

The least singular value of a random symmetric matrix

Campos, Marcelo; Jenssen, Matthew; Michelen, Marcus; Sahasrabudhe, Julian

DOI:

[10.48550/arXiv.2203.06141](https://doi.org/10.48550/arXiv.2203.06141)

License:

Creative Commons: Attribution (CC BY)

Document Version

Other version

Citation for published version (Harvard):

Campos, M, Jenssen, M, Michelen, M & Sahasrabudhe, J 2022 'The least singular value of a random symmetric matrix' arXiv. <https://doi.org/10.48550/arXiv.2203.06141>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

THE LEAST SINGULAR VALUE OF A RANDOM SYMMETRIC MATRIX

MARCELO CAMPOS, MATTHEW JENSSEN, MARCUS MICHELEN, JULIAN SAHASRABUDHE

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

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

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.

1. INTRODUCTION

Let A be a $n \times n$ random symmetric matrix whose entries on and above the diagonal $(A_{i,j})_{i \leq j}$ are 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 [44], who established the famous “semi-circular 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

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

for all $\varepsilon > 0$ (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 [40].

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

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

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 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 conjectured that the adjacency matrix of the binomial random graph $G(n, 1/2)$ has no repeated eigenvalues with probability $1 - o(1)$ (see [38]). Tao and Vu [38] proved this conjecture in 2014 and, in subsequent work on the topic with Nguyen [25], 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 a $n \times n$ random symmetric matrix where $(A_{i,j})_{i \leq j}$ are independent and distributed according to ζ . Then A has no repeated eigenvalues with probability $1 - e^{-cn}$, 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 $\geq e^{-(3+o(1))pn}$.

We in fact prove a more refined version 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 and an important step in the proof of Theorem 1.1. For this, we let $\lambda_1(A) \geq \dots \geq \lambda_n(A)$ denote the eigenvalues of the $n \times n$ real symmetric matrix A .

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

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

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 [42] 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 [33] and proved by Szarek [35] and Edelman [9] in the case that

$B_n = G_n$ is a random matrix with i.i.d. *standard gaussian* entries. Edelman proved the stronger result that

$$\mathbb{P}(\sigma_{\min}(G_n) \leq \varepsilon n^{-1/2}) \leq \varepsilon, \quad (3)$$

for all $\varepsilon \geq 0$ and gave an asymptotic expansion for this probability when ε is fixed and $n \rightarrow \infty$. While this gives a very satisfying understanding of the gaussian case, one encounters serious difficulties when trying to extend this result to other distributions, as Edelman's proof relies crucially on the special tools available only in the gaussian case. 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 (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 non-zero probability. This singularity event will affect the quantity (1) for very small ε and thus estimating the probability that $\sigma_{\min}(B_n) = 0$ is a crucial aspect of generalizing (3). This is reflected in the famous and influential Spielman–Teng conjecture [34] which stipulates the bound

$$\mathbb{P}(\sigma_{\min}(B_n) \leq \varepsilon n^{-1/2}) \leq \varepsilon + 2e^{-cn}, \quad (4)$$

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 [29] who proved that if B_n has i.i.d. subgaussian entries then

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

This result was extended in a series of works [31, 36, 39, 43] ultimately terminating in the influential work of Rudelson and Vershynin [30] who showed the “up-to-constants” version of Spielman–Teng:

$$\mathbb{P}(\sigma_{\min}(B_n) \leq \varepsilon n^{-1/2}) \leq C\varepsilon + e^{-cn}, \quad (5)$$

where B_n is a matrix with i.i.d. entries that follow any subgaussian distribution and $C, c > 0$ depend only on ζ .

While falling just short of the Spielman–Teng conjecture, the work of Rudelson and Vershynin [30] and its subsequent refinement by Tikhomirov [23] and Livshyts, Tikhomirov and Vershynin [23] (see also [22, 28]) 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 non-singular with probability $1 - o(1)$ was not resolved until the important 2006 paper of Costello, Tao and Vu [8].

Progress on the symmetric version of Spielman–Teng continued with Nguyen [26] and, independently, Vershynin [40]. Nguyen proved that for any $B > 0$ there exists an $A > 0$ for

which¹

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

Vershynin [40] proved that if A_n is a matrix with subgaussian entries then, for all $\varepsilon > 0$, we have

$$\mathbb{P}(\sigma_{\min}(A_n) \leq \varepsilon n^{-1/2}) \leq C_\eta \varepsilon^{1/8-\eta} + 2e^{-n^c}, \quad (6)$$

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 $\varepsilon^{1/8}$ as the correct order of magnitude and that e^{-cn} should replace e^{-n^c} .

After Vershynin, a series of works [3, 6, 17, 18, 20] 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 $\{-1, 1\}$:

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

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

However, for general ε the state of the art is due to Jain, Sah and Sawhney [20], who improved on Vershynin's bound (6) by showing

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

under the subgaussian hypothesis on A_n .

For large ε , for example $\varepsilon \geq n^{-c}$, another very different and powerful set of techniques have been developed, which in fact apply more generally to the distribution of other “bulk” eigenvalues. The works of Tao and Vu [36, 37], Erdős, Schlein and Yau [11, 12, 14], Erdős, Ramírez, Schlein, Tao, Vu, Yau [10], and specifically Bourgade, Erdős, Yau and Yin [2] tell us that

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

thus obtaining the correct dependence on ε asymptotically. 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

¹We note that this result is actually proved for all matrices of the form $A_n + F$, where F is any fixed $n \times n$ matrix and the entries of A_n have mean 0, but need not be identically distributed and may have large variances.

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

Here we state and discuss two *approximate* negative correlation results: one of which is from our paper [4], but is used in a 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, which will be sufficient for our purposes:

$$\mathbb{P}_X(|\langle X, v \rangle| \leq \varepsilon \text{ and } \langle X, u \rangle > t) \leq \mathbb{P}_X(|\langle X, v \rangle| \leq \varepsilon) \mathbb{P}_X(\langle X, u \rangle > t), \quad (7)$$

where u, v are unit vectors and $t, \varepsilon > 0$ and $X = (X_1, \dots, X_n)$ 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 (7). 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 [41])

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

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

We now turn to understand the more complicated small-ball event $|\langle X, v \rangle| \leq \varepsilon$ appearing in (7). 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

$$\mathbb{P}_X(|\langle X, v \rangle| \leq \varepsilon) \leq C\varepsilon, \quad (8)$$

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 $v = n^{-1/2}(1, \dots, 1)$ and the X_i are uniform in $\{-1, 1\}$ then

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

for any $0 < \varepsilon < n^{-1/2}$. This, of course, stands in contrast to (8) 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 [30]. For parameters $\alpha, \gamma \in (0, 1)$ define the *Least Common Denominator* (LCD) of $v \in \mathbb{R}^n$ to be

$$D_{\alpha, \gamma}(v) := \inf \left\{ \phi > 0 : \|\phi v\|_{\mathbb{T}} \leq \min \{ \gamma \phi \|v\|_2, \sqrt{\alpha n} \} \right\}, \quad (9)$$

where $\|v\|_{\mathbb{T}} := \text{dist}(v, \mathbb{Z}^n)$, for all $v \in \mathbb{R}^n$. What makes this definition useful is the important “inverse Littlewood-Offord theorem” of Rudelson and Vershynin [30], which tells us (roughly speaking) that one has (8) whenever $D_{\alpha, \gamma}(v) = \Omega(\varepsilon^{-1})$. We may now state our approximate version of (7), which uses $D_{\alpha, \gamma}(v)^{-1}$ as a proxy for $\mathbb{P}(|\langle X, v \rangle| \leq \varepsilon)$.

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

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

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 $|\langle X, v \rangle| \leq \varepsilon$ is replaced with a small-ball event of the form

$$|f(X_1, \dots, X_n)| \leq \varepsilon,$$

where f is a quadratic polynomial in variables X_1, \dots, X_n . The proof of this result is carried out in Section 5 and is an important aspect of this paper. Theorem 1.4, on the other hand, is only 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 A for completeness and to suggest further inquiry into inequalities of the form (7).

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

$$|\langle X, v \rangle| \leq \varepsilon \quad \text{and} \quad |\langle X, w_1 \rangle| \ll 1, \dots, |\langle X, w_k \rangle| \ll 1, \quad (10)$$

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

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

though in a more technical form.

To understand our result, again it makes sense to consider the two events in (10) in isolation. Since we have already discussed the subtle event $|\langle X, v \rangle| \leq \varepsilon$, it remains only to consider the event on the right of (10). Returning to the gaussian case, we note that if X has independent standard gaussian entries, then one may compute directly that

$$\mathbb{P}_X(|\langle X, w_1 \rangle| \ll 1, \dots, |\langle X, w_k \rangle| \ll 1) = \mathbb{P}(|X_1| \ll 1, \dots, |X_k| \ll 1) \leq e^{-\Omega(k)}, \quad (11)$$

by rotational invariance of the gaussian. Here the generalization to other random variables is not as subtle, and the well-known Hanson-Wright inequality tells us that (11) 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 $\mathbb{P}(|\langle X, v \rangle| \leq \varepsilon)$.

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

Theorem 1.5. *Let $\gamma \in (0, 1)$, $d \in \mathbb{N}$, $\alpha \in (0, 1)$, $0 \leq k \leq c_1 \alpha d$ and $\varepsilon \geq \exp(-c_1 \alpha d)$. Let $v \in \mathbb{S}^{d-1}$, let $w_1, \dots, w_k \in \mathbb{S}^{d-1}$ be orthogonal and let W be the matrix with rows w_1, \dots, w_k . If $X \in \{-1, 0, 1\}^d$ is a $1/4$ -lazy random vector and $D_{\alpha, \gamma}(v) > 16/\varepsilon$ then*

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

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 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 a $n \times d$ random $\{-1, 0, 1\}$ -matrix conditioned on the event $\|Hv\|_2 \leq \varepsilon n^{1/2}$, 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 $\leq e^{-ckn}$.

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

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

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 (21) 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, 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 supplementary paper [5].

Finally we note that it may be interesting to investigate to what extent one may sharpen these approximate negative correlation theorems in the direction of their idealized forms.

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, (13) and (14). 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 $X := X_1, \dots, X_{n+1}$ be the columns of A_{n+1} , let

$$V = \text{Span}\{X_2, \dots, X_{n+1}\}$$

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 [30] that allows for a geometric perspective

on the least singular value problem²

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

Here our first significant challenge presents itself: X and V are not independent and thus the event $\text{dist}(X, V) \leq \varepsilon$ is hard to understand directly. However, one can establish a formula for $\text{dist}(X, V)$ that is a rational function in the vector X with coefficients that depend only on V . This brings us to the useful inequality³ due to Vershynin [40],

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2}) \lesssim \sup_{r \in \mathbb{R}} \mathbb{P}_{A_n, X}(|\langle A_n^{-1} X, X \rangle - r| \leq \varepsilon \|A_n^{-1} X\|_2), \quad (12)$$

where we are ignoring the possibility of A_n being singular for now. We thus arrive at our main technical focus of this paper, which is bounding the quantity on the right-hand-side of (12).

We now make our two simplifying assumptions that shall allow us to give the overall shape of our proof without any added complexity. We 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 (12) satisfies

$$\|A_n^{-1} X\|_2 \approx \|A_n^{-1}\|_{\text{HS}}, \quad (13)$$

where we point out that $\|A_n^{-1}\|_{\text{HS}}^2 = \mathbb{E}_X \|A_n^{-1} X\|_2^2$. This seems to be a fairly innocent assumption at first, as the Hanson-Wright inequality tells us that $\|A_n^{-1} X\|_2$ is concentrated about its mean, for all reasonable A^{-1} . However, as we will see, mere concentration is not enough for us here.

As a second assumption, let us assume that the relevant A_n in the right-hand-side of (12) satisfies

$$\|A_n^{-1}\|_{\text{HS}} \approx cn^{1/2}. \quad (14)$$

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 $\|A_n^{-1}\|_{\text{HS}} = \Theta_\delta(n^{1/2})$ with probability $1 - \delta$. This, for example, follows from Vershynin's theorem [40] along with Corollary 8.5, which is based on the work of [14].

With these assumptions, our task reduces to proving

$$\min_r \mathbb{P}_X(|\langle A^{-1} X, X \rangle - r| \leq \varepsilon n^{1/2}) \lesssim \varepsilon, \quad (15)$$

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 (15) is $\geq cn^{-1/2}$ for *all* $\varepsilon > 0$, falling vastly short of our desired (15).

²Here and throughout we understand $A \lesssim B$ to mean that there exists an absolute constant $C > 0$ for which $A \leq CB$.

³In 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} .”

Thus, we need to introduce a collection of fairly strong “quasi-randomness properties” of A that hold with probability $1 - e^{-cn}$. These will ensure that A^{-1} is sufficiently “non-structured” to make our goal (15) possible. The most important and difficult of these quasi-randomness conditions is to show that all of the eigenvectors v of A satisfy

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

for some appropriate α, γ , where $D_{\alpha,\gamma}(v)$ is the *least common denominator* of v defined at (9). Roughly this means that none of the eigenvectors of A “correlate” with a re-scaled copy of the integer lattice $t\mathbb{Z}^n$, for any $e^{-cn} \leq t \leq 1$.

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 a separate, supplementary paper [5].

With these quasi-randomness conditions in tow, we can return to (15) and apply Esseen’s inequality to bound the left-hand-side of (15) in terms of the characteristic function $\varphi(\theta)$ of the random variable $\langle A^{-1}X, X \rangle$,

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

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 [7, 24]. 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 $\varphi(\theta)$ point-wise in terms of an average over a related collection of characteristic functions of *linear* sums of independent random variables

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

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

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

for all but exponentially few Y , allowing us to show

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

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 (13), (14). 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 (14), whose resolution governs the overall structure of the paper.

2.2.1. Removing the assumption (14). What is most concerning about making the assumption $\|A_n^{-1}\|_{\text{HS}} \approx n^{-1/2}$ is that it is, in a sense, *circular*: If we assume the modest-looking hypothesis $\mathbb{E}\|A^{-1}\|_{\text{HS}} \lesssim n^{1/2}$, we would be able to deduce

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

by Markov. In other words, showing that $\|A^{-1}\|_{\text{HS}}$ is concentrated about $n^{-1/2}$ (in the above sense) actually *implies* Theorem 1.1. However this is not as worrisome as one might first suspect: 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

$$\mathbb{E}_A \|A^{-1}\|_{\text{HS}}^\alpha \mathbf{1}(\sigma_{\min}(A_n) \geq e^{-cn}) \approx n^{\alpha/2},$$

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 non-zero probability.

2.2.2. Base case. In the first step of our iteration, we prove a “base case” of

$$\mathbb{P}(\sigma_{\min}(A_n) \leq \varepsilon/\sqrt{n}) \lesssim \varepsilon^{1/4} + e^{-cn} \quad (16)$$

without the assumption (14) which is equivalent to

$$\mathbb{E}_{A_n} \|A_n^{-1}\|_{\text{HS}}^{1/4} \mathbf{1}(\sigma_{\min}(A_n) \geq e^{-cn}) \approx n^{1/8}.$$

To prove this “base case” we upgrade (12) to

$$\mathbb{P}\left(\sigma_{\min}(A_{n+1}) \leq \frac{\varepsilon}{\sqrt{n}}\right) \lesssim \varepsilon + \sup_{r \in \mathbb{R}} \mathbb{P}\left(\frac{|\langle A_n^{-1}X, X \rangle - r|}{\|A_n^{-1}X\|_2} \leq C\varepsilon, \|A_n^{-1}\|_{\text{HS}} \leq \frac{n^{1/2}}{\varepsilon}\right). \quad (17)$$

In other words, we can intersect with the event

$$\|A_n^{-1}\|_{\text{HS}} \leq n^{1/2}/\varepsilon \quad (18)$$

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 (16). For this, we first use the Hanson-Wright inequality to give a weak version of (13), and then use (18) as a weak version of our assumption(14). We note that this base step (16) 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 (16) to the near-optimal result

$$\mathbb{P}(\sigma_{\min}(A_n) \leq \varepsilon/\sqrt{n}) \lesssim \varepsilon\sqrt{\log 1/\varepsilon} + e^{-cn}. \quad (19)$$

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

$$\mathbb{P}_X(\|A^{-1}X\|_2 \leq s), \quad (20)$$

for $s \in (\varepsilon, n^{-1/2})$, 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 (20), 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

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

While this leads to a decent first bound of $O(\delta s)$ on the probability (20) (after using the quasi-randomness properties of A), it is not enough for our purposes and we have to use 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 (20) is bounded above by

$$\mathbb{P}_X\left(|\langle X, v_1 \rangle| \leq s\delta \text{ and } |\langle X, v_i \rangle| \leq \sigma_i s \text{ for all } i = 2, \dots, n-1\right), \quad (21)$$

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 (21) is at most

$$O(\delta s) \exp(-cN_A(-c/s, c/s)) \quad (22)$$

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 $|\langle X, v_1 \rangle| \leq s\delta$ and the second factor comes from approximating

$$\mathbb{P}_X\left(|\langle X, w_i \rangle| < c \text{ for all } i \text{ s.t. } s\sigma_i < c\right) \approx \exp(-cN_A(-c/s, c/s)). \quad (23)$$

This bound is now sufficiently strong for our purposes, *provided* the spectrum of A adheres sufficiently closely to the typical spectrum of A .

This now leads us to understand the rest of the spectrum of A_n and, in particular, the next smallest singular values $\sigma_{n-1}, \sigma_{n-2}, \dots$. This might seem like a step backwards as we are now forced 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 (for the most part) on scales of size $\Omega(n^{-1/2})$, which is now well understood due to the important work of Erdős, Schlein and Yau [14]. Although some additional information is needed about how the eigenvalues cluster on much smaller scales. Here we can use the work of Nguyen [27] and our own Theorem 1.3 on the crowding of the spectrum.

These results ultimately allow us to derive sufficiently strong results on quantities of the form (20), which in-turn allow us to prove our “bootstrapping lemma”. We then use this lemma to prove the near-optimal result

$$\mathbb{P}(\sigma_{\min}(A_n) \leq \varepsilon/\sqrt{n}) \lesssim \varepsilon\sqrt{\log 1/\varepsilon} + e^{-cn}. \quad (24)$$

2.2.4. Removing the assumption (13) and the last jump to Theorem 1.1. We now turn to discuss how to remove our simplifying assumption (13), made above, which will allow us to close the gap between (24) and Theorem 1.1.

To achieve this, we need to consider how $\|A^{-1}X\|_2$ varies about $\|A^{-1}\|_{\text{HS}}$. Now, the Hanson-Wright inequality tells us that indeed $\|A^{-1}X\|_2$ is concentrated about $\|A^{-1}\|_{\text{HS}}$, on the scale of $\lesssim \|A^{-1}\|_{\text{op}}$. 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

$$\{|\langle A^{-1}X, X \rangle - r| < K\varepsilon\|A^{-1}\|_{\text{HS}}\} \text{ and } \{\|A^{-1}X\|_2 > K\|A^{-1}\|_{\text{HS}}\} \quad (25)$$

for all $K > 0$. Our first step towards understanding (25) is to replace the *quadratic* large deviation event $\|A^{-1}X\|_2 > K\|A^{-1}\|_{\text{HS}}$ with a collection of *linear* large deviation events:

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

where w_n, w_{n-1}, \dots, w_1 are the eigenvectors of A corresponding to singular values $\sigma_n \leq \sigma_{n-1} \leq \dots \leq \sigma_1$ 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 (although we don’t go into details here): if the “bulk” spectrum of A^{-1} is behaving erratically this replacement step will be too lossy for our purposes. Thus we are lead to prove another result showing that the spectrum of A^{-1} adheres sufficiently to its typical spectrum. 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 works of Erdős, Schlein and Yau [14], Nguyen [27] and our own Theorem 1.3 on the crowding of the spectrum.

Thus we have reduced the task of proving the approximate independence of the events at (25) to proving the approximate independence of the collection of events

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

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

$$\mathbb{P}(|\langle A^{-1}X, X \rangle - t| < \delta, \langle X, u \rangle \geq s) \lesssim \delta e^{-s} \int_{-1/\delta}^{1/\delta} \left| \mathbb{E} e^{2\pi i \theta \langle A^{-1}X, X \rangle + \langle X, u \rangle} \right| d\theta. \quad (26)$$

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 (24) 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 $1 - e^{-\Omega(n)}$ and that we condition on for most of the paper. In Section 5 we detail our Fourier decoupling argument and establish an inequality of the form (26). 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 (17), 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 (20) and (23). In Section 10 we establish our base step (16) and bootstrap this to prove the near optimal bound (24). 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-zero, variance 1 random variable. We define the *subgaussian moment* of ζ to be

$$\|\zeta\|_{\psi_2} := \sup_{p \geq 1} \frac{1}{\sqrt{p}} (\mathbb{E}|\zeta|^p)^{1/p}.$$

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

For $\zeta \in \Gamma$, define $\text{Sym}_n(\zeta)$ to be the probability space on $n \times n$ symmetric matrices A for which $(A_{i,j})_{i \geq j}$ are independent and distributed according to ζ . Similarly, we write $X \sim \text{Col}_n(\zeta)$ 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 $\{\text{Sym}_n(\zeta)\}_n$ as coupled in the natural way: the matrix $A_{n+1} \sim \text{Sym}_{n+1}(\zeta)$ can be sampled by first sampling $A_n \sim \text{Sym}_n(\zeta)$, which we think of as the principle 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 $X \sim \text{Col}_n(\zeta)$. 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 principle minor A_n of A_{n+1} , due to the bound (12) 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 [30].

For parameters $\rho, \delta \in (0, 1)$, we define the set of compressible vectors $\text{Comp}(\delta, \rho)$ 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, i.e. $\text{Incomp}(\delta, \rho) := \mathbb{S}^{n-1} \setminus \text{Comp}(\delta, \rho)$. The following basic fact about incompressible vectors from [30] 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 $v \in \text{Incomp}(\delta, \rho)$ we have that $|v_j| \sqrt{n} \in [c_{\rho, \delta}, c_{\rho, \delta}^{-1}]$ 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 [40], we will safely be able to ignore compressible vectors. In particular, [40, Proposition 4.2] implies the following.

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

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

and

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

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.

We deduce Lemma 3.2 from [40, Proposition 4.2] with a very simple net argument in Appendix B.

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 i th 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 $X \leq CY$. 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. QUASIRANDOMNESS 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 principle 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 a supplementary paper [5].

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 $A = A_n \sim \text{Sym}_n(\zeta)$ and let $X \sim \text{Col}_n(\zeta)$ 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

$$\mathcal{E}_1 = \{\|A\|_{op} \leq 4\sqrt{n}\}. \quad (27)$$

For the next property we need a definition. Let $X, X' \sim \text{Col}_n(\zeta)$ 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, i.e. 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

$$\mathcal{E}_2 = \left\{ \mathbb{P}_{\tilde{X}} \left(A^{-1} \tilde{X} / \|A^{-1} \tilde{X}\|_2 \in \text{Comp}(\delta, \rho) \right) \leq e^{-c_2 n} \right\}. \quad (28)$$

We remind the reader that $\text{Comp}(\delta, \rho)$ 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 $\tilde{X} = 0$, we interpret $\mathbb{P}_{\tilde{X}}(A^{-1} \tilde{X} / \|A^{-1} \tilde{X}\|_2 \in \text{Comp}(\delta, \rho)) = 1$.

Recalling the least common denominator defined at (9), we now define the event \mathcal{E}_3 by

$$\mathcal{E}_3 = \{D_{\alpha, \gamma}(u) \geq e^{c_3 n} \text{ for every unit eigenvector } u \text{ of } A\}. \quad (29)$$

The next condition tells us that the random vector $A^{-1} \tilde{X}$ 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

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

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

$$\mathcal{E}_4 = \left\{ \mathbb{P}_{\tilde{X}} \left(\hat{D}_{\alpha,\gamma,\mu}(v) < e^{c_4 n} \right) \leq e^{-c_4 n} \right\} . \quad (30)$$

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

We now define our main quasirandomness event \mathcal{E} to be the intersection of these events:

$$\mathcal{E} := \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4 . \quad (31)$$

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 (28), (29) and (30) so that*

$$\mathbb{P}_A(\mathcal{E}^c) \leq 2e^{-\Omega(n)} . \quad (32)$$

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 $|I| \geq (1 - 2\mu)n$ satisfies

$$\|w\|_2 \leq c_{\rho,\delta}^{-2} \|w_I\|_2, \quad (33)$$

for every $w \in \text{Incomp}(\delta, \rho)$. 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

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

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

$$\Sigma = \Sigma_{\alpha,\gamma,\mu} := \{v \in \mathbb{S}^{n-1} : \hat{D}_{\alpha,\gamma,\mu}(v) \leq e^{c_\Sigma n}\} . \quad (34)$$

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

$$q_n(w) = q_n(w; \alpha, \gamma, \mu) := \mathbb{P}_A(\exists v \in \Sigma \text{ and } \exists s, t \in [-4\sqrt{n}, 4\sqrt{n}] : Av = sv + tw) \quad (35)$$

and set

$$q_n = q_n(\alpha, \gamma, \mu) := \sup_{w \in \mathbb{S}^{n-1}} q_n(w) . \quad (36)$$

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

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*

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

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 supplementary paper [5]. 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 pseudorandom 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 [16] we may deduce⁴ the following concentration bound

$$\mathbb{P}(\|A\|_{op} \geq (3+t)\sqrt{n}) \lesssim e^{-ct^{3/2}n}, \quad (37)$$

which holds for all $t \geq 0$. Thus, by (37), the event \mathcal{E}_1 at (27) 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

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

Applying Markov's inequality shows

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

and so the event in (28) fails with probability at most $O(e^{-\Omega(n)})$, under the event $\tilde{X} \neq 0$. By Theorem 3.1.1 in [41] we have that

$$\mathbb{P}_{\tilde{X}}(\tilde{X} = 0) \leq e^{-\Omega(\mu n)}. \quad (38)$$

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

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

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

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

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

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

The event \mathcal{E}_4 : Note first that by (38), we may assume $\tilde{X} \neq 0$. For a fixed instance of $\tilde{X} \neq 0$, we have

$$\mathbb{P}_A\left(\hat{D}_{\alpha, \gamma, \mu}\left(A^{-1}\tilde{X}/\|\tilde{X}\|_2\right) < e^{c_4 n}\right) \leq \mathbb{P}_A(\exists v \in \Sigma : Av = \tilde{X}/\|\tilde{X}\|_2) \leq q_n\left(\tilde{X}/\|\tilde{X}\|_2\right), \quad (39)$$

⁴Technically, the result of [16] is sharper and for random matrices whose entries are symmetric random variables. We deduce (37) for random variables from [16], using a symmetrization trick. This easy deduction is in Appendix C.

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

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

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

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 (39).

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. \square

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 $X \sim \text{Col}_n(\zeta)$. Let M be an $n \times n$ symmetric matrix, $u \in \mathbb{R}^n$, $t \in \mathbb{R}$ and $s, \delta \geq 0$. Then*

$$\mathbb{P}(|\langle MX, X \rangle - t| < \delta, \langle X, u \rangle \geq s) \lesssim \delta e^{-s} \int_{-1/\delta}^{1/\delta} |\mathbb{E} e^{2\pi i \theta \langle MX, X \rangle + \langle X, u \rangle}| d\theta. \quad (40)$$

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 [21] (a similar technique appears in [1]). 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 (40). To state this lemma, let us recall that \mathcal{E} (defined at (31)) 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 fixed constant throughout this paper.

Lemma 5.2. *For $B > 0$, let $\zeta \in \Gamma_B$, $X \sim \text{Col}_n(\zeta)$ 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 $s \geq 0, \delta > e^{-cn}$ and $u \in \mathbb{S}^{n-1}$. Then*

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

where

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

$X' \sim \text{Col}_n(\zeta)$ 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 right direction.

All works bounding the singularity probability for random symmetric matrices contain a related decoupling step [3, 4, 6, 17, 20, 26, 40], starting with Costello, Tao and Vu's breakthrough [8] building off of Costello's earlier work [7] on anticoncentration of bilinear and quadratic forms. A subtle difference in the decoupling approach from [21] 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 [15]: if Z is a random variable taking values in \mathbb{R} with characteristic function $\varphi_Z(\theta) := \mathbb{E}_Z e^{2\pi i \theta Z}$, then for all $t \in \mathbb{R}$ we have

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

We shall also use the following basic fact about subgaussian random vectors (see, for example, [41, Prop. 2.6.1]): If $\zeta \in \Gamma_B$ and $Y \sim \text{Col}_n(\zeta)$ then for every vector $u \in \mathbb{R}^n$ we have

$$\mathbb{E}_Y e^{\langle Y, u \rangle} \leq \exp(2B^2 \|u\|_2^2). \quad (41)$$

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

$$\mathbb{P}_X(|\langle MX, X \rangle - t| < \delta, \langle X, u \rangle \geq s) \leq e^{-s} \mathbb{E} [\mathbf{1}\{|\langle MX, X \rangle - t| < \delta\} e^{\langle X, u \rangle}]. \quad (42)$$

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

$$\mathbb{P}(Y \in U) = (\mathbb{E} e^{\langle X, u \rangle})^{-1} \mathbb{E} [\mathbf{1}_U e^{\langle X, u \rangle}], \quad (43)$$

for all open $U \subseteq \mathbb{R}^n$. Note that the expectation $\mathbb{E} e^{\langle X, u \rangle}$ is finite by (41). We now use this definition to rewrite the expectation on the right-hand-side of (42),

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

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

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

By the definition of Y we have

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

completing the lemma. \square

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 [21].

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

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

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

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

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

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

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

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

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

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 work [40].

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*

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

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 $\zeta \in \Gamma_B$, $X \sim \text{Col}_n(\zeta)$ and let $u, v \in \mathbb{R}^n$. Then*

$$\left| \mathbb{E}_X e^{2\pi i \langle X, v \rangle + \langle X, u \rangle} \right| \leq \exp \left(-c \min_{r \in [1, c^{-1}]} \|rv\|_{\mathbb{T}}^2 + c^{-1} \|u\|_2^2 \right), \quad (44)$$

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

$$|\mathbb{E}_X e^{2\pi i \langle X, v \rangle + \langle X, u \rangle}|^2 \leq \mathbb{E} e^{\langle \tilde{Y}, u \rangle} \prod_j \cos(2\pi \tilde{X}_j v_j) \leq \left(\mathbb{E}_{\tilde{Y}} e^{2\langle \tilde{Y}, u \rangle} \right)^{1/2} \left(\prod_j \mathbb{E}_\xi |\cos(2\pi \xi v_j)| \right)^{1/2}, \quad (45)$$

where we have applied the Cauchy-Schwarz inequality along with the bound $|\cos(x)|^2 \leq |\cos(x)|$ to obtain the last inequality. By (41), the first expectation on the right-hand-side of (45) is at most $\exp(O(\|u\|_2^2))$. Applying Fact 5.4 completes the Lemma. \square

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 a $n \times n$ 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 $\tilde{X} := X_J - X'_J$.

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

$$\mathcal{F}_1 := \{|J| \in [\mu n/2, 2\mu n]\} \quad (46)$$

$$\mathcal{F}_2 := \{\|\tilde{X}\|_2 n^{-1/2} \in [c, c^{-1}]\} \quad (47)$$

$$\mathcal{F}_3(A) := \{A^{-1}\tilde{X}/\|A^{-1}\tilde{X}\|_2 \in \text{Incomp}(\delta, \rho)\}. \quad (48)$$

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

$$\mathcal{F}_4(A) := \left\{ D_{\alpha, \gamma} \left(\frac{v_I}{\|v_I\|} \right) > e^{cn} \right\}. \quad (49)$$

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 (46), (47), (48), (49) hold in all that follows. We recall that the constants $\delta, \rho, \mu, \alpha, \gamma$ were chosen in Lemma 3.2 and Lemma 4.1 as a function of the subgaussian moment B . Thus the only new parameter in \mathcal{F} is the constant c in lines (47) and (49).

Lemma 5.6. *For $B > 0$, let $\zeta \in \Gamma_B$, let $X, X' \sim \text{Col}_n(\zeta)$ be independent and let $J \subseteq [n]$ be a μ -random subset. Let A be a $n \times n$ real symmetric matrix with $A \in \mathcal{E}$. We may choose the constant $c \in (0, 1)$ appearing in (47) and (49) 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 \tilde{X} are independent, subgaussian, and have variance 2μ , and so $\tilde{X}/(\sqrt{2\mu})$ 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 [41] we have

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

For $\mathcal{F}_3(A), \mathcal{F}_4(A)$, recall that $A \in \mathcal{E}$ means that (28) and (30) hold, thus exponential bounds on $\mathbb{P}(\mathcal{F}_3^c)$ and $\mathbb{P}(\mathcal{F}_4^c)$ follow from Markov's inequality. \square

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

$$\mathbb{P}_X \left(|\langle A^{-1}X, X \rangle - t| \leq \delta \mu_1, \langle X, u \rangle \geq s \right) \lesssim \delta e^{-s} \int_{-1/\delta}^{1/\delta} \left| \mathbb{E}_X e^{2\pi i \theta \frac{\langle A^{-1}X, X \rangle}{\mu_1} + \langle X, u \rangle} \right| d\theta, \quad (50)$$

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

$$\left| \mathbb{E}_X e^{2\pi i \theta \frac{\langle A^{-1}X, X \rangle}{\mu_1} + \langle X, u \rangle} \right|^2 \leq \mathbb{E}_J \mathbb{E}_{X_J, X'_J} e^{\langle (X+X')_J, u \rangle} \cdot \left| \mathbb{E}_{X_I} e^{4\pi i \theta \langle \frac{A^{-1}\tilde{X}}{\mu_1}, X_I \rangle + 2\langle X_I, u \rangle} \right|, \quad (51)$$

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

We first consider the contribution to the expectation on the right-hand-side of (51) 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

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

By Cauchy-Schwarz, (41) and Lemma 5.6, we have

$$\mathbb{E}_{J, X, X'}^{\mathcal{F}^c} e^{\langle Y, u \rangle} \leq \mathbb{E}_{J, X, X'} [e^{\langle Y, 2u \rangle}]^{1/2} \mathbb{P}_{J, X_J, X'_J}(\mathcal{F}^c)^{1/2} \lesssim e^{-\Omega(n)}. \quad (52)$$

We now consider the contribution to the expectation on the right-hand-side of (51) from triples $(J, X_J, X'_J) \in \mathcal{F}$. For this, let $w = w(X) := \frac{A^{-1}\tilde{X}}{\mu_1}$ and assume $(J, X_J, X'_J) \in \mathcal{F}$. By Lemma 5.5, we have

$$\left| \mathbb{E}_{X_I} e^{4\pi i \theta \langle X_I, w \rangle + \langle X_I, 2u \rangle} \right| \lesssim \exp \left(-c \min_{r \in [1, c^{-1}]} \|2r\theta w_I\|_{\mathbb{T}}^2 \right). \quad (53)$$

Note that $\|w_I\|_2 \leq \|\tilde{X}\|_2 \leq c^{-1}\sqrt{n}$, by the definition of $\mu_1 = \sigma_{\max}(A^{-1})$ and line (47) in the definition of $\mathcal{F}(A)$.

Now, from property (49) in that definition and by the hypothesis $\delta > e^{-c_1 n}$, we may choose $c_1 > 0$ small enough so that

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

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

$$\min_{r \in [1, c^{-1}]} \|2r\theta w_I\|_{\mathbb{T}} = \min_{r \in [1, c^{-1}]} \left\| 2r\theta \|w_I\|_2 \cdot \frac{w_I}{\|w_I\|_2} \right\|_{\mathbb{T}} \geq \min \left\{ \gamma\theta \|w_I\|_2, \sqrt{\alpha|I|} \right\}. \quad (54)$$

So for $|\theta| \leq 1/\delta$ we use (54) in (53) to bound the right-hand-side of (51) as

$$\mathbb{E}_{J, X_J, X'_J}^{\mathcal{F}} e^{\langle (X+X')_J, u \rangle} \cdot \left| \mathbb{E}_{X_I} e^{4\pi i \theta \langle w, X_I \rangle + 2\langle X_I, u \rangle} \right| \lesssim \mathbb{E}_{J, X_J, X'_J}^{\mathcal{F}} e^{\langle (X+X')_J, u \rangle} e^{-c \min \{ \gamma^2 \theta^2 \|w_I\|_2^2, \alpha |I| \}}. \quad (55)$$

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

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

for some $C > 0$ (see (33)). Thus the right-hand-side of (55) is

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

Combining this with (55), (51) obtains the desired bound in the case in the case $(J, X_J, X'_J) \in \mathcal{F}$. Combining this with (52) completes the proof of Lemma 5.2. \square

6. PREPARATION FOR THE “BASE STEP” OF THE ITERATION

As we mentioned at (12), Vershynin [40], gave a natural way of bounding the least singular value of a random symmetric matrix:

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

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 $\sigma_{\min}(A_n) \geq \varepsilon/n^{1/2}$ in the probability on the right-hand-side at a loss of only $C\varepsilon$. This will be crucial for the base step in our iteration, since the bound we obtain on $\mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon/n^{1/2})$ deteriorates as $\sigma_{\min}(A_n)$ decreases.

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

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

where $C > 0$ depends only on B .

We deduce this lemma from a geometric form of the lemma. For this, we let X_j denote the j th column of A_{n+1} , let H_j be the linear span of $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_{n+1}$, and let $d_j(A_{n+1}) := \text{dist}(X_j, H_j)$.

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

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon/\sqrt{n}) \lesssim \varepsilon + \mathbb{P}(d_1(A_{n+1}) \leq C\varepsilon \text{ and } \sigma_{\min}(A_n) \geq \varepsilon/\sqrt{n}) + e^{-\Omega(n)},$$

where $C > 0$ depends only on B .

6.1. Preparations. We require an elementary, but extremely useful, fact from linear algebra. This fact is actually a key step in the work of Nguyen, Tao and Vu on eigenvalue repulsion in random matrices (see [25, Section 4]) and we reproduce their short proof here for completeness. If M is a $n \times n$ matrix and $j \in [n]$, let $M^{(j)}$ denote the j th principle minor of M , i.e. M with the j th row and column removed.

Fact 6.3. *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 eigenvalue of the minor $M^{(j)}$ with corresponding unit eigenvector v . Then*

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

where $X^{(j)}$ is the j th column of M with the j th 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

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

□

We will apply Fact 6.3 to see that when both $\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2}$ and $\sigma_{\min}(A^{(j)}) \leq \varepsilon n^{-1/2}$ hold we have $|\langle v, X^{(j)} \rangle| \lesssim \varepsilon$, assuming that $|u_j| \approx n^{-1/2}$. We then show that this latter event holds, subject to appropriate pseudo-random conditions, with probably $O(\varepsilon)$. The only wrinkle in this line of thinking is that we cannot rule out the possibility that for a *given* j we have $|u_j| \ll n^{-1/2}$. We *can*, however, rule out the possibility that many such $|u_j|$ are small, which will be enough for us. For this, we use a theorem of Rudelson and Vershynin [32] which we state here in a specialized form.

Theorem 6.4 (Theorem 1.5 of [32]). *For $B > 0$, $\zeta \in \Gamma_B$, let $A \sim \text{Sym}_n(\zeta)$ 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*

$$\mathbb{P}(|\{j : |v_j| \leq (c_2 c_1)^6 n^{-1/2}\}| \geq c_1 n) \leq e^{-c_1 n},$$

for n sufficiently large.

To understand the event that $|\langle v, X^{(j)} \rangle| \lesssim \varepsilon$ (mentioned above), we need the Littlewood-Offord theorem of Rudelson and Vershynin [30], which we state here in a specialized form. Recall that $D_{\alpha, \gamma}(v)$ is the least common denominator of the vector v , as defined at (9).

Theorem 6.5. *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 $X \sim \text{Col}_n(\zeta)$, where $\zeta \in \Gamma_B$. Then*

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

Here $c > 0$ depends only on B .

The final ingredient in the proof of Theorem 6.1 is the observation that the event

$$\{\sigma_{\min}(A_{n+1}^{(1)}) \leq \varepsilon n^{-1/2}\} \cap \{\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2}\}$$

(as in Lemma 6.1) is roughly equivalent to the event

$$\{\text{there exist } \geq cn \text{ values of } j \text{ for which } \sigma_{\min}(A_{n+1}^{(j)}) \leq \varepsilon n^{-1/2}\} \cap \{\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2}\}.$$

Before we make this rigorous we prove this latter event has probability $\leq \varepsilon + e^{-\Omega(n)}$.

Lemma 6.6. For $B > 0$, $\zeta \in \Gamma_B$, let $A_{n+1} \sim \text{Sym}_{n+1}(\zeta)$. Then, for $\varepsilon > 0$, we have

$$\mathbb{P}\left(\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2} \text{ and } |\{j : \sigma_{\min}(A_{n+1}^{(j)}) \leq \varepsilon n^{-1/2}\}| \geq cn\right) \lesssim \varepsilon + e^{-\Omega(n)}, \quad (56)$$

where $c > 0$ depends only on B .

Proof. Let \mathcal{A}_1 denote the event on left-hand-side of (56). Let v be a unit eigenvector corresponding to the least singular value of A_{n+1} . We first show that if \mathcal{A}_1 holds then with probability $1 - e^{-\Omega(n)}$, we can find $\geq cn/2$ values of $j \in [n]$ so that $\sigma_{\min}(A_{n+1}^{(j)}) < \varepsilon n^{-1/2}$ and $|v_j| \gtrsim n^{-1/2}$. With this in mind we let

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

where c_2 is the constant from Theorem 6.4. We let \mathcal{A}_2 denote the event that $|S_2| < cn/2$ and apply Theorem 6.4 with $c_1 = c/2$ to see that $\mathbb{P}(\mathcal{A}_2^c) \leq e^{-cn/2}$. Now set $S := S_1 \cap ([n] \setminus S_2)$ and note that if $\mathcal{A}_1 \cap \mathcal{A}_2$ holds then $|S| \geq cn/2$.

Now, for $j \in [n]$, let $w_j = w(A_{n+1}^{(j)})$ denote a unit eigenvector of $A_{n+1}^{(j)}$ corresponding to the least singular value of $A_{n+1}^{(j)}$. Note that if $j \in S_1 \cap ([n] \setminus S_2)$ then, by Fact 6.3, we have

$$|\langle w_j, X^{(j)} \rangle| \leq 2\varepsilon/(c_2 c/2)^6 =: C\varepsilon. \quad (57)$$

Now let \mathcal{Q}_j be the event that w_j satisfies $D_{\alpha, \gamma}(w_j) \geq e^{c_3 n}$ where α, γ, c_3 are chosen according to Lemma 4.1 and set $\mathcal{Q} = \cap_j \mathcal{Q}_j$. By Lemma 4.1 we have $\mathbb{P}(\mathcal{Q}^c) \lesssim e^{-\Omega(n)}$.

Putting this all together, we define the random variable

$$R := n^{-1} \sum_{j=1}^n \mathbf{1}(|\langle w_j, X^{(j)} \rangle| \leq C\varepsilon \text{ and } \mathcal{Q}_j),$$

and then observe that

$$\mathbb{P}(\mathcal{A}_1) \leq \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{Q}) + e^{-\Omega(n)} \leq \mathbb{P}(R \geq c/2) + e^{-\Omega(n)}.$$

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

$$\mathbb{P}(R \geq c/2) \lesssim n^{-1} \sum_{i=1}^n \mathbb{E}_{A_{n+1}^{(j)}} \mathbb{P}_{X^{(j)}}(|\langle w_j, X^{(j)} \rangle| \leq C\varepsilon \cap \mathcal{Q}_j) \lesssim \varepsilon$$

where the last inequality follows from the fact that $X^{(j)}$ is independent of the event \mathcal{Q}_j and w_j and therefore we may put the property \mathcal{Q}_j to use by applying Theorem 6.5. \square

To prove Lemma 6.2, we will also use a basic fact which is at the heart of the geometric approach of Rudelson and Vershynin (see, e.g., [30, Lemma 3.5]).

Fact 6.7. Let M be an $n \times n$ matrix and v be a unit vector satisfying $\|Mv\|_2 = \sigma_{\min}(M)$. Then

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

Proof. Let X_j denote the j th column of M and let H_j denote the span of the remaining columns. Then

$$\sigma_{\min}(M) = \|Mv\|_2 \geq \text{dist}(Mv, H_j) = \text{dist}(v_j X_j, H_j) = |v_j| d_j(M).$$

□

6.2. Proofs of Lemma 6.2 and Lemma 6.1. With these preliminaries in-hand, we are now in a position to prove Lemma 6.2.

Proof of Lemma 6.2. We look to bound the quantity

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2}).$$

Let v denote a unit eigenvector corresponding to the least singular value of A_{n+1} . Let \mathcal{A} denote the event that $v \in \text{Incomp}(\delta, \rho)$: at least $c_{\rho, \delta} n$ coordinates of v have absolute value at least $c_{\rho, \delta} n^{-1/2}$. By Lemma 3.2, $\mathbb{P}(\mathcal{A}^c) \lesssim e^{-\Omega(n)}$ and so

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2}) \leq \mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2} \text{ and } \mathcal{A}) + e^{-\Omega(n)}.$$

Now let $c > 0$ denote the constant from Lemma 6.6 and let \mathcal{B} denote the event that at most cn principal minors of A_{n+1} satisfy $\sigma_{\min}(A_{n+1}^{(j)}) \leq \varepsilon n^{-1/2}$. Also note we may assume $c \leq c_{\rho, \delta}/2$. By Lemma 6.6 we have

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2} \text{ and } \mathcal{B}^c) \lesssim \varepsilon + e^{-\Omega(n)}$$

and so

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2}) \leq \mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon n^{-1/2} \text{ and } \mathcal{A} \cap \mathcal{B}) + C\varepsilon + e^{-\Omega(n)}.$$

Now let

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

Observe that if $\sigma_{\min}(A_{n+1}) \leq \varepsilon/\sqrt{n}$ and $j \in [n]$ is such that $|v_j| \geq c_{\rho, \delta} n^{-1/2}$, then $d_j(A_{n+1}) \leq \varepsilon/c_{\rho, \delta}$, by Fact 6.7. Thus if $\sigma_{\min}(A_{n+1}) \leq \varepsilon/\sqrt{n}$ and \mathcal{A} hold, then there are $\geq c_{\rho, \delta} n$ values of j for which $d_j(A_{n+1}) < \varepsilon/c_{\rho, \delta}$. If \mathcal{B} holds in addition to $\sigma_{\min}(A_{n+1}) \leq \varepsilon/\sqrt{n}$ and \mathcal{A} , then at most $c_{\rho, \delta} n/2$ of these values of j have $\sigma_{\min}(A_{n+1}^{(j)}) < \varepsilon n^{-1/2}$. In other words,

$$\mathcal{A} \cap \mathcal{B} \cap \{\sigma_{\min}(A_{n+1}) \leq \varepsilon/\sqrt{n}\} \subseteq \{|S| \geq c_{\rho, \delta} n/2\}. \quad (58)$$

Using (58) along with Markov's inequality tells us that

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon/\sqrt{n} \text{ and } \mathcal{A} \cap \mathcal{B}) \leq \mathbb{P}(|S| \geq c_{\rho, \delta} n/2) \leq \frac{2}{c_{\rho, \delta} n} \mathbb{E}|S|. \quad (59)$$

If we write

$$|S| = \sum_j \mathbf{1}(d_j(A_{n+1}) \leq \varepsilon/c_{\rho, \delta}, \sigma_{\min}(A_{n+1}^{(j)}) \geq \varepsilon n^{-1/2}),$$

then we see that

$$\mathbb{E}|S| = \mathbb{P}\left(d_1(A_{n+1}) \leq \varepsilon/c_{\rho, \delta}, \sigma_{\min}(A_{n+1}^{(1)}) \geq \varepsilon/\sqrt{n}\right), \quad (60)$$

by symmetry. Putting (58), (59), (60) together gives us our desired conclusion. □

Lemma 6.1 now follows.

Proof of Lemma 6.1. If we set $a_{1,1}$ to be the first entry of $A = A_{n+1}$ then, by [40, Prop. 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 [40, Prop. 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. \square

7. EIGENVALUE CROWDING (AND THE PROOFS OF THEOREM 1.2 AND THEOREM 1.3)

The main purpose of this section is to prove the following theorem which gives an upper-bound on the probability that $k \geq 2$ eigenvalues of a random matrix fall in an interval of length ε . This will be key in our work on the “bulk” of the spectrum of A^{-1} in Section 8. This result is of independent interest as the $\varepsilon = 0$ case 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, Tao and Vu [25].

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

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

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

where $C, c > 0$ are constants depending on B and c additionally depends on j .

We suspect that the bound in Lemma 1.3 is actually *far* from the truth, for $\varepsilon > e^{-cn}$ and $j \geq 1$. In fact, one expects *quadratic* dependence on j in the exponent of ε . This type of dependence was recently confirmed by Nguyen [27] for $\varepsilon > e^{-n^c}$. As we shall also need Nguyen’s result, we discuss it further in Section 8.

For the proof of Lemma 1.3, we remind the reader that if $u \in \mathbb{R}^n \cap \text{Incomp}(\rho, \delta)$ 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 a $n \times n$ symmetric matrix A , we use the notation $A^{(i_1, \dots, i_r)}$ to refer to the minor of A for which the rows and columns indexed by i_1, \dots, 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 $1 \leq k + j < n$, let A be a $n \times n$ symmetric matrix for which*

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

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

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

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

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

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

then Cauchy's interlacing theorem implies

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

for all $r = 1, \dots, j$. So let $w^{(r)}$ denote a unit eigenvector of M_r corresponding to eigenvalue $\lambda_k(M_r)$. Thus, by Fact 6.3, we see that

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

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

Proof of Theorem 1.3. Note 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 $A_n^{(i_1, \dots, i_j)}$ lie in $\text{Incomp}(\rho, \delta)$ 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 Lemma 4.1 and Lemma 3.2, we have $\mathbb{P}(\mathcal{A}^c) \leq n^{j+1} e^{-\Omega(n)} \lesssim e^{-cn}$ by taking c small enough.

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

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

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

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

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

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

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

$$\mathbb{P}(|\lambda_{k+j}(A_n) - \lambda_k(A_n)| \leq \varepsilon / \sqrt{n}) \leq \mathbb{P}(N \geq (c_{\rho, \delta} n / 2)^j \text{ and } \mathcal{A}) + O(e^{-cn}) \quad (61)$$

$$\leq \left(\frac{2}{c_{\rho, \delta}} \right)^j \mathbb{P}(\mathcal{E}_{1, \dots, j} \cap \mathcal{A}) + O(e^{-cn}) \quad (62)$$

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

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

$$\mathbb{P}(\mathcal{E}_{1,\dots,j} \cap \mathcal{A}) \leq \max_{w_1,\dots,w_j: D_{\alpha,\gamma}(w_i) > e^{c_3 n}} \mathbb{P}_{X_1,\dots,X_r} \left(|\langle w_r, X_r \rangle| \leq \varepsilon/c_{\rho,\delta} \quad \text{for all } r \in [j] \right) \quad (63)$$

$$\leq \max_{w_1,\dots,w_j: D_{\alpha,\gamma}(w_i) > e^{c_3 n}} \prod_{r=1}^j \mathbb{P}_{X_r} \left(|\langle w_r, X_r \rangle| \leq \varepsilon/c_{\rho,\delta} \right) \leq (C\varepsilon)^j, \quad (64)$$

where the penultimate inequality follows from the independence of the X_r 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 6.5.

Putting (62) and (64) together completes the proof of Theorem 1.3. \square

Of course, the proof of Theorem 1.2 follows immediately.

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

8. PROPERTIES OF THE SPECTRUM

In this section we describe and deduce Lemma 8.1 and Lemma 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 $\|A^{-1}\|_{\text{HS}}$. Thus the most delicate and important aspect of our analysis amounts to studying the *smallest* singular values of A .

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

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

$$\sigma_{n-k+1} \approx k/\sqrt{n}, \quad (65)$$

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

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

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

for all k .

We shall deduce Lemma 8.1 from the “local semicircular law” of Erdős, Schlein and Yau [14], which gives us good control of the bulk of the spectrum at “scales” of size $\gg n^{-1/2}$. The next result is a type of “reverse” of Lemma 8.1 and will follow from a result of Nguyen [27] along with our Theorem 1.3.

Lemma 8.2. *For $p > 1$, $B > 0$ and $\zeta \in \Gamma_B$, let $A \sim \text{Sym}_n(\zeta)$. There exist constants $C_p, c_p > 0$ depending on B, p so that for all $k \in [C_p, n]$ we have*

$$\mathbb{E} \left[\left(\frac{\mu_k k}{\sqrt{n}} \right)^p \mathbf{1}_{\{\mu_1 \leq e^{c_p n}\}} \right] \leq C_p. \quad (66)$$

We point out that the condition $k \geq C_p$ is a very important assumption in Lemma 8.2 and the above statement with $k = 1$ and $p = 1$ would imply our main theorem. Thus one might think of Lemma 8.2 as a weaker relative of our main Theorem.

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

$$\|M\|_*^2 = \sum_{k=1}^n \sigma_k(M)^2 (\log(1+k))^2. \quad (67)$$

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 (65). 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 8.3. *For $p > 1$, $B > 0$ and $\zeta \in \Gamma_B$, let $A \sim \text{Sym}_n(\zeta)$. Then there exists constants $C_p, c_p > 0$ depending on B, p such that*

$$\mathbb{E} \left[\left(\frac{\|A^{-1}\|_*}{\mu_1} \right)^p \mathbf{1}_{\{\mu_1 \leq e^{c_p n}\}} \right] \leq C_p.$$

In the remainder of this section we describe the results of Erdős, Schlein and Yau [14] and of Nguyen [27] and show how to use them to deduce Lemma 8.1 and Lemma 8.2 respectively. We then deduce Corollary 8.3.

8.1. The local semi-circular 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 *semi-circular law* which says that

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

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

We use a powerful “local” version of the semi-circle law developed by Erdős, Schlein and Yau in a series of important papers [11, 12, 14]. Their results show that the spectrum of a random symmetric matrix actually adheres surprisingly closely to the semi-circular 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 semi-circular law predicts that

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

Theorem 1.11 of [13] makes this prediction rigorous.

Theorem 8.4. *Let $B > 0$, $\zeta \in \Gamma_B$, and let $A \sim \text{Sym}_n(\zeta)$. Then for all $t \in [Cn^{-1/2}, n^{1/2}]$ we have*

$$\mathbb{P}\left(|N_A(-t, t)/(n^{1/2}t) - 2\pi^{-1}| > \pi\right) \lesssim \exp(-c_1(t^2n)^{1/4}) \quad (68)$$

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

Lemma 8.1 follows quickly from Theorem 8.4. In fact we shall only use the follow corollary.

Corollary 8.5. *Let $B > 0$, $\zeta \in \Gamma_B$, and let $A \sim \text{Sym}_n(\zeta)$. Then for all $s \geq C$ and $k \in \mathbb{N}$ satisfying $sk \leq n$ we have*

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

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

Proof. Let C be the maximum of the constant C from Lemma 8.4 and π . If $\frac{\sqrt{n}}{\mu_k k} \geq s$ then $N_A(-skn^{-1/2}, skn^{-1/2}) \leq k$. We now apply Lemma 8.4 with $t = skn^{-1/2} \geq sn^{-1/2} \geq Cn^{-1/2}$ to see that this event occurs with probability $\lesssim e^{-c\sqrt{sk}}$. \square

Proof of Lemma 8.1. Let C be the constant from Corollary 8.5. From bounds on the upper tail of $\|A\|_{op}$ (like (37)), we immediately see that for all $k \geq n/C$ we have

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

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

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

We may bound

$$\mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \leq C^p + \mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbf{1}_{E_2} + \mathbb{E}\left(\frac{\sqrt{n}}{\mu_k k}\right)^p \mathbf{1}_{E_3}. \quad (69)$$

To deal with the second term in (69), we use Corollary 8.5 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 (69), we note that since $n/k \geq C$ we may apply Corollary 8.5, with $s = n/k$, to conclude that $\mathbb{P}(E_3) \lesssim e^{-c\sqrt{n}}$. Thus, by Cauchy-Schwarz, we have

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

where we have used the upper tail estimate at (37) to see $\mathbb{E}\sigma_1^{2p} = O_p(n^p)$. \square

8.2. Eigenvalue crowding and Lemma 8.2. In [27], Nguyen proved the following result which gives good estimates on the probability that k th smallest singular value is much smaller than typical, where $k \gg 1$. In fact he proved a more general result which bounds the probability that k eigenvalues fall into any interval of length ε . Here we need only the following less general result, which comes from Theorem 1.12 of [27].

Theorem 8.6. *For $B > 0$, $\zeta \in \Gamma_B$, let $A \sim \text{Sym}_n(\zeta)$. Then for all $k \in [b_1^{-1}, b_1 n]$ and $\varepsilon > 0$ we have*

$$\mathbb{P}(\sigma_{n-k+1}(A) \leq \varepsilon n^{-1/2}) \leq (C\varepsilon/k)^{k^2/4} + O(e^{-n^{b_2}}),$$

where $C, b_1, b_2 > 0$ are absolute constants.

With Nguyen's result in hand, we quickly take care of the proof of Lemma 8.2.

Proof of Lemma 8.2. Let b_1, b_2 denote the constants b_1, b_2 from Theorem 8.6 and note that we may assume $k \leq b_1 n$ since for $k > b_1 n$ we may bound $\mu_k \leq \mu_{b_1 n}$. We now may assume that C_p , the constant in the statement of Theorem 8.2, satisfies $C_p \geq \max\{b_1^{-1}, 3p\}$. Thus we may restrict our attention to k for which $k \geq \max\{b_1^{-1}, 3p\}$. We let $c_p > 0$ be a constant to be determined later. We set $R = e^{n^{b_2}}$ and integrate to see that

$$I := \mathbb{E} \left(\frac{\mu_k k}{\sqrt{n}} \right)^p \mathbf{1}_{\{\mu_1 \leq e^{c_p n}\}}$$

is at most

$$\int_0^R k^p p s^{p-1} \mathbb{P}(\sigma_{n-k+1} \leq n^{-1/2}/s) ds + \int_R^{e^{c_p n}} p k^p s^{p-1} \mathbb{P}(\sigma_{n-k+1} \leq n^{-1/2}/s) ds =: I_0 + I_1.$$

Here we could truncate the integral I_1 at $e^{c_p n}$ since $\sigma_{n-k+1} \geq \sigma_n = 1/\mu_n \geq e^{-c_p n}$.

We bound these two ranges by applying different results. To bound I_0 we use Theorem 8.6 with $\varepsilon = 1/s$ to see

$$\mathbb{P}(\sigma_{n-k+1} \leq n^{-1/2}/s) \leq (C/ks)^{2p} + O(e^{-n^{c_2}}),$$

since $k \geq 2\sqrt{p}$. Thus integrating gives $I_0 = O_p(1)$.

To bound I_1 , we use our Theorem 1.3 with $j = k \geq 3p$ and $\varepsilon = 1/s$ to see that

$$\mathbb{P}(\sigma_{n-k+1} \leq n^{-1/2}/s) \leq \sup_i \mathbb{P}(|\lambda_{i+j} - \lambda_i| \leq 2n^{-1/2}/s) \lesssim (C/s)^{3p} + \exp(-b_3 n)$$

for some $c_3 > 0$. Assuming that $c_p > 0$ is small enough relative to b_3 allows us to bound $I_1 = O_p(1)$. Thus we see $I = I_0 + I_1 = O_p(1)$, completing the proof of Lemma 8.2. \square

8.3. Deduction of Corollary 8.3. We now conclude this section by deducing Corollary 8.3 from Lemma 8.1 and Lemma 8.2.

Proof of Corollary 8.3. Let c be the constant from Lemma 8.2 and $C = C_{2p}$ be the maximum of the two constants from Lemmas 8.1 and 8.2. Now define the event $\mathcal{E}_0 = \{\mu_1 \leq e^{c_p n}\}$ and express

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

Note that we may omit the first C terms in this sum, as we can (trivially) bound $\mu_k^2 \leq \mu_1^2$. Further, by Hölder's inequality we may assume without loss of generality that $p \geq 2$. Applying the triangle inequality for the $L^{p/2}$ norm gives

$$\left[\mathbb{E}^{\mathcal{E}_0} \left(\sum_{k>C} \frac{\mu_k^2 (\log 2(1+k))^2}{\mu_1^2} \right)^{p/2} \right]^{2/p} \leq \sum_{k>C} (\log(1+k))^2 \mathbb{E}^{\mathcal{E}_0} \left[\frac{\mu_k^p}{\mu_1^p} \right]^{2/p}$$

which is

$$\sum_{k>C} \frac{(\log(1+k))^2}{k^2} \left(\mathbb{E}^{\mathcal{E}_0} \left(\frac{\mu_k k}{\sqrt{n}} \right)^p \left(\frac{\sqrt{n}}{\mu_1} \right)^p \right)^{2/p} \leq \sum_{k>C} \frac{(\log(1+k))^2}{k^2} \left(\mathbb{E}^{\mathcal{E}_0} \left(\frac{\mu_k k}{\sqrt{n}} \right)^{2p} \right)^{1/p} \left(\mathbb{E} \left(\frac{\sqrt{n}}{\mu_1} \right)^{2p} \right)^{1/p},$$

by Cauchy-Schwarz. Thus Lemmas 8.1 and 8.2 tell us that this is $O_p(1)$, completing the proof of Corollary 8.3. \square

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 $A = A_n \sim \text{Sym}_n(\zeta)$ and let $X \sim \text{Col}_n(\zeta)$. Let $u \in \mathbb{R}^{n-1}$ be a random vector with $\|u\|_2 \leq 1$ and depending only on A . Then, for $\delta, \varepsilon > e^{-cn}$ and $s \geq 0$, we have*

$$\begin{aligned} \mathbb{E}_A \sup_r \mathbb{P}_X \left(\frac{|\langle A^{-1}X, X \rangle - r|}{\|A^{-1}\|_*} \leq \delta, \langle X, u \rangle \geq s, \frac{\mu_1}{\sqrt{n}} \leq \varepsilon^{-1} \right) \\ \lesssim \delta e^{-s} \left[\mathbb{E}_A \left(\frac{\mu_1}{\sqrt{n}} \right)^{7/9} \mathbf{1} \left\{ \frac{\mu_1}{\sqrt{n}} \leq \varepsilon^{-1} \right\} \right]^{6/7} + e^{-cn}, \end{aligned} \quad (70)$$

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

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 $0 \leq k \leq c\alpha n$ and $\varepsilon \geq \exp(-c\alpha n)$, let $v \in \mathbb{S}^{n-1}$, and let $w_1, \dots, w_k \in \mathbb{S}^{n-1}$ 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 $\tilde{X} \in \mathbb{R}^n$ 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

$$\mathbb{P}_X \left(|\langle \tilde{X}, v \rangle| \leq \varepsilon \text{ and } \sum_{j=1}^k \langle w_j, \tilde{X} \rangle^2 \leq ck \right) \leq R\varepsilon \cdot e^{-ck}. \quad (71)$$

The proof of Theorem 9.2 is provided in the supplement to this paper [5]. We now prove Lemma 9.3.

Lemma 9.3. *Let A be a $n \times n$ 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 $X, X' \sim \text{Col}_n(\zeta)$ be independent, let $J \subseteq [n]$ be a μ -random subset with $\mu \in (0, 2^{-15})$, and set $\tilde{X} := (X - X')_J$. If $k \in [1, cn]$ is such that $s \in (e^{-cn}, \mu_k/\mu_1)$ then*

$$\mathbb{P}_{\tilde{X}} \left(\|A^{-1}\tilde{X}\|_2 \leq s\mu_1 \right) \lesssim se^{-ck}, \quad (72)$$

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

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

and thus

$$\mathbb{P}_{\tilde{X}} \left(\|A^{-1}\tilde{X}\|_2 \mu_1^{-1} \leq s \right) \leq \mathbb{P}_{\tilde{X}} \left(|\langle v_1, \tilde{X} \rangle| \leq s \text{ and } \sum_{j=2}^k \frac{\mu_j^2}{\mu_1^2} \langle v_j, \tilde{X} \rangle^2 \leq s^2 \right). \quad (73)$$

We now use that $s \leq 1$ and $\mu_k/\mu_1 \leq 1$ in (73) to obtain

$$\mathbb{P}_{\tilde{X}} \left(\|A^{-1}\tilde{X}\|_2 \mu_1^{-1} \leq s \right) \leq \mathbb{P}_{\tilde{X}} \left(|\langle v_1, \tilde{X} \rangle| \leq s \text{ and } \sum_{j=2}^k \langle v_j, \tilde{X} \rangle^2 \leq 1 \right). \quad (74)$$

We now carefully observe that we are in a position to apply Theorem 1.5 to the right-hand-side of (74). The coordinates of \tilde{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, \dots, v_k are orthogonal and, importantly, we use that $A \in \mathcal{E}$ to learn that⁵ $D_{\alpha,\gamma}(v_1) > 1/s$ by property (29), provided we choose the constant $c > 0$ (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.

⁵Recall here that the constants $\alpha, \gamma > 0$ are implicit in the definition of \mathcal{E} and are chosen so that Lemma 4.1 holds.

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

$$\sup_r \mathbb{P}_X(|\langle A^{-1}X, X \rangle - r| \leq \delta \mu_1, \langle X, u \rangle \geq s) \lesssim \delta e^{-s} \sum_{k=2}^{cn} e^{-ck} \left(\frac{\mu_1}{\mu_k} \right)^{2/3} + e^{-cn}, \quad (75)$$

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

Proof. We apply Lemma 5.2 to the left-hand-side of (75) to get

$$\sup_r \mathbb{P}_X(|\langle A^{-1}X, X \rangle - r| \leq \delta \mu_1, \langle X, u \rangle \geq s) \lesssim \delta e^{-s} \int_{-1/\delta}^{1/\delta} I(\theta)^{1/2} d\theta + e^{-\Omega(n)}, \quad (76)$$

where

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

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

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

and apply Hölder's inequality

$$I(\theta) = \mathbb{E}_{J, X_J, X'_J} \left[e^{\langle (X+X')_J, u \rangle} e^{-c' \theta^2 \|v\|_2^2 / \mu_1^2} \right] \lesssim \left(\mathbb{E}_{\tilde{X}} e^{-c'' \theta^2 \|v\|_2^2 / \mu_1^2} \right)^{8/9} \left(\mathbb{E}_{J, X_J, X'_J} e^{9 \langle (X+X')_J, u \rangle} \right)^{1/9}. \quad (77)$$

Thus we apply (41) to see that the second term on the right-hand-side of (77) is $O(1)$. Thus, for each $\theta > 0$ we have

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

As a result, we have

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

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

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

while

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

Summing over all k and plugging the result into (76) completes the lemma. \square

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

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

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

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

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

$$\mathbb{E}_A^{\mathcal{E}_0} \left(\frac{\|A^{-1}\|_*}{\mu_1} \right) \left(\frac{\mu_1}{\mu_k} \right)^{2/3} \lesssim k \cdot \mathbb{E}_A^{\mathcal{E}_0} \left[\left(\frac{\mu_1}{\sqrt{n}} \right)^{7/9} \right]^{6/7}, \quad (78)$$

for each $k \in [2, cn]$. For this, apply Hölder's inequality to the left-hand-side of (78) to get

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

We now apply Corollary 8.3 to see the first term is $O(1)$ and Lemma 8.1 to see that the second term is $O(k)$. This establishes (78) and thus Lemma 9.1. \square

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 $A_n \sim \text{Sym}_n(\zeta)$. Then for all $\varepsilon > 0$*

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

To prove Lemma 10.1, we first prove the following “base step” (Lemma 10.3) which we then improve upon in three steps, ultimately arriving at Lemma 10.1. This “base step” is an easy consequence of Lemma 6.2 and Lemma 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 [19,45] (see also [41, Theorem 6.2.1]).

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

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

where $c > 0$ is absolute constant.

We now prove the base step of our iteration.

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

$$\mathbb{P}(\sigma_{\min}(A_{n+1}) \leq \varepsilon/\sqrt{n}) \lesssim \varepsilon^{1/4} + e^{-\Omega(n)}.$$

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

$$\mathbb{P}_{A,X} \left(\frac{|\langle A^{-1}X, X \rangle - r|}{\|A^{-1}X\|_2} \leq C\varepsilon, \sigma_n(A) \geq \frac{\varepsilon}{\sqrt{n}} \right) \lesssim \varepsilon^{1/4} + e^{-\Omega(n)}. \quad (79)$$

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

$$\mathbb{P}_X(\|A^{-1}X\|_2 \geq C' \sqrt{\log 1/\varepsilon} \|A^{-1}\|_{\text{HS}}) \leq \varepsilon \quad (80)$$

and so the left-hand-side of (79) is bounded above by

$$\varepsilon + \mathbb{P}_{A,X} \left(\frac{|\langle A^{-1}X, X \rangle - r|}{\|A^{-1}\|_{\text{HS}}} \leq \delta, \sigma_n(A) \geq \varepsilon/\sqrt{n} \right),$$

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

$$\mathbb{P}_{A,X} \left(\frac{|\langle A^{-1}X, X \rangle - r|}{\|A^{-1}\|_{\text{HS}}} \leq \delta, \sigma_n(A) \geq \frac{\varepsilon}{\sqrt{n}} \right) \lesssim \delta \varepsilon^{-2/3} + e^{-\Omega(n)} \lesssim \varepsilon^{1/4} + e^{-\Omega(n)}, \quad (81)$$

where we have used that $\|A^{-1}\|_* \geq \|A^{-1}\|_{\text{HS}}$. 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 (85). Since $\sup_r \mathbb{P}_{A,X}(\mathcal{A}) \leq \mathbb{E}_A \sup_r \mathbb{P}_X(\mathcal{A})$, the bound (81), and thus Lemma 10.3, follows. \square

The next lemma is our “bootstrapping step”: given bounds of the form

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

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 $A_{n+1} \sim \text{Sym}_{n+1}(\zeta)$ and let $\kappa \in (0, 1) \setminus \{7/10\}$. If for all $\varepsilon > 0$, and all n we have*

$$\mathbb{P}(\sigma_{\min}(A_n) \leq \varepsilon/\sqrt{n}) \lesssim \varepsilon^\kappa + e^{-\Omega(n)}, \quad (82)$$

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

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

Proof. Let $c > 0$ denote the implicit constant in the exponent on the right-hand-side of (82). Note that if $0 < \varepsilon < e^{-cn}$, by the assumption of the lemma, then we have

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

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 Lemma 6.2 and Lemma 9.1 in sequence. For this we write $A = A_n$ and bound (70) as in the conclusion of Lemma 9.1

$$\mathbb{E}_A \left(\frac{\mu_1}{\sqrt{n}} \right)^{7/9} \mathbf{1} \left\{ \frac{\mu_1}{\sqrt{n}} \leq \varepsilon^{-1} \right\} \leq \int_0^{\varepsilon^{-7/9}} \mathbb{P}(\sigma_{\min}(A) \leq x^{-9/7} n^{-1/2}) dx, \quad (83)$$

where we used that $\sigma_{\min}(A) = 1/\mu_1(A)$. Now use assumption (82) to see the right-hand-side of (83) is

$$\lesssim 1 + \int_1^{\varepsilon^{-7/9}} (x^{-9\kappa/7} + e^{-cn}) dx \lesssim \max\{1, \varepsilon^{\kappa-7/9}\}. \quad (84)$$

Now we apply Lemma 9.1 with $\delta = C\varepsilon\sqrt{\log 1/\varepsilon}$, $s = 0$ and $u = 0$ to see that

$$\sup_r \mathbb{P}_{A,X} \left(\frac{|\langle A^{-1}X, X \rangle - r|}{\|A^{-1}\|_{\text{HS}}} \leq \delta, \frac{\mu_1}{\sqrt{n}} \leq \varepsilon^{-1} \right) \lesssim \max\{\varepsilon, \varepsilon^{6\kappa/7+1/3}\} \sqrt{\log 1/\varepsilon} + e^{-\Omega(n)}, \quad (85)$$

where we have used that $\|A^{-1}\|_{\text{HS}} \leq \|A^{-1}\|_*$.

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

$$\mathbb{P}_X(\|A^{-1}X\|_2 \geq C'\|A^{-1}\|_{\text{HS}}\sqrt{\log 1/\varepsilon}) \leq \varepsilon.$$

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

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

Lemma 6.1 now completes the proof of Lemma 10.4. \square

Lemma 10.1 now follows by iterating Lemma 10.4 three times.

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

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

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. \square

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 Lemma 31 to see that that it is enough to prove

$$\mathbb{P}^{\mathcal{E}} \left(\frac{|\langle A^{-1}X, X \rangle - r|}{\|A^{-1}X\|_2} \leq C\varepsilon, \text{ and } \sigma_n(A) \geq \varepsilon n^{-1/2} \right) \lesssim \varepsilon + e^{-\Omega(n)}, \quad (86)$$

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 $\mathcal{E}_0 := \mathcal{E} \cap \{\sigma_{\min}(A) \geq \varepsilon/\sqrt{n}\}$ 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 (86) as

$$\mathbb{P}^{\mathcal{E}_0}(Q(A, X) \leq C\varepsilon) \leq \mathbb{P}^{\mathcal{E}_0}(Q_*(A, X) \leq 2C\varepsilon) + \mathbb{P}^{\mathcal{E}_0}\left(Q(A, X) \leq C\varepsilon, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geq 2\right). \quad (87)$$

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

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

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

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

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

By Lemma 10.1 and the calculation at (84), the expectation on the right is bounded by a constant. \square

We now focus on the latter term on the right-hand-side of (87). By considering the dyadic partition $2^j \leq \|A^{-1}X\|_2 / \|A^{-1}\|_* \leq 2^{j+1}$ we have

$$\mathbb{P}^{\mathcal{E}_0}\left(Q(A, X) \leq C\varepsilon, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geq 2\right) \lesssim \sum_{j=1}^{\log n} \mathbb{P}^{\mathcal{E}_0}\left(Q_*(A, X) \leq 2^{j+1}C\varepsilon, \frac{\|A^{-1}X\|_2}{\|A^{-1}\|_*} \geq 2^j\right) + e^{-\Omega(n)}. \quad (88)$$

Here we have dealt with $j \geq \log n$ by using Hanson-Wright (Theorem 10.2) and the fact that $\|A^{-1}\|_* \geq \|A^{-1}\|_{\text{HS}}$ to see

$$\mathbb{P}_X(\|A^{-1}X\|_2 \geq \sqrt{n}\|A^{-1}\|_*) \lesssim e^{-\Omega(n)}.$$

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

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

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

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

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

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

Thus

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

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. \square

Proof of Theorem 1.1. Recall that it suffices to establish (86). Combining (87) with Lemma 11.2 and Lemma 11.1 tells us that

$$\mathbb{P}^{\mathcal{E}_0}(Q(A, X) \leq C\varepsilon) \lesssim \varepsilon + 2 \sum_{j=1}^{\log n} \sum_{k=1}^n \mathbb{P}^{\mathcal{E}_0}(Q_*(A, X) \leq 2^{j+1}C\varepsilon, \langle X, v_k \rangle \geq 2^j \log(1+k)) + e^{-\Omega(n)}. \quad (89)$$

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,

$$\mathbb{P}^{\mathcal{E}_0}(Q_*(A, X) \leq 2Ct\varepsilon, \langle X, v_k \rangle \geq t \log(1+k)) \lesssim \varepsilon t (k+1)^{-t} \cdot I^{6/7} + e^{-\Omega(n)}. \quad (90)$$

where

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

Using (90) in (89) yields

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

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

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

and apply Lemma 10.1 to see

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

Thus, Lemma 6.1 completes the proof of Theorem 1.1. \square

ACKNOWLEDGMENTS

The authors thank Rob Morris for comments on a previous draft. Marcelo Campos is partially supported by CNPq. Marcus Michelen is supported in part by NSF grant DMS-2137623.

APPENDIX A. 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

$$\mathbb{P}(|\langle X, v \rangle| \leq \varepsilon \text{ and } \langle X, u \rangle > t) \leq \mathbb{E} [\mathbf{1}\{|\langle X, v \rangle| \leq \varepsilon\} e^{\lambda \langle X, u \rangle - \lambda t}], \quad (91)$$

where $\lambda \geq 0$ will be optimized later. Now apply Esseen's inequality in a similar way to Lemma 5.1 to bound

$$\mathbb{E} [\mathbf{1}\{|\langle X, v \rangle| \leq \varepsilon\} e^{\lambda \langle X, u \rangle - \lambda t}] \lesssim \varepsilon e^{-\lambda t} \int_{-1/\varepsilon}^{1/\varepsilon} |\mathbb{E} e^{2\pi i \theta \langle X, v \rangle + \lambda \langle X, u \rangle}| d\theta. \quad (92)$$

Applying Lemma 5.5 bounds

$$|\mathbb{E} e^{2\pi i \theta \langle X, v \rangle + \lambda \langle X, u \rangle}| \lesssim \exp \left(-c \min_{r \in [1, c^{-1}]} \|\theta r v\|_{\mathbb{T}}^2 + c^{-1} \lambda^2 \right) + e^{-c\alpha n}. \quad (93)$$

Combining the lines (91), (92) and (93) and choosing C large enough gives the bound

$$\begin{aligned} \mathbb{P}(|\langle X, v \rangle| \leq \varepsilon \text{ and } \langle X, u \rangle > t) &\lesssim \varepsilon e^{-\lambda t + c^{-1} \lambda^2} \int_{-1/\varepsilon}^{1/\varepsilon} (e^{-c\gamma^2 \theta^2} + e^{-c\alpha n}) d\theta \\ &\lesssim \varepsilon e^{-\lambda t + c^{-1} \lambda^2} \gamma^{-1} + e^{-c\alpha n - \lambda t + c^{-1} \lambda^2}. \end{aligned}$$

Choosing $\lambda = ct/2$ completes the proof. \square

APPENDIX B. PROOF OF LEMMA 3.2

We deduce the second part of Lemma 3.2 from the following special case of a Proposition of Vershynin [40, Prop. 4.2].

Proposition B.1. *For $B > 0$, let $\zeta \in \Gamma_B$, let $A_n \sim \text{Sym}_n(\zeta)$ and let $K \geq 1$. 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*

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

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 $|\mathcal{N}| \leq 8/c$. 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

$$\begin{aligned} \mathbb{P}(\exists x \in \text{Comp}(\delta, \rho), \exists t \in [-4\sqrt{n}, 4\sqrt{n}] : A_n x = tu) \\ \leq \sum_{t_0 \in \mathcal{N}} \mathbb{P}(\exists x \in \text{Comp}(\delta, \rho) : \|A_n x - t_0 u\|_2 \leq c\sqrt{n}), \end{aligned}$$

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 $\|A_n x - t_0 u\|_2 \leq c\sqrt{n}$. Now to bound each term in the sum take $\lambda = 0$, $K = 4$, $w = t_0 u$ in Proposition B.1 and notice we may assume $\|A_n\|_{op} \leq 4\sqrt{n}$ again. For the second conclusion, it is sufficient to show

$$\begin{aligned} \mathbb{P}(\exists x \in \text{Comp}(\delta, \rho), \exists t \in [-4\sqrt{n}, 4\sqrt{n}] : \|(A_n - tI)x\|_2 = 0 \text{ and } \|A_n - tI\|_{op} \leq 8\sqrt{n}) \\ \lesssim e^{-\Omega(n)}, \end{aligned} \quad (94)$$

since we have $\mathbb{P}(\|A_n\|_{op} \geq 4\sqrt{n}) \lesssim e^{-\Omega(n)}$, by (37), so we may assume that all eigenvalues of A_n lie in $[-4\sqrt{n}, 4\sqrt{n}]$ and $\|A_n - tI\|_{op} \leq |t| + \|A_n\|_{op} \leq 8\sqrt{n}$, for all $t \in [-4\sqrt{n}, 4\sqrt{n}]$.

For this, we apply Proposition B.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 $|\mathcal{N}| \leq 8/c$. 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 $|t - t_0| \leq c\sqrt{n}$ and

$$\|(A_n - t_0 I)x\|_2 \leq |t - t_0| \|x\|_2 \leq c\sqrt{n}.$$

Thus the left hand side of (94) is at most

$$\sum_{t_0 \in \mathcal{N}} \mathbb{P}(\exists x \in \text{Comp}(\delta, \rho) : \|(A_n - t_0 I)x\|_2 \leq c\sqrt{n} \text{ and } \|A_n - t_0 I\|_{op} \leq 8\sqrt{n}) \lesssim e^{-cn},$$

where the last line follows from Proposition B.1. \square

APPENDIX C. PROOF OF EQUATION (37)

We deduce (37) from the following result of Feldheim and Sodin [16, Cor. V.2.1] for symmetric random variables. Recall that a random variable ζ is *symmetric* if $-\zeta$ has the same distribution as ζ .

Theorem C.1. *For $B > 0$, let $\zeta \in \Gamma_B$ be a symmetric random variable and let $A \sim \text{Sym}_n(\zeta)$. For all $s \geq 0$, we have*

$$\mathbb{P}(\|A_n\|_{op} \geq 2\sqrt{n}(1+s)) \lesssim \exp(-cns^{3/2}).$$

The bound (37) follows from a basic symmetrization argument, where we deduce the bound for $\|A\|_{op}$ based on an application of Theorem C.1 to the random matrix $A - A'$, where A, A' are independent copies of A .

Proof of (37). Let $\zeta \in \Gamma_B$ and $A, A' \sim \text{Sym}_n(\zeta)$ be independent. We bound

$$\mathbb{P}(\|A\|_{op} \geq (3+s)\sqrt{n}) \leq \mathbb{P}(\|A - A'\|_{op} \geq (3+s/2)\sqrt{n}) + \mathbb{P}(\mathcal{A}), \quad (95)$$

where

$$\mathcal{A} := \{\|A\|_{op} \geq (3+s)\sqrt{n}, \|A - A'\|_{op} \leq (3+s/2)\sqrt{n}\}.$$

We take care of the first term on the right-hand-side of (95) by applying Theorem C.1 to the matrix $(A - A')/\sqrt{2}$ and noting that $2\sqrt{2} \leq 3$. Thus

$$\mathbb{P}(\|A_n - A'_n\|_{op} \geq (3+s/2)\sqrt{n}) \lesssim \exp(-cns^{3/2}).$$

Turning our attention to the event \mathcal{A} , let A be a matrix satisfying $\|A\|_{op} \geq (3+s)\sqrt{n}$. Then there is a vector $x := x(A) \in \mathbb{S}^{n-1}$ with $\|A\|_{op} = \langle Ax, x \rangle$. Thus, if we also have $\|A - A'\|_{op} \leq (3+s/2)\sqrt{n}$, then we must have $\langle A'x, x \rangle \geq s\sqrt{n}/2$. In other words

$$\mathbb{P}(\mathcal{A}) \leq \sup_{x \in \mathbb{S}^{n-1}} \mathbb{P}(\langle A'x, x \rangle \geq s\sqrt{n}/2) \leq 2e^{-cs^2n},$$

where the last line holds by a large deviation inequality for sums of subgaussian random variables: see Theorem 2.6.2 in Section 2.6 of [41]. This completes the proof of (37). \square

REFERENCES

- [1] R. Berkowitz. A local limit theorem for cliques in $G(n, p)$. *arXiv preprint arXiv:1811.03527*, 2018.
- [2] P. Bourgade, L. Erdős, H.-T. Yau, and J. Yin. Fixed energy universality for generalized Wigner matrices. *Comm. Pure Appl. Math.*, 69(10):1815–1881, 2016.
- [3] M. Campos, M. Jensen, M. Michelen, and J. Sahasrabudhe. Singularity of random symmetric matrices revisited. *Proc. of Amer. Math. Soc.*, 2021.
- [4] M. Campos, M. Jensen, M. Michelen, and J. Sahasrabudhe. The singularity probability of a random symmetric matrix is exponentially small. <https://arxiv.org/abs/2105.11384>, 2021.
- [5] M. Campos, M. Jensen, M. Michelen, and J. Sahasrabudhe. Supplementary paper to the least singular value of a random symmetric matrix. <https://arxiv.org/abs/2203.06141>, 2022.
- [6] M. Campos, L. Mattos, R. Morris, and N. Morrison. On the singularity of random symmetric matrices. *Duke Mathematical Journal*, 170(5):881–907, 2021.
- [7] K. P. Costello. Bilinear and quadratic variants on the Littlewood-Offord problem. *Israel Journal of Mathematics*, 194(1):359–394, 2013.
- [8] K. P. Costello, T. Tao, and V. Vu. Random symmetric matrices are almost surely nonsingular. *Duke Math. J.*, 135(2):395–413, 2006.
- [9] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM journal on matrix analysis and applications*, 9(4):543–560, 1988.
- [10] 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):667–674, 2010.
- [11] L. Erdős, B. Schlein, and H.-T. Yau. Local semicircle law and complete delocalization for Wigner random matrices. *Comm. Math. Phys.*, 287(2):641–655, 2009.
- [12] 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):815–852, 2009.
- [13] L. Erdős. Universality of Wigner random matrices: a survey of recent results. *Russian Mathematical Surveys*, 66(3):507, 2011.
- [14] L. Erdős, B. Schlein, and H.-T. Yau. Wegner estimate and level repulsion for Wigner random matrices. *International Mathematics Research Notices*, 2010(3):436–479, 2010.
- [15] C. Esseen. On the Kolmogorov-Rogozin inequality for the concentration function. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 5(3):210–216, 1966.
- [16] O. N. Feldheim and S. Sodin. A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Funct. Anal.*, 20(1):88–123, 2010.
- [17] A. Ferber and V. Jain. Singularity of random symmetric matrices—a combinatorial approach to improved bounds. *Forum Math. Sigma*, 7:Paper No. e22, 29, 2019.
- [18] A. Ferber, V. Jain, K. Luh, and W. Samotij. On the counting problem in inverse Littlewood-Offord theory. *Journal of the London Mathematical Society*.
- [19] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.
- [20] V. Jain, A. Sah, and M. Sawhney. On the smallest singular value of symmetric random matrices. *arXiv preprint arXiv:2011.02344*, 2020.
- [21] M. Kwan and L. Saueremann. An algebraic inverse theorem for the quadratic Littlewood-Offord problem, and an application to Ramsey graphs. *Discrete Analysis*, Paper No. 12, 34, 2020.
- [22] G. V. Livshyts. The smallest singular value of heavy-tailed not necessarily iid random matrices via random rounding. *Journal d’Analyse Mathématique*, pages 1–50, 2021.
- [23] G. V. Livshyts, K. Tikhomirov, and R. Vershynin. The smallest singular value of inhomogeneous square random matrices. *The Annals of Probability*, 49(3):1286 – 1309, 2021.
- [24] R. Meka, O. Nguyen, and V. Vu. Anti-concentration for polynomials of independent random variables. *Theory of Computing*, 12(11):1–17, 2016.

- [25] H. Nguyen, T. Tao, and V. Vu. Random matrices: tail bounds for gaps between eigenvalues. *Probability Theory and Related Fields*, 167(3):777–816, 2017.
- [26] H. H. Nguyen. Inverse Littlewood-Offord problems and the singularity of random symmetric matrices. *Duke Math. J.*, 161(4):545–586, 2012.
- [27] H. H. Nguyen. Random matrices: Overcrowding estimates for the spectrum. *Journal of functional analysis*, 275(8):2197–2224, 2018.
- [28] E. Rebrova and K. Tikhomirov. Coverings of random ellipsoids, and invertibility of matrices with iid heavy-tailed entries. *Israel Journal of Mathematics*, 227(2):507–544, 2018.
- [29] M. Rudelson. Invertibility of random matrices: norm of the inverse. *Ann. of Math. (2)*, 168(2):575–600, 2008.
- [30] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.
- [31] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62(12):1707–1739, 2009.
- [32] M. Rudelson and R. Vershynin. No-gaps delocalization for general random matrices. *Geometric and Functional Analysis*, 26(6):1716–1776, 2016.
- [33] S. Smale. On the efficiency of algorithms of analysis. *Bulletin (New Series) of The American Mathematical Society*, 13(2):87–121, 1985.
- [34] D. Spielman and S.-H. Teng. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. In *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, pages 296–305. ACM, New York, 2001.
- [35] S. a. J. Szarek. Spaces with large distance to l_∞^n and random matrices. *Amer. J. Math.*, 112(6):899–942, 1990.
- [36] T. Tao and V. Vu. Random matrices: A general approach for the least singular value problem. *preprint*, 2008.
- [37] T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics. *Acta Math.*, 206(1):127–204, 2011.
- [38] T. Tao and V. Vu. Random matrices have simple spectrum. *Combinatorica*, 37(3):539–553, 2017.
- [39] T. Tao and V. H. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math. (2)*, 169(2):595–632, 2009.
- [40] R. Vershynin. Invertibility of symmetric random matrices. *Random Structures Algorithms*, 44(2):135–182, 2014.
- [41] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
- [42] J. Von Neumann. *Design of computers, theory of automata and numerical analysis*, volume 5. Pergamon Press, 1963.
- [43] V. H. Vu and T. Tao. The condition number of a randomly perturbed matrix. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC ’07, page 248–255, New York, NY, USA, 2007. Association for Computing Machinery.
- [44] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958.
- [45] F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *The Annals of Probability*, 1(6):1068–1070, 1973.

SUPPLEMENTARY PAPER TO THE LEAST SINGULAR VALUE OF A RANDOM SYMMETRIC MATRIX

MARCELO CAMPOS, MATTHEW JENSSEN, MARCUS MICHELEN, JULIAN SAHASRABUDHE

ABSTRACT. This is a supplementary article to our paper “The least singular value of a random symmetric matrix”. Here, we lay out the details of the “master quasirandomness theorem”, which we left unproved in that paper. The proof is a technical adaptation of the authors’ previous work on the singularity of random symmetric matrices.

1. INTRODUCTION

We start with a few definitions from [2] 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*

$$\|\zeta\|_{\psi_2} := \sup_{p \geq 1} \frac{1}{\sqrt{p}} (\mathbb{E}|\zeta|^p)^{1/p}$$

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

For $\zeta \in \Gamma$, let $\text{Sym}_n(\zeta)$ denote the probability space of $n \times n$ symmetric matrices with $(A_{i,j})_{i \leq j}$ i.i.d. distributed according to ζ . Let $\text{Col}_n(\zeta)$ 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

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

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

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

Remark 1.1. We note that in [2] we deal 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 2.6 below eliminates those v for which $\|v_I\|_2$ is less than a constant. Thus, we work with the slightly simpler definition (2) throughout.

We define the set of “structured direction on the sphere”

$$\Sigma = \Sigma_{\alpha, \gamma, \mu} := \{v \in \mathbb{S}^{n-1} : \hat{D}_{\alpha, \gamma, \mu}(v) \leq e^{c_{\Sigma} n}\}. \quad (3)$$

The first named author is partially supported by CNPq. The third named author is partially supported by NSF grant DMS-2137623.

Now for $\zeta \in \Gamma$, $A \sim \text{Sym}_n(\zeta)$ and a given vector $w \in \mathbb{R}^n$, we define the quantity (defined in Section 4 in [2])

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

We then define

$$q_n := \max_{w \in \mathbb{S}^{n-1}} q_n(w). \quad (5)$$

The first goal of this supplement is to prove the following theorem.

Theorem 1.2 (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*

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

The second main goal of this supplement is to prove the following negative correlation result—or inverse Littlewood-Offord Theorem—that will be of use in [2]. In fact, we will prove it along the way to proving Theorem 1.2 as it is an important step in the proof that theorem.

Theorem 1.3. *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 $0 \leq k \leq c_0 \alpha d$ and $t \geq \exp(-c_0 \alpha d)$; let $v \in \mathbb{S}^{n-1}$ and let $w_1, \dots, w_k \in \mathbb{S}^{d-1}$ 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

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

The proofs of Theorem 1.2 and Theorem 1.3 follow the same path as [1] where the authors proved analogous statement for the case where the entries of A are uniform in $\{-1, 1\}$. We refer the reader to the following Section 1.1 for a discussion of how this paper is structured relative to [1].

1.1. A Reader's Guide. The proof of Theorem 1.2 is similar to the proof of the main theorem in [1], with only a few tweaks and additions required to make the adaptation go through. In several places, we need only to update the constants and will be satisfied in pointing the interested reader to [1] for more detail. Elsewhere, more significant adaptations are required, and we will not hesitate to outline these changes in full detail. As such, parts of this paper will bore restless experts. 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.

Here we describe the correspondence between sections in this paper and sections in [1] and point out the key changes that come up.

In Section 2 we set up many of the basic notions that we will need for the proof of Theorem 1.2. 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 3, we turn to prove Theorem 1.2, while assuming several key results that we either import from [1] or prove in later sections. This section is the analogue of Section 9 in [1] and the main difference between these sections arises from the different definitions of q_n in these two papers (see (5)). Here q_n is defined in terms of the least common denominator $D_{\alpha,\gamma}$, rather than the threshold \mathcal{T}_L (see (12)). 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 [1], but another net argument is required to exclude vectors with \mathcal{T}_L small but $D_{\alpha,\gamma}$ large.

In Section 4, we define many of the key Fourier-related notions that we will need to prove the remaining results, including Theorem 1.3. The main differences between the two papers in these sections comes from the different definition of the sublevel sets S_W (see (23)). This new definition requires us to reprove a few of our basic Lemmas from [1], however the proofs go through easily.

In Section 4.2, we state our main inverse Littlewood-Offord Theorem for conditioned random walks and deduce Theorem 1.3 from it. Lemma 4.3 in this section is also one of the main ingredients that goes into Theorem 3.2. This section corresponds to Section 3 of [1].

Section 5 deals with Fourier replacement and is the analogue of Appendix B in [1]. 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 6 we prove Lemma 4.3. This corresponds to Sections 4 and 5 of [1], from which several key geometric facts are imported wholesale, making our task significantly lighter here. The difference in the definitions from Section 4 are salient here, but the majority of the proof is the same as in [1, Section 5], up to the constants involved.

The next three sections, Sections 7, 8 and 9, correspond to Sections 6, 7, and 8 respectively of [1]. Here the adaptation to this paper requires little more than updating constants. These three sections amount to converting Lemma 4.3 into the main net bound Theorem 3.2.

Finally, in Section 10 we deduce the Hanson-Wright inequality, Lemma 6.7, from Talagrand's inequality; this corresponds to Appendix E of [1] where the difference again is only up to constants.

1.2. Asymptotic notation. For notational ease, we allow constants implicit in big-O notation to *depend on the subgaussian constant B* , if this constant is relevant in the context.

2. PREPARATIONS

2.1. 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 4, 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 non-degenerate and not-too-large LCD in Section 6. 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 .

Lemma 2.1. $p \geq \frac{1}{2^7 B^4}$.

Proof. By the Paley-Zygmund inequality

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

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

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

Combining the bounds completes the proof. \square

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

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

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 2.1 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

$$\phi_{\xi_\nu}(t) = 1 - \nu p + \nu p \mathbb{E}_{\bar{\zeta}} \cos(2\pi t\bar{\zeta}). \quad (6)$$

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

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

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

$$\exp(-32\nu p \cdot \mathbb{E}_{\bar{\zeta}} \|t\bar{\zeta}\|_{\mathbb{T}}^2) \leq \phi_{\xi_\nu}(t) \leq \exp(-\nu p \cdot \mathbb{E}_{\bar{\zeta}} \|t\bar{\zeta}\|_{\mathbb{T}}^2). \quad (7)$$

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

$$\phi_{\tilde{\zeta} Z_\nu}(t) \leq 1 - \nu + \nu(1 - p) + \nu p \mathbb{E}_{\tilde{\zeta}}[\cos(2\pi t\tilde{\zeta}) \mid |\tilde{\zeta}| \in I_B] = \phi_{\xi_\nu}(t). \quad (8)$$

2.2. 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. [9, Prop. 2.6.1])

Fact 2.2. *For $B > 0$, let $Y = (Y_1, \dots, Y_d)$ be a random vector with $Y_1, \dots, Y_d \in \Gamma_B$. 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 [2, Appendix C])

Fact 2.3. *For $B > 0$, let $\zeta \in \Gamma$ and $A \sim \text{Sym}_n(\zeta)$. Then*

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

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

$$\mathbb{P}^{\mathcal{K}}(\mathcal{E}) = \mathbb{P}(\mathcal{K} \cap \mathcal{E}), \quad (9)$$

for every event \mathcal{E} .

2.3. Compressibility and eliminating non-flat vectors. As in [1], 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 [5].

Following [5], 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 $\text{Comp}(\delta, \rho)$ denote the set of all such compressible vectors in \mathbb{S}^{n-1} . Proposition 4.2 from Vershynin’s paper [8] takes care of all compressible vectors.

Lemma 2.4. *For $B > 0$, let $\zeta \in \Gamma_B$, let $A_n \sim \text{Sym}_n(\zeta)$ and let $K \geq 1$. 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*

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

For the remainder of the paper, we let δ, ρ be the constants given in Lemma 2.4. Define

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

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 [5].

Lemma 2.5. *Let $v \in \text{Incomp}(\delta, \rho)$. Then*

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

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 2.4. For $D \subseteq [n]$ define the set of directions in \mathbb{S}^{n-1} that are “flat on D ”:

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

and let

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

Applying Lemmas 2.4 and 2.5 in tandem will allow us to eliminate vectors outside of \mathcal{I} .

Lemma 2.6. *Let $\delta, \rho, c > 0$ be the constants defined in Lemma 2.4 and let $d < \rho^2 \delta n / 2$. Then*

$$\max_{w \in \mathbb{S}^{n-1}} \mathbb{P}_A \left(\exists v \in \mathbb{S}^{n-1} \setminus \mathcal{I} \text{ and } \exists s, t \in [-4\sqrt{n}, 4\sqrt{n}] : \|Av - sv - tw\|_2 \leq c\sqrt{n}/2 \right) \leq 2e^{-\Omega(n)}. \quad (10)$$

Proof. Lemma 2.5, along with the definitions of κ_0, κ_1 and \mathcal{I} , implies that

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

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 $\|Av - sv - tw\|_2 \leq c\sqrt{n}/2$ implies that there exists $(s', t') \in \mathcal{N}$ for which

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

Thus the left-hand-side of (10) is

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

where the final inequality follows by first intersecting each term in the sum with the event $\mathcal{E} := \{\|A - s'I\|_{op} \leq 16n^{1/2}\}$ (noting that $\mathbb{P}(\mathcal{E}^c) \leq 2e^{-\Omega(n)}$, by Fact 2.3) and applying Lemma 2.4 to each term in the sum with $\lambda = -s'$ and $K = 16$. \square

2.4. 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 [1] 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

$$M = \begin{bmatrix} \mathbf{0}_{[d] \times [d]} & H_1^T \\ H_1 & \mathbf{0}_{[d+1, n] \times [d+1, n]} \end{bmatrix}, \quad (11)$$

where H_1 is a $(n - d) \times d$ random matrix with 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 $\|Av\|_2 \leq \varepsilon n^{1/2}$, 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

$$\mathcal{T}_L(v) := \sup\{t \in [0, 1] : \mathbb{P}(\|Mv\|_2 \leq t\sqrt{n}) \geq (4Lt)^n\}. \quad (12)$$

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”.

3. PROOF OF THEOREM 1.2

Here we recall some key notions from [1], state analogous lemmas, and prove Theorem 1.2 assuming these lemmas.

3.1. Efficient nets. Our goal is to obtain an exponential bound on the quantity

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

defined at (5), where

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

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

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

This will be the trickier region and will depend on the net construction from [1]. We also need to take care of the region

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

which we take care of using the nets constructed by Rudelson and Vershynin in [5]. We recall that \mathcal{T}_L is defined at (12).

We also note that since the event $\mathcal{K} := \{\|A\|_{\text{op}} \geq 4n^{1/2}\}$ fails with probability $2e^{-cn}$ (Fact 2.3) and we only need to deal with incompressible vectors $v \in \mathcal{I}$ (by Lemma 2.6), it is enough to show

$$\sup_{w \in \mathbb{S}^{n-1}} \mathbb{P}_A^{\mathcal{K}} \left(\exists v \in \mathcal{I} \cap S, s, t \in [-4\sqrt{n}, 4\sqrt{n}] : Av = sv + tw \right) \leq e^{-\Omega(n)}, \quad (13)$$

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 [1]. For a random variable $Y \in \mathbb{R}^d$ and $\varepsilon > 0$, we define the Lévy concentration of Y by

$$\mathcal{L}(Y, \varepsilon) = \sup_{w \in \mathbb{R}^d} \mathbb{P}(\|Y - w\|_2 \leq \varepsilon). \quad (14)$$

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

$$\mathcal{L}_{A, \text{op}}(v, \varepsilon\sqrt{n}) := \sup_{w \in \mathbb{R}^n} \mathbb{P}^{\mathcal{K}}(\|Av - w\|_2 \leq \varepsilon\sqrt{n}). \quad (15)$$

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

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

Define the (trivial) net

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

3.1.1. *Definition of net for $v \in S$.* To deal with vectors in S , for $\varepsilon \geq \exp(-2c_\Sigma n)$ define

$$\Sigma_\varepsilon := \{v \in \mathcal{I}([d]) : \mathcal{T}_L(v) \in [\varepsilon, 2\varepsilon]\}. \quad (16)$$

If $v \in \Sigma_\varepsilon$, for some $\varepsilon \geq \exp(-2c_\Sigma n)$ then the proof will be basically the same as in [1]. As such, we approximate Σ_ε by \mathcal{N}_ε , where we define

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

and show that \mathcal{N}_ε is appropriately small.

First the following lemma allows us to approximate Σ_ε by \mathcal{N}_ε .

Lemma 3.1. *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 $\|u - v\|_\infty \leq 4\varepsilon n^{-1/2}$.*

This lemma is analogous to Lemma 8.2 in [1] and we postpone its proof to Section 9. The main difficulty faced in [1] is to prove an appropriate bound on $|\mathcal{N}_\varepsilon|$. In our case we have an analogous bound.

Theorem 3.2. *For $L \geq 2$ and $0 < c_0 \leq 2^{-50}B^{-4}$, let $n \geq L^{64/c_0^2}$, $d \in [c_0^2 n/4, c_0^2 n]$ and $\varepsilon > 0$ be so that $\log \varepsilon^{-1} \leq nL^{-32/c_0^2}$. Then*

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

where $C > 0$ is an absolute constant.

The proof of Theorem 3.2 will follow mostly from Lemma 4.3, with the rest of the deduction following exactly the same path as in [1], which we present in Sections 7 and 8.

3.1.2. *Definition of net for $v \in S'$.* We now need to tackle the vectors in S' ; that is, those with

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

Here we construct the nets using only the second condition using a construction of Rudelson and Vershynin [5]. Then the condition $\mathcal{T}_L(v) \leq \exp(-2c_\Sigma n)$ will come in when we union bound over nets. With this in mind, let

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

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

$$G_\varepsilon := \bigcup_{|I| \geq (1-2\mu)n} \left\{ \frac{p}{\|p\|_2} : p \in (\mathbb{Z}^I \oplus \sqrt{\alpha}\mathbb{Z}^{I^c}) \cap B_n(0, \varepsilon^{-1}) \setminus \{0\} \right\}. \quad (17)$$

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 3.3. *Let $\varepsilon > 0$ satisfy $\varepsilon \leq \gamma(\alpha n)^{-1/2}/4$. If $v \in \Sigma'_\varepsilon$, then there exists $u \in G_\varepsilon$ such that $\|u - v\|_2 \leq 16\varepsilon\sqrt{\alpha n}$.*

Proof. Set $D = \min_{|I| \geq (1-2\mu)n} D_{\alpha, \gamma}(v_I)$, 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

$$\|Dv_I - p_I\|_2 < \min\{\gamma D \|v_I\|_2, \sqrt{\alpha n}\} \leq \sqrt{\alpha n},$$

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

$$\|Dv_{I^c} - p_{I^c}\|_2 \leq \sqrt{\alpha n}.$$

Thus, if we set $p = p_I \oplus p_{I^c}$, by the triangle inequality we have

$$\left\|v - \frac{p}{\|p\|_2}\right\|_2 \leq \frac{1}{D}(\|Dv - p\|_2 + |D - \|p\|_2|) \leq 4D^{-1}\sqrt{\alpha n} \leq 16\varepsilon\sqrt{\alpha n},$$

as desired. \square

We also note that this net is sufficiently small for our purposes (see [5]).

Fact 3.4. For $\alpha, \mu \in (0, 1)$, $K \geq 1$ and $\varepsilon \leq Kn^{-1/2}$ we have

$$|G_\varepsilon| \leq \left(\frac{32K}{\alpha^{2\mu}\varepsilon\sqrt{n}}\right)^n,$$

where G_ε is as defined at (17).

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 3.5. For $\alpha, \mu \in (0, 1)$, $K \geq 1$ and $\varepsilon \leq Kn^{-1/2}$ there is a $32\varepsilon\sqrt{\alpha n}$ -net G'_ε for Σ'_ε with $G'_\varepsilon \subseteq \Sigma'_\varepsilon$ and

$$|G'_\varepsilon| \leq \left(\frac{32K}{\alpha^{2\mu}\varepsilon\sqrt{n}}\right)^n.$$

This follows from a standard argument.

3.2. Proof of Theorem 1.2. We need the following easy observation to make sure we can use Corollary 3.5.

Fact 3.6. Let $v \in \mathcal{I}$, $\mu < d/4n$ and $\gamma < \kappa_0\sqrt{d/2n}$, then $\hat{D}_{\alpha, \gamma, \mu}(v) \geq (2\kappa_1)^{-1}\sqrt{n}$.

Proof. Since $v \in \mathcal{I}$ there is $D \subset [n]$ such that $|D| = d$ and $\kappa_0 n^{-1/2} \leq |v_i| \leq \kappa_1 n^{-1/2}$ for all $i \in D$. Now write $\hat{D}(v) = \min_{|I| \geq (1-2\mu)n} D_{\alpha, \gamma}(v_I)$, and let I be a set attaining the minimum. Since $|I| \geq (1-2\mu)n \geq n - d/2$, we have $|I \cap D| \geq d/2$. So put $D' := I \cap D$ and note that for all $t \leq (2\kappa_1)^{-1}\sqrt{n}$, we have

$$\min_I d(tv_I, \mathbb{Z}^n) \geq d(tv_{D'}, \mathbb{Z}^{D'}) = t\|v_{D'}\|_2 \geq t\kappa_0\sqrt{d/2n} > \gamma t.$$

Therefore, $D_{\alpha, \gamma}(v_I) \geq (2\kappa_1)^{-1}\sqrt{n}$, by definition. \square

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 $\mathcal{T}_L(v) \leq \varepsilon$.

Lemma 3.7. *Let $\nu \leq 2^{-8}$. For $v \in \mathbb{R}^n$ and $t \geq \mathcal{T}_L(v)$ we have*

$$\mathcal{L}(Av, t\sqrt{n}) \leq (50Lt)^n.$$

We prove this lemma in Section 5 using a fairly straight-forward argument on the Fourier-side. We now prove our main theorem, Theorem 1.2.

Proof of Theorem 1.2. We pick up from (13) and look to show that

$$q_{n,S} := \sup_{w \in \mathbb{S}^{n-1}} \mathbb{P}_A^{\mathcal{K}}(\exists v \in \mathcal{I} \cap S, s, t \in [-4\sqrt{n}, +4\sqrt{n}] : Av = sv + tw) \leq e^{-\Omega(n)}, \quad (18)$$

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 3.2 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 3.6 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

$$q_{n,S} \leq 2^n \sup_{w \in \mathbb{S}^{n-1}} \mathbb{P}_A^{\mathcal{K}}(\exists v \in \mathcal{I}([d]) \cap S, s, t \in [-4\sqrt{n}, +4\sqrt{n}] : Av = sv + tw). \quad (19)$$

Thus it is enough to show that the supremum at (19) 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 $|s - s'| \leq 2^{-n}$ and $|t - t'| \leq 2^{-n}$. So, union-bounding over all (s', t') , the supremum term in (19) is at most

$$\leq 8^n \sup_{w \in \mathbb{R}^n, |s| \leq 4\sqrt{n}} \mathbb{P}_A^{\mathcal{K}}(\exists v \in \mathcal{I}([d]) \cap (S \cup S') : \|Av - sv - w\|_2 \leq 2^{-n+1})$$

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

$$\eta \leq \mathcal{T}_L(v) \leq 1/L \leq \kappa_0/8, \quad (20)$$

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 (16), we may write

$$\mathcal{I}([d]) \cap S \subseteq \bigcup_{j=0}^n \{v \in \mathcal{I} : \mathcal{T}_L(v) \in [2^j \eta, 2^{j+1} \eta]\} = \bigcup_{j=0}^{j_0} \Sigma_{2^j \eta},$$

where j_0 is the largest integer such that $2^{j_0} \eta \leq \kappa_0/2$. Thus, by the union bound, it is enough to show

$$Q_\varepsilon := \max_{w \in \mathbb{R}^n, |s| \leq 4\sqrt{n}} \mathbb{P}_A^{\mathcal{K}}(\exists v \in \Sigma_\varepsilon : \|Av - sv - w\|_2 \leq 2^{-n+1}) \leq 2^{-4n}, \quad (21)$$

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

$$(2\kappa_1)^{-1}\sqrt{n} \leq \hat{D}_{\alpha,\gamma,\mu}(v) \leq \exp(c_\Sigma n) = \eta^{-1/2},$$

by Fact 3.6. 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 $2^{j_1} \sqrt{\eta} \geq \kappa_1/(2\sqrt{n})$. Union-bounding over j shows that it is sufficient to show

$$Q'_\varepsilon := \max_{w \in \mathbb{R}^n, |s| \leq 4\sqrt{n}} \mathbb{P}_A^\mathcal{K}(\exists v \in \Sigma'_\varepsilon : \|Av - sv - w\|_2 \leq 2^{-n+1}) \leq 2^{-6n}, \quad (22)$$

for all $\varepsilon \in [\sqrt{\eta}, \kappa_1/\sqrt{n}]$.

Step 2: A Bound on Q_ε : Take $w \in \mathbb{R}^n$ and $|s| \leq 4\sqrt{n}$; 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 3.1, to find a $u \in \mathcal{N}_\varepsilon = \mathcal{N}_\varepsilon([d])$ so that $\|v - u\|_2 \leq 4\varepsilon$. So if $\|A\|_{op} \leq 4\sqrt{n}$, we see that

$$\begin{aligned} \|Au - su - w\|_2 &\leq \|Av - sv - w\|_2 + \|A(v - u)\|_2 + |s|\|v - u\|_2 \\ &\leq \|Av - sv - w\|_2 + 8\sqrt{n}\|v - u\|_2 \\ &\leq 33\varepsilon\sqrt{n} \end{aligned}$$

and thus

$$\{\exists v \in \Sigma_\varepsilon : \|Av - sv - w\|_2 \leq 2^{-n+1}\} \cap \{\|A\| \leq 4\sqrt{n}\} \subseteq \{\exists u \in \mathcal{N}_\varepsilon : \|Au - su - w\| \leq 33\varepsilon\sqrt{n}\}.$$

So, by union bounding over our net \mathcal{N}_ε , we see that

$$\begin{aligned} Q_\varepsilon &\leq \mathbb{P}_A^\mathcal{K}(\exists v \in \mathcal{N}_\varepsilon : \|Av - sv - w\| \leq 33\varepsilon\sqrt{n}) \leq \sum_{u \in \mathcal{N}_\varepsilon} \mathbb{P}_A^\mathcal{K}(\|Au - s'u - w\|_2 \leq 33\varepsilon\sqrt{n}) \\ &\leq \sum_{u \in \mathcal{N}_\varepsilon} \mathcal{L}_{A,op}(u, 33\varepsilon\sqrt{n}), \end{aligned}$$

where $\mathcal{L}_{A,op}$ is defined at (15).

Note that for any u we have that $\mathcal{L}_{A,op}(u, 33\varepsilon\sqrt{n}) \leq (67)^n \mathcal{L}_{A,op}(u, \varepsilon\sqrt{n})$ (see, e.g., Fact 6.2 in [1]); as such, for any $u \in \mathcal{N}_\varepsilon$ we have $\mathcal{L}_{A,op}(u, 33\varepsilon\sqrt{n}) \leq (2^{17}L\varepsilon)^n$. Using this bound gives

$$Q_\varepsilon \leq |\mathcal{N}_\varepsilon|(2^{17}L\varepsilon)^n \leq \left(\frac{C}{L^2\varepsilon}\right)^n (2^{17}L\varepsilon)^n \leq 2^{-4n},$$

where the penultimate inequality follows from our Theorem 3.2 and the last inequality holds for the choice of L large enough relative to the universal constant C and so that (20) holds. To see that the application of Theorem 3.2 is valid, note that

$$\log 1/\varepsilon \leq \log 1/\eta = 2c_\Sigma n \leq nL^{-32/c_0^2},$$

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 $|s| \leq 4\sqrt{n}$. Since we chose μ, γ to be sufficiently small so that Fact 3.6 holds, we have that

$$\varepsilon \leq \kappa_1/\sqrt{n}.$$

Thus we may apply Corollary 3.5 with $K = \kappa_1$ for each $v \in \Sigma'_\varepsilon$ to get $u \in G'_\varepsilon \subset \Sigma'_\varepsilon$ such that $\|v - u\|_2 \leq 32\varepsilon\sqrt{\alpha n}$. Now since

$$\{\exists v \in \Sigma'_\varepsilon : \|Av - sv - w\|_2 \leq 2^{-n+1}\} \cap \{\|A\| \leq 4\sqrt{n}\} \subseteq \{\exists u \in G'_\varepsilon : \|Au - su - w\| \leq 2^9\varepsilon\sqrt{\alpha n}\}$$

and since $2^9\varepsilon\sqrt{\alpha n} \geq \exp(-2c_\Sigma n) \geq \mathcal{T}_L(u)$, by Lemma 3.7 we have

$$Q'_\varepsilon \leq \left(\frac{32\kappa_1}{\alpha^\mu \varepsilon \sqrt{n}} \right)^n \sup_{u \in G'_\varepsilon} \mathcal{L}(Au, 2^9\varepsilon\sqrt{\alpha n}) \leq (2^{20}L\kappa_1\alpha^{1/4})^n \leq 2^{-4n},$$

assuming that α is chosen to be sufficiently small relative to $L\kappa_1$. This completes the proof of Theorem 1.2. \square

4. FOURIER PREPARATIONS FOR THEOREM 1.3

4.1. Concentration, level sets, and Esseen-type inequalities. One of the main differences between this work and [1] is the notion of a “level set” of the Fourier transform, an adjustment 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 (14) by

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

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 $t \geq 0$ to be

$$S_W(t) := \{\theta \in \mathbb{R}^\ell : \mathbb{E}_{\bar{\zeta}} \|\bar{\zeta} W \theta\|_{\mathbb{T}}^2 \leq t\}. \quad (23)$$

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, i.e. $\gamma_d(S) = \mathbb{P}_g(g \in S)$ for every Borel set $S \subset \mathbb{R}^d$. We first upper bound the concentration via an Esseen-like inequality.

Lemma 4.1. *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*

$$\mathcal{L}(W^T \tau, \beta\sqrt{\ell}) \leq 2 \exp(2\beta^2\ell - \nu pm/2) \gamma_\ell(S_W(m)).$$

Proof. For $w \in \mathbb{R}^\ell$, apply Markov's inequality to obtain

$$\mathbb{P}_\tau(\|W^T \tau - w\|_2 \leq \beta\sqrt{\ell}) \leq \exp\left(\frac{\pi}{2}\beta^2\ell\right) \mathbb{E}_\tau \exp\left(-\frac{\pi\|W^T \tau - w\|_2^2}{2}\right).$$

Using the Fourier transform of a Gaussian, we compute

$$\mathbb{E}_\tau \exp\left(-\frac{\pi\|W^T \tau - w\|_2^2}{2}\right) = \mathbb{E}_g e^{-2\pi i \langle w, g \rangle} \mathbb{E}_\tau e^{2\pi i g^T W^T \tau}. \quad (24)$$

Now denote the rows of W as w_1, \dots, w_{2d} and write

$$\mathbb{E}_\tau e^{2\pi i g^T W^T \tau} = \prod_{i=1}^{2d} \mathbb{E}_{\tau_i} e^{2\pi i \sum \tau_i \langle g, w_i \rangle} = \prod_{i=1}^{2d} \phi_\tau(\langle g, w_i \rangle),$$

where $\phi_\tau(\theta)$ is the characteristic function of τ . Now apply (8) and then (7) to see the right-hand-side of (24) is

$$\leq \left| \mathbb{E}_g e^{-2\pi i \langle w, g \rangle} \mathbb{E}_\tau e^{2\pi i g^T W^T \tau} \right| \leq \mathbb{E}_g \exp(-\nu p \mathbb{E}_{\bar{\zeta}} \|\bar{\zeta} W g\|_{\mathbb{T}}^2).$$

We rewrite this as

$$\begin{aligned} \int_0^1 \mathbb{P}_g(\exp(-\nu p \mathbb{E}_{\bar{\zeta}} \|\bar{\zeta} W g\|_{\mathbb{T}}^2) \geq t) dt &= \nu p \int_0^\infty \mathbb{P}_g(\mathbb{E}_{\bar{\zeta}} \|\bar{\zeta} W g\|_{\mathbb{T}}^2 \leq u) e^{-\nu p u} du \\ &= \nu p \int_0^\infty \gamma_\ell(S_W(u)) e^{-\nu p u} du, \end{aligned}$$

where for the first equality we made the change of variable $t = e^{-\nu p u}$. Choosing m to maximize $\gamma_\ell(S_W(u))e^{-\nu p u/2}$ as a function of u yields

$$\nu p \int_0^\infty \gamma_\ell(S_W(u)) e^{-\nu p u} du \leq \nu p \gamma_\ell(S_W(m)) e^{-\nu p m/2} \int_0^\infty e^{-\nu p u/2} du = 2\gamma_\ell(S_W(m)) e^{-\nu p m/2}.$$

Putting everything together, we obtain

$$\mathbb{P}_\tau(\|W^T \tau - w\|_2 \leq 2\beta\sqrt{\ell}) \leq 2e^{2\beta^2\ell} e^{-\nu p m/2} \gamma_\ell(S_W(m)).$$

□

We also prove a comparable lower bound.

Lemma 4.2. *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 $t \geq 0$ we have*

$$\gamma_\ell(S_W(t)) e^{-32\nu p t} \leq \mathbb{P}_\tau(\|W^T \tau\|_2 \leq \beta\sqrt{\ell}) + \exp(-\beta^2\ell).$$

Proof. Set $X = \|W^T \tau\|_2$ and write

$$\mathbb{E}_X e^{-\pi X^2/2} = \mathbb{E}_X \mathbb{1}(X \leq \beta\sqrt{\ell}) e^{-\pi X^2/2} + \mathbb{E}_X \mathbb{1}(X \geq \beta\sqrt{\ell}) e^{-\pi X^2/2} \leq \mathbb{P}_X(X \leq \beta\sqrt{\ell}) + e^{-\pi\beta^2\ell/2}.$$

Bounding $\exp(-\pi\beta^2\ell/2) \leq \exp(-\beta^2\ell)$ implies

$$\mathbb{E}_\tau \exp\left(\frac{-\pi\|W^T \tau\|_2^2}{2}\right) \leq \mathbb{P}_\tau(\|W^T \tau\|_2 \leq \beta\sqrt{\ell}) + e^{-\beta^2\ell}.$$

As in the proof of Lemma 4.1 above, use the Fourier transform of the Gaussian and (7) to lower bound

$$\mathbb{E}_\tau \exp \left(-\frac{\pi \|W^T \tau\|_2^2}{2} \right) \geq \mathbb{E}_g [\exp(-32\nu p \mathbb{E}_\xi \|\zeta W g\|_{\mathbb{T}}^2)].$$

Similar to the proof of Lemma 4.1, write

$$\mathbb{E}_g [\exp(-32\nu p \mathbb{E}_\xi \|W g\|_{\mathbb{T}}^2)] = 32\nu p \int_0^\infty \gamma_\ell(S_W(u)) e^{-32\nu p u} du \geq 32\nu p \gamma_\ell(S_W(t)) \int_t^\infty e^{-32\nu p u} du,$$

where we have used that $\gamma_\ell(S_W(b)) \geq \gamma_\ell(S_W(a))$ for all $b \geq a$. This completes the proof of Lemma 4.2. \square

4.2. Inverse Littlewood-Offord for conditioned random walks. First we need a generalization of our important Lemma 3.1 from [1]. Given a $2d \times \ell$ matrix W and a vector $Y \in \mathbb{R}^d$, we define the Y -augmented matrix W_Y as

$$W_Y = \begin{bmatrix} W & \begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix} \end{bmatrix}. \quad (25)$$

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 (1)

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

The following is our generalization of Lemma 3.1 from [1].

Lemma 4.3. *For any $0 < \nu \leq 2^{-15}$, $c_0 \leq 2^{-35} B^{-4} \nu$, $d \in \mathbb{N}$, $\alpha \in (0, 1)$ and $\gamma \in (0, 1)$, let $k \leq 2^{-32} B^{-4} \nu \alpha d$ and $t \geq \exp(-2^{-32} B^{-4} \nu \alpha d)$. Let $Y \in \mathbb{R}^d$ satisfy $\|Y\|_2 \geq 2^{-10} c_0 \gamma^{-1} t^{-1}$, let W be a $2d \times k$ matrix with $\|W\| \leq 2$, $\|W\|_{\text{HS}} \geq \sqrt{k}/2$, and let $\tau \sim \Phi_\nu(2d; \zeta)$.*

If $D_{\alpha, \gamma}(Y) > 2^{10} B^2$ then

$$\mathcal{L} \left(W_Y^T \tau, c_0^{1/2} \sqrt{k+1} \right) \leq (Rt)^2 \exp(-c_0 k), \quad (26)$$

where $R = 2^{35} B^2 \nu^{-1/2} c_0^{-2}$.

We present the proof of Lemma 4.3 in Section 6, and deduce our standalone “inverse Littlewood-Offord theorem” Theorem 1.3 here:

Proof of Theorem 1.3. Let $c_0 = 2^{-35} B^{-4} \gamma^2 \nu$. First note that

$$\mathbb{P} \left(|\langle v, \tau \rangle| \leq t \text{ and } \sum_{i=1}^k \langle w_i, \tau \rangle^2 \leq c_0 k \right)^2 \leq \mathbb{P} \left(|\langle v, \tau \rangle| \leq t, |\langle v, \tau' \rangle| \leq t \text{ and } \sum_{i=1}^k \langle w_i, \tau \rangle^2 \leq c_0 k \right)$$

where $\tau, \tau' \sim \Phi_\nu(d; \zeta)$ are independent. We now look to bound the probability on the right-hand-side using Lemma 4.3.

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 $|\langle v, \tau \rangle| \leq t$, $|\langle v, \tau' \rangle| \leq t$ and $\sum_{i=1}^k \langle w_i, \tau \rangle^2 \leq c_0 k$ then $\|W_Y^T(\tau, \tau')\|_2 \leq c_0^{1/2} \sqrt{k+1}$. Therefore

$$\mathbb{P} \left(|\langle v, \tau \rangle| \leq t, |\langle v, \tau' \rangle| \leq t \text{ and } \sum_{i=1}^k \langle w_i, \tau \rangle^2 \leq c_0 k \right) \leq \mathcal{L} \left(W_Y^T(\tau, \tau'), c_0^{1/2} \sqrt{k+1} \right).$$

Now, $\|Y\|_2 = \sqrt{c_0/2}t^{-1} > 2^{-10}c_0\gamma^{-1}t^{-1}$, $\|W\| = 1$, $\|W\|_{\text{HS}} = \sqrt{k}$ and

$$D_{\alpha, \gamma}(Y) \geq t c_0^{-1/2} D_{\alpha, \gamma}(v) > 2^{10} B^2.$$

We may therefore apply Lemma 4.3 to bound

$$\mathcal{L} \left(W_Y^T(\tau, \tau'), c_0^{1/2} \sqrt{k+1} \right) \leq (Rt)^2 \exp(-c_0 k).$$

The result follows. \square

5. FOURIER REPLACEMENT

The goal of this section is to prove Lemma 3.7, which relates the “zeroed out and lazy” matrix M , defined at (11), to our original matrix A . We will need a few inequalities on the Fourier side first.

Lemma 5.1. *For every $t \in \mathbb{R}$ and $\nu \leq 1/4$ we have*

$$|\phi_\zeta(t)| \leq \phi_{\tilde{\zeta}Z_\nu}(t).$$

Proof. Note $|\phi_\zeta(t)|^2 = \mathbb{E}_{\tilde{\zeta}} \cos(2\pi t \tilde{\zeta})$. Use the elementary inequality $\cos(a) \leq 1 - 2\nu(1 - \cos(a))$, for $\nu \leq 1/4$, and that $\sqrt{1-x} \leq 1 - x/2$ to bound

$$|\phi_\zeta(t)| = \sqrt{\mathbb{E}_{\tilde{\zeta}} \cos(2\pi t \tilde{\zeta})} \leq \sqrt{1 - 2\nu \mathbb{E}_{\tilde{\zeta}}(1 - \cos(2\pi t \tilde{\zeta}))} \leq \phi_{\tilde{\zeta}Z_\nu}(t).$$

\square

We also need a bound on a Gaussian-type moment for $\|Mv\|_2$. On a somewhat technical point, we notice that $\mathcal{T}_L(v) \geq 2^n$, since the definition of \mathcal{T}_L (12) depends on the definition of M at (11), which trivially satisfies

$$\mathbb{P}_M(Mv = 0) \geq \mathbb{P}_M(M = 0) = (1 - \nu)^{\binom{n+1}{2}},$$

for all v and $\nu < 1/2$.

Fact 5.2. *For $v \in \mathbb{R}^n$, and $t \geq \mathcal{T}_L(v)$, we have*

$$\mathbb{E} \exp(-\pi \|Mv\|_2^2 / 2t^2) \leq (9Lt)^n.$$

Proof. Bound

$$\mathbb{E} \exp(-\pi \|Mv\|_2^2 / 2t^2) \leq \mathbb{P}(\|Mv\|_2 \leq t\sqrt{n}) + \sqrt{n} \int_t^\infty e^{-s^2 n / t^2} \mathbb{P}(\|Mv\|_2 \leq s\sqrt{n}) ds. \quad (27)$$

Since $t \geq \mathcal{T}_L(v)$, we have $\mathbb{P}(\|Mv\|_2 \leq s\sqrt{n}) \leq (4Ls)^n$ for all $s \geq t$. Thus we may bound

$$\sqrt{n} \int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right) \mathbb{P}(\|Mv\|_2 \leq s\sqrt{n}) ds \leq \sqrt{n} (8Lt)^n \int_t^\infty \exp\left(-\frac{s^2 n}{t^2}\right) (s/t)^n ds.$$

Changing variables $u = s/t$, we may bound the right-hand side by

$$t^{-1}\sqrt{n}(4Lt)^n \int_1^\infty \exp(-u^2n)u^n du \leq t^{-1}\sqrt{n}(4Lt)^n \int_1^\infty \exp(-u^2/2) du \leq (9Lt)^n,$$

as desired. Note here that we used that $t \geq 2^{-n}$. \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

$$\chi_v(x) := \mathbb{E}_M e^{2\pi i \langle Mv, x \rangle} = \prod_{j=1}^d \prod_{k=d+1}^n \phi_{\tilde{\zeta}_{Z_\nu}}(x_j v_k + x_k v_j).$$

Replacement now goes through:

Proof of Lemma 3.7. By Markov, we have

$$\mathbb{P}(\|Av - w\|_2 \leq t\sqrt{n}) \leq \exp(\pi n/2) \mathbb{E} \exp(-\pi \|Av - w\|_2^2 / (2t^2)). \quad (28)$$

Then use Fourier inversion to write

$$\mathbb{E}_A \exp(-\pi \|Av - w\|_2^2 / (2t^2)) = \int_{\mathbb{R}^n} e^{-\pi \|\xi\|_2^2} \cdot e^{-2\pi i t^{-1} \langle w, \xi \rangle} \psi_v(t^{-1} \xi) d\xi. \quad (29)$$

Now apply the triangle inequality, Lemma 5.1 and the non-negativity of χ_v yields that the right-hand side of (29) is

$$\leq \int_{\mathbb{R}^n} e^{-\pi \|\xi\|_2^2} \chi_v(t^{-1} \xi) d\xi = \mathbb{E}_M \exp(-\pi \|Mv\|_2^2 / 2t^2).$$

Now use Fact 5.2 along with the assumption $t \geq \mathcal{T}_L(v)$ to bound

$$\mathbb{E}_M \exp(-\pi \|Mv\|_2^2 / 2t^2) \leq (9Lt)^n,$$

as desired. \square

6. PROOF OF LEMMA 4.3

In this section we prove Lemma 4.3, which generalizes the Littlewood-Offord theorem for conditioned random walks from [1] to the case of subgaussian random variables. Fortunately much of the geometry needed to prove this theorem can be pulled from the proof of the $\{-1, 0, 1\}$ -case in [1], and so the deduction of the theorem becomes relatively straightforward.

6.1. Properties of Gaussian space and level sets. For $r, s > 0$ and $k \in \mathbb{N}$ define the cylinder $\Gamma_{r,s}$ by

$$\Gamma_{r,s} := \{\theta \in \mathbb{R}^{k+2} : \|\theta_{[k]}\|_2 \leq r, |\theta_{k+1}| \leq s \text{ and } |\theta_{k+2}| \leq s\}. \quad (30)$$

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 [1]:

Lemma 6.1. *Let $S \subset \mathbb{R}^{k+2}$ and $s > 0$ satisfy*

$$8s^2 e^{-k/8} + 32s^2 \max_{a,b,y} (\gamma_k(F_y(S; a, b) - F_y(S; a, b)))^{1/4} \leq \gamma_{k+2}(S). \quad (31)$$

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 4.3. The following basic fact will help explain the use of the difference appearing in Lemma 6.1.

Fact 6.2. *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 $\mathbb{E}_{\bar{\zeta}} \|\bar{\zeta} W x\|_{\mathbb{T}}^2, \mathbb{E}_{\bar{\zeta}} \|\bar{\zeta} W y\|_{\mathbb{T}}^2 \leq m$. The triangle inequality implies

$$\|\bar{\zeta} W(x - y)\|_{\mathbb{T}}^2 \leq 2\|\bar{\zeta} W x\|_{\mathbb{T}}^2 + 2\|\bar{\zeta} W y\|_{\mathbb{T}}^2.$$

Taking $\mathbb{E}_{\bar{\zeta}}$ on both sides completes the fact. \square

6.2. Proof of Lemma 6.3. The following is our main step towards Lemma 4.3.

Lemma 6.3. *For $d \in \mathbb{N}$, $\gamma, \alpha \in (0, 1)$ and $0 < \nu \leq 2^{-15}$, let $k \leq 2^{-17} B^{-4} \nu \alpha d$ and $t \geq \exp(-2^{-17} B^{-4} \nu \alpha d)$. For $c_0 \in (0, 2^{-50} B^{-4})$, let $Y \in \mathbb{R}^d$ satisfy $\|Y\| \geq 2^{-10} c_0 \gamma^{-1} / t$ and let W be a $2d \times k$ matrix with $\|W\| \leq 2$.*

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

$$\mathcal{L}(W_Y^T \tau, \beta \sqrt{k+1}) \geq (Rt)^2 \exp(4\beta^2 k) \left(\mathbb{P}(\|W^T \tau'\|_2 \leq \beta' \sqrt{k}) + \exp(-\beta'^2 k) \right)^{1/4} \quad (32)$$

then $D_{\alpha, \gamma}(Y) \leq 2^{10} B^2$. Here we have set $R = 2^{35} \nu^{-1/2} B^2 / c_0^2$.

Proof. By Lemma 4.1 we may find an m for which the level set $S = S_{W_Y}(m)$ satisfies

$$\mathcal{L}(W_Y^T \tau, \beta \sqrt{k+1}) \leq 4e^{-\nu pm/2+2\beta^2 k} \gamma_{k+2}(S). \quad (33)$$

Combining (33) with the assumption (32) provides a lower bound of

$$\gamma_{k+2}(S) \geq \frac{1}{4} e^{\nu pm/2+2\beta^2 k} (Rt)^2 \left(\mathbb{P}(\|W^T \tau'\|_2 \leq \beta' \sqrt{k}) + \exp(-\beta'^2 k) \right)^{1/4}. \quad (34)$$

Now, preparing for an application of Lemma 6.1, define

$$r_0 := \sqrt{k} \quad \text{and} \quad s_0 := 2^{16} c_0^{-1} (\sqrt{m} + 8B^2 \sqrt{k}) t. \quad (35)$$

Recalling the definition of our *cylinders* from (30), we state the following Claim:

Claim 6.4. *There exists $x \in S \subseteq \mathbb{R}^{k+2}$ so that*

$$(\Gamma_{2r_0, 16} \setminus \Gamma_{2r_0, s_0} + x) \cap S \neq \emptyset. \quad (36)$$

Proof of Claim 6.4. We will use Lemma 6.1 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

$$\theta'' := (\theta_1 - \theta'_1, \dots, \theta_k - \theta'_k, 0, 0) \in S_{W_Y}(4m)$$

by Fact 6.2. This shows that for any y, a, b we have

$$F_y(S; a, b) - F_y(S; a, b) \subset S_{W_Y}(4m) \cap \{\theta \in \mathbb{R}^{k+2} : \theta_{k+1} = \theta_{k+2} = 0\} = S_W(4m), \quad (37)$$

where the equality holds by definition of W_Y and the level set S_{W_Y} . Thus we may apply Lemma 4.2 to obtain

$$\gamma_k(S_W(4m)) \leq e^{128\nu pm} \left(\mathbb{P}(\|W^T \tau'\|_2 \leq \beta' \sqrt{k}) + \exp(-\beta'^2 k) \right). \quad (38)$$

Combining lines (34), (37) and (38), we note that in order to apply Lemma 6.1, it is sufficient to check

$$\begin{aligned} 8s_0^2 e^{-k/8} + 32s_0^2 e^{32\nu pm} \left(\mathbb{P}(\|W^T \tau'\|_2 \leq \beta' \sqrt{k}) + \exp(-\beta'^2 k) \right)^{1/4} \\ < \frac{1}{4} e^{\nu pm/2+2\beta^2 k} (Rt)^2 \left(\mathbb{P}(\|W^T \tau'\|_2 \leq \beta' \sqrt{k}) + \exp(-\beta'^2 k) \right)^{1/4}. \end{aligned} \quad (39)$$

We will show that each term on the left-hand-side of (39) is at most half of the right-hand side. Bound

$$s_0^2 = 2^{32} c_0^{-2} (\sqrt{m} + 8B^2 \sqrt{k})^2 t^2 < 2^{33} (m + 64B^4 k) (t/c_0)^2 \leq 2^{-20} \nu (c_0^2 k + (2B)^{-6} m) (Rt)^2 \quad (40)$$

since $R = 2^{35} B^2 \nu^{-1/2} c_0^{-2}$. By Lemma 2.1 we have that $p \geq 2^{-7} B^{-4}$ and so we may bound

$$8s_0^2 e^{-k/8} \leq e^{-k/8} 2^{-17} \nu (c_0^2 k + (2B)^{-4} m) (Rt)^2 \leq \frac{1}{8} e^{\nu pm/2} (Rt)^2 e^{-\beta'^2 k/4}.$$

Similarly, use (40), $c_0 \leq \beta$ and $\nu = 2^{-7} \nu$ to bound

$$32s_0^2 e^{32\nu pm} \leq 2^{-15} (c_0^2 k + (2B)^{-4} m) (Rt)^2 \exp(\nu pm/4) \leq \frac{1}{8} (Rt)^2 e^{\nu pm/2+\beta^2 k}$$

thus showing (39). Applying Lemma 6.1 completes the claim. \square

The following basic consequence of Claim 6.4 will bring us closer to the construction of our LCD:

Claim 6.5. *We have that $S_{W_Y}(4m) \cap (\Gamma_{2r_0,16} \setminus \Gamma_{2r_0,s_0}) \neq \emptyset$.*

Proof of Claim 6.5. Claim 6.4 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 6.2. \square

We now complete the proof of Lemma 6.3 by showing that an element of the non-empty intersection above provides an LCD:

Claim 6.6. *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*

$$\|\bar{\zeta}_0 \phi_i Y\|_{\mathbb{T}} < \min\{\gamma \bar{\zeta}_0 \phi_i \|Y\|_2, \sqrt{\alpha d}\}.$$

Proof of Claim 6.6. Note that since $\phi \in S_{W_Y}(4m)$ we have

$$\mathbb{E}_{\bar{\zeta}} \|\bar{\zeta} W_Y \phi\|_{\mathbb{T}}^2 \leq 4m.$$

Thus there is some instance $\bar{\zeta}_0 \in (1, 16B^2)$ of $\bar{\zeta}$ so that

$$\|\bar{\zeta}_0 W_Y \phi\|_{\mathbb{T}}^2 \leq 4m. \quad (41)$$

For simplicity, define $\psi =: \bar{\zeta}_0 \phi$.

By (41), 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

$$\psi_{k+1} \begin{bmatrix} Y \\ \mathbf{0}_d \end{bmatrix} + \psi_{k+2} \begin{bmatrix} \mathbf{0}_d \\ Y \end{bmatrix} \in B_{2d}(z, 2\sqrt{m}) - W \psi_{[k]} \subseteq B_{2d}(z, 2\sqrt{m} + 2^6 B^2 \sqrt{k}), \quad (42)$$

where the last inclusion holds because

$$\|W \psi_{[k]}\|_2 \leq \|W\|_{op} \|\psi_{[k]}\|_2 \leq 2|\bar{\zeta}_0| \|\phi_{[k]}\|_2 \leq 32\sqrt{k}B^2,$$

since $\phi \in \Gamma_{2r_0,16}$, $|\bar{\zeta}_0| \leq 16B^2$ and $\|W\|_{op} \leq 2$.

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 (42) onto the first d coordinates yields

$$\psi_{k+1} Y \in B_d(z_{[d]}, 2\sqrt{m} + 2^6 B^2 \sqrt{k}). \quad (43)$$

Now we show that $\|\psi_{k+1} Y\|_{\mathbb{T}} < \gamma \psi_{k+1} \|Y\|_2$. Indeed,

$$\psi_{k+1} \|Y\|_2 \gamma \geq s_0 \|Y\|_2 \gamma > \left(\frac{2^{15}(\sqrt{m} + 8B^2 \sqrt{k})t}{c_0} \right) \left(2^{-10} \frac{c_0}{t} \right) \geq (2\sqrt{m} + 2^6 B^2 \sqrt{k}), \quad (44)$$

where we used the definition of s_0 and that $\|Y\|_2 > 2^{-10} c_0 \gamma^{-1} / t$.

We now need to show

$$2\sqrt{m} + 2^6 B^2 \sqrt{k} \leq \sqrt{\alpha d} \quad (45)$$

Note that since $k \leq 2^{-32}\alpha d/B^4$ we have $2^8 B^2 \sqrt{k} \leq \sqrt{\alpha d}/2$. We claim that $m \leq 2^{-4}\alpha d$. To show this, apply the lower bound (34) and $\gamma_{k+2}(S) \leq 1$ to see

$$e^{-2^{-11}\nu m/B^4} \geq e^{-\nu pm/2} \geq \gamma_{k+2}(S)e^{-\nu pm/2} \geq (Rt)^2 e^{-2\beta'^2 k} \geq t^2 e^{-k} \geq e^{-2^{-15}\nu \alpha d/B^4},$$

where we have used $k \leq 2^{-17}\nu \alpha d/B^4$ and $t \geq e^{-2^{-17}\nu \alpha d/B^4}$. Therefore $m \leq 2^{-4}\alpha d$, i.e. $2\sqrt{m} \leq \sqrt{\alpha d}/2$. Combining this with (43) and (44) we see

$$\|\psi_{k+1}Y\|_{\mathbb{T}} \leq \sqrt{\alpha d},$$

as desired. This completes the proof of the Claim 6.6. \square

Let ϕ , $\bar{\zeta}_0$ and $i \in \{k+1, k+2\}$ be as guaranteed by Claim 6.6. Then $\bar{\zeta}_0 \phi_i \leq 2^{10}B^2$, and

$$\|\bar{\zeta}_0 \phi_i Y\|_{\mathbb{T}} < \min\{\|\bar{\zeta}_0 \phi_i Y\|_2 \gamma, \sqrt{\alpha d}\},$$

and so $D_{\alpha, \gamma}(Y) \leq 2^{10}B^2$ thus completing the proof of Lemma 6.3. \square

6.3. Proof of Lemma 4.3. In order to bridge the gap between Lemmas 6.3 and 4.3, 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 [3, 6]:

Lemma 6.7. *Let $\nu \in (0, 1)$ and $\beta' \in (0, 2^{-7}B^{-2}\sqrt{\nu})$. Let W be a $2d \times k$ matrix satisfying $\|W\|_{\text{HS}} \geq \sqrt{k}/2$ and $\|W\| \leq 2$ and $\tau' \sim \Xi_{\nu}(2d; \zeta)$. Then*

$$\mathbb{P}(\|W^T \tau'\|_2 \leq \beta' \sqrt{k}) \leq 4 \exp(-2^{-20}B^{-4}\nu k).$$

We derive Lemma 6.7 from Talagrand's inequality in Section 10, (see [6] or [3] for more context). From here, we are ready to prove Lemma 4.3.

Proof of Lemma 4.3. Recalling that $c_0 \leq 2^{-35}B^{-4}\nu$, and that our given W satisfies $\|W\|_{\text{HS}} \geq \sqrt{k}/2$ and $\|W\| \leq 2$, we apply Lemma 6.7, 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

$$\mathbb{P}(\|W^T \tau'\|_2 \leq \beta' \sqrt{k}) \leq 4 \exp(-2^{-27}B^{-4}\nu k) \leq 4 \exp(-32c_0 k).$$

We now consider the right-hand-side of (32) in Lemma 6.3: if $\beta \leq \sqrt{c_0}$ we have

$$\begin{aligned} e^{4\beta^2 k} \left(\mathbb{P}(\|W^T \tau'\|_2 \leq \beta' \sqrt{k}) + \exp(-\beta'^2 k) \right)^{1/4} &\leq \exp(4c_0 k - 8c_0 k) + \exp(4c_0 k - 16c_0 k) \\ &\leq 2 \exp(-c_0 k). \end{aligned}$$

We now note that the hypotheses in Lemma 4.3 align with the hypotheses in Lemma 6.3 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 6.3 to obtain

$$\begin{aligned} \mathcal{L}(W_Y^T \tau, \beta \sqrt{k+1}) &\leq (2^{35}B^2\nu^{-1/2}c_0^{-2}t/2)^2 e^{4\beta^2 k} \left(\mathbb{P}(\|W^T \tau'\|_2 \leq 2\beta' \sqrt{k}) + e^{-\beta'^2 k} \right)^{1/4} \\ &\leq (Rt)^2 \exp(-c_0 k), \end{aligned}$$

as desired. \square

7. 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 [1].

Theorem 7.1. *For $n \in \mathbb{N}$ and $0 < c_0 \leq 2^{-50} B^{-4}$, let $d \leq c_0^2 n$, and for $\alpha, \gamma \in (0, 1)$, let $0 \leq k \leq 2^{-32} B^{-4} \alpha d$ and $N \leq \exp(2^{-32} B^{-4} \alpha d)$. Let $X \in \mathbb{R}^d$ satisfy $\|X\|_2 \geq c_0 2^{-10} \gamma^{-1} n^{1/2} N$, 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*

$$\mathbb{P}_H(\sigma_{2d-k+1}(H) \leq c_0 2^{-4} \sqrt{n} \text{ and } \|H_1 X\|_2, \|H_2 X\|_2 \leq n) \leq e^{-c_0 n k / 3} \left(\frac{R}{N} \right)^{2n-2d}, \quad (46)$$

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 $R := 2^{43} B^2 c_0^{-3}$.

7.1. Tensorization and random rounding step. We import the following tensorization Lemma from [1]

Lemma 7.2. *For $d < n$ and $k \geq 0$, 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*

$$\mathbb{P}_H(\|HW\|_{\text{HS}} \leq \beta^2 \sqrt{(k+1)(n-d)}) \leq \left(2^5 e^{2\beta^2 k} \mathcal{L}(W^T \tau, \beta \sqrt{k+1}) \right)^{n-d}.$$

Similarly we use net for the singular vectors of H , constructed in [1]. Let $\mathcal{U}_{2d,k} \subset \mathbb{R}^{[2d] \times [k]}$ be the set of $2d \times k$ matrices with orthonormal columns.

Lemma 7.3. *For $k \leq d$ and $\delta \in (0, 1/2)$, there exists $\mathcal{W} = \mathcal{W}_{2d,k} \subset \mathbb{R}^{[2d] \times [k]}$ with $|\mathcal{W}| \leq (2^6/\delta)^{2dk}$ 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) $\|A(W - U)\|_{\text{HS}} \leq \delta(k/2d)^{1/2} \|A\|_{\text{HS}},$
- (2) $\|W - U\|_{\text{HS}} \leq \delta \sqrt{k}$ and
- (3) $\|W - U\|_{\text{op}} \leq 8\delta.$

7.2. Proof of Theorem 7.1. We also use the following standard fact from linear algebra.

Fact 7.4. *For $3d < n$, let H be a $(n-d) \times 2d$ matrix. If $\sigma_{2d-k+1}(H) \leq x$ then there exist k orthogonal unit vectors $w_1, \dots, w_k \in \mathbb{R}^{2d}$ so that $\|Hw_i\|_2 \leq x$. In particular, there exists $W \in \mathcal{U}_{2d,k}$ so that $\|HW\|_{\text{HS}} \leq x\sqrt{k}$.*

We will also need a bound on $\|H\|_{\text{HS}}$:

Fact 7.5. *Let H be the random $(n-d) \times (2d)$ matrix whose rows are i.i.d. samples of $\Phi_\nu(2d; \zeta)$. Then*

$$\mathbb{P}(\|H\|_{\text{HS}} \geq 2\sqrt{d(n-d)}) \leq 2 \exp(-2^{-21} B^{-4} nd)$$

We are now ready to prove Theorem 7.1.

Proof of Theorem 7.1. Let $Y := \frac{c_0}{32\sqrt{n}} \cdot X$. We may upper bound the left-hand side of (46) by Fact 7.4

$$\begin{aligned} \mathbb{P}(\sigma_{2d-k+1}(H) \leq c_0 2^{-4} \sqrt{n} \text{ and } \|H_1 X\|_2, \|H_2 X\|_2 \leq n) \\ \leq \mathbb{P}(\exists U \in \mathcal{U}_{2d,k} : \|HU_Y\|_{\text{HS}} \leq c_0 \sqrt{n(k+1)}/8). \end{aligned}$$

Set $\delta := c_0/16$, and let \mathcal{W} be as in Lemma 7.3.

For each fixed H , if we have $\|H\|_{\text{HS}} \leq 2\sqrt{d(n-d)}$ and there is some $U \in \mathcal{U}_{2d,k}$ so that $\|HU_Y\|_{\text{HS}} \leq c_0 \sqrt{n(k+1)}/8$, we may apply Lemma 7.3 to find $W \in \mathcal{W}$ so that

$$\|HW_Y\|_{\text{HS}} \leq \|H(W_Y - U_Y)\|_{\text{HS}} + \|HU_Y\|_{\text{HS}} \leq \delta(k/2d)^{1/2} \|H\|_{\text{HS}} + c_0 \sqrt{n(k+1)}/8$$

which is at most $c_0 \sqrt{n(k+1)}/4$. This shows the bound

$$\mathbb{P}_H \left(\exists U \in \mathcal{U}_{2d,k} : \|HU_Y\|_{\text{HS}} \leq c_0 \sqrt{n(k+1)}/8 \right) \leq \mathbb{P}_H \left(\exists W \in \mathcal{W} : \|HW_Y\|_{\text{HS}} \leq c_0 \sqrt{n(k+1)}/4 \right).$$

Conditioning on the event that $\|H\|_{\text{HS}} \leq 2\sqrt{d(n-d)}$, applying Fact 7.5, and union bounding over \mathcal{W} shows that the right-hand-side of the above is at most

$$\sum_{W \in \mathcal{W}} \mathbb{P}_H \left(\|HW_Y\|_2 \leq c_0 \sqrt{n(k+1)}/4 \right) + 2 \exp(-2^{-21} B^{-4} nd).$$

Bound

$$|\mathcal{W}| \leq (2^6/\delta)^{2dk} \leq \exp(32dk \log c_0^{-1}) \leq \exp(c_0 k(n-d)/6),$$

where the last inequality holds since $d \leq c_0^2 n$. Thus

$$\sum_{W \in \mathcal{W}} \mathbb{P}_H(\|HW_Y\|_2 \leq c_0 \sqrt{n(k+1)}/4) \leq \exp(c_0 k(n-d)/6) \max_{W \in \mathcal{W}} \mathbb{P}_H(\|HW\|_2 \leq c_0 \sqrt{n(k+1)}/4). \quad (47)$$

For each $W \in \mathcal{W}$ apply Lemma 7.2 with $\beta := \sqrt{c_0/3}$ (noting that $\sqrt{n-d}/3 \geq \sqrt{n}/4$) to obtain

$$\mathbb{P}_H(\|HW_Y\|_2 \leq c_0 \sqrt{n(k+1)}/4) \leq \left(2^5 e^{2c_0 k/3} \mathcal{L}(W_Y^T \tau, c_0^{1/2} \sqrt{k+1}) \right)^{n-d}. \quad (48)$$

Preparing to apply Lemma 4.3, define $t := (c_0 N/32)^{-1} \geq \exp(-2^{-32} B^{-4} \alpha d)$ 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

$$\|Y\|_2 = c_0 \|X\|_2 / (32n^{1/2}) \geq 2^{-15} c_0^2 N \gamma^{-1} = 2^{-10} c_0 \gamma^{-1} / t.$$

Since $W \in \mathcal{W}$, we have $\|W\|_{\text{op}} \leq 2$ and $\|W\|_{\text{HS}} \geq \sqrt{k}/2$. We also note the bounds $k \leq 2^{-32} B^{-4} \alpha d$, $D_{\alpha, \gamma}(\frac{c_0}{32\sqrt{n}} X) = D_{\alpha, \gamma}(Y) > 2^{10} B^2$. Thus, we may apply Lemma 4.3 to see that

$$\mathcal{L}(W_Y^T \tau, c_0^{1/2} \sqrt{k+1}) \leq (R_0 t)^2 e^{-c_0 k} \leq \left(\frac{R}{8N} \right)^2 e^{-c_0 k}.$$

Substituting this bound into (48) gives

$$\max_{W \in \mathcal{W}} \mathbb{P}_H(\|HW_Y\|_2 \leq c_0 \sqrt{n(k+1)}/4) \leq \frac{1}{2} \left(\frac{R}{N} \right)^{2n-2d} e^{-c_0 k(n-d)/3}.$$

Combining with the previous bounds and noting

$$2 \exp(-2^{-21} B^{-4} n d) \leq \frac{1}{2} \left(\frac{R}{N} \right)^{2n-2d} e^{-c_0 k(n-d)/3}$$

shows

$$\mathbb{P}(\sigma_{2d-k+1}(H) \leq c_0 \sqrt{n}/16 \text{ and } \|H_1 X\|_2, \|H_2 X\|_2 \leq n) \leq \left(\frac{R}{N} \right)^{2n-2d} e^{-c_0 k(n-d)/3}.$$

This completes the proof of Theorem 7.1. \square

8. NETS FOR STRUCTURED VECTORS: SIZE OF THE NET

The goal of this subsection is to prove Theorem 3.2. We follow the same path as Section 7 of [1]. As such, we work with the intersection of \mathcal{N}_ε with a selection of “boxes” which cover a rescaling of the trivial net Λ_ε . We recall the definition of the relevant boxes from [1].

Definition 8.1. Define a (N, κ, d) -box to be a set of the form $\mathcal{B} = B_1 \times \dots \times B_n \subset \mathbb{Z}^n$ where $|B_i| \geq N$ for all $i \geq 1$; $B_i = [-\kappa N, -N] \cup [N, \kappa N]$, for $i \in [d]$; and $|\mathcal{B}| \leq (\kappa N)^n$.

We now interpret these boxes probabilistically and seek to understand the probability that we have

$$\mathbb{P}_M(\|MX\|_2 \leq n) \geq \left(\frac{L}{N} \right)^n,$$

where X is chosen uniformly at random from \mathcal{B} . Theorem 3.2 will follow quickly from the following “box” version:

Lemma 8.2. For $L \geq 2$ and $0 < c_0 \leq 2^{-50} B^{-4}$, let $n > L^{64/c_0^2}$ and let $\frac{1}{4} c_0^2 n \leq d \leq c_0^2 n$. For $N \geq 2$, satisfying $N \leq \exp(c_0 L^{-8n/d} d)$, and $\kappa \geq 2$, let \mathcal{B} be a (N, κ, d) -box. If X is chosen uniformly at random from \mathcal{B} then

$$\mathbb{P}_X \left(\mathbb{P}_M(\|MX\|_2 \leq n) \geq \left(\frac{L}{N} \right)^n \right) \leq \left(\frac{R}{L} \right)^{2n},$$

where $R := C c_0^{-3}$ and $C > 0$ is an absolute constant.

8.1. Counting with the LCD and anti-concentration 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 8.3. For $\alpha \in (0, 1)$, $K \geq 1$ and $\kappa \geq 2$, let $n \geq d \geq K^2/\alpha$ and let $N \geq 2$ 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

$$\mathbb{P}_X(D_{\alpha, \gamma}(r_n X) \leq K) \leq (2^{20} \alpha)^{d/4}, \quad (49)$$

where we have set $r_n := c_0 2^{-5} n^{-1/2}$.

Proof. Writing $\phi = \psi r_n$, note that

$$\mathbb{P}_X(D_{\alpha,\gamma}(r_n X) \leq K) = \mathbb{P}(\exists \phi \in (0, Kr_n] : \|\phi X\|_{\mathbb{T}} < \min\{\gamma\phi\|X\|_2, \sqrt{\alpha d}\}) .$$

We note that any such ϕ must have $|\phi| \geq (2\kappa N)^{-1}$, since if we had $\phi < (2\kappa N)^{-1}$ then each coordinate of ϕX would lie in $(-1/2, 1/2)$, implying $\|\phi X\|_{\mathbb{T}} = \phi\|X\|_2$, i.e. $\|\phi X\|_{\mathbb{T}} > \gamma\phi\|X\|_2$. The proof of Lemma 7.4 in [1] shows that

$$\mathbb{P}_X(\exists \phi \in [(2\kappa N)^{-1}, r_n K] : \|\phi X\|_{\mathbb{T}} < \sqrt{\alpha d}) \leq (2^{20}\alpha)^{d/4}$$

completing the Lemma. \square

We also import from [1, Lemma 7.5] a result showing anti-concentration for random vectors AX , where A is a fixed matrix and X is a random vector with independent entries. As noted in [1], this is essentially a rephrasing of Corollary 1.4 and Remark 2.3 in Rudelson and Vershynin's paper [7]:

Lemma 8.4. *Let $N \in \mathbb{N}$, $n, d, k \in \mathbb{N}$ be such that $n - d \geq 2d > 2k$, H be a $2d \times (n - d)$ matrix with $\sigma_{2d-k}(H) \geq c_0\sqrt{n}/16$ and $B_1, \dots, B_{n-d} \subset \mathbb{Z}$ with $|B_i| \geq N$. If X is taken uniformly at random from $\mathcal{B} := B_1 \times \dots \times B_{n-d}$, then*

$$\mathbb{P}_X(\|HX\|_2 \leq n) \leq \left(\frac{Cn}{dc_0N} \right)^{2d-k},$$

where $C > 0$ is an absolute constant.

8.2. Proof of Theorem 8.2. 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

$$\begin{aligned} \mathcal{A}_1 &:= \{H : \|H_1 X_{[d]}\|_2 \leq n \text{ and } \|H_2 X_{[d]}\|_2 \leq n\} \\ \mathcal{A}_2 &:= \{H : \|H^T X_{[d+1,n]}\|_2 \leq 2n\} . \end{aligned}$$

We now note a simple bound on $\mathbb{P}_M(\|MX\|_2 \leq n)$ in terms of \mathcal{A}_1 and \mathcal{A}_2 .

Fact 8.5. *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*

$$(\mathbb{P}_M(\|MX\|_2 \leq n))^2 \leq \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2).$$

This fact is a straightforward consequence of Fubini's theorem, the details of which are in [1, Fact 7.7]. We shall also need the robust" notion of the rank of the matrix H used in [1]: for $k = 0, \dots, 2k$ define \mathcal{E}_k to be the event

$$\mathcal{E}_k := \{H : \sigma_{2d-k}(H) \geq c_0\sqrt{n}/16 \text{ and } \sigma_{2d-k+1}(H) \leq c_0\sqrt{n}/16\}$$

and note that always at least one of the events $\mathcal{E}_0, \dots, \mathcal{E}_{2d}$ holds.

We now define

$$\alpha := 2^{13} L^{-8n/d} \quad (50)$$

and for a given box \mathcal{B} we define the set of typical vectors $T(\mathcal{B}) \subseteq \mathcal{B}$ by

$$T = T(\mathcal{B}) := \{X \in \mathcal{B} : D_\alpha(c_0 X_{[d]}/(32\sqrt{n})) > 2^{10} B^2\}.$$

Now set $K := 2^{10} B^2$ and note the following implication of Lemma 8.3: if X is chosen uniformly from \mathcal{B} and $n \geq L^{64/c_0^2} \geq 2^{10} B^2/\alpha$ then we have that

$$\mathbb{P}_X(X \notin T) = \mathbb{P}_X(D_\alpha(c_0 X_{[d]}/(32\sqrt{n})) \leq 2^{10} B^2) \leq (2^{33} L^{-8n/d})^{d/4} \leq \left(\frac{2}{L}\right)^{2n}. \quad (51)$$

Proof of Lemma 8.2. 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

$$\mathcal{E} := \{X \in \mathcal{B} : \mathbb{P}_M(\|MX\|_2 \leq n) \geq (L/N)^n\}$$

and bound

$$\mathbb{P}_X(\mathcal{E}) \leq \mathbb{P}_X(\mathcal{E} \cap \{X \in T\}) + \mathbb{P}_X(X \notin T).$$

For each X define

$$f(X) := \mathbb{P}_M(\|MX\|_2 \leq n) \mathbb{1}(X \in T)$$

and apply (51) to bound

$$\mathbb{P}_X(\mathcal{E}) \leq \mathbb{P}_X(f(X) \geq (L/N)^n) + (2/L)^{2n} \leq (N/L)^{2n} \mathbb{E}_X f(X)^2 + (2/L)^{2n}, \quad (52)$$

where the last inequality follows from Markov's inequality. Thus, in order to prove Lemma 8.2 it is enough to prove $\mathbb{E}_X f(X)^2 \leq 2(R/N)^{2n}$.

Apply Fact 8.5 to write

$$\mathbb{P}_M(\|MX\|_2 \leq n)^2 \leq \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{A}_2) = \sum_{k=0}^d \mathbb{P}_H(\mathcal{A}_2 | \mathcal{A}_1 \cap \mathcal{E}_k) \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \quad (53)$$

and so

$$f(X)^2 \leq \sum_{k=0}^d \mathbb{P}_H(\mathcal{A}_2 | \mathcal{A}_1 \cap \mathcal{E}_k) \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \mathbb{1}(X \in T). \quad (54)$$

We will now apply Theorem 7.1 to upper bound $\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k)$ for $X \in T$. For this, note that $d \leq c_0^2 n$, $N \leq \exp(c_0 L^{-8n/d} d) \leq \exp(2^{-32} B^{-4} \alpha n)$ 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 $d \geq \frac{1}{4} c_0^2 n$, we have that $\|X_{[d]}\|_2 \geq d^{1/2} N \geq c_0 2^{-10} \sqrt{n} N$. Now set $\alpha' := 2^{-32} B^{-4} \alpha$ and apply Theorem 7.1 to see that for $X \in T$ and $0 \leq k \leq \alpha' d$, we have

$$\mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \leq e^{-c_0 n k / 3} \left(\frac{R_0}{N}\right)^{2n-2d}.$$

Additionally by Theorem 7.1 we may bound the tail sum:

$$\sum_{k \geq \alpha' d} \mathbb{P}_H(\mathcal{A}_1 \cap \mathcal{E}_k) \leq \mathbb{P}_H(\{\sigma_{2d-\alpha' d}(H) \leq c_0 \sqrt{n}/16\} \cap \mathcal{A}_1) \leq e^{-c_0 \alpha' d n / 4}.$$

Thus, for all $X \in \mathcal{B}$, the previous two equations bound

$$f(X)^2 \leq \sum_{k=0}^{\alpha'd} \mathbb{P}_H(\mathcal{A}_2 | \mathcal{A}_1 \cap \mathcal{E}_k) e^{-c_0 n k / 3} \left(\frac{R_0}{N} \right)^{2n-2d} + e^{-c_0 \alpha' d n / 3}. \quad (55)$$

Seeking to bound the right-hand side of (55), define $g_k(X) := \mathbb{P}_H(\mathcal{A}_2 | \mathcal{A}_1 \cap \mathcal{E}_k)$. Write

$$\mathbb{E}_X[g_k(X)] = \mathbb{E}_X \mathbb{E}_H[\mathcal{A}_2 | \mathcal{A}_1 \cap \mathcal{E}_k] = \mathbb{E}_{X_{[d]}} \mathbb{E}_H \left[\mathbb{E}_{X_{[d+1,n]}} \mathbb{1}[\mathcal{A}_2] | \mathcal{A}_1 \cap \mathcal{E}_k \right].$$

Let $k \leq \alpha'd$. Note that each $H \in \mathcal{A}_1 \cap \mathcal{E}_k$ has $\sigma_{2d-k}(H) \geq c_0 \sqrt{n}/16$ and thus we may apply Lemma 8.4 to bound

$$\mathbb{E}_{X_{[d+1,n]}} \mathbb{1}[\mathcal{A}_2] = \mathbb{P}_{X_{[d+1,n]}}(\|H^T X_{[d+1,n]}\|_2 \leq n) \leq \left(\frac{C'n}{c_0 d N} \right)^{2d-k} \leq \left(\frac{4C'}{c_0^3 N} \right)^{2d-k}$$

for an absolute constant $C' > 0$, where we used that $d \geq \frac{1}{4} c_0^2 n$. Thus, for each $0 \leq k \leq \alpha'd$, if we define $R := \max\{8C'c_0^{-3}, 2R_0\}$ then we have

$$\mathbb{E}_X[g_k(X)] \leq \left(\frac{R}{2N} \right)^{2d-k}. \quad (56)$$

Applying \mathbb{E}_X to (55) using (56) shows

$$\mathbb{E}_X f(X)^2 \leq \left(\frac{R}{2N} \right)^{2n} \sum_{k=0}^{\alpha'd} \left(\frac{2N}{R} \right)^k e^{-c_0 n k / 3} + e^{-c_0 \alpha' d n / 3}.$$

Using that $N \leq e^{c_0 L^{-8n/d} d} = e^{c_0 \alpha' d / 8}$ and $N \leq e^{c_0 n / 3}$ bounds

$$\mathbb{E}_X f(X)^2 \leq 2 \left(\frac{R}{2N} \right)^{2n}. \quad (57)$$

Combining (57) with (52) completes the proof of Lemma 8.2. \square

8.3. Proof of Theorem 3.2. The main work of proving Theorem 3.2 is now complete with the proof of Lemma 8.2. In order to complete it, we need to cover the sphere with a suitable set of boxes. Recall the definitions from Section 3.1:

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

and

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

and that the constants κ_0, κ_1 satisfy $0 < \kappa_0 < 1 < \kappa_1$ and are defined in Section 2.3.

We import the following simple covering lemma from [1, Lemma 7.8]

Lemma 8.6. *For all $\varepsilon \in [0, 1]$, $\kappa \geq \max\{\kappa_1/\kappa_0, 2^8 \kappa_0^{-4}\}$, there exists a family \mathcal{F} of (N, κ, d) -boxes with $|\mathcal{F}| \leq \kappa^n$ so that*

$$\Lambda_\varepsilon \subseteq \bigcup_{\mathcal{B} \in \mathcal{F}} (4\varepsilon n^{-1/2}) \cdot \mathcal{B}, \quad (58)$$

where $N = \kappa_0/(4\varepsilon)$.

Combining Lemma 8.6 with Lemma 8.2 will imply Theorem 3.2.

Proof of Theorem 3.2. Apply Lemma 8.6 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

$$\mathcal{N}_\varepsilon \subseteq \bigcup_{\mathcal{B} \in \mathcal{F}} ((4\varepsilon n^{-1/2}) \cdot \mathcal{B}) \cap \mathcal{N}_\varepsilon$$

and so

$$|\mathcal{N}_\varepsilon| \leq \sum_{\mathcal{B} \in \mathcal{F}} |(4\varepsilon n^{-1/2}) \cdot \mathcal{B} \cap \mathcal{N}_\varepsilon| \leq |\mathcal{F}| \cdot \max_{\mathcal{B} \in \mathcal{F}} |(4\varepsilon n^{-1/2}) \cdot \mathcal{B} \cap \mathcal{N}_\varepsilon|.$$

Rescaling by $\sqrt{n}/(4\varepsilon)$ and applying Lemma 8.2 bounds

$$|(4\varepsilon n^{-1/2}) \cdot \mathcal{B} \cap \mathcal{N}_\varepsilon| \leq |\{X \in \mathcal{B} : \mathbb{P}_M(\|MX\|_2 \leq n) \geq (L\varepsilon)^n\}| \leq \left(\frac{R}{L}\right)^{2n} |\mathcal{B}|.$$

To see that the application of Lemma 8.2 is justified, note that $0 < c_0 \leq 2^{-50}B^{-4}$, $c_0^2 n/2 \leq d \leq c_0^2 n$, $\kappa \geq 2$, and $\log 1/\varepsilon \leq n/L^{64/c_0^2}$ and so

$$\log N = \log \kappa_0/(4\varepsilon) \leq n/L^{64/c_0^2} \leq c_0 L^{-8n/d} d,$$

as required by Lemma 8.2, since $\kappa_0 < 1$, $d \geq L^{-1/c_0^2} n$, $c_0 \geq L^{-1/c_0^2}$ and $8n/d \leq 16/c_0^2$. Using that $|\mathcal{F}| \leq \kappa^n$ and $|\mathcal{B}| \leq (\kappa N)^n$ for each $\mathcal{B} \in \mathcal{F}$ bounds

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

where we set $C := \kappa^2 R^2 c_0^6$. This completes the proof of Theorem 3.2. \square

9. NETS FOR STRUCTURED VECTORS: APPROXIMATING WITH THE NET

In this section we prove Lemma 3.1 which tells us that \mathcal{N}_ε is a net for Σ_ε . The proof uses the random rounding technique developed by Livshyts [4] in the same way as in [1].

Proof of Lemma 3.1. 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 $|r_i| \leq 4\varepsilon n^{-1/2}$ 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, $\|r\|_\infty = \|u - v\|_\infty \leq 4\varepsilon n^{-1/2}$ for all u . Also, $u \in \mathcal{I}'([d])$ for all u , since $v \in \mathcal{I}([d])$ and $\|u - v\|_\infty \leq 4\varepsilon/\sqrt{n} \leq \kappa_0/(2\sqrt{n})$. Thus, from the definition of \mathcal{N}_ε , we need only show that with positive probability u satisfies

$$\mathbb{P}(\|Mu\|_2 \leq 4\varepsilon\sqrt{n}) \geq (L\varepsilon)^n \text{ and } \mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) \leq (2^{10}L\varepsilon)^n. \quad (59)$$

We first show that *all* u satisfy the upper bound at (59). To see this, recall $\mathcal{K} = \{\|A\|_{\text{op}} \leq 4\sqrt{n}\}$ and let $w(u) \in \mathbb{R}^n$, be such that

$$\begin{aligned} \mathcal{L}_{A,op}(u, \varepsilon\sqrt{n}) &= \mathbb{P}^{\mathcal{K}}(\|Av - Ar - w(u)\| \leq \varepsilon\sqrt{n}) \\ &\leq \mathbb{P}^{\mathcal{K}}(\|Av - w(u)\| \leq 17\varepsilon\sqrt{n}) \\ &\leq \mathcal{L}_{A,op}(v, 17\varepsilon\sqrt{n}) \leq \mathcal{L}(Av, 17\varepsilon\sqrt{n}). \end{aligned}$$

Since $v \in \Sigma_\varepsilon$, Lemma 3.7 bounds

$$\mathcal{L}(Av, 17\varepsilon\sqrt{n}) \leq (2^{10}L\varepsilon)^n. \quad (60)$$

We now show that

$$\mathbb{E}_u \mathbb{P}_M(\|Mu\|_2 \leq 4\varepsilon\sqrt{n}) \geq (1/2)\mathbb{P}_M(\|Mv\|_2 \leq 2\varepsilon\sqrt{n}) \geq (1/4)(2\varepsilon L)^n, \quad (61)$$

where the last inequality holds by the fact $v \in \Sigma_\varepsilon$. From (61), it then follows that there is some $u \in \Lambda_\varepsilon$ satisfying (59).

To prove the first inequality in (59), define the event

$$\mathcal{E} := \{M : \|Mv\|_2 \leq 2\varepsilon\sqrt{n} \text{ and } \|M\|_{\text{HS}} \leq n/4\}$$

and note that for all u , we have

$$\mathbb{P}_M(\|Mu\|_2 \leq 4\varepsilon\sqrt{n}) = \mathbb{P}_M(\|Mv - Mr\|_2 \leq 4\varepsilon\sqrt{n}) \geq \mathbb{P}_M(\|Mr\|_2 \leq 2\varepsilon\sqrt{n} \text{ and } \mathcal{E}).$$

Since by Bernstein inequality $\mathbb{P}(\|M\|_{\text{HS}}^2 \geq n^2/16) \leq 2\exp(-cn^2)$ and the fact that

$$\varepsilon \geq \exp(-2c_\Sigma n) \geq \exp(-cn)$$

we have

$$\mathbb{P}(\mathcal{E}) \geq (2L\varepsilon)^n - 2\exp(-cn^2) \geq (1/2)(2L\varepsilon)^n,$$

assuming that c_Σ is chosen appropriately small compared to this absolute constant. Thus

$$\begin{aligned} \mathbb{P}_M(\|Mu\|_2 \leq 4\varepsilon\sqrt{n}) &\geq \mathbb{P}_M(\|Mr\|_2 \leq 2\varepsilon\sqrt{n} \mid \mathcal{E})\mathbb{P}(\mathcal{E}) \\ &\geq (1 - \mathbb{P}_M(\|Mr\|_2 > 2\varepsilon\sqrt{n} \mid \mathcal{E})) (1/2)(2L\varepsilon)^n. \end{aligned}$$

Taking expectations gives

$$\mathbb{E}_u \mathbb{P}_M(\|Mu\|_2 \leq 4\varepsilon\sqrt{n}) \geq (1 - \mathbb{E}_u \mathbb{P}_M(\|Mr\|_2 > 2\varepsilon\sqrt{n} \mid \mathcal{E})) (1/2)(2L\varepsilon)^n. \quad (62)$$

Exchanging the expectations and rearranging, we see that it is enough to show

$$\mathbb{E}_M [\mathbb{P}_r(\|Mr\|_2 > 2\varepsilon\sqrt{n}) \mid \mathcal{E}] \leq 1/2.$$

We will show that $\mathbb{P}_r(\|Mr\|_2 > 2\varepsilon\sqrt{n}) \leq 1/4$ for all $M \in \mathcal{E}$, by Markov's inequality. Note that

$$\mathbb{E}_r \|Mr\|_2^2 = \sum_{i,j} \mathbb{E} (M_{i,j} r_i)^2 = \sum_i \mathbb{E} r_i^2 \sum_j M_{i,j}^2 \leq 16\varepsilon^2 \|M\|_{\text{HS}}^2 / n \leq \varepsilon^2 n,$$

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 $\|r\|_\infty \leq 4\varepsilon/\sqrt{n}$; and for the final inequality we used $\|M\|_{\text{HS}} \leq n/4$. Thus by Markov's inequality gives

$$\mathbb{P}_r(\|Mr\|_2 \geq 2\varepsilon\sqrt{n}) \leq (2\varepsilon\sqrt{n})^{-2} \mathbb{E}_r \|Mr\|_2^2 \leq 1/4. \quad (63)$$

Putting (63) together with (62) proves (61), completing the proof of (59). \square

10. PROOF OF LEMMA 6.7

We will derive Lemma 6.7 from Talagrand's inequality:

Theorem 10.1 (Talagrand's Inequality). *Let $F : \mathbb{R}^n \rightarrow \mathbb{R}$ be a convex 1-Lipschitz function and $\sigma = (\sigma_1, \dots, \sigma_n)$ where the σ_i are i.i.d. random variables such that $|\sigma_i| \leq 1$. Then for any $t \geq 0$ we have*

$$\mathbb{P}(|F(\sigma) - m_F| \geq t) \leq 4 \exp(-t^2/16),$$

where m_F is the median of $F(\sigma)$.

Proof of Lemma 6.7. Note the theorem is trivial if $k \leq 2^{20}B^4/\nu$, 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 $|\sigma_i| \leq 2^{-4}B^{-2}|\tau_i| \leq 1$ and the σ_i are i.i.d., Theorem 10.1 tells us that $F(\sigma)$ is concentrated about the median m_F and so we only need to estimate m_F . For this, write

$$m := \mathbb{E} \|W^T \sigma\|_2^2 = \sum_{i,j} W_{ij}^2 \mathbb{E} \sigma_i^2 = \mathbb{E} \sigma_i^2 \|W\|_{\text{HS}}^2,$$

and

$$m_2 := \mathbb{E} \|W^T \sigma\|_2^4 - (\mathbb{E} \|W^T \sigma\|_2^2)^2 = \sum_{i,j} W_{ij}^2 (\mathbb{E} \sigma_i^4 - (\mathbb{E} \sigma_i^2)^2) \leq \mathbb{E} \sigma_i^2 \|W\|_{\text{HS}}^2,$$

where for the final inequality we used that $\mathbb{E} \sigma_i^4 \leq \mathbb{E} \sigma_i^2$, since $|\sigma_i| \leq 1$. For $t > 0$, Markov's inequality bounds

$$\mathbb{P}(\|W^T \sigma\|_2^2 \leq m - t) \leq t^{-2} \mathbb{E} (\|W^T \sigma\|_2^2 - m)^2 = t^{-2} m_2 \leq t^{-2} \mathbb{E} \sigma_i^2 \|W\|_{\text{HS}}^2.$$

Setting $t = \mathbb{E} \sigma_i^2 \|W\|_{\text{HS}}^2 / 2$ gives

$$\mathbb{P}(\|W^T \sigma\|_2^2 \leq \mathbb{E} \sigma_i^2 \|W\|_{\text{HS}}^2 / 2) \leq 4(\mathbb{E} \sigma_i^2 \|W\|_{\text{HS}}^2)^{-1} < 1/2$$

since $\mathbb{E} \sigma_i^2 = 2^{-8}B^{-4}\mathbb{E} \tau_i'^2 \geq 2^{-8}B^{-4}\nu$ and $\|W\|_{\text{HS}}^2 \geq k/4 > 2^{11}\nu^{-1}B^4$ (by assumption). It follows that

$$m_F \geq \sqrt{\mathbb{E} \sigma_i^2 / 2} \|W\|^{-1} \|W\|_{\text{HS}} \geq 2^{-6} \|W\|^{-1} B^{-2} \sqrt{\nu k},$$

since $\|W\|_{\text{HS}} \geq \sqrt{k}/2$. Now we may apply Talagrand's Inequality (Theorem 10.1) with $t = m_F - \beta' \sqrt{k} \|W\|^{-1}$ to obtain

$$\mathbb{P}(\|W^T \sigma\|_2 \leq \beta' \sqrt{k}) \leq 4 \exp(-2^{-20} B^{-4} \nu k)$$

as desired. □

REFERENCES

- [1] M. Campos, M. Jensen, M. Michelen, and J. Sahasrabudhe. The singularity probability of a random symmetric matrix is exponentially small. <https://arxiv.org/abs/2105.11384>, 2021.
- [2] M. Campos, M. Jensen, M. Michelen, and J. Sahasrabudhe. The least singular value of a random symmetric matrix. <https://arxiv.org/abs/2203.06141>, 2022.
- [3] D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.
- [4] G. V. Livshyts. The smallest singular value of heavy-tailed not necessarily iid random matrices via random rounding. *arXiv preprint arXiv:1811.07038*, 2018.
- [5] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.
- [6] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-Gaussian concentration. *Electronic Communications in Probability*, 18, 2013.
- [7] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices*, 2015(19):9594–9617, 2015.
- [8] R. Vershynin. Invertibility of symmetric random matrices. *Random Structures Algorithms*, 44(2):135–182, 2014.
- [9] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

INSTITUTO DE MATEMÁTICA PURA E APLICADA (IMPA).

Email address: `marcelo.campos@impa.br`

UNIVERSITY OF BIRMINGHAM, SCHOOL OF MATHEMATICS.

Email address: `m.jenssen@bham.ac.uk`

UNIVERSITY OF ILLINOIS AT CHICAGO. DEPARTMENT OF MATHEMATICS, STATISTICS AND COMPUTER SCIENCE.

Email address: `michelen.math@gmail.com`

UNIVERSITY OF CAMBRIDGE. DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICS STATISTICS.

Email address: `jdrs2@cam.ac.uk`