : \|Av - sv - w\| \leqslant 33\varepsilon\sqrt{n}\right) &\leqslant \sum_{u \in \mathcal{N}_\varepsilon} \mathbb{P}_A^{\mathcal{K}}(\|Au - s'u - w\|_2 \leqslant 33\varepsilon\sqrt{n})\\ &\leqslant \sum_{u \in \mathcal{N}_\varepsilon} \mathcal{L}_{A,op}\left(u, 33\varepsilon\sqrt{n}\right),\end{aligned}$$

where $\mathcal{L}_{A,op}$ is defined at (III.3).

Note that for any u, we have that $\mathcal{L}_{A,op}\left(u, 33\varepsilon\sqrt{n}\right) \leqslant (67)^n \mathcal{L}_{A,op}(u, \varepsilon\sqrt{n})$ (see, e.g., Fact 6.2 in [4]); as such, for any $u \in \mathcal{N}_\varepsilon$, we have $\mathcal{L}_{A,op}\left(u, 33\varepsilon\sqrt{n}\right) \leqslant (2^{17}L\varepsilon)^n$. Using this bound gives

$$Q_\varepsilon \leqslant |\mathcal{N}_\varepsilon|(2^{17}L\varepsilon)^n \leqslant \left(\frac{C}{L^2\varepsilon}\right)^n (2^{17}L\varepsilon)^n \leqslant 2^{-4n},$$

where the penultimate inequality follows from our Theorem III.2 and the last inequality holds for the choice of L large enough relative to the universal constant C and so that (III.8) holds. To see that the application of Theorem III.2 is valid, note that

$$\log 1/\varepsilon \leqslant \log 1/\eta = 2c_\Sigma n \leqslant nL^{-32/c_0^2},$$

where the last inequality holds for c_Σ small compared to L^{-1}.

Step 3: A Bound on Q'_ε. To deal with Q'_ε, we employ a similar strategy. Fix $w \in \mathbb{R}^n$ and $|s| \leqslant 4\sqrt{n}$. Since we chose μ, γ to be sufficiently small so that Fact III.6 holds, we have that

$$\varepsilon \leqslant \kappa_1/\sqrt{n}.$$

Thus, we may apply Corollary III.5 with $K = \kappa_1$ for each $v \in \Sigma'_\varepsilon$ to get $u \in G'_\varepsilon \subset \Sigma'_\varepsilon$, such that $\|v-u\|_2 \leqslant 32\varepsilon\sqrt{\alpha n}$. Now, since

$$\{\exists v \in \Sigma'_\varepsilon :\ \|Av - sv - w\|_2 \leqslant 2^{-n+1}\} \cap \{\|A\| \leqslant 4\sqrt{n}\} \subseteq \{\exists u \in G'_\varepsilon : \|Au - su - w\| \leqslant 2^9\varepsilon\sqrt{\alpha n}\}$$

and since $2^9\varepsilon\sqrt{\alpha n} \geqslant \exp(-2c_\Sigma n) \geqslant \mathcal{T}_L(u)$, by Lemma III.7, we have

$$Q'_\varepsilon \leqslant \left(\frac{32\kappa_1}{\alpha^\mu\varepsilon\sqrt{n}}\right)^n \sup_{u \in G'_\varepsilon} \mathcal{L}(Au, 2^9\varepsilon\sqrt{\alpha n}) \leqslant (2^{20}L\kappa_1\alpha^{1/4})^n \leqslant 2^{-4n},$$

assuming that α is chosen to be sufficiently small relative to $L\kappa_1$. This completes the proof of Theorem I.2. □

IV. Fourier preparations for Theorem I.3

IV.1. Concentration, level sets, and Esseen-type inequalities

One of the main differences between this work and [4] is the notion of a "level set" of the Fourier transform, an change that requires us to make a fair number of small adjustments throughout. Here, we set up this definition along with a few related definitions.

For a random variable $Y \in \mathbb{R}^d$ and $\varepsilon > 0$, we recall that Lévy concentration of Y was defined at (III.2) by

$$\mathcal{L}(Y,\varepsilon) = \sup_{w\in\mathbb{R}^d} \mathbb{P}(\|Y-w\|_2 \leqslant \varepsilon).$$

Our goal is to compare the concentration of certain random vectors to the gaussian measure of associated (sub-)level sets. Given a $2d\times\ell$ matrix W, define the *W-level* set for $t \geqslant 0$ to be

$$S_W(t) := \{\theta\in\mathbb{R}^\ell : \mathbb{E}_{\bar\zeta}\,\|\bar\zeta W\theta\|_{\mathbb{T}}^2 \leqslant t\}\,. \tag{IV.1}$$

Let $g = g_d$ denote the gaussian random variable in dimension d with mean 0 and covariance matrix $(2\pi)^{-1}I_{d\times d}$. Define γ_d to be the corresponding measure, that is $\gamma_d(S) = \mathbb{P}_g(g\in S)$ for every Borel set $S\subset\mathbb{R}^d$. We first upper bound the concentration via an Esseen-like inequality.

Lemma IV.1. *Let $\beta > 0, \nu\in(0,1/4)$, let W be a $2d\times\ell$ matrix and $\tau\sim\Phi_\nu(2d;\zeta)$. Then there is an $m > 0$ so that*

$$\mathcal{L}(W^T\tau,\beta\sqrt{\ell}) \leqslant 2\exp\left(2\beta^2\ell - \nu pm/2\right)\gamma_\ell(S_W(m))\,.$$

Proof. For $w\in\mathbb{R}^\ell$, apply Markov's inequality to obtain

$$\mathbb{P}_\tau(\|W^T\tau - w\|_2 \leqslant \beta\sqrt{\ell}) \leqslant \exp\left(\frac{\pi}{2}\beta^2\ell\right)\mathbb{E}_\tau\exp\left(-\frac{\pi\|W^T\tau-w\|_2^2}{2}\right).$$

Using the Fourier transform of a gaussian, we compute

$$\mathbb{E}_\tau\exp\left(-\frac{\pi\|W^T\tau-w\|_2^2}{2}\right) = \mathbb{E}_g\, e^{-2\pi i\langle w,g\rangle}\mathbb{E}_\tau e^{2\pi i g^TW^T\tau}. \tag{IV.2}$$

Now, denote the rows of W as $w_1,\ldots,w_{2d}$ and write

$$\mathbb{E}_\tau e^{2\pi i g^TW^T\tau} = \prod_{i=1}^{2d}\mathbb{E}_{\tau_i}e^{2\pi i\sum\tau_i\langle g,w_i\rangle} = \prod_{i=1}^{2d}\phi_\tau(\langle g,w_i\rangle),$$

where $\phi_\tau(\theta)$ is the characteristic function of τ. Now, apply (II.3) and then (II.2) to see the right-hand side of (IV.2) is

$$\leqslant \left|\mathbb{E}_g\, e^{-2\pi i\langle w,g\rangle}\mathbb{E}_\tau e^{2\pi i g^TW^T\tau}\right| \leqslant \mathbb{E}_g\,\exp(-\nu p\mathbb{E}_{\bar\zeta}\|\bar\zeta Wg\|_{\mathbb{T}}^2).$$

We rewrite this as

$$\begin{aligned}\int_0^1 \mathbb{P}_g(\exp(-\nu p\mathbb{E}_{\bar\zeta}\|\bar\zeta Wg\|_{\mathbb{T}}^2) \geqslant t)\,dt &= \nu p\int_0^\infty \mathbb{P}_g(\mathbb{E}_{\bar\zeta}\|\bar\zeta Wg\|_{\mathbb{T}}^2 \leqslant u)e^{-\nu pu}\,du\\ &= \nu p\int_0^\infty \gamma_\ell(S_W(u))e^{-\nu pu}\,du\,,\end{aligned}$$

where for the first equality, we made the change of variable $t = e^{-\nu p u}$. Choosing m to maximize $\gamma_\ell(S_W(u))e^{-\nu pu/2}$ as a function of u yields

$$\nu p \int_0^\infty \gamma_\ell(S_W(u))e^{-\nu pu}du \leqslant \nu p \gamma_\ell(S_W(m))e^{-\nu pm/2} \int_0^\infty e^{-\nu pu/2}du = 2\gamma_\ell(S_W(m))e^{-\nu pm/2}.$$

Putting everything together, we obtain

$$\mathbb{P}_\tau(\|W^T\tau - w\|_2 \leqslant 2\beta\sqrt{\ell}) \leqslant 2e^{2\beta^2\ell}e^{-\nu pm/2}\gamma_\ell(S_W(m)).$$ □

We also prove a comparable lower bound.

Lemma IV.2. *Let $\beta > 0$, $\nu \in (0, 1/4)$, let W be a $2d \times \ell$ matrix, and let $\tau \sim \Xi_\nu(2d;\zeta)$. Then for all $t \geqslant 0$, we have*

$$\gamma_\ell(S_W(t))e^{-32\nu pt} \leqslant \mathbb{P}_\tau(\|W^T\tau\|_2 \leqslant \beta\sqrt{\ell}) + \exp\left(-\beta^2\ell\right).$$

Proof. Set $X = \|W^T\tau\|_2$, and write

$$\mathbb{E}_X e^{-\pi X^2/2} = \mathbb{E}_X \mathbf{1}(X \leqslant \beta\sqrt{\ell})e^{-\pi X^2/2} + \mathbb{E}_X \mathbf{1}\big(X \geqslant \beta\sqrt{\ell}\big)e^{-\pi X^2/2} \leqslant \mathbb{P}_X(X \leqslant \beta\sqrt{\ell}) + e^{-\pi\beta^2\ell/2}.$$

Bounding $\exp(-\pi\beta^2\ell/2) \leqslant \exp(-\beta^2\ell)$ implies

$$\mathbb{E}_\tau \exp\left(\frac{-\pi\|W^T\tau\|_2^2}{2}\right) \leqslant \mathbb{P}_\tau(\|W^T\tau\|_2 \leqslant \beta\sqrt{\ell}) + e^{-\beta^2\ell}.$$

As in the proof of Lemma IV.1 above, use the Fourier transform of the gaussian and (II.2) to lower bound

$$\mathbb{E}_\tau \exp\left(-\frac{\pi\|W^T\tau\|_2^2}{2}\right) \geqslant \mathbb{E}_g[\exp(-32\nu p\mathbb{E}_{\bar\zeta}\|\bar\zeta Wg\|_{\mathbb{T}}^2)].$$

Similar to the proof of Lemma IV.1, write

$$\mathbb{E}_g[\exp(-32\nu p\mathbb{E}_{\bar\zeta}\|Wg\|_{\mathbb{T}}^2)] = 32\nu p \int_0^\infty \gamma_\ell(S_W(u))e^{-32\nu pu}du \geqslant 32\nu p\gamma_\ell(S_W(t)) \int_t^\infty e^{-32\nu pu}\,du,$$

where we have used that $\gamma_\ell(S_W(b)) \geqslant \gamma_\ell(S_W(a))$ for all $b \geqslant a$. This completes the proof of Lemma IV.2. □

IV.2. Inverse Littlewood-Offord for conditioned random walks

First, we need a generalization of our important Lemma 3.1 from [4]. Given a $2d \times \ell$ matrix W and a vector $Y \in \mathbb{R}^d$, we define the Y-augmented matrix W_Y as

$$W_Y = \left[\ W\ ,\begin{bmatrix}\mathbf{0}_d\\ Y\end{bmatrix},\begin{bmatrix}Y\\ \mathbf{0}_d\end{bmatrix}\right]. \tag{IV.3}$$

When possible, we are explicit with the many necessary constants and "pin" several to a constant c_0, which we treat as a parameter to be taken sufficiently small. We also recall the definition of "least common denominator" $D_{\alpha,\gamma}$ from (I.1)

$$D_{\alpha,\gamma}(v) := \inf\left\{t > 0 : \|tv\|_{\mathbb{T}} < \min\{\gamma\|tv\|_2, \sqrt{\alpha n}\}\right\}.$$

The following is our generalization of Lemma 3.1 from [4].

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Lemma IV.3. *For any $0 < \nu \leqslant 2^{-15}$, $c_0 \leqslant 2^{-35}B^{-4}\nu$, $d \in \mathbb{N}$, $\alpha \in (0,1)$, and $\gamma \in (0,1)$, let $k \leqslant 2^{-32}B^{-4}\nu\alpha d$ and $t \geqslant \exp\left(-2^{-32}B^{-4}\nu\alpha d\right)$. Let $Y \in \mathbb{R}^d$ satisfy $\|Y\|_2 \geqslant 2^{-10}c_0\gamma^{-1}t^{-1}$, let W be a $2d \times k$ matrix with $\|W\| \leqslant 2$, $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$, and let $\tau \sim \Phi_\nu(2d;\zeta)$.*

If $D_{\alpha,\gamma}(Y) > 2^{10}B^2$, then

$$\mathcal{L}\left(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}\right) \leqslant (Rt)^2\exp\left(-c_0k\right)\,, \tag{IV.4}$$

where $R = 2^{35}B^2\nu^{-1/2}c_0^{-2}$.

We present the proof of Lemma IV.3 in Section VI, and deduce our standalone “inverse Littlewood-Offord theorem” Theorem I.3 here:

Proof of Theorem I.3. Let $c_0 = 2^{-35}B^{-4}\gamma^2\nu$. First, note that

$$\mathbb{P}\left(|\langle v,\tau\rangle| \leqslant t \text{ and } \sum_{i=1}^k \langle w_i,\tau\rangle^2 \leqslant c_0k\right)^2 \leqslant \mathbb{P}\left(|\langle v,\tau\rangle| \leqslant t\,, |\langle v,\tau'\rangle| \leqslant t \text{ and } \sum_{i=1}^k \langle w_i,\tau\rangle^2 \leqslant c_0k\right),$$

where $\tau,\tau' \sim \Phi_\nu(d;\zeta)$ are independent. We now look to bound the probability on the right-hand side using Lemma IV.3.

Let W be the $2d \times k$ matrix

$$W = \begin{bmatrix} w_1 & \dots & w_k \\ \mathbf{0_d} & \dots & \mathbf{0_d} \end{bmatrix}$$

and $Y = \sqrt{c_0/2}\,vt^{-1}$. Note that if $|\langle v,\tau\rangle| \leqslant t$, $|\langle v,\tau'\rangle| \leqslant t$ and $\sum_{i=1}^k\langle w_i,\tau\rangle^2 \leqslant c_0k$, then $\|W_Y^T(\tau,\tau')\|_2 \leqslant c_0^{1/2}\sqrt{k+1}$. Therefore

$$\mathbb{P}\left(|\langle v,\tau\rangle| \leqslant t\,, |\langle v,\tau'\rangle| \leqslant t \text{ and } \sum_{i=1}^k \langle w_i,\tau\rangle^2 \leqslant c_0k\right) \leqslant \mathcal{L}\left(W_Y^T(\tau,\tau'), c_0^{1/2}\sqrt{k+1}\right).$$

Now, $\|Y\|_2 = \sqrt{c_0/2}\,t^{-1} > 2^{-10}c_0\gamma^{-1}t^{-1}$, $\|W\| = 1$, $\|W\|_{\mathrm{HS}} = \sqrt{k}$, and

$$D_{\alpha,\gamma}(Y) \geqslant tc_0^{-1/2}D_{\alpha,\gamma}(v) > 2^{10}B^2.$$

We may therefore apply Lemma IV.3 to bound

$$\mathcal{L}\left(W_Y^T(\tau,\tau'), c_0^{1/2}\sqrt{k+1}\right) \leqslant (Rt)^2\exp\left(-c_0k\right).$$

The result follows. □

V. Fourier replacement

The goal of this section is to prove Lemma III.7, which relates the “zeroed out and lazy” matrix M, defined at (II.6), to our original matrix A. We will need a few inequalities on the Fourier side first.

Lemma V.1. *For every $t \in \mathbb{R}$ and $\nu \leqslant 1/4$, we have*

$$|\phi_\zeta(t)| \leqslant \phi_{\tilde{\zeta}Z_\nu}(t)\,.$$

Proof. Note $|\phi_\zeta(t)|^2 = \mathbb{E}_{\tilde{\zeta}}\cos(2\pi t\tilde{\zeta})$. Use the elementary inequality

$$\cos(a) \leqslant 1 - 2\nu(1-\cos(a)) \qquad \text{for } \nu \leqslant 1/4,$$

and that $\sqrt{1-x} \leqslant 1 - x/2$ to bound

$$|\phi_\zeta(t)| = \sqrt{\mathbb{E}_{\tilde{\zeta}} \cos(2\pi t \tilde{\zeta})} \leqslant \sqrt{1 - 2\nu \mathbb{E}_{\tilde{\zeta}}(1 - \cos(2\pi t \tilde{\zeta}))} \leqslant \phi_{\tilde{\zeta} Z_\nu}(t)\,. \qquad \square$$

We also need a bound on a gaussian-type moment for $\|Mv\|_2$. On a somewhat technical point, we notice that $\mathcal{T}_L(v) \geqslant 2^n$, since the definition of $\mathcal{T}_L$ (II.7) depends on the definition of M at (II.6), which trivially satisfies

$$\mathbb{P}_M(Mv = 0) \geqslant \mathbb{P}_M(M = 0) = (1-\nu)^{\binom{n+1}{2}},$$

for all v and $\nu < 1/2$.

Fact V.2. For $v \in \mathbb{R}^n$, and $t \geqslant \mathcal{T}_L(v)$, we have

$$\mathbb{E} \exp(-\pi \|Mv\|_2^2 / 2t^2) \leqslant (9Lt)^n.$$

Proof. Bound

$$\mathbb{E} \exp(-\pi \|Mv\|_2^2/2t^2) \leqslant \mathbb{P}(\|Mv\|_2 \leqslant t\sqrt{n}) + \sqrt{n} \int_t^\infty e^{-s^2 n/t^2} \mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n})\, ds\,. \tag{V.1}$$

Since $t \geqslant \mathcal{T}_L(v)$, we have $\mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n}) \leqslant (4Ls)^n$ for all $s \geqslant t$. Thus, we may bound

$$\sqrt{n} \int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right) \mathbb{P}(\|Mv\|_2 \leqslant s\sqrt{n})\, ds \leqslant \sqrt{n}(8Lt)^n \int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right) (s/t)^n\, ds\,.$$

Changing variables $u = s/t$, we may bound the right-hand side by

$$t^{-1}\sqrt{n}(4Lt)^n \int_1^\infty \exp(-u^2 n) u^n\, du \leqslant t^{-1}\sqrt{n}(4Lt)^n \int_1^\infty \exp(-u^2/2)\, du \leqslant (9Lt)^n,$$

as desired. Note, here, that we used that $t \geqslant 2^{-n}$. $\square$

For $v, x \in \mathbb{R}^n$ and $\nu \in (0, 1/4)$, define the characteristic functions of Av and Mv, respectively, ψ_v and $\chi_{v,\nu}$, by

$$\psi_v(x) := \mathbb{E}_A\, e^{2\pi i \langle Av, x\rangle} = \left(\prod_{k=1}^n \phi_\zeta(v_k x_k)\right)\left(\prod_{j<k} \phi_\zeta(x_j v_k + x_k v_j)\right)$$

and

$$\chi_v(x) := \mathbb{E}_M\, e^{2\pi i \langle Mv, x\rangle} = \prod_{j=1}^d \prod_{k=d+1}^n \phi_{\tilde{\zeta} Z_\nu}(x_j v_k + x_k v_j)\,.$$

Our “replacement” now goes through.

Proof of Lemma III.7. By Markov, we have

$$\mathbb{P}(\|Av - w\|_2 \leqslant t\sqrt{n}) \leqslant \exp(\pi n/2) \mathbb{E} \exp\left(-\pi \|Av - w\|_2^2/(2t^2)\right)\,. \tag{V.2}$$

Then use Fourier inversion to write

$$\mathbb{E}_A \exp\left(-\pi \|Av - w\|_2^2/(2t^2)\right) = \int_{\mathbb{R}^n} e^{-\pi \|\xi\|_2^2} \cdot e^{-2\pi i t^{-1} \langle w, \xi\rangle} \psi_v(t^{-1}\xi)\, d\xi\,. \tag{V.3}$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Now, apply the triangle inequality, Lemma V.1 and the nonnegativity of χ_v yield that the right-hand side of (V.3) is

$$\leqslant \int_{\mathbb{R}^n} e^{-\pi\|\xi\|_2^2}\chi_v(t^{-1}\xi)\,d\xi = \mathbb{E}_M \exp(-\pi\|Mv\|_2^2/2t^2)\,.$$

Now, use Fact V.2 along with the assumption $t \geqslant \mathcal{T}_L(v)$ to bound

$$\mathbb{E}_M \exp(-\pi\|Mv\|_2^2/2t^2) \leqslant (9Lt)^n,$$

as desired. □

VI. Proof of Lemma IV.3

In this section, we prove the crucial Lemma IV.3. Fortunately, much of the geometry needed to prove this theorem can be pulled from the proof of the $\{-1,0,1\}$-case in [4], and so the deduction of the theorem becomes relatively straightforward.

VI.1. Properties of gaussian space and level sets

For $r, s > 0$ and $k \in \mathbb{N}$, define the *cylinder* $\Gamma_{r,s}$ by

$$\Gamma_{r,s} := \left\{\theta \in \mathbb{R}^{k+2} : \left\|\theta_{[k]}\right\|_2 \leqslant r, |\theta_{k+1}| \leqslant s \text{ and } |\theta_{k+2}| \leqslant s\right\}. \tag{VI.1}$$

For a measurable set $S \subset \mathbb{R}^{k+2}$ and $y \in \mathbb{R}^{k+2}$, define the set

$$F_y(S; a, b) := \{\theta_{[k]} = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k : (\theta_1, \dots, \theta_k, a, b) \in S - y\}\,.$$

Recall that γ_k is the k-dimensional gaussian measure defined by $\gamma_k(S) = \mathbb{P}(g \in S)$, where $g \sim \mathcal{N}(0, (2\pi)^{-1}I_k)$, and where I_k denotes the $k \times k$ identity matrix. The following is a key geometric lemma from [4].

Lemma VI.1. *Let $S \subset \mathbb{R}^{k+2}$ and $s > 0$ satisfy*

$$8s^2e^{-k/8} + 32s^2 \max_{a,b,y} \left(\gamma_k(F_y(S;a,b) - F_y(S;a,b))\right)^{1/4} \leqslant \gamma_{k+2}(S)\,. \tag{VI.2}$$

Then there is an $x \in S$ so that

$$(\Gamma_{2\sqrt{k},16} \setminus \Gamma_{2\sqrt{k},s} + x) \cap S \neq \emptyset\,.$$

This geometric lemma will be of crucial importance for identifying the LCD. Indeed, we will take S to be a representative level set, on the Fourier side, for the probability implicit on the left-hand side of Lemma IV.3. The following basic fact will help explain the use of the difference appearing in Lemma VI.1.

Fact VI.2. For any $2d \times \ell$ matrix W and $m > 0$, we have

$$S_W(m) - S_W(m) \subseteq S_W(4m)\,.$$

Proof. For any $x, y \in S_W(m)$, we have $\mathbb{E}_{\bar\zeta}\|\bar\zeta Wx\|_{\mathbb{T}}^2, \mathbb{E}_{\bar\zeta}\|\bar\zeta Wy\|_{\mathbb{T}}^2 \leqslant m$. The triangle inequality implies

$$\|\bar\zeta W(x-y)\|_{\mathbb{T}}^2 \leqslant 2\|\bar\zeta Wx\|_{\mathbb{T}}^2 + 2\|\bar\zeta Wy\|_{\mathbb{T}}^2\,.$$

Taking $\mathbb{E}_{\bar\zeta}$ on both sides completes the fact. □

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

VI.2. Proof of Lemma IV.3

The following is our main step toward Lemma IV.3.

Lemma VI.3. *For $d \in \mathbb{N}$, $\gamma, \alpha \in (0,1)$ and $0 < \nu \leqslant 2^{-15}$, let $k \leqslant 2^{-17}B^{-4}\nu\alpha d$ and $t \geqslant \exp(-2^{-17}B^{-4}\nu\alpha d)$. For $c_0 \in (0, 2^{-50}B^{-4})$, let $Y \in \mathbb{R}^d$ satisfy $\|Y\| \geqslant 2^{-10}c_0\gamma^{-1}/t$ and let W be a $2d \times k$ matrix with $\|W\| \leqslant 2$.*

Let $\tau \sim \Xi_\nu(2d;\zeta)$ and $\tau' \sim \Xi_{\nu}(2d;\zeta)$ with $\nu = 2^{-7}\nu$, and let $\beta \in [c_0/2^{10}, \sqrt{c_0}]$ and $\beta' \in (0, 1/2)$. If

$$\mathcal{L}(W_Y^T\tau, \beta\sqrt{k+1}) \geqslant (Rt)^2 \exp(4\beta^2 k)\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4}, \tag{VI.3}$$

then $D_{\alpha,\gamma}(Y) \leqslant 2^{10}B^2$. Here, we have set $R = 2^{35}\nu^{-1/2}B^2/c_0^2$.

Proof. By Lemma IV.1, we may find an m for which the level set $S = S_{W_Y}(m)$ satisfies

$$\mathcal{L}(W_Y^T\tau, \beta\sqrt{k+1}) \leqslant 4e^{-\nu pm/2+2\beta^2 k}\gamma_{k+2}(S). \tag{VI.4}$$

Combining (VI.4) with the assumption (VI.3) provides a lower bound of

$$\gamma_{k+2}(S) \geqslant \frac{1}{4}e^{\nu pm/2+2\beta^2 k}\,(Rt)^2\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4}. \tag{VI.5}$$

Now, preparing for an application of Lemma VI.1, define

$$r_0 := \sqrt{k} \qquad \text{and} \qquad s_0 := 2^{16}c_0^{-1}(\sqrt{m} + 8B^2\sqrt{k})t\,. \tag{VI.6}$$

Recalling the definition of our *cylinders* from (VI.1), we state the following claim:

Claim VI.4. There exists $x \in S \subseteq \mathbb{R}^{k+2}$ so that

$$\left(\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0} + x\right) \cap S \neq \emptyset\,. \tag{VI.7}$$

Proof of Claim VI.4. We will use Lemma VI.1 with $s = s_0$, and so we check the hypotheses. We first observe that for any y, a, b, if $\theta_{[k]}, \theta'_{[k]} \in F_y(S; a, b)$, then we have

$$\theta'' := (\theta_1 - \theta'_1, \ldots, \theta_k - \theta'_k, 0, 0) \in S_{W_Y}(4m)$$

by Fact VI.2. This shows that for any y, a, b, we have

$$F_y(S; a, b) - F_y(S; a, b) \subset S_{W_Y}(4m) \cap \{\theta \in \mathbb{R}^{k+2} : \theta_{k+1} = \theta_{k+2} = 0\} = S_W(4m)\,, \tag{VI.8}$$

where the equality holds by definition of W_Y and the level set S_{W_Y}. Thus, we may apply Lemma IV.2 to obtain

$$\gamma_k(S_W(4m)) \leqslant e^{128\nu pm}\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right). \tag{VI.9}$$

Combining lines (VI.5), (VI.8), and (VI.9), we note that in order to apply Lemma VI.1, it is sufficient to check

$$\begin{aligned} 8s_0^2e^{-k/8} + 32s_0^2e^{32\nu pm}&\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4} \\ &< \frac{1}{4}e^{\nu pm/2+2\beta^2 k}\,(Rt)^2\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4}. \end{aligned} \tag{VI.10}$$

We will show that each term on the left-hand side of (VI.10) is at most half of the right-hand side. Bound

$$s_0^2 = 2^{32}c_0^{-2}(\sqrt{m} + 8B^2\sqrt{k})^2t^2 < 2^{33}(m + 64B^4k)(t/c_0)^2 \leqslant 2^{-20}\nu(c_0^2k + (2B)^{-6}m)(Rt)^2 \qquad \text{(VI.11)}$$

since $R = 2^{35}B^2\nu^{-1/2}c_0^{-2}$. By Lemma II.1, we have that $p \geqslant 2^{-7}B^{-4}$ and so we may bound

$$8s_0^2e^{-k/8} \leqslant e^{-k/8}2^{-17}\nu(c_0^2k + (2B)^{-4}m)(Rt)^2 \leqslant \frac{1}{8}e^{\nu pm/2}(Rt)^2e^{-\beta'^2k/4}.$$

Similarly, use (VI.11), $c_0 \leqslant \beta$ and $\nu = 2^{-7}\nu$ to bound

$$32s_0^2e^{32\nu pm} \leqslant 2^{-15}(c_0^2k + (2B)^{-4}m)(Rt)^2\exp(\nu pm/4) \leqslant \frac{1}{8}(Rt)^2e^{\nu pm/2+\beta^2k},$$

thus showing (VI.10). Applying Lemma VI.1 completes the claim. □

The following basic consequence of Claim VI.4 will bring us closer to the construction of our LCD:

Claim VI.5. We have that $S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0}) \neq \emptyset$.

Proof of Claim VI.5. Claim VI.4 shows that there exists $x, y \in S = S_{W_Y}(m)$ so that $y \in (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0} + x)$. Now define $\phi := y - x$, and note that $\phi \in S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0})$ due to Fact VI.2. □

We now complete the proof of Lemma VI.3 by showing that an element of the nonempty intersection above provides an LCD.

Claim VI.6. If $\phi \in S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0})$, then there is a $\bar{\zeta}_0 \in (1, 16B^2)$ and $i \in \{k+1, k+2\}$ so that

$$\|\bar{\zeta}_0\phi_iY\|_{\mathbb{T}} < \min\{\gamma\bar{\zeta}_0\phi_i\|Y\|_2, \sqrt{\alpha d}\}.$$

Proof of Claim VI.6. Note that since $\phi \in S_{W_Y}(4m)$, we have

$$\mathbb{E}_{\bar{\zeta}}\|\bar{\zeta}W_Y\phi\|_{\mathbb{T}}^2 \leqslant 4m.$$

Thus, there is some instance $\bar{\zeta}_0 \in (1, 16B^2)$ of $\bar{\zeta}$ so that

$$\|\bar{\zeta}_0W_Y\phi\|_{\mathbb{T}}^2 \leqslant 4m. \qquad \text{(VI.12)}$$

For simplicity, define $\psi =: \bar{\zeta}_0\phi$.

By (VI.12), there is a $z \in \mathbb{Z}^{2d}$ so that $W_Y\psi \in B_{2d}(z, 2\sqrt{m})$. Expand

$$W_Y\psi = W\psi_{[k]} + \psi_{k+1}\begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} + \psi_{k+2}\begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix},$$

and note that

$$\psi_{k+1}\begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} + \psi_{k+2}\begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix} \in B_{2d}(z, 2\sqrt{m}) - W\psi_{[k]} \subseteq B_{2d}(z, 2\sqrt{m} + 2^6B^2\sqrt{k}), \qquad \text{(VI.13)}$$

where the last inclusion holds because

$$\|W\psi_{[k]}\|_2 \leqslant \|W\|_{op}\|\psi_{[k]}\|_2 \leqslant 2|\bar{\zeta}_0|\|\phi_{[k]}\|_2 \leqslant 32\sqrt{k}B^2,$$

since $\phi \in \Gamma_{2r_0,16}$, $|\bar{\zeta}_0| \leqslant 16B^2$, and $\|W\|_{op} \leqslant 2$.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Since $\phi \notin \Gamma_{2r_0,s_0}$ and $\bar{\zeta}_0 > 1$, we have $\max\{|\psi_{k+1}|, |\psi_{k+2}|\} > s_0$, and so we assume, without loss, that $|\psi_{k+1}| > s_0$. Projecting (VI.13) onto the first d coordinates yields

$$\psi_{k+1} Y \in B_d(z_{[d]}, 2\sqrt{m} + 2^6 B^2 \sqrt{k}). \tag{VI.14}$$

Now, we show that $\|\psi_{k+1} Y\|_{\mathbb{T}} < \gamma \psi_{k+1} \|Y\|_2$. Indeed,

$$\psi_{k+1}\|Y\|_2\gamma \geqslant s_0\|Y\|_2\gamma > \left(\frac{2^{15}(\sqrt{m} + 8B^2\sqrt{k})t}{c_0}\right)\left(2^{-10}\frac{c_0}{t}\right) \geqslant (2\sqrt{m} + 2^6 B^2\sqrt{k}), \tag{VI.15}$$

where we used the definition of s_0 and that $\|Y\|_2 > 2^{-10} c_0 \gamma^{-1}/t$.

We now need to show

$$2\sqrt{m} + 2^6 B^2 \sqrt{k} \leqslant \sqrt{\alpha d}. \tag{VI.16}$$

Note that since $k \leqslant 2^{-32}\alpha d/B^4$, we have $2^8 B^2 \sqrt{k} \leqslant \sqrt{\alpha d}/2$. We claim that $m \leqslant 2^{-4}\alpha d$. To show this, apply the lower bound (VI.5) and $\gamma_{k+2}(S) \leqslant 1$ to see

$$e^{-2^{-11}\nu m/B^4} \geqslant e^{-\nu p m/2} \geqslant \gamma_{k+2}(S) e^{-\nu p m/2} \geqslant (Rt)^2 e^{-2\beta'^2 k} \geqslant t^2 e^{-k} \geqslant e^{-2^{-15}\nu\alpha d/B^4},$$

where we have used $k \leqslant 2^{-17}\nu\alpha d/B^4$ and $t \geqslant e^{-2^{-17}\nu\alpha d/B^4}$. Therefore, $m \leqslant 2^{-4}\alpha d$, that is $2\sqrt{m} \leqslant \sqrt{\alpha d}/2$. Combining this with (VI.14) and (VI.15), we see

$$\|\psi_{k+1} Y\|_{\mathbb{T}} \leqslant \sqrt{\alpha d},$$

as desired. This completes the proof of the Claim VI.6. □

Let ϕ, $\bar{\zeta}_0$, and $i \in \{k+1, k+2\}$ be as guaranteed by Claim VI.6. Then $\bar{\zeta}_0 \phi_i \leqslant 2^{10} B^2$, and

$$\|\bar{\zeta}_0 \phi_i Y\|_{\mathbb{T}} < \min\{\|\bar{\zeta}_0 \phi_i Y\|_2 \gamma, \sqrt{\alpha d}\},$$

and so $D_{\alpha,\gamma}(Y) \leqslant 2^{10} B^2$, thus completing the proof of Lemma VI.3. □

VI.3. Proof of Lemma IV.3

In order to bridge the gap between Lemmas VI.3 and IV.3, we need an anticoncentration lemma for $\|W\sigma\|_2$ when σ is random and W is fixed. We will use the following bound, which is a version of the Hanson-Wright inequality [18, 33].

Lemma VI.7. *Let $\nu \in (0,1)$ and $\beta' \in (0, 2^{-7} B^{-2}\sqrt{\nu})$. Let W be a $2d \times k$ matrix satisfying $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$ and $\|W\| \leqslant 2$ and $\tau' \sim \Xi_\nu(2d;\zeta)$. Then*

$$\mathbb{P}(\|W^T \tau'\|_2 \leqslant \beta'\sqrt{k}) \leqslant 4\exp\left(-2^{-20} B^{-4}\nu k\right).$$

We derive Lemma VI.7 from Talagrand's inequality in Section X, (see [33] or [18] for more context). From here, we are ready to prove Lemma IV.3.

Proof of Lemma IV.3. Recalling that $c_0 \leqslant 2^{-35} B^{-4}\nu$, and that our given W satisfies $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$ and $\|W\| \leqslant 2$, we apply Lemma VI.7, with $\beta' = 2^6\sqrt{c_0}$ and the ν-lazy random vector $\tau' \sim \Xi_\nu(2d;\zeta)$, where $\nu = 2^{-7}\nu$, to see

$$\mathbb{P}(\|W^T \tau'\|_2 \leqslant \beta'\sqrt{k}) \leqslant 4\exp\left(-2^{-27} B^{-4}\nu k\right) \leqslant 4\exp(-32 c_0 k).$$

We now consider the right-hand side of (VI.3) in Lemma VI.3: if $\beta \leqslant \sqrt{c_0}$, we have

$$e^{4\beta^2 k}\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant \beta'\sqrt{k}) + \exp(-\beta'^2 k)\right)^{1/4} \leqslant \exp(4c_0k - 8c_0k) + \exp(4c_0k - 16c_0k)$$
$$\leqslant 2\exp(-c_0k)\,.$$

We now note that the hypotheses in Lemma IV.3 align with the hypotheses in Lemma VI.3 with respect to the selection of $\beta, \alpha, t, R, Y, W$; if we additionally assume $D_{\alpha,\gamma}(Y) > 2^{10}B^2$, we may apply the contrapositive of Lemma VI.3 to obtain

$$\mathcal{L}\left(W_Y^T\tau, \beta\sqrt{k+1}\right) \leqslant (2^{35}B^2\nu^{-1/2}c_0^{-2}t/2)^2 e^{4\beta^2 k}\left(\mathbb{P}(\|W^T\tau'\|_2 \leqslant 2\beta'\sqrt{k}) + e^{-\beta'^2 k}\right)^{1/4}$$
$$\leqslant (Rt)^2\exp(-c_0k)\,,$$

as desired. □

VII. Inverse Littlewood-Offord for conditioned matrix walks

In this section, we prove an inverse Littlewood-Offord theorem for matrices conditioned on their robust rank. Everything in this section will be analogous to Section 6 of [4].

Theorem VII.1. *For $n \in \mathbb{N}$ and $0 < c_0 \leqslant 2^{-50}B^{-4}$, let $d \leqslant c_0^2 n$, and for $\alpha, \gamma \in (0,1)$, let $0 \leqslant k \leqslant 2^{-32}B^{-4}\alpha d$ and $N \leqslant \exp(2^{-32}B^{-4}\alpha d)$. Let $X \in \mathbb{R}^d$ satisfy $\|X\|_2 \geqslant c_0 2^{-10}\gamma^{-1}n^{1/2}N$, and let H be a random $(n-d)\times 2d$ matrix with i.i.d. rows sampled from $\Phi_\nu(2d;\zeta)$ with $\nu = 2^{-15}$. If $D_{\alpha,\gamma}(r_n \cdot X) > 2^{10}B^2$, then*

$$\mathbb{P}_H\left(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n} \text{ and } \|H_1X\|_2, \|H_2X\|_2 \leqslant n\right) \leqslant e^{-c_0 nk/3}\left(\frac{R}{N}\right)^{2n-2d}, \tag{VII.1}$$

where we have set $H_1 := H_{[n-d]\times[d]}$, $H_2 := H_{[n-d]\times[d+1,2d]}$, $r_n := \frac{c_0}{32\sqrt{n}}$ and $R := 2^{43}B^2c_0^{-3}$.

VII.1. Tensorization and random rounding step

We import the following tensorization lemma from [4].

Lemma VII.2. *For $d < n$ and $k \geqslant 0$, let W be a $2d \times (k+2)$ matrix and let H be a $(n-d)\times 2d$ random matrix with i.i.d. rows. Let $\tau \in \mathbb{R}^{2d}$ be a random vector with the same distribution as the rows of H. If $\beta \in (0, 1/8)$, then*

$$\mathbb{P}_H\left(\|HW\|_{\mathrm{HS}} \leqslant \beta^2\sqrt{(k+1)(n-d)}\right) \leqslant \left(2^5 e^{2\beta^2 k}\mathcal{L}(W^T\tau, \beta\sqrt{k+1})\right)^{n-d}.$$

Similarly, we use net for the singular vectors of H, constructed in [4]. Let $\mathcal{U}_{2d,k} \subset \mathbb{R}^{[2d]\times[k]}$ be the set of $2d \times k$ matrices with orthonormal columns.

Lemma VII.3. *For $k \leqslant d$ and $\delta \in (0, 1/2)$, there exists $\mathcal{W} = \mathcal{W}_{2d,k} \subset \mathbb{R}^{[2d]\times[k]}$ with $|\mathcal{W}| \leqslant (2^6/\delta)^{2dk}$ so that for any $U \in \mathcal{U}_{2d,k}$, any $r \in \mathbb{N}$, and $r \times 2d$ matrix A, there exists $W \in \mathcal{W}$ so that*

1. $\|A(W-U)\|_{\mathrm{HS}} \leqslant \delta(k/2d)^{1/2}\|A\|_{\mathrm{HS}}$,
2. $\|W-U\|_{\mathrm{HS}} \leqslant \delta\sqrt{k}$, *and*
3. $\|W-U\|_{op} \leqslant 8\delta$.

VII.2. Proof of Theorem VII.1

We also use the following standard fact from linear algebra.

Fact VII.4. For $3d < n$, let H be a $(n-d)\times 2d$ matrix. If $\sigma_{2d-k+1}(H) \leqslant x$, then there exist k orthogonal unit vectors $w_1, \dots, w_k \in \mathbb{R}^{2d}$ so that $\|Hw_i\|_2 \leqslant x$. In particular, there exists $W \in \mathcal{U}_{2d,k}$ so that $\|HW\|_{\mathrm{HS}} \leqslant x\sqrt{k}$.

We will also need a bound on $\|H\|_{\mathrm{HS}}$:

Fact VII.5. Let H be the random $(n-d)\times(2d)$ matrix whose rows are i.i.d. samples of $\Phi_\nu(2d;\zeta)$. Then

$$\mathbb{P}(\|H\|_{\mathrm{HS}} \geqslant 2\sqrt{d(n-d)}) \leqslant 2\exp\left(-2^{-21}B^{-4}nd\right).$$

We are now ready to prove Theorem VII.1.

Proof of Theorem VII.1. Let $Y := \frac{c_0}{32\sqrt{n}} \cdot X$. We may upper bound the left-hand side of (VII.1) by Fact VII.4

$$\mathbb{P}(\sigma_{2d-k+1}(H) \leqslant c_0 2^{-4}\sqrt{n} \text{ and } \|H_1X\|_2, \|H_2X\|_2 \leqslant n)$$
$$\leqslant \mathbb{P}(\exists U \in \mathcal{U}_{2d,k} : \|HU_Y\|_{\mathrm{HS}} \leqslant c_0\sqrt{n(k+1)}/8).$$

Set $\delta := c_0/16$, and let $\mathcal{W}$ be as in Lemma VII.3.

For each fixed H, if we have $\|H\|_{\mathrm{HS}} \leqslant 2\sqrt{d(n-d)}$ and there is some $U \in \mathcal{U}_{2d,k}$ so that $\|HU_Y\|_{\mathrm{HS}} \leqslant c_0\sqrt{n(k+1)}/8$, we may apply Lemma VII.3 to find $W \in \mathcal{W}$ so that

$$\|HW_Y\|_{\mathrm{HS}} \leqslant \|H(W_Y - U_Y)\|_{\mathrm{HS}} + \|HU_Y\|_{\mathrm{HS}} \leqslant \delta(k/2d)^{1/2}\|H\|_{\mathrm{HS}} + c_0\sqrt{n(k+1)}/8$$

which is at most $c_0\sqrt{n(k+1)}/4$. This shows the bound

$$\mathbb{P}_H\left(\exists U \in \mathcal{U}_{2d,k} :\ \|HU_Y\|_{\mathrm{HS}} \leqslant c_0\sqrt{n(k+1)}/8\right) \leqslant \mathbb{P}_H\left(\exists W \in \mathcal{W} : \|HW_Y\|_{\mathrm{HS}} \leqslant c_0\sqrt{n(k+1)}/4\right).$$

Conditioning on the event that $\|H\|_{\mathrm{HS}} \leqslant 2\sqrt{d(n-d)}$, applying Fact VII.5, and union bounding over $\mathcal{W}$ show that the right-hand side of the above is at most

$$\sum_{W\in\mathcal{W}} \mathbb{P}_H\left(\|HW_Y\|_2 \leqslant c_0\sqrt{n(k+1)}/4\right) + 2\exp\left(-2^{-21}B^{-4}nd\right).$$

Bound

$$|\mathcal{W}| \leqslant (2^6/\delta)^{2dk} \leqslant \exp(32dk\log c_0^{-1}) \leqslant \exp(c_0k(n-d)/6),$$

where the last inequality holds since $d \leqslant c_0^2 n$. Thus

$$\sum_{W\in\mathcal{W}} \mathbb{P}_H(\|HW_Y\|_2 \leqslant c_0\sqrt{n(k+1)}/4) \leqslant \exp(c_0k(n-d)/6)\max_{W\in\mathcal{W}} \mathbb{P}_H(\|HW\|_2 \leqslant c_0\sqrt{n(k+1)}/4). \tag{VII.2}$$

For each $W \in \mathcal{W}$, apply Lemma VII.2 with $\beta := \sqrt{c_0/3}$ (noting that $\sqrt{n-d}/3 \geqslant \sqrt{n}/4$) to obtain

$$\mathbb{P}_H(\|HW_Y\|_2 \leqslant c_0\sqrt{n(k+1)}/4) \leqslant \left(2^5 e^{2c_0k/3}\mathcal{L}(W_Y^T\tau, c_0^{1/2}\sqrt{k+1})\right)^{n-d}. \tag{VII.3}$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Preparing to apply Lemma IV.3, define $t := (c_0 N/32)^{-1} \geqslant \exp(-2^{-32}B^{-4}\alpha d)$ and $R_0 := 2^{-8}c_0 R = 2^{-8}c_0(2^{43}B^2 c_0^{-3}) = 2^{35}B^2 c_0^{-2}$ so that we have

$$\|Y\|_2 = c_0\|X\|_2/(32n^{1/2}) \geqslant 2^{-15}c_0^2 N\gamma^{-1} = 2^{-10}c_0\gamma^{-1}/t\,.$$

Since $W \in \mathcal{W}$, we have $\|W\|_{op} \leqslant 2$ and $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$. We also note the bounds $k \leqslant 2^{-32}B^{-4}\alpha d$, $D_{\alpha,\gamma}(\frac{c_0}{32\sqrt{n}}X) = D_{\alpha,\gamma}(Y) > 2^{10}B^2$. Thus, we may apply Lemma IV.3 to see that

$$\mathcal{L}(W_Y^T\tau, c_0^{1/2}\sqrt{k+1}) \leqslant (R_0 t)^2 e^{-c_0 k} \leqslant \left(\frac{R}{8N}\right)^2 e^{-c_0 k}\,.$$

Substituting this bound into (VII.3) gives

$$\max_{W\in\mathcal{W}} \mathbb{P}_H(\|HW_Y\|_2 \leqslant c_0\sqrt{n(k+1)}/4) \leqslant \frac{1}{2}\left(\frac{R}{N}\right)^{2n-2d} e^{-c_0 k(n-d)/3}\,.$$

Combining with the previous bounds and noting

$$2\exp\left(-2^{-21}B^{-4}nd\right) \leqslant \frac{1}{2}\left(\frac{R}{N}\right)^{2n-2d} e^{-c_0 k(n-d)/3}$$

show

$$\mathbb{P}(\sigma_{2d-k+1}(H) \leqslant c_0\sqrt{n}/16 \text{ and } \|H_1X\|_2, \|H_2X\|_2 \leqslant n) \leqslant \left(\frac{R}{N}\right)^{2n-2d} e^{-c_0 k(n-d)/3}\,.$$

This completes the proof of Theorem VII.1. □

VIII. Nets for structured vectors: Size of the net

The goal of this subsection is to prove Theorem III.2. We follow the same path as Section 7 of [4]. As such, we work with the intersection of $\mathcal{N}_\varepsilon$ with a selection of "boxes" which cover a rescaling of the trivial net Λ_ε. We recall the definition of the relevant boxes from [4].

Definition VIII.1. Define a (N,κ,d)-*box* to be a set of the form $\mathcal{B} = B_1 \times \ldots \times B_n \subset \mathbb{Z}^n$, where $|B_i| \geqslant N$ for all $i \geqslant 1$; $B_i = [-\kappa N, -N] \cup [N, \kappa N]$, for $i \in [d]$; and $|\mathcal{B}| \leqslant (\kappa N)^n$.

We now interpret these boxes probabilistically and seek to understand the probability that we have

$$\mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant \left(\frac{L}{N}\right)^n,$$

where X is chosen uniformly at random from $\mathcal{B}$. Theorem III.2 will follow quickly from the following "box" version:

Lemma VIII.2. *For $L \geqslant 2$ and $0 < c_0 \leqslant 2^{-50}B^{-4}$, let $n > L^{64/c_0^2}$ and let $\frac{1}{4}c_0^2 n \leqslant d \leqslant c_0^2 n$. For $N \geqslant 2$, satisfying $N \leqslant \exp(c_0 L^{-8n/d}d)$, and $\kappa \geqslant 2$, let $\mathcal{B}$ be a (N,κ,d)-box. If X is chosen uniformly at random from $\mathcal{B}$, then*

$$\mathbb{P}_X\left(\mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant \left(\frac{L}{N}\right)^n\right) \leqslant \left(\frac{R}{L}\right)^{2n},$$

where $R := Cc_0^{-3}$ and $C > 0$ is an absolute constant.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

VIII.1. Counting with the LCD and anticoncentration for linear projections of random vectors

We first show that if we choose $X \in \mathcal{B}$ uniformly at random, then it typically has a large LCD.

Lemma VIII.3. *For $\alpha \in (0,1)$, $K \geqslant 1$, and $\kappa \geqslant 2$, let $n \geqslant d \geqslant K^2/\alpha$ and let $N \geqslant 2$ be so that $KN < 2^d$. Let $\mathcal{B} = ([-\kappa N, -N] \cup [N, \kappa N])^d$, and let X be chosen uniformly at random from $\mathcal{B}$. Then*

$$\mathbb{P}_X\left(D_{\alpha,\gamma}(r_n X) \leqslant K\right) \leqslant (2^{20}\alpha)^{d/4}, \tag{VIII.1}$$

where we have set $r_n := c_0 2^{-5} n^{-1/2}$.

Proof. Writing $\phi = \psi r_n$, note that

$$\mathbb{P}_X\left(D_{\alpha,\gamma}(r_n X) \leqslant K\right) = \mathbb{P}\left(\exists\, \phi \in (0, K r_n] : \|\phi X\|_{\mathbb{T}} < \min\{\gamma\phi\|X\|_2, \sqrt{\alpha d}\}\right).$$

We note that any such ϕ must have $|\phi| \geqslant (2\kappa N)^{-1}$, since if we had $\phi < (2\kappa N)^{-1}$, then each coordinate of ϕX would lie in $(-1/2, 1/2)$, implying $\|\phi X\|_{\mathbb{T}} = \phi\|X\|_2$, that is $\|\phi X\|_{\mathbb{T}} > \gamma\phi\|X\|_2$. The proof of Lemma 7.4 in [4] shows that

$$\mathbb{P}_X\left(\exists\, \phi \in [(2\kappa N)^{-1}, r_n K] : \|\phi X\|_{\mathbb{T}} < \sqrt{\alpha d}\right) \leqslant (2^{20}\alpha)^{d/4},$$

completing the Lemma. □

We also import from [4, Lemma 7.5] a result showing anticoncentration for random vectors AX, where A is a fixed matrix and X is a random vector with independent entries. As noted in [4], this is essentially a rephrasing of Corollary 1.4 and Remark 2.3 in Rudelson and Vershynin's paper [34]:

Lemma VIII.4. *Let $N \in \mathbb{N}$, $n, d, k \in \mathbb{N}$ be such that $n - d \geqslant 2d > 2k$, H be a $2d \times (n-d)$ matrix with $\sigma_{2d-k}(H) \geqslant c_0\sqrt{n}/16$ and $B_1, \ldots, B_{n-d} \subset \mathbb{Z}$ with $|B_i| \geqslant N$. If X is taken uniformly at random from $\mathcal{B} := B_1 \times \ldots \times B_{n-d}$, then*

$$\mathbb{P}_X(\|HX\|_2 \leqslant n) \leqslant \left(\frac{Cn}{dc_0 N}\right)^{2d-k},$$

where $C > 0$ is an absolute constant.

VIII.2. Proof of Theorem VIII.2

Recall that the matrix M is defined as

$$M = \begin{bmatrix} \mathbf{0}_{[d]\times[d]} & H_1^T \\ H_1 & \mathbf{0}_{[n-d]\times[n-d]} \end{bmatrix},$$

where H_1 is a $(n-d) \times d$ random matrix with whose entries are i.i.d. copies of $\tilde{\zeta} Z_\nu$. Let H_2 be an independent copy of H_1, and define H to be the $(n-d) \times 2d$ matrix

$$H := \begin{bmatrix} H_1 & H_2 \end{bmatrix}.$$

For a vector $X \in \mathbb{R}^n$, we define the events $\mathcal{A}_1 = \mathcal{A}_1(X)$ and $\mathcal{A}_2 = \mathcal{A}_2(X)$ by

$$\mathcal{A}_1 := \left\{H : \|H_1 X_{[d]}\|_2 \leqslant n \text{ and } \|H_2 X_{[d]}\|_2 \leqslant n\right\}$$
$$\mathcal{A}_2 := \left\{H : \|H^T X_{[d+1,n]}\|_2 \leqslant 2n\right\}.$$

We now note a simple bound on $\mathbb{P}_M(\|MX\|_2 \leqslant n)$ in terms of $\mathcal{A}_1$ and $\mathcal{A}_2$.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Fact VIII.5. For $X \in \mathbb{R}^n$, let $\mathcal{A}_1 = \mathcal{A}_1(X)$, $\mathcal{A}_2 = \mathcal{A}_2(X)$ be as above. We have

$$(\mathbb{P}_M(\|MX\|_2 \leqslant n))^2 \leqslant \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2).$$

This fact is a straightforward consequence of Fubini's theorem, the details of which are in [4, Fact 7.7]. We shall also need the "robust" notion of the rank of the matrix H used in [4]: for $k = 0, \ldots, 2k$, define $\mathcal{E}_k$ to be the event

$$\mathcal{E}_k := \left\{H : \sigma_{2d-k}(H) \geqslant c_0\sqrt{n}/16 \text{ and } \sigma_{2d-k+1}(H) \leqslant c_0\sqrt{n}/16\right\},$$

and note that always at least one of the events $\mathcal{E}_0, \ldots, \mathcal{E}_{2d}$ holds.

We now define

$$\alpha := 2^{13} L^{-8n/d}, \tag{VIII.2}$$

and for a given box $\mathcal{B}$, we define the set of typical vectors $T(\mathcal{B}) \subseteq \mathcal{B}$ by

$$T = T(\mathcal{B}) := \left\{X \in \mathcal{B} : D_\alpha(c_0 X_{[d]}/(32\sqrt{n})) > 2^{10}B^2\right\}.$$

Now, set $K := 2^{10}B^2$ and note the following implication of Lemma VIII.3: if X is chosen uniformly from $\mathcal{B}$ and $n \geqslant L^{64/c_0^2} \geqslant 2^{10}B^2/\alpha$, then we have that

$$\mathbb{P}_X(X \notin T) = \mathbb{P}_X(D_\alpha(c_0 X_{[d]}/(32\sqrt{n})) \leqslant 2^{10}B^2) \leqslant \left(2^{33}L^{-8n/d}\right)^{d/4} \leqslant \left(\frac{2}{L}\right)^{2n}. \tag{VIII.3}$$

Proof of Lemma VIII.2. Let $M, H_1, H_2, H, \mathcal{A}_1, \mathcal{A}_2, \mathcal{E}_k, \alpha$, and $T := T(\mathcal{B})$ be as above. Define

$$\mathcal{E} := \{X \in \mathcal{B} : \mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant (L/N)^n\}$$

and bound

$$\mathbb{P}_X(\mathcal{E}) \leqslant \mathbb{P}_X(\mathcal{E} \cap \{X \in T\}) + \mathbb{P}_X(X \notin T).$$

For each X, define

$$f(X) := \mathbb{P}_M(\|MX\|_2 \leqslant n)\mathbf{1}(X \in T)$$

and apply (VIII.3) to bound

$$\mathbb{P}_X(\mathcal{E}) \leqslant \mathbb{P}_X(f(X) \geqslant (L/N)^n) + (2/L)^{2n} \leqslant (N/L)^{2n}\mathbb{E}_X f(X)^2 + (2/L)^{2n}, \tag{VIII.4}$$

where the last inequality follows from Markov's inequality. Thus, in order to prove Lemma VIII.2, it is enough to prove $\mathbb{E}_X f(X)^2 \leqslant 2(R/N)^{2n}$.

Apply Fact VIII.5 to write

$$\mathbb{P}_M(\|MX\|_2 \leqslant n)^2 \leqslant \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2) = \sum_{k=0}^{d} \mathbb{P}_H(\mathcal{A}_2|\mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \tag{VIII.5}$$

and so

$$f(X)^2 \leqslant \sum_{k=0}^{d} \mathbb{P}_H(\mathcal{A}_2|\mathcal{A}_1 \cap \mathcal{E}_k)\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k)\mathbf{1}(X \in T). \tag{VIII.6}$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

We will now apply Theorem VII.1 to upper bound $\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k)$ for $X \in T$. For this, note that $d \leqslant c_0^2 n$, $N \leqslant \exp(c_0 L^{-8n/d} d) \leqslant \exp(2^{-32} B^{-4} \alpha n)$ and set $R_0 := 2^{43} B^2 c_0^{-3}$. Also note that by the definition of a (N, κ, d)-box and the fact that $d \geqslant \frac{1}{4} c_0^2 n$, we have that $\|X_{[d]}\|_2 \geqslant d^{1/2} N \geqslant c_0 2^{-10} \sqrt{n} N$. Now, set $\alpha' := 2^{-32} B^{-4} \alpha$ and apply Theorem VII.1 to see that for $X \in T$ and $0 \leqslant k \leqslant \alpha' d$, we have

$$\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \leqslant e^{-c_0 n k/3} \left(\frac{R_0}{N}\right)^{2n-2d}.$$

Additionally, by Theorem VII.1, we may bound the tail sum:

$$\sum_{k \geqslant \alpha' d} \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \leqslant \mathbb{P}_H\left(\{\sigma_{2d-\alpha' d}(H) \leqslant c_0 \sqrt{n}/16\} \cap \mathcal{A}_1\right) \leqslant e^{-c_0 \alpha' d n/4}.$$

Thus, for all $X \in \mathcal{B}$, the previous two equations bound

$$f(X)^2 \leqslant \sum_{k=0}^{\alpha' d} \mathbb{P}_H(\mathcal{A}_2 \,|\, \mathcal{A}_1 \cap \mathcal{E}_k) e^{-c_0 n k/3} \left(\frac{R_0}{N}\right)^{2n-2d} + e^{-c_0 \alpha' d n/3}. \tag{VIII.7}$$

Seeking to bound the right-hand side of (VIII.7), define $g_k(X) := \mathbb{P}_H(\mathcal{A}_2 \,|\, \mathcal{A}_1 \cap \mathcal{E}_k)$. Write

$$\mathbb{E}_X[g_k(X)] = \mathbb{E}_X \mathbb{E}_H\left[\mathcal{A}_2 \,|\, \mathcal{A}_1 \cap \mathcal{E}_k\right] = \mathbb{E}_{X_{[d]}} \mathbb{E}_H\left[\mathbb{E}_{X_{[d+1,n]}} \mathbf{1}[\mathcal{A}_2] \,\middle|\, \mathcal{A}_1 \cap \mathcal{E}_k\right].$$

Let $k \leqslant \alpha' d$. Note that each $H \in \mathcal{A}_1 \cap \mathcal{E}_k$ has $\sigma_{2d-k}(H) \geqslant c_0 \sqrt{n}/16$, and thus we may apply Lemma VIII.4 to bound

$$\mathbb{E}_{X_{[d+1,n]}} \mathbf{1}[\mathcal{A}_2] = \mathbb{P}_{X_{[d+1,n]}}(\|H^T X_{[d+1,n]}\|_2 \leqslant n) \leqslant \left(\frac{C' n}{c_0 d N}\right)^{2d-k} \leqslant \left(\frac{4C'}{c_0^3 N}\right)^{2d-k}$$

for an absolute constant $C' > 0$, where we used that $d \geqslant \frac{1}{4} c_0^2 n$. Thus, for each $0 \leqslant k \leqslant \alpha' d$, if we define $R := \max\{8C' c_0^{-3}, 2R_0\}$, then we have

$$\mathbb{E}_X[g_k(X)] \leqslant \left(\frac{R}{2N}\right)^{2d-k}. \tag{VIII.8}$$

Applying $\mathbb{E}_X$ to (VIII.7) using (VIII.8) shows

$$\mathbb{E}_X f(X)^2 \leqslant \left(\frac{R}{2N}\right)^{2n} \sum_{k=0}^{\alpha' d} \left(\frac{2N}{R}\right)^k e^{-c_0 n k/3} + e^{-c_0 \alpha' d n/3}.$$

Using that $N \leqslant e^{c_0 L^{-8n/d} d} = e^{c_0 \alpha' d/8}$ and $N \leqslant e^{c_0 n/3}$ bounds

$$\mathbb{E}_X f(X)^2 \leqslant 2\left(\frac{R}{2N}\right)^{2n}. \tag{VIII.9}$$

Combining (VIII.9) with (VIII.4) completes the proof of Lemma VIII.2. □

VIII.3. Proof of Theorem III.2

The main work of proving Theorem III.2 is now complete with the proof of Lemma VIII.2. In order to complete it, we need to cover the sphere with a suitable set of boxes. Recall the definitions from

Section III.1:

$$\mathcal{I}'([d]) := \left\{ v \in \mathbb{R}^n : \kappa_0 n^{-1/2} \leqslant |v_i| \leqslant \kappa_1 n^{-1/2} \text{ for all } i \in [d] \right\},$$

and

$$\Lambda_\varepsilon := B_n(0,2) \cap \left(4\varepsilon n^{-1/2} \cdot \mathbb{Z}^n\right) \cap \mathcal{I}'([d]),$$

and that the constants κ_0, κ_1 satisfy $0 < \kappa_0 < 1 < \kappa_1$ and are defined in Section II.3.

We import the following simple covering lemma from [4, Lemma 7.8]

Lemma VIII.6. *For all $\varepsilon \in [0,1]$, $\kappa \geqslant \max\{\kappa_1/\kappa_0, 2^8\kappa_0^{-4}\}$, there exists a family $\mathcal{F}$ of (N,κ,d)-boxes with $|\mathcal{F}| \leqslant \kappa^n$ so that*

$$\Lambda_\varepsilon \subseteq \bigcup_{\mathcal{B} \in \mathcal{F}} (4\varepsilon n^{-1/2}) \cdot \mathcal{B}, \tag{VIII.10}$$

where $N = \kappa_0/(4\varepsilon)$.

Combining Lemma VIII.6 with Lemma VIII.2 will imply Theorem III.2.

Proof of Theorem III.2. Apply Lemma VIII.6 with $\kappa = \max\{\kappa_1/\kappa_0, 2^8\kappa_0^{-4}\}$ and use the fact that $\mathcal{N}_\varepsilon \subseteq \Lambda_\varepsilon$ to write

$$\mathcal{N}_\varepsilon \subseteq \bigcup_{\mathcal{B} \in \mathcal{F}} \left((4\varepsilon n^{-1/2}) \cdot \mathcal{B}\right) \cap \mathcal{N}_\varepsilon$$

and so

$$|\mathcal{N}_\varepsilon| \leqslant \sum_{\mathcal{B} \in \mathcal{F}} |(4\varepsilon n^{-1/2} \cdot \mathcal{B}) \cap \mathcal{N}_\varepsilon| \leqslant |\mathcal{F}| \cdot \max_{\mathcal{B} \in \mathcal{F}} |(4\varepsilon n^{-1/2} \cdot \mathcal{B}) \cap \mathcal{N}_\varepsilon|.$$

Rescaling by $\sqrt{n}/(4\varepsilon)$ and applying Lemma VIII.2 bound

$$|(4\varepsilon n^{-1/2} \cdot \mathcal{B}) \cap \mathcal{N}_\varepsilon| \leqslant |\{X \in \mathcal{B} : \mathbb{P}_M(\|MX\|_2 \leqslant n) \geqslant (L\varepsilon)^n\}| \leqslant \left(\frac{R}{L}\right)^{2n} |\mathcal{B}|.$$

To see that the application of Lemma VIII.2 is justified, note that $0 < c_0 \leqslant 2^{-50}B^{-4}$, $c_0^2 n/2 \leqslant d \leqslant c_0^2 n$, $\kappa \geqslant 2$, and $\log 1/\varepsilon \leqslant n/L^{64/c_0^2}$ and so

$$\log N = \log \kappa_0/(4\varepsilon) \leqslant n/L^{64/c_0^2} \leqslant c_0 L^{-8n/d} d,$$

as required by Lemma VIII.2, since $\kappa_0 < 1$, $d \geqslant L^{-1/c_0^2} n$, $c_0 \geqslant L^{-1/c_0^2}$, and $8n/d \leqslant 16/c_0^2$. Using that $|\mathcal{F}| \leqslant \kappa^n$ and $|\mathcal{B}| \leqslant (\kappa N)^n$ for each $\mathcal{B} \in \mathcal{F}$ bound

$$|\mathcal{N}_\varepsilon| \leqslant \kappa^n \left(\frac{R}{L}\right)^{2n} |\mathcal{B}| \leqslant \kappa^n \left(\frac{R}{L}\right)^{2n} (\kappa N)^n \leqslant \left(\frac{C}{c_0^6 L^2 \varepsilon}\right)^n,$$

where we set $C := \kappa^2 R^2 c_0^6$. This completes the proof of Theorem III.2. □

IX. Nets for structured vectors: Approximating with the net

In this section, we prove Lemma III.1, which tells us that $\mathcal{N}_\varepsilon$ is a net for Σ_ε. The proof uses the random rounding technique developed by Livshyts [21] in the same way as in [4].

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Proof of Lemma III.1. Given $v \in \Sigma_\varepsilon$, we define a random variable $r = (r_1, \dots, r_n)$, where the r_i are independent and satisfy $\mathbb{E}\, r_i = 0$ as well as the deterministic properties $|r_i| \leqslant 4\varepsilon n^{-1/2}$ and $v - r \in 4\varepsilon n^{-1/2}\mathbb{Z}^n$. We then define the random variable $u := v - r$. We will show that with positive probability that $u \in \mathcal{N}_\varepsilon$.

By definition, $\|r\|_\infty = \|u - v\|_\infty \leqslant 4\varepsilon n^{-1/2}$ for all u. Also, $u \in \mathcal{I}'([d])$ for all u, since $v \in \mathcal{I}([d])$ and $\|u - v\|_\infty \leqslant 4\varepsilon/\sqrt{n} \leqslant \kappa_0/(2\sqrt{n})$. Thus, from the definition of $\mathcal{N}_\varepsilon$, we need only show that with positive probability u satisfies

$$\mathbb{P}(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (L\varepsilon)^n \text{ and } \mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) \leqslant (2^{10}L\varepsilon)^n. \tag{IX.1}$$

We first show that *all* u satisfy the upper bound at (IX.1). To see this, recall $\mathcal{K} = \{\|A\|_{\text{op}} \leqslant 4\sqrt{n}\}$ and let $w(u) \in \mathbb{R}^n$ be such that

$$\begin{aligned}
\mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) &= \mathbb{P}^{\mathcal{K}}\left(\|Av - Ar - w(u)\| \leqslant \varepsilon\sqrt{n}\right) \\
&\leqslant \mathbb{P}^{\mathcal{K}}\left(\|Av - w(u)\| \leqslant 17\varepsilon\sqrt{n}\right) \\
&\leqslant \mathcal{L}_{A,op}(v, 17\varepsilon\sqrt{n}) \leqslant \mathcal{L}(Av, 17\varepsilon\sqrt{n}).
\end{aligned}$$

Since $v \in \Sigma_\varepsilon$, Lemma III.7 bounds

$$\mathcal{L}(Av, 17\varepsilon\sqrt{n}) \leqslant (2^{10}L\varepsilon)^n\,. \tag{IX.2}$$

We now show that

$$\mathbb{E}_u\, \mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant (1/2)\mathbb{P}_M(\|Mv\|_2 \leqslant 2\varepsilon\sqrt{n}) \geqslant (1/4)(2\varepsilon L)^n\,, \tag{IX.3}$$

where the last inequality holds by the fact $v \in \Sigma_\varepsilon$. From (IX.3), it then follows that there is some $u \in \Lambda_\varepsilon$ satisfying (IX.1). To prove the first inequality in (IX.1), define the event

$$\mathcal{E} := \{M : \|Mv\|_2 \leqslant 2\varepsilon\sqrt{n} \text{ and } \|M\|_{\text{HS}} \leqslant n/4\}$$

and note that for all u, we have

$$\mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) = \mathbb{P}_M(\|Mv - Mr\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant \mathbb{P}_M(\|Mr\|_2 \leqslant 2\varepsilon\sqrt{n} \text{ and } \mathcal{E})\,.$$

Since by the Bernstein inequality, $\mathbb{P}(\|M\|_{\text{HS}}^2 \geqslant n^2/16) \leqslant 2\exp(-cn^2)$ and the fact that

$$\varepsilon \geqslant \exp(-2c_\Sigma n) \geqslant \exp(-cn),$$

we have

$$\mathbb{P}(\mathcal{E}) \geqslant (2L\varepsilon)^n - 2\exp(-cn^2) \geqslant (1/2)(2L\varepsilon)^n,$$

assuming that c_Σ is chosen appropriately small compared to this absolute constant. Thus

$$\begin{aligned}
\mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) &\geqslant \mathbb{P}_M(\|Mr\|_2 \leqslant 2\varepsilon\sqrt{n}\,\big|\mathcal{E})\mathbb{P}(\mathcal{E}) \\
&\geqslant \left(1 - \mathbb{P}_M(\|Mr\|_2 > 2\varepsilon\sqrt{n}\,\big|\mathcal{E})\right)(1/2)(2L\varepsilon)^n\,.
\end{aligned}$$

Taking expectations gives

$$\mathbb{E}_u\mathbb{P}_M(\|Mu\|_2 \leqslant 4\varepsilon\sqrt{n}) \geqslant \left(1 - \mathbb{E}_u\mathbb{P}_M(\|Mr\|_2 > 2\varepsilon\sqrt{n}\,\big|\mathcal{E})\right)(1/2)(2L\varepsilon)^n\,. \tag{IX.4}$$

Exchanging the expectations and rearranging, we see that it is enough to show

$$\mathbb{E}_M\left[\mathbb{P}_r(\|Mr\|_2 > 2\varepsilon\sqrt{n})\,\middle|\,\mathcal{E}\right] \leqslant 1/2\,.$$

We will show that $\mathbb{P}_r(\|Mr\|_2 > 2\varepsilon\sqrt{n}) \leqslant 1/4$ for all $M \in \mathcal{E}$, by Markov's inequality. Note that

$$\mathbb{E}_r\,\|Mr\|_2^2 = \sum_{i,j}\mathbb{E}\left(M_{i,j}r_i\right)^2 = \sum_i \mathbb{E}\,r_i^2\sum_j M_{i,j}^2 \leqslant 16\varepsilon^2\|M\|_{\mathrm{HS}}^2/n \leqslant \varepsilon^2 n,$$

where for the second equality, we have used that the r_i are mutually independent and $\mathbb{E}\,r_i = 0$; for the third inequality, we used $\|r\|_\infty \leqslant 4\varepsilon/\sqrt{n}$; and for the final inequality, we used $\|M\|_{\mathrm{HS}} \leqslant n/4$. Thus, by Markov's inequality gives

$$\mathbb{P}_r(\|Mr\|_2 \geqslant 2\varepsilon\sqrt{n}) \leqslant (2\varepsilon\sqrt{n})^{-2}\mathbb{E}_r\,\|Mr\|_2^2 \leqslant 1/4\,. \tag{IX.5}$$

Putting (IX.5) together with (IX.4) proves (IX.3), completing the proof of (IX.1). □

X. Proof of Lemma VI.7

We will derive Lemma VI.7 from Talagrand's inequality:

Theorem X.1 (Talagrand's inequality). *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a convex 1-Lipschitz function and $\sigma = (\sigma_1, \ldots, \sigma_n)$, where the σ_i are i.i.d. random variables, such that $|\sigma_i| \leqslant 1$. Then for any $t \geqslant 0$, we have*

$$\mathbb{P}\left(|F(\sigma) - m_F| \geqslant t\right) \leqslant 4\exp\left(-t^2/16\right),$$

where m_F is the median of $F(\sigma)$.

Proof of Lemma VI.7. Note the theorem is trivial if $k \leqslant 2^{20}B^4/\nu$, so assume that $k > 2^{20}B^4/\nu$. Set $\sigma = 2^{-4}B^{-2}\tau'$, define

$$F(x) := \|W\|^{-1}\|W^T x\|_2,$$

and note that F is convex and 1-Lipschitz. Since $|\sigma_i| \leqslant 2^{-4}B^{-2}|\tau_i| \leqslant 1$ and the σ_i are i.i.d., Theorem X.1 tells us that $F(\sigma)$ is concentrated about the median m_F and so we only need to estimate m_F. For this, write

$$m := \mathbb{E}\,\|W^T\sigma\|_2^2 = \sum_{i,j}W_{ij}^2\mathbb{E}\,\sigma_i^2 = \mathbb{E}\sigma_i^2\|W\|_{\mathrm{HS}}^2,$$

and

$$m_2 := \mathbb{E}\,\|W^T\sigma\|_2^4 - (\mathbb{E}\,\|W^T\sigma\|_2^2)^2 = \sum_{i,j}W_{ij}^2\left(\mathbb{E}\,\sigma_i^4 - (\mathbb{E}\,\sigma_i^2)^2\right) \leqslant \mathbb{E}\,\sigma_i^2\|W\|_{\mathrm{HS}}^2,$$

where for the final inequality, we used that $\mathbb{E}\,\sigma_i^4 \leqslant \mathbb{E}\,\sigma_i^2$, since $|\sigma_i| \leqslant 1$. For $t > 0$, Markov's inequality bounds

$$\mathbb{P}(\|W^T\sigma\|_2^2 \leqslant m - t) \leqslant t^{-2}\mathbb{E}\left(\|W^T\sigma\|_2^2 - m\right)^2 = t^{-2}m_2 \leqslant t^{-2}\mathbb{E}\,\sigma_i^2\|W\|_{\mathrm{HS}}^2.$$

Setting $t = \mathbb{E}\,\sigma_i^2\|W\|_{\mathrm{HS}}^2/2$ gives

$$\mathbb{P}(\|W^T\sigma\|_2^2 \leqslant \mathbb{E}\,\sigma_i^2\|W\|_{\mathrm{HS}}^2/2) \leqslant 4(\mathbb{E}\sigma_i^2\|W\|_{\mathrm{HS}}^2)^{-1} < 1/2,$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

since $\mathbb{E}\sigma_i^2 = 2^{-8}B^{-4}\mathbb{E}\tau_i'^2 \geqslant 2^{-8}B^{-4}\nu$ and $\|W\|_{\mathrm{HS}}^2 \geqslant k/4 > 2^{11}\nu^{-1}B^4$ (by assumption). It follows that

$$m_F \geqslant \sqrt{\mathbb{E}\,\sigma_i^2/2}\|W\|^{-1}\|W\|_{\mathrm{HS}} \geqslant 2^{-6}\|W\|^{-1}B^{-2}\sqrt{\nu k}\,,$$

since $\|W\|_{\mathrm{HS}} \geqslant \sqrt{k}/2$. Now, we may apply Talagrand's inequality (Theorem X.1) with $t = m_F - \beta'\sqrt{k}\|W\|^{-1}$ to obtain

$$\mathbb{P}\left(\|W^T\sigma\|_2 \leqslant \beta'\sqrt{k}\right) \leqslant 4\exp\left(-2^{-20}B^{-4}\nu k\right)$$

as desired. □

XI. Proof of Theorem 1.4

Here, we deduce Theorem 1.4, which shows negative correlation between a small ball and large deviation event. The proof is similar in theme to those in Section 5 but is, in fact, quite a bit simpler due to the fact we are working with a linear form rather than a quadratic form.

Proof of Theorem 1.4. We first write

$$\mathbb{P}(|\langle X, v\rangle| \leqslant \varepsilon \text{ and } \langle X, u\rangle > t) \leqslant \mathbb{E}\left[\mathbf{1}\{|\langle X, v\rangle| \leqslant \varepsilon\}e^{\lambda\langle X,u\rangle-\lambda t}\right], \tag{XI.1}$$

where $\lambda \geqslant 0$ will be optimized later. Now, apply Esseen's inequality in a similar way to Lemma 5.1 to bound

$$\mathbb{E}\left[\mathbf{1}\{|\langle X, v\rangle| \leqslant \varepsilon\}e^{\lambda\langle X,u\rangle-\lambda t}\right] \lesssim \varepsilon e^{-\lambda t}\int_{-1/\varepsilon}^{1/\varepsilon}\left|\mathbb{E}e^{2\pi i\theta\langle X,v\rangle+\lambda\langle X,u\rangle}\right|\,d\theta\,. \tag{XI.2}$$

Applying Lemma 5.5 bounds

$$\left|\mathbb{E}e^{2\pi i\theta\langle X,v\rangle+\lambda\langle X,u\rangle}\right| \lesssim \exp\left(-c\min_{r\in[1,c^{-1}]}\|\theta r v\|_{\mathbb{T}}^2 + c^{-1}\lambda^2\right) + e^{-c\alpha n}\,. \tag{XI.3}$$

Combining the lines (XI.1),(XI.2), and (XI.3) and choosing C large enough give the bound

$$\begin{aligned}\mathbb{P}(|\langle X, v\rangle| \leqslant \varepsilon \text{ and } \langle X, u\rangle > t) &\lesssim \varepsilon e^{-\lambda t+c^{-1}\lambda^2}\int_{-1/\varepsilon}^{1/\varepsilon}\left(e^{-c\gamma^2\theta^2} + e^{-c\alpha n}\right)\,d\theta \\ &\lesssim \varepsilon e^{-\lambda t+c^{-1}\lambda^2}\gamma^{-1} + e^{-c\alpha n-\lambda t+c^{-1}\lambda^2}\,.\end{aligned}$$

Choosing $\lambda = ct/2$ completes the proof. □

XII. Proof of Lemma 3.2

We deduce the second part of Lemma 3.2 from the following special case of a proposition of Vershynin [46, Proposition 4.2].

Proposition XII.1. *For $B > 0$, let $\zeta \in \Gamma_B$, let $A_n \sim \mathrm{Sym}_n(\zeta)$, and let $K \geqslant 1$. Then there exist $\rho, \delta, c > 0$ depending only on K, B so that for every $\lambda \in \mathbb{R}$ and $w \in \mathbb{R}^n$, we have*

$$\mathbb{P}\Big(\inf_{x\in\mathrm{Comp}\,(\delta,\rho)}\|(A_n + \lambda I)x - w\|_2 \leqslant c\sqrt{n} \text{ and } \|A_n + \lambda I\|_{op} \leqslant K\sqrt{n}\Big) \leqslant 2e^{-cn}\,.$$

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

Proof of Lemma 3.2. To get the first conclusion of Lemma 3.2, we may assume without loss of generality that $u \in \mathbb{S}^{n-1}$. So first let $\mathcal{N}$ be a $c\sqrt{n}$-net for $[-4\sqrt{n}, 4\sqrt{n}]$, with $|\mathcal{N}| \leqslant 8/c$. Note that $\mathbb{P}(\|A_n\|_{op} > 4\sqrt{n}) \lesssim e^{-\Omega(n)}$ so if $A_n x = tu$, then we may assume $t \in [-4\sqrt{n}, 4\sqrt{n}]$. So

$$\mathbb{P}\big(\exists\, x \in \mathrm{Comp}\,(\delta,\rho), \exists t \in [-4\sqrt{n}, 4\sqrt{n}] : A_n x = tu\big) \leqslant \sum_{t_0 \in \mathcal{N}} \mathbb{P}\left(\exists\, x \in \mathrm{Comp}\,(\delta,\rho) : \|A_n x - t_0 u\|_2 \leqslant c\sqrt{n}\right),$$

since for each $t \in [-4\sqrt{n}, 4\sqrt{n}]$ there's $t_0 \in \mathcal{N}$, such that if $A_n x = tu$, then $\|A_n x - t_0 u\|_2 \leqslant c\sqrt{n}$. Now to bound each term in the sum, take $\lambda = 0$, $K = 4$, $w = t_0 u$ in Proposition XII.1 and notice we may assume $\|A_n\|_{op} \leqslant 4\sqrt{n}$ again. For the second conclusion, it is sufficient to show

$$\mathbb{P}\big(\exists\, x \in \mathrm{Comp}\,(\delta,\rho), \exists t \in [-4\sqrt{n}, 4\sqrt{n}] : \|(A_n - tI)x\|_2 = 0 \text{ and } \|A_n - tI\|_{op} \leqslant 8\sqrt{n}\big) \lesssim e^{-\Omega(n)}, \tag{XII.1}$$

since we have $\mathbb{P}(\|A_n\|_{op} \geqslant 4\sqrt{n}) \lesssim e^{-\Omega(n)}$, by (4.11), so we may assume that all eigenvalues of A_n lie in $[-4\sqrt{n}, 4\sqrt{n}]$ and $\|A_n - tI\|_{op} \leqslant |t| + \|A_n\|_{op} \leqslant 8\sqrt{n}$, for all $t \in [-4\sqrt{n}, 4\sqrt{n}]$.

For this, we apply Proposition XII.1 with $K = 8$ to obtain ρ, δ, c. Again, let $\mathcal{N}$ be a $c\sqrt{n}$-net for the interval $[-4\sqrt{n}, 4\sqrt{n}]$ with $|\mathcal{N}| \leqslant 8/c$. So, if $t \in [-4\sqrt{n}, 4\sqrt{n}]$ satisfies $A_n x = tx$ for some $x \in \mathbb{S}^{n-1}$, then there is a $t_0 \in \mathcal{N}$ with $|t - t_0| \leqslant c\sqrt{n}$ and

$$\|(A_n - t_0 I)x\|_2 \leqslant |t - t_0|\|x\|_2 \leqslant c\sqrt{n}\,.$$

Thus, the left-hand side of (XII.1) s at most

$$\sum_{t_0 \in \mathcal{N}} \mathbb{P}\left(\exists\, x \in \mathrm{Comp}\,(\delta,\rho) : \|(A_n - t_0 I)x\|_2 \leqslant c\sqrt{n} \text{ and } \|A_n - t_0 I\|_{op} \leqslant 8\sqrt{n}\right) \lesssim e^{-cn},$$

where the last line follows from Proposition XII.1. □

Acknowledgments. The authors thank Rob Morris for comments on the presentation of this paper. The authors also thank the anonymous referees for many useful comments and a simplification to the proof of Corollary 8.2. Marcelo Campos is partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico. Matthew Jenssen is supported by a UK Research and Innovation Future Leaders Fellowship MR/W007320/1. Marcus Michelen is supported in part by National Science Foundation grants Division of Mathematical Sciences (DMS)-2137623 and DMS-2246624.

Competing interest. The authors have no competing interest to declare.

References

[1] R. Berkowitz, 'A local limit theorem for cliques in G(n,p)', Preprint, 2018, arXiv:1811.03527.

[2] P. Bourgade, L. Erdős, H.-T. Yau and J. Yin, 'Fixed energy universality for generalized Wigner matrices', *Comm. Pure Appl. Math.* **69**(10) (2016), 1815–1881.

[3] M. Campos, M. Jenssen, M. Michelen and J. Sahasrabudhe, 'Singularity of random symmetric matrices revisited', *Proc. of Amer. Math. Soc.* **150**(757) (2021), 3147–3159.

[4] M. Campos, M. Jenssen, M. Michelen and J. Sahasrabudhe, 'The singularity probability of a random symmetric matrix is exponentially small', Preprint, 2021, arxiv:2105.11384.

[5] M. Campos, L. Mattos, R. Morris and N. Morrison, 'On the singularity of random symmetric matrices', *Duke Math. J.* **170**(5) (2021), 881–907.

[6] K. P. Costello, 'Bilinear and quadratic variants on the Littlewood-Offord problem', *Isr. J. Math.* **194**(1) (2013), 359–394.

[7] K. P. Costello, T. Tao and V. Vu, 'Random symmetric matrices are almost surely nonsingular', *Duke Math. J.* **135**(2) (2006), 395–413.

[8] A. Edelman, 'Eigenvalues and condition numbers of random matrices', *SIAM J. Matrix Anal. Appl.* **9**(4) (1988), 543–560.

[9] L. Erdős, J. Ramírez, B. Schlein, T. Tao, V. Vu and H.-T. Yau, 'Bulk universality for Wigner Hermitian matrices with subexponential decay', *Math. Res. Lett.* **17**(4) (2010), 667–674.

[10] L. Erdős, B. Schlein and H.-T. Yau, 'Local semicircle law and complete delocalization for Wigner random matrices', *Comm. Math. Phys.* **287**(2) (2009), 641–655.
[11] L. Erdős, B. Schlein and H.-T. Yau, 'Semicircle law on short scales and delocalization of eigenvectors for Wigner random matrices', *Ann. Probab.* **37**(3) (2009), 815–852.
[12] L. Erdős, 'Universality of Wigner random matrices: A survey of recent results', *Russ. Math. Surv.* **66**(3) (2011), 507.
[13] L. Erdős, B. Schlein and H.-T. Yau, 'Wegner estimate and level repulsion for Wigner random matrices', *Int. Math. Res. Not.* 2010(3) (**2010**), 436–479.
[14] C. Esseen, 'On the Kolmogorov-Rogozin inequality for the concentration function', *Z. Wahrscheinlichkeitstheor. Verw. Geb.* **5**(3) (1966), 210–216.
[15] O. N. Feldheim and S. Sodin, 'A universality result for the smallest eigenvalues of certain sample covariance matrices', *Geom. Funct. Anal.* **20**(1) (2010), 88–123.
[16] A. Ferber and V. Jain, 'Singularity of random symmetric matrices—A combinatorial approach to improved bounds', *Forum Math. Sigma* **7** (2019), Paper No. e22, 29.
[17] A. Ferber, V. Jain, K. Luh and W. Samotij, 'On the counting problem in inverse Littlewood–Offord theory', *J. London Math. Soc.* **103**(4) (2021), 1333–1362.
[18] D. L. Hanson and F. T. Wright, 'A bound on tail probabilities for quadratic forms in independent random variables', *Ann. Math. Stat.* **42**(3) (1971), 1079–1083.
[19] V. Jain, A. Sah and M. Sawhney, 'On the smallest singular value of symmetric random matrices', *Comb. Probab. Comput.* **31**(4) (2022), 662–683.
[20] M. Kwan and L. Sauermann, 'An algebraic inverse theorem for the quadratic Littlewood-Offord problem, and an application to Ramsey graphs', *Discret. Anal.* (2020), Paper No. 12, 34, arXiv:1909.02089.
[21] G. V. Livshyts, 'The smallest singular value of heavy-tailed not necessarily i.i.d. random matrices via random rounding', *J. Anal. Math.* **145** (2021), 257–306.
[22] G. V. Livshyts, K. Tikhomirov and R. Vershynin, 'The smallest singular value of inhomogeneous square random matrices', *Ann. Probab.* **49**(3) (2021), 1286–1309.
[23] R. Meka, O. Nguyen and V. Vu, 'Anti-concentration for polynomials of independent random variables', *Theory Comput.* **12**(11) (2016), 1–17.
[24] H. Nguyen, T. Tao and V. Vu, 'Random matrices: Tail bounds for gaps between eigenvalues', *Probab. Theory Relat. Fields* **167**(3) (2017), 777–816.
[25] H. H. Nguyen, 'Inverse Littlewood-Offord problems and the singularity of random symmetric matrices', *Duke Math. J.* **161**(4) (2012), 545–586.
[26] H. H. Nguyen, 'On the least singular value of random symmetric matrices', *Electron. J. Probab.* **17**(53) (2012), 19.
[27] H. H. Nguyen, 'Random matrices: Overcrowding estimates for the spectrum', *J. Funct. Anal.* **275**(8) (2018), 2197–2224.
[28] H. H. Nguyen and V. H. Vu, 'Small probability, inverse theorems, and applications', *in Erdös centennial, Bolyai Society Mathematical Studies*, vol. 25 (János Bolyai Mathematical Society, Budapest, 2013), 409–463.
[29] E. Rebrova and K. Tikhomirov, 'Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries', *Isr. J. Math.* **227**(2) (2018), 507–544.
[30] M. Rudelson, 'Invertibility of random matrices: Norm of the inverse', *Ann. of Math. (2)* **168**(2) (2008), 575–600.
[31] M. Rudelson and R. Vershynin, 'The Littlewood-Offord problem and invertibility of random matrices', *Adv. Math.* **218**(2) (2008), 600–633.
[32] M. Rudelson and R. Vershynin, 'Smallest singular value of a random rectangular matrix', *Comm. Pure Appl. Math.* **62**(12) (2009), 1707–1739.
[33] M. Rudelson and R. Vershynin, 'Hanson-Wright inequality and sub-Gaussian concentration', Preprint, 2013, arXiv:1306.2872.
[34] M. Rudelson and R. Vershynin, 'Small ball probabilities for linear images of high-dimensional distributions', *Int. Math. Res. Not.* 2015(19) (**2015**), 9594–9617.
[35] M. Rudelson and R. Vershynin, 'No-gaps delocalization for general random matrices', *Geom. Funct. Anal.* **26**(6) (2016), 1716–1776.
[36] S. Smale, 'On the efficiency of algorithms of analysis', *Bull. Am. Math. Soc.* **13**(2) (1985), 87–121.
[37] D. Spielman and S.-H. Teng, 'Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time', *in Proceedings of the Annual ACM Symposium on Theory of Computing* (ACM, New York, 2001), 296–305.
[38] D. A. Spielman and S.-H. Teng, 'Smoothed analysis of algorithms', in *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)* (Higher Education Press, Beijing, 2002), 597–606.
[39] S. J. Szarek, 'Spaces with large distance to l_∞^n and random matrices', *Amer. J. Math.* **112**(6) (1990), 899–942.
[40] T. Tao and V. Vu, 'Random matrices: A general approach for the least singular value problem', Preprint, 2008, arXiv:0805.3167.
[41] T. Tao and V. Vu, 'Random matrices: The distribution of the smallest singular values', *Geom. Funct. Anal.* **20**(1) (2010), 260–297.
[42] T. Tao and V. Vu, 'Random matrices: Universality of local eigenvalue statistics', *Acta Math.* **206**(1), (2011), 127–204.
[43] T. Tao and V. Vu, 'Random matrices have simple spectrum', *Combinatorica* **37**(3) (2017) 539–553.

https://doi.org/10.1017/fmp.2023.29 Published online by Cambridge University Press

[44] T. Tao and V. H. Vu, 'Inverse Littlewood-Offord theorems and the condition number of random discrete matrices', *Ann. of Math. (2)* **169**(2) (2009), 595–632.

[45] K. Tikhomirov, 'Singularity of random Bernoulli matrices', *Ann. of Math.* **191**(2) (2020), 593–634.

[46] R. Vershynin, 'Invertibility of symmetric random matrices', *Rand. Struct. Algorithms* **44**(2) (2014), 135–182.

[47] R. Vershynin, *High-Dimensional Probability: An Introduction With Applications in Data Science*, vol. 47 (Cambridge University Press, Cambridge, 2018).

[48] J. Von Neumann, *Design of Computers, Theory of Automata and Numerical Analysis*, vol. 5 (Pergamon Press, London, 1963).

[49] V. H. Vu and T. Tao, 'The condition number of a randomly perturbed matrix', *in Proceedings of the Annual ACM Symposium on Theory of Computing*, STOC '07 (Association for Computing Machinery, New York, NY, USA, 2007), 248–255.

[50] E. P. Wigner, 'On the distribution of the roots of certain symmetric matrices', *Ann. of Math. (2)* **67**(2) (1958), 325–327.

[51] F. T. Wright, 'A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric', *Ann. Probab.* **1**(6) (1973), 1068–1070.