Invertibility of symmetric random matrices

Roman Vershynin*
University of Michigan
romanv@umich.edu

February 1, 2011; last revised March 16, 2012

Abstract

We study $n \times n$ symmetric random matrices H, possibly discrete, with iid above-diagonal entries. We show that H is singular with probability at most $\exp(-n^c)$, and $\|H^{-1}\| = O(\sqrt{n})$. Furthermore, the spectrum of H is delocalized on the optimal scale $o(n^{-1/2})$. These results improve upon a polynomial singularity bound due to Costello, Tao and Vu, and they generalize, up to constant factors, results of Tao and Vu, and Erdös, Schlein and Yau.

Keywords: Symmetric random matrices, invertibility problem, singularity probability.

Contents

*Partially supported by NSF grant DMS 1001829

1 Introduction

1.1 Invertibility problem

This work is motivated by the invertibility problem for $n \times n$ random matrices H. This problem consists of two questions:

1. What is the singularity probability $\mathbb{P}\{H \text{ is singular}\}$?
2. What is the typical value of the spectral norm of the inverse, $\|H^{-1}\|$?

A motivating example is for random Bernoulli matrices B whose entries are ± 1 valued symmetric random variables. If all entries are independent, it is conjectured that the singularity probability of B is $(\frac{1}{2}+o(1))^n$, while the best current bound $(\frac{1}{\sqrt{2}}+o(1))^n$ is due to Bourgain, Vu and Wood [2]. The typical norm of the inverse in this case is $\|B^{-1}\| = O(\sqrt{n})$ [13, 20], see [15]. Moreover, the following inequality due to Rudelson and the author [13] simultaneously establishes the exponentially small singularity probability and the correct order for the norm of the inverse:

$$\mathbb{P}\Big\{\min_k s_k(B) \le \varepsilon n^{-1/2}\Big\} \le C\varepsilon + 2e^{-cn}, \tag{1.1}$$

where $C, c > 0$ are absolute constants. Here $s_k(B)$ denote the singular values of B, so the matrix B is singular iff $\min_k s_k(B) = 0$; otherwise $\min_k s_k(B) = 1/\|B^{-1}\|$.

Less is known about the invertibility problem for *symmetric* Bernoulli matrices H, where the entries on and above the diagonal are independent ± 1 valued symmetric random variables. As is the previous case of iid entries, it is even difficult to show that the singularity probability converges to zero as $n \to \infty$. This was done by Costello, Tao and Vu [4] who showed that

$$\mathbb{P}\big\{H \text{ is singular}\big\} = O(n^{-1/8+\delta}) \tag{1.2}$$

for every $\delta > 0$. They conjectured that the optimal singularity probability bound is for symmetric Bernoulli matrices is again $(\frac{1}{2} + o(1))^n$.

1.2 Main result

In this paper, we establish a version of (1.1) for symmetric random matrices. To give a simple specific example, our result will yield both an exponential bound on the singularity probability and the correct order of the norm of the inverse for symmetric Bernoulli matrices:

$$\mathbb{P}\big\{H \text{ is singular}\big\} \le 2e^{-n^c}; \qquad \mathbb{P}\big\{\|H^{-1}\| \le C\sqrt{n}\big\} \ge .99$$

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

Our results will apply not just for Bernoulli matrices, but also for general matrices H that satisfy the following set of assumptions:

(H) $H = (h_{ij})$ is a real symmetric matrix. The above-diagonal entries h_{ij}, $i < j$, are independent and identically distributed random variables with zero mean and unit variance. The diagonal entries h_{ii} can be arbitrary numbers (either non-random, or random but independent of the off-diagonal entries).

The eigenvalues of H in a non-decreasing order are denoted by $\lambda_k(H)$.

Theorem 1.1 (Main). *Let H be an $n \times n$ symmetric random matrix satisfying* **(H)** *and whose off-diagonal entries have finite fourth moment. Let $K > 0$. Then for every $z \in \mathbb{R}$ and $\varepsilon \ge 0$, one has*

$$\mathbb{P}\Big\{\min_k |\lambda_k(H) - z| \le \varepsilon n^{-1/2} \text{ and } \max_k |\lambda_k(H)| \le K\sqrt{n}\Big\} \le C\varepsilon^{1/9} + 2e^{-n^c}. \tag{1.3}$$

Here $C, c > 0$ depend only on the fourth moment of the entries of H and on K.

The bound on the spectral norm $\|H\| = \max_k |\lambda_k(H)|$ can often be removed from (1.3) at no cost, as one always has $\|H\| = O(\sqrt{n})$ with high probability under the four moment assumptions of Theorem 1.1; see Theorem 1.5 for a general result.

Moreover, for some ensembles of random matrices one has $\|H\| = O(\sqrt{n})$ with exponentially high probability. This holds under the higher moment assumption that

$$\mathbb{E}\exp(h_{ij}^2/M^2) \leq e, \quad i \neq j \tag{1.4}$$

for some number $M > 0$. Such random variables h_{ij} are called *sub-gaussian* random variables, and the minimal number M is called the sub-gaussian moment of h_{ij}. The class of sub-gaussian random variables contains standard normal, Bernoulli, and generally all bounded random variables, see [24] for more information. For matrices with subgaussian entries, it is known that $\|H\| = O(\sqrt{n})$ with probability at least $1 - 2e^{-n}$, see Lemma 2.3. Thus Theorem 1.1 implies:

Theorem 1.2 (Subgaussian)**.** *Let H be an $n \times n$ symmetric random matrix satisfying* **(H)***, whose off-diagonal entries are subgaussian random variables, and whose diagonal entries satisfy $|h_{ii}| \leq K\sqrt{n}$ for some K. Then for every $z \in \mathbb{R}$ and $\varepsilon \geq 0$, one has*

$$\mathbb{P}\Big\{ \min_k |\lambda_k(H) - z| \leq \varepsilon n^{-1/2} \Big\} \leq C\varepsilon^{1/9} + 2e^{-n^c}. \tag{1.5}$$

Here $c > 0$ and C depend only on the sub-gaussian moment M and on K.

Singularity and invertibility. For $\varepsilon = 0$, Theorem 1.2 yields an exponential bound on singularity probability:

$$\mathbb{P}\big\{H \text{ is singular}\big\} \leq 2e^{-n^c}.$$

Furthermore, since $\min_k |\lambda_k(H) - z| = \|(H - zI)^{-1}\|$, (1.5) can be stated as a bound on the spectral norm of the resolvent,

$$\mathbb{P}\Big\{ \|(H - zI)^{-1}\| \geq \frac{\sqrt{n}}{\varepsilon} \Big\} \leq C\varepsilon^{1/9} + 2e^{-n^c}.$$

This estimate is valid for all $z \in \mathbb{R}$ and all $\varepsilon \geq 0$. In particular, we have

$$\|(H - zI)^{-1}\| = O(\sqrt{n}) \text{ with high probability.} \tag{1.6}$$

For $z = 0$ this yields the bound on the norm of the inverse, and on the *condition number* of H:

$$\|H^{-1}\| = O(\sqrt{n}), \quad \kappa(H) := \|H\|\|H^{-1}\| = O(n) \text{ with high probability.} \tag{1.7}$$

In these estimates, the constants implicit in $O(\cdot)$ depend only on M, K and the desired probability level.

Delocalization of eigenvalues. Theorem 1.2 is a statement about delocalization of eigenvalues of H. It states that, for any fixed short interval $I \subseteq \mathbb{R}$ of length $|I| = o(n^{-1/2})$, there are no eigenvalues in I with high probability. This is consistent with the simple heuristics about eigenvalue spacings. According to the spectral norm bound, all n eigenvalues of H lie in the interval of length $O(\sqrt{n})$. So the average spacing between the eigenvalues is of the order $n^{-1/2}$. Theorem 1.2 states that, indeed, any interval of smaller length $o(n^{-1/2})$ is likely to fall in a gap between consequtive eigenvalues. For results in the converse direction, on good localization of eigenvalues around their means, see [23] and the references therein.

Related results. A result of the type of Theorem 1.2 was known for random matrices H whose entries have continuous distributions with certain smoothness properties, and in the bulk of spectrum, i.e. for $|z| \le (2-\delta)\sqrt{n}$ (and assuming that the diagonal entries of H are independent random variables with zero mean and unit variance). A result of Erdös, Schlein and Yau [5] (stated for complex Hermitian matrices) is that

$$\mathbb{P}\Big\{ \min_k |\lambda_k(H) - z| \le \varepsilon n^{-1/2} \Big\} \le C\varepsilon. \tag{1.8}$$

This estimate does not have a singularity probability term $2e^{-n^c}$ that appears in (1.5), which is explained by the fact that matrices with continuous distributions are almost surely non-singular. In particular, this result does not hold for discrete distributions.

Some related results which apply for discrete distributions are due to Tao and Vu. Theorem 1.14 in [22] states that for every $\delta > 0$ and $1 \le k \le n$, one has

$$\mathbb{P}\Big\{ \lambda_{k+1}(H) - \lambda_k(H) \le n^{-\frac{1}{2}-\delta} \Big\} \le n^{-c(\delta)}. \tag{1.9}$$

This result does not assume a continuous distribution of the entries of H, just appropriate (exponential) moment assumptions. In particular, the eigenvalue gaps $\lambda_{k+1}(H) - \lambda_k(H)$ are of the order at least $n^{-\frac{1}{2}-\delta}$ with high probability. This order is optimal up to δ in the exponent, but the polynomial probability bound $n^{-c(\delta)}$ is not. Furthermore, (1.2) and (1.9) are results of somewhat different nature: (1.2) establishes absolute delocalization of eigenvalues with respect to a given point z, while (1.9) gives a relative delocalization with respect to the neighboring eigenvalues.

Finally, recent universality results due to Tao and Vu [21, 22] allow to compare the distribution of $\lambda_k(H)$ to the distribution of $\lambda_k(G)$ where G is a symmetric matrix with independent $N(0,1)$ entries. These results also apply for matrices H with discrete distributions, although one has to assume that the first few moments (such as three or four) of the entries of H and of G are equal (so it does not seem that this approach can be used for symmetric Bernoulli matrices). Also, such comparisons come at a cost of a polynomial, rather than exponential, probability error:

$$\begin{aligned}\mathbb{P}\Big\{\min_k |\lambda_k(G)| \le \varepsilon n^{-1/2} - n^{-c-1/2}\Big\} - O(n^{-c}) \\ \le \mathbb{P}\Big\{\min_k |\lambda_k(H)| \le \varepsilon n^{-1/2}\Big\} \\ \le \mathbb{P}\Big\{\min_k |\lambda_k(G)| \le \varepsilon n^{-1/2} + n^{-c-1/2}\Big\} + O(n^{-c}). \quad (1.10)\end{aligned}$$

(See Corollary 24 in [21] and its proof.)

Remark 1.3. After the results of this paper had been obtained, the author was informed of an independent work by Nguyen [10], which improved Costello-Tao-Vu's singularity probability bound (1.2) for symmetric Bernoulli matrices to

$$\mathbb{P}\big\{H \text{ is singular}\big\} = O(n^{-M})$$

for every $M > 0$, where a constant implicit in $O(\cdot)$ depends only on M. The even more recent work by Nguyen [11], which was announced a few days after the current paper had been posted, demonstrated that for every $M > 0$ there exists $K > 0$ such that

$$\mathbb{P}\Big\{\min_k |\lambda_k(H)| \le n^{-K}\Big\} \le n^{-M}.$$

While Nguyen's results give weaker conclusions than the results in this paper, they hold under somewhat weaker conditions on the distribution than **(H)** (for example, the entries of H do not to have mean zero); see [11] for precise statements.

Remark 1.4 (Optimality). Although the magnitude of the gap $n^{-1/2}$ in Theorem 1.1 is optimal, the form of (1.1) and (1.8) suggests that the exponent 1/9 is not optimal. Indeed, our argument automatically yields $\varepsilon^{1/8+\delta}$ for every $\delta > 0$ (with constants C, c depending also on δ). Some further improvement of the exponent may be possible with a more accurate argument, but the technique of this paper would still not reach the optimal exponent 1 (in particular, due to losses in decoupling). Furthermore, we conjecture that the singularity probability term $2e^{-n^c}$ in (1.5) may be improved to $2e^{-cn}$.

1.3 Four moments

Even without subgaussian assumption (1.4) on the entries of H, the bound on the spectral norm $\|H\| = \max_k |\lambda_k(H)|$ can be removed from (1.3), however this will lead to a weaker probability bound than in Theorem 1.2:

Theorem 1.5 (Four moments)**.** *Let H be an $n \times n$ symmetric random matrix satisfying* **(H)***, whose off-diagonal entries have finite fourth moment M_4^4, and whose diagonal entries satisfy $|h_{ii}| \le K\sqrt{n}$ for some K. For every $p > 0$ there exist $n_0, \varepsilon > 0$ that depend only on the fourth moment of entries, K and p, and such that for all $n \ge n_0$ one has*

$$\mathbb{P}\Big\{ \min_k |\lambda_k(H) - z| \le \varepsilon n^{-1/2} \Big\} \le p.$$

To see how this result follows from Theorem 1.1, note that a result of Latala implies a required bound on the spectral norm. Indeed, Lemma 2.4 and Markov's inequality yield $\|H\| = \max_k |\lambda_k(H)| \le (CM_4 + K)\sqrt{n}$ with high probability. Using this together with (1.1) implies Theorem 1.5.

An immediate consequence of Theorem 1.5 is that such matrices H are asymptotically almost surely non-singular:

$$\mathbb{P}\big\{ H \text{ is singular} \big\} \le p_n(M_4, K) \to 0 \quad \text{as } n \to \infty.$$

Like Theorem 1.2, Theorem 1.5 also establishes the delocalization of eigenvalues on the optimal scale $n^{-1/2}$ and the bounds on the resolvent (1.6), on the norm of the inverse and on the condition number (1.7) – all these hold under just the fourth moment assumption as in Theorem 1.5.

1.4 Overview of the argument

Decomposition into compressible and incompressible vectors. Let us explain the heuristics of the proof of Theorem 1.1. Consider the matrix $A = H - zI$. Note that $\min_k |\lambda_k(H) - z| = \min_k |\lambda_k(A)| = \min_{x \in S^{n-1}} \|Ax\|_2$ where S^{n-1} denotes the Euclidean sphere in $\mathbb{R}^n$. So our task is to bound above the probability

$$\mathbb{P}\Big\{ \min_{x \in S^{n-1}} \|Ax\|_2 \le \varepsilon n^{-1/2} \Big\}.$$

In other words, we need to prove the lower bound $\|Ax\|_2 \gtrsim n^{-1/2}$ uniformly for all vectors $x \in S^{n-1}$, and with high probability.

Our starting point is the method developed in [13] for a similar invertibility problem for matrices A with all independent entries, see also [15]. We decompose the sphere $S^{n-1} = \mathrm{Comp} \cup \mathrm{Incomp}$ into the classes of compressible and incompressible vectors. A vector x is in Comp if x is within distance, say, 0.1 from the set of vectors of support $0.1n$. We seek to establish invertibility of A separately for the two classes, our goal being

$$\min_{x \in \mathrm{Comp}} \|Ax\|_2 \gtrsim n^{1/2}, \qquad \min_{x \in \mathrm{Incomp}} \|Ax\|_2 \gtrsim n^{-1/2}. \tag{1.11}$$

(The first estimate is even stronger than we need.) Each of the two classes, compressible and incompressible, has its own advantages.

Invertibility for compressible vectors. The class Comp has small metric entropy, which makes it amenable to covering arguments. This essentially reduces the invertibility problem for Comp to proving the lower bound $\|Ax\|_2 \gtrsim n^{1/2}$ with high probability for one (arbitrary) vector $x \in \text{Comp}$. If A had all independent entries (as in [13]) then we could express $\|Ax\|_2^2$ as a sum of independent random variables $\sum_{k=1}^n \langle A_k, x\rangle^2$ where A_k denote the rows of A, and finish by showing that each $\langle A_k, x\rangle$ is unlikely to be $o(1)$. But in our case, A is symmetric, so A_k are not independent. Nevertheless, we can extract from A a minor G with all independent entries. To this end, consider a subset $I \subset [n]$ with $|I| = \lambda n$ where $\lambda \in (0,1)$ is a small number. We decompose

$$A = \begin{pmatrix} D & G \\ G^* & E \end{pmatrix}, \quad x = \begin{pmatrix} y \\ z \end{pmatrix} \tag{1.12}$$

where D is a $I^c \times I^c$ matrix, G is a $I^c \times I$ matrix, $y \in I^c$, $z \in I$. Then $\|Ax\|_2 \geq \|Dy + Gz\|_2$. Conditioning on the entries in D and denoting the fixed vector $-Dy$ by v, we reduced the problem to showing that

$$\|Ax\|_2 \geq \|Gz - v\|_2 \gtrsim n^{1/2} \quad \text{with high probability.} \tag{1.13}$$

Now G is a matrix with all independent entries, so the previous reasoning yields (1.13) with probability at least $1 - 2e^{-cn}$. This establishes the first part of our goal (1.11), i.e. the good invertibility of A on the class of compressible vectors.

Concentration of quadratic forms. The second part of our goal (1.11) is more difficult. A very general observation from [13] reduces the invertibility problem for incompressible vectors to a *distance problem* for a random vector and a random hyperplane (Section 3.3). Specifically, we need to show that

$$\text{dist}(X_1, H_1) \gtrsim 1 \quad \text{with high probability,} \tag{1.14}$$

where X_1 denotes the first column of A and H_1 denotes the span of the other $n-1$ columns. An elementary observation (Proposition 5.1) is that

$$\text{dist}(A_1, H_1) = \frac{\left|\langle B^{-1}Z, Z\rangle - a_{11}\right|}{\sqrt{1 + \|B^{-1}Z\|_2^2}}, \quad \text{where } A = \begin{pmatrix} a_{11} & Z \\ Z^* & B \end{pmatrix}.$$

Obviously the random vector $Z \in \mathbb{R}^{n-1}$ and the $(n-1)\times(n-1)$ symmetric random matrix B are independent, and B has the same structure as A (its above-diagonal entries are independent). So lifting the problem back into dimension n, we arrive at the following problem for *quadratic forms.* Let X be a random vector in $\mathbb{R}^n$ with iid coordinates with mean zero and bounded fourth moment. Show that for every fixed $u \in \mathbb{R}$,

$$\left|\langle A^{-1}X, X\rangle - u\right| \gtrsim \|A^{-1}\|_{\text{HS}} \quad \text{with high probability,} \tag{1.15}$$

where $\|\cdot\|_{\mathrm{HS}}$ denotes the Hilbert-Schmidt norm. In other words, we need to show that the distribution of the quadratic form $\langle A^{-1}X, X\rangle$ is spread on the real line.

The spread of a general random variable S is measured by the *Lévy concentration function*

$$\mathcal{L}(S,\varepsilon) := \sup_{u\in\mathbb{R}} \mathbb{P}\big\{|S-u| \le \varepsilon\big\}, \quad \varepsilon \ge 0.$$

So our problem becomes to estimate Lévy concentration function of quadratic forms of the type $\langle A^{-1}X, X\rangle$ where A is a symmetric random matrix, and X is an independent random vector with iid coordinates.

Littlewood-Offord theory. A decoupling argument allows one to replace $\langle A^{-1}X, X\rangle$ by the bilinear form $\langle A^{-1}Y, X\rangle$ where Y is an independent copy of X. (This is an ideal situation; a realistic decoupling argument will incur some losses which we won't discuss here, see Section 8.2.) Using that $\mathbb{E}\|A^{-1}Y\|_2^2 = \|A^{-1}\|_{\mathrm{HS}}^2$, we reduce the problem to showing that for every $u \in \mathbb{R}$ one has

$$\big|\langle x_0, X\rangle - u\big| \gtrsim 1 \quad \text{with high probability,} \quad \text{where } x_0 = \frac{A^{-1}Y}{\|A^{-1}Y\|_2}. \tag{1.16}$$

By conditioning on A and X we can consider x_0 as a fixed vector. The product

$$S := \langle x_0, X\rangle = \sum_{k=1}^{n} x_0(k)X(k)$$

is a sum of independent random variables. So our problem reduces to estimating Lévy concentration function for general sums of independent random variables with given coefficients $x_0(k)$.

It turns out that the concentration function depends not only on the magnitude of the coefficients $x_0(k)$, but also on their *additive structure*. A vector x_0 with less 'commensurate' coefficients tends to produce better estimates for $\mathcal{L}(S,\varepsilon)$. Many researchers including Littlewood, Offord, Erdös, Moser, Sárközi, Szemerédi and Halasz produced initial findings of this type; Kahn, Komlós and Szemerédi [7] found applications to the invertibility problem for random matrices. Recently this phenomenon was termed the (inverse) *Littlewood-Offord theory* by Tao and Vu [19]. They initiated a systematic study of the effect the additive structure of the coefficient vector x_0 has on the concentration function; see a general discussion in [18, 15] with a view toward random matrix theory.

In [13, 14], Rudelson and the author of this paper proposed to quantify the amount of additive structure of a vector $x \in S^{n-1}$ by the *least common denominator (LCD)*; the version of LCD we use here (due to Rudelson) is

$$D(x) = \inf\Big\{\theta > 0 :\ \mathrm{dist}(\theta x, \mathbb{Z}^n) \lesssim \sqrt{\log_+ \theta}\Big\}. \tag{1.17}$$

The larger $D(x)$, the less structure x has, the smaller $\mathcal{L}(S,\varepsilon)$ is expected to be. Indeed, a variant of the Littlewood-Offord theory developed in [13, 14] states that

$$\mathcal{L}(S,\varepsilon) \lesssim \varepsilon + \frac{1}{D(x_0)}, \quad \varepsilon \geq 0. \tag{1.18}$$

The actual, more accurate, definition of LCD and the precise statement of (1.18) is given in Section 6.1.

Additive structure. In order to use Littlewood-Offord theory, one has to show that $D(x_0)$ is large for the vector x_0 in (1.16). This is the main difficulty in this paper, coming from the symmetry restrictions in the matrix A. We believe that the action of A^{-1} on an (arbitrary) vector Y should make the random vector x_0 completely unstructured, so it is plausible that $D(x_0) \geq e^{cn}$ with high probability, where $c > 0$ is a constant. If so, the singularity probability term in (1.3) would improve to e^{-cn}. Unfortunately, we can not even prove that $D(x_0) \geq e^{c\sqrt{n}}$.

The main losses occur in the process of decoupling and conditioning, which is performed to reduce the symmetric matrix A to a matrix with all independent entries. In order to resist such losses, we propose in this paper to work with an alternative (but essentially equivalent) robust version of LCD which we call the *regularized LCD*. It is designed to capture the most unstructured part of x of a given size. So, for a parameter $\lambda \in (0,1)$, we consider

$$\widehat{D}(x,\lambda) = \max\Big\{D\big(x_I/\|x_I\|_2\big) : I \subseteq [n],\, |I| = \lceil \lambda n \rceil\Big\} \tag{1.19}$$

where $x_I \in \mathbb{R}^I$ denotes the restriction of vector x onto the subset I. The actual, more accurate, definition of regularized LCD is given in Section 6.2.

On the one hand, if $\widehat{D}(x,\lambda)$ is large, then x has some unstructured part x_I, so we can still apply the linear Littlewood-Offord theory (restricted to I) to produce good bounds on the Lévy concentration function for linear forms (Proposition 6.9), and extend this for quadratic forms by decoupling. On the other hand, if $\widehat{D}(x,\lambda)$ is small, then not only x_I but *all* restrictions of x onto arbitrary $\lceil \lambda n \rceil$ coordinates are nicely structured, so in fact the entire x is highly structured. This yields a good control of the metric entropy of the set of vectors with small $\widehat{D}(x,\lambda)$. Ultimately, this approach (explained in more detail below) leads us to the desired *structure theorem*, which states that for $\lambda \geq n^{-c}$, one has

$$\widehat{D}(x_0,\lambda) \gtrsim n^{c/\lambda} \quad \text{with high probability.} \tag{1.20}$$

See Theorem 7.1 for the actual statement. In other words, the structure theorem that the regularized LCD is larger than any polynomial in n. As we explained, this estimate is then used in combination with the Littlewood-Offord theory (1.18) to deduce estimate (1.15) for quadratic forms (after optimization in λ);

see Theorem 8.1 for the actual result on concentration of quadratic forms. This in turn yields a solution of the distance problem (1.14), see Corollary 9.1. Ultimately, this solves the second part of invertibility problem (1.11), i.e. for the incompressible vectors, and completes the proof of Theorem 1.1.

The structure theorem. The proof of structure theorem (1.20) is the main technical ingredient of the paper. We shall explain heuristics of this argument in some more detail here. Let us condition on the independent vector Y in (1.16).

By definition of x_0, the vector Ax_0 is co-linear with the fixed vector Y, so (apart from the normalization issue, which we ignore now) we can assume that Ax_0 equals some fixed vector $u \in \mathbb{R}^n$. Then structure theorem (1.20) will follow if we can show that, with high probability, all vectors $x \in S^{n-1}$ with $\widehat{D}(x, \lambda) \ll n^{c/\lambda}$ satisfy $Ax \neq u$.

To this end, fix some value $D \ll n^{c/\lambda}$ and consider the level set

$$S_D = \big\{x \in S^{n-1} : \ \widehat{D}(x, \lambda) \sim D\big\}.$$

Our goal is to show that, with high probability, $Ax \neq u$ for all $x \in S_D$. This will be done by a covering argument.

First we show an individual estimate, that for an arbitrary given $x \in S_D$, $Ax \neq u$ with high probability. So let us fix $x \in S_D$ and assume that $Ax = u$. We choose the most unstructured subset of indices I of x, i.e. let I be the maximizing set in definition (1.19) of the regularized LCD. The decomposition $[n] = I^c \cup I$ induces the decomposition of matrix A we considered earlier in (1.12). Conditioning on the minor D, we estimate

$$0 = \|Ax - u\|_2 \geq \|Gz - v\|_2 = \sum_{k \in I^c} \big(\langle G_k, x_I\rangle - v_k\big)^2$$

where $v = (v_1, \dots, v_n)$ denotes some fixed vector (which depends on u the entries of D, which are now fixed), and G_k denote the rows of the minor G. It follows that $\langle G_k, x_I\rangle - v_k = 0$ for all $k \in I^c$. Since G has independent entries, the probability of these equalities can be estimated using a Littlewood-Offord estimate (1.18) as

$$\mathbb{P}\Big\{\langle G_k, x_I\rangle - v_k = 0\Big\} \lesssim \frac{1}{D(x_I)} \sim \frac{1}{\widehat{D}(x, \lambda)} \sim \frac{1}{D}, \quad k \in I^c.$$

Therefore, by independence we have

$$\mathbb{P}\big\{Ax = u\big\} \lesssim \Big(\frac{1}{D}\Big)^{|I^c|} = \Big(\frac{1}{D}\Big)^{n - \lambda n} \quad \text{for all } x \in S_D. \tag{1.21}$$

On the other hand, the level set S_D has small metric entropy. To see this, first consider the level set of the usual LCD in (1.17):

$$T_D = \big\{x \in S^{n-1} : \ D(x) \sim D\big\}.$$

Since the number of integer points in a Euclidean ball of radius D in $\mathbb{R}^n$ is about $(D/\sqrt{n})^n$, the definition of LCD implies that there exists an β-net $\mathcal{M}$ of T_D in the Euclidean metric with

$$\beta \sim \frac{\sqrt{\log D}}{D}, \quad |\mathcal{M}| \lesssim \Big(\frac{D}{\sqrt{n}}\Big)^n.$$

Now consider an arbitrary $x \in S_D$. By definition of the regularized LCD, the restriction x_I of any set I of λn coordinates has $D(x_I/\|x_I\|_2) \lesssim D$. So we can decompose $[n]$ into $1/\lambda$ sets of indices I_j, $|I_j| = \lambda n$, and for the restriction of x onto each I_j construct a β-net $\mathcal{M}_j$ in $\mathbb{R}^{I_j}$ with $|\mathcal{M}_j| \lesssim (D/\sqrt{\lambda n})^{\lambda n}$ as above. The product of these nets $\mathcal{M}_j$ obviously forms a $\beta/\sqrt{\lambda}$-net $\mathcal{N}$ of S_D with

$$|\mathcal{N}| \lesssim \Big(\Big(\frac{D}{\sqrt{\lambda n}}\Big)^{\lambda n}\Big)^{1/\lambda} = \Big(\frac{D}{\sqrt{\lambda n}}\Big)^n.$$

Finally, we take a union bound of probability estimates (1.21) over all x in the net $\mathcal{N}$ of S_D. This gives

$$\mathbb{P}\Big\{\exists x \in \mathcal{N} : \ Ax = u\Big\} \lesssim \Big(\frac{1}{D}\Big)^{n-\lambda n}\Big(\frac{D}{\sqrt{\lambda n}}\Big)^n = \Big(\frac{D^\lambda}{\sqrt{\lambda n}}\Big)^n.$$

Therefore, if $D \ll (\lambda n)^{2/\lambda}$ then the probability bound is exponentially small. An approximation argument (using the bound $\|A\| = O(\sqrt{n})$) extends this from the net $\mathcal{N}$ to the entire sub-level set S_D, and a simple union bound over all $D \ll (\lambda n)^{2/\lambda}$ finally yields

$$\mathbb{P}\Big\{\exists x \in S^{n-1}, \ \widehat{D}(x,\lambda) \ll (\lambda n)^{2/\lambda} : \ Ax = u\Big\} \lesssim e^{-n}.$$

As we said, this implies that with (exponentially) large probability, $\widehat{D}(x,\lambda) \gtrsim (\lambda n)^{2/\lambda}$, which is essentially the statement of structure theorem (1.20).

2 Notation and initial reductions of the problem

2.1 Notation

Throughout this paper $C, C_1, C_2, c, c_1, c_2, \ldots$ will denote positive constants. When it does not create confusion, the same letter (say, C) may denote different constants in different parts of the proof. The value of the constants may depend on some natural parameters such as the fourth moment of the entries of H, but it will never depend on the dimension n. Whenever possible, we will state which parameters the constant depends on.

The discrete interval is denoted $[n] = \{1, \ldots, n\}$. The logarithms $\log a$ are natural unless noted otherwise.

$\mathbb{P}\{\mathcal{E}\} = \mathbb{P}_{X,Y}\{\mathcal{E}\}$ stands for the probability of an event $\mathcal{E}$ that depends on the values of random variables, say, X and Y. Similarly, $\mathbb{E}f(X,Y) = \mathbb{E}_{X,Y}f(X,Y)$ stands for the expected value of a certain function $f(X,Y)$ of random variables X and Y.

For a vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, the Euclidean norm is $\|x\|_2 = \big(\sum_{k=1}^n |x_k|^2\big)^{1/2}$ and the sup-norm is $\|x\|_\infty = \max_k |x_k|$. The unit Euclidean sphere is $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ and the unit Euclidean ball is $B_2^n = \{x \in \mathbb{R}^n : \|x\|_2 \le 1\}$. The Euclidean distance from a point $x \in \mathbb{R}^n$ to a subset $D \subset \mathbb{R}^n$ is denoted $\mathrm{dist}(x,T) = \inf\{\|x-t\|_2 : t \in T\}$.

Consider a subset $I \subseteq [n]$. The unit Euclidean ball in $\mathbb{R}^I$ is denoted B_2^I. The orthogonal projection in $\mathbb{R}^n$ onto $\mathbb{R}^I$ is denoted $P_I : \mathbb{R}^n \to \mathbb{R}^n$. The restriction of a vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ onto the coordinates in I is denoted x_I. Thus $P_I x$ is a vector in $\mathbb{R}^n$ (with zero coordinates outside I), while $x_I = (x_k)_{k \in I}$ is a vector in $\mathbb{R}^I$.

Let A be an $n \times n$ symmetric matrix. The eigenvalues of A arranged in a non-decreasing order are denoted $\lambda_k(A)$. The spectral norm of A is

$$\max_k |\lambda_k(A)| = \max_{x \in S^{n-1}} \|Ax\|_2 = \|A\|. \tag{2.1}$$

The eigenvalue of the smallest magnitude determines the norm of the inverse:

$$\min_k |\lambda_k(A)| = \min_{x \in S^{n-1}} \|Ax\|_2 = 1/\|A^{-1}\|. \tag{2.2}$$

The transpose of A is denotes A^*. The Hilbert-Schmidt norm of A is denoted

$$\|A\|_{\mathrm{HS}} = \Big(\sum_{k=1}^n \lambda_k(A)^2\Big)^{1/2}.$$

2.2 Nets and bounds on the spectral norm

Consider a compact set $T \in \mathbb{R}^n$ and $\varepsilon > 0$. A subset $\mathcal{N} \subseteq T$ is called an ε-net of T if for every point $t \in T$ one has $\mathrm{dist}(t, \mathcal{N}) \le \varepsilon$. The minimal cardinality of an ε-net of T is called the *covering number* of T (for a given ε), and is denoted $N(T,\varepsilon)$. Equivalently, $N(T,\varepsilon)$ is the minimal number of closed Euclidean balls of radii ε and centered in points of T, whose union covers T.

Remark 2.1 (Centering). Suppose T can be covered with N balls of radii ε, but their centers are not necessarily in T. Then enlarging the radii by the factor of 2, we can place the centers in T. So $N(T, 2\varepsilon) \le N$.

Lemma 2.2 (See e.g. [24], Lemma 2). *For every subset $T \subseteq S^{n-1}$ and every $\varepsilon \in (0,1]$, one has*

$$N(T,\varepsilon) \le (3/\varepsilon)^n.$$

The following known lemma was used to deduce Theorem 1.2 for subgaussian matrices from our general result, Theorem 1.1.

Lemma 2.3 (Spectral norm: subgaussian)**.** *Let H be a symmetric random matrix as in Theorem 1.2. Then*

$$\mathbb{P}\Big\{\|H\| \le (C_{2.3}M + K)\sqrt{n}\Big\} \ge 1 - 2e^{-n},$$

where $C_{2.3}$ is an absolute constant.

Proof. Let us decompose the matrix as $H = D + B + B^*$ where D is the diagonal part of H, and B is the above-diagonal part of H. Since $\|D\| \le K\sqrt{n}$ by assumption and $\|B\| = \|B^*\|$, we have $\|H\| \le K\sqrt{n} + 2\|B\|$. Furthermore, since the entries of B on and below the diagonal are zero, all n^2 entries of B are independent mean zero random variables with subgaussian moments bounded by M. Proposition 2.4 of [15] then implies a required bound on $\|B\|$:

$$\mathbb{P}\big\{\|B\| \le CM\sqrt{n}\big\} \ge 1 - 2e^{-n},$$

where C is an absolute constant. This completes the proof. □

A similar spectral bound holds just under the fourth moment assumption, although only in expectation.

Lemma 2.4 (Spectral norm: four moments)**.** *Let H be a symmetric random matrix as in Theorem 1.5. Then*

$$\mathbb{E}\|H\| \le (C_{2.4}M_4 + K)\sqrt{n},$$

Here $C_{2.4}$ is an absolute constant.

Proof. We use the same decomposition $H = D + B + B^*$ as in the proof of Lemma 2.3. A result of Latala [8] implies that $\mathbb{E}\|B\| \le CM_4$ where C is an absolute constant. Thus

$$\mathbb{E}\|H\| \le \|D\| + 2\mathbb{E}\|B\| \le (K + 2CM_4)\sqrt{n}.$$

The lemma is proved. □

2.3 Initial reductions of the problem

We are going to prove Theorem 1.1. Without loss of generality, we can assume that $K \ge 1$ by increasing this value. Also we can assume that the constant c in this theorem is sufficiently small, depending on the value of the fourth moment and on K. Consequently, we can assume that $n \ge n_0$ where n_0 is a sufficiently large number that depends on the fourth moment and on K. (For $n < n_0$ the probability bound in (1.3) will be larger than 1, which is trivially true.) By a similar reasoning, we can assume that $\varepsilon \in (0, \varepsilon_0)$ for a sufficiently small number $\varepsilon_0 > 0$ which depends on the fourth moment and on K.

So we can assume that $K\sqrt{n} \geq \varepsilon n^{-1/2}$. Therefore, for $|z| > 2K\sqrt{n}$ the probability in question is automatically zero. So we can assume that $|z| \leq 2K\sqrt{n}$.

We shall work with the random matrix

$$A = H - zI.$$

If $\|H\| = \max_k |\lambda_k(H)| \leq K\sqrt{n}$ as in (1.3) then $\|A\| \leq \|H\| + |z| \leq 3K\sqrt{n}$. Therefore, the probability of the desired event in (1.3) is bounded above by

$$p := \mathbb{P}\Big\{ \min_k |\lambda_k(A)| \leq \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\}$$

where $\mathcal{E}_K$ denotes the event

$$\mathcal{E}_K = \big\{ \|A\| \leq 3K\sqrt{n} \big\}. \tag{2.3}$$

Using (2.2), we see that Theorem 1.1 would follow if we prove that

$$p := \mathbb{P}\Big\{ \min_{x \in S^{n-1}} \|Ax\|_2 \leq \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\} \leq C\varepsilon^{1/9} + 2e^{-n^c}. \tag{2.4}$$

We do this under the following assumptions on the random matrix A:

(A) $A = (a_{ij})$ is an $n \times n$ real symmetric matrix. The above-diagonal entries a_{ij}, $i < j$, are independent and identically distributed random variables with

$$\mathbb{E}a_{ij} = 0, \quad \mathbb{E}a_{ij}^2 = 1, \quad \mathbb{E}a_{ij}^4 \leq M_4^4 \quad \text{for } j > i, \tag{2.5}$$

where M_4 is some finite number. The diagonal entries a_{ii} are arbitrary fixed numbers.

The constants C and $c > 0$ in (2.4) will have to depend only on K and M_4.

By a small perturbation of the entries of A (e.g. adding independent normal random variables with zero means and small variances), we can assume that the distribution of the entries a_{ij} is absolutely continuous. In particular, the columns of A are in a general position almost surely. So the matrix A as well as all of its square minors are invertible almost surely; this allows us to ignore some technicalities that can arise in degenerate cases.

3 Preliminaries: small ball probabilities, compressible and incompressible vectors

In this section we recall some preliminary material from [13, 14].

3.1 Small ball probabilities, Lévy concentration function

Definition 3.1 (Small ball probabilities)**.** *Let Z be a random vector in $\mathbb{R}^n$. The* Lévy concentration function *of Z is defined as*

$$\mathcal{L}(Z,\varepsilon) = \sup_{u\in\mathbb{R}^n} \mathbb{P}\big\{\|Z-u\|_2 \le \varepsilon\big\}.$$

The Lévy concentration function bounds the *small ball probabilities* for Z, which are the probabilities that Z falls in a Euclidean ball of radius ε.

A simple but rather weak bound on Lévy concentration function follows from Paley-Zygmund inequality.

Lemma 3.2 ([14], Lemma 3.2)**.** *Let Z be a random variable with unit variance and with finite fourth moment, and put $M_4^4 := \mathbb{E}(Z-\mathbb{E}Z)^4$. Then for every $\varepsilon \in (0,1)$ there exists $p = p(M_4,\varepsilon) \in (0,1)$ such that*

$$\mathcal{L}(\xi,\varepsilon) \le p.$$

There has been a significant interest in bounding Lévy concentration function for sums of independent random variables; see [13, 14, 15] for discussion. The following simple but weak bound was essentially proved in [13], Lemma 2.6 (up to centering).

Lemma 3.3 (Lévy concentration function for sums)**.** *Let $\xi_1,\dots,\xi_n$ be independent random variables with unit variances and $\mathbb{E}(\xi_k-\mathbb{E}\xi_k)^4 \le M_4^4$, where M_4 is some finite number. Then for every $\varepsilon\in(0,1)$ there exists $p = p(M_4,\varepsilon)\in(0,1)$ such that the following holds.*

For every vector $x=(x_1,\dots,x_n)\in S^{n-1}$, the sum $S=\sum_{k=1}^n x_k\xi_k$ satisfies

$$\mathcal{L}(S,\varepsilon)\le p.$$

Proof. Clearly S has unit variance. Furthermore, since $S-\mathbb{E}S = \sum_{k=1}^n x_k(\xi_k - \mathbb{E}\xi_k)$, an application of Khinchine inequality yields

$$\mathbb{E}(S-\mathbb{E}S)^4 \le CM_4^4,$$

where C is an absolute constant (see [13], proof of Lemma 2.6). The desired concentration bound then follows from Lemma 3.2 with $Z = S-\mathbb{E}S$. $\square$

The following tensorization lemma can be used to transfer bounds for the Lévy concentration function from random variables to random vectors. This result follows from [13], Lemma 2.2 with $\xi_k = |x_k-u_k|$, where $u=(u_1,\dots,u_n)\in\mathbb{R}^n$.

Lemma 3.4 (Tensorization)**.** *Let $X=(X_1,\dots,X_n)$ be a random vector in $\mathbb{R}^n$ with independent coordinates X_k.*

1. Suppose there exists numbers $\varepsilon_0 \geq 0$ and $L \geq 0$ such that

$$\mathcal{L}(X_k, \varepsilon) \leq L\varepsilon \text{ for all } \varepsilon \geq \varepsilon_0 \text{ and all } k.$$

Then

$$\mathcal{L}(X, \varepsilon\sqrt{n}) \leq (C_{3.4} L\varepsilon)^n \text{ for all } \varepsilon \geq \varepsilon_0,$$

where $C_{3.4}$ is an absolute constant.

2. Suppose there exists numbers $\varepsilon > 0$ and $p \in (0,1)$ such that

$$\mathcal{L}(X_k, \varepsilon) \leq p \text{ for all } k.$$

There exists numbers $\varepsilon_1 = \varepsilon_1(\varepsilon, p) > 0$ and $p_1 = p_1(\varepsilon, p) \in (0,1)$ such that

$$\mathcal{L}(X, \varepsilon_1\sqrt{n}) \leq p_1^n.$$

Remark 3.5. A useful equivalent form of Lemma 3.4 (part 1) is the following one. Suppose there exist numbers $a, b \geq 0$ such that

$$\mathcal{L}(X_k, \varepsilon) \leq a\varepsilon + b \text{ for all } \varepsilon \geq 0 \text{ and all } k.$$

Then

$$\mathcal{L}(X, \varepsilon) \leq \big[C_{3.5} a\varepsilon + b)\big]^n \text{ for all } \varepsilon \geq 0,$$

where $C_{3.5}$ is an absolute constant.

3.2 Compressible and incompressible vectors

Let $c_0, c_1 \in (0,1)$ be two numbers. We will choose their values later as small constants that depend only on the parameters K and M_4 from (2.4) and **(A)**, see Remark 4.3 below.

Definition 3.6 ([14], Definition 2.4)**.** *A vector $x \in \mathbb{R}^n$ is called* sparse *if* $|\operatorname{supp}(x)| \leq c_0 n$. *A vector $x \in S^{n-1}$ is called* compressible *if x is within Euclidean distance c_1 from the set of all sparse vectors. A vector $x \in S^{n-1}$ is called* incompressible *if it is not compressible.*

The sets of compressible and incompressible vectors in S^{n-1} will be denoted by $\operatorname{Comp}(c_0, c_1)$ *and* $\operatorname{Incomp}(c_0, c_1)$ *respectively.*

The classes of compressible and incompressible vectors each have their own advantages. The set of compressible vectors has small covering numbers, which are exponential in $c_0 n$ rather than in n:

Lemma 3.7 (Covering compressible vectors)**.** *One has*

$$N\big(\operatorname{Comp}(c_0, c_1), 2c_1\big) \leq (9/c_0 c_1)^{c_0 n}.$$

Proof. Let $s = \lfloor c_0 n \rfloor$. By Lemma 2.2, the unit sphere S^{s-1} of $\mathbb{R}^s$ can be covered with at most $(3/c_1)^s$ Euclidean balls of radii c_1. Therefore, the set S of sparse vectors in $\mathbb{R}^n$ can be covered with at most $\binom{n}{s}(3/c_1)^s$ Euclidean balls of radii c_1 centered in S. Enlarging the radii of these balls we conclude that $\mathrm{Comp}(c_0, c_1)$ can be covered with at most $\binom{n}{s}(3/c_1)^s$ Euclidean balls of radii $2c_1$ centered in S. The conclusion of the lemma follows by estimating $\binom{n}{s} \le (en/s)^s$, which is a consequence of Stirling's approximation. □

The set of incompressible vectors have a different advantage. Each incompressible vector x has a set of coordinates of size proportional to n, whose magnitudes are all of the same order $n^{-1/2}$. We can say that an incompressible vector is *spread* over this set:

Lemma 3.8 (Incompressible vectors are spread, [13], Lemma 3.4)**.** *For every $x \in \mathrm{Incomp}(c_0, c_1)$, one has*

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

for at least $\frac{1}{2} c_0 c_1^2 n$ coordinates x_k of x.

Since S^{n-1} can be decomposed into two disjoint sets $\mathrm{Comp}(c_0, c_1)$ and $\mathrm{Incomp}(c_0, c_1)$, the problem of proving (2.4) reduces to establishing the good invertibility of the matrix A on these two classes separately:

$$\mathbb{P}\Big\{ \min_{x \in S^{n-1}} \|Ax\|_2 \le \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\} \le \mathbb{P}\Big\{ \inf_{x \in \mathrm{Comp}(c_0,c_1)} \|Ax\|_2 \le \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\} + \mathbb{P}\Big\{ \inf_{x \in \mathrm{Incomp}(c_0,c_1)} \|Ax\|_2 \le \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\}. \quad (3.1)$$

3.3 Invertibility for incompressible vectors via the distance problem

The first part of the invertibility problem (3.1), for compressible vectors, will be settled in Section 4. The second part, for incompressible vectors, quickly reduces to a *distance problem* for a random vector and a random hyperplane:

Lemma 3.9 (Invertibility via distance, [13], Lemma 3.5)**.** *Let A be any $n \times n$ random matrix. Let $A_1, \ldots, A_n$ denote the columns of A, and let H_k denote the span of all columns except the k-th. Then for every $c_0, c_1 \in (0,1)$ and every $\varepsilon \ge 0$, one has*

$$\mathbb{P}\Big\{ \inf_{x \in \mathrm{Incomp}(c_0,c_1)} \|Ax\|_2 \le \varepsilon n^{-1/2} \Big\} \le \frac{1}{c_0 n} \sum_{k=1}^{n} \mathbb{P}\big\{ \mathrm{dist}(A_k, H_k) \le c_1^{-1} \varepsilon \big\}. \quad (3.2)$$

This reduces our task to finding a lower bound for $\mathrm{dist}(A_k, H_k)$. This distance problem will be studied in the second half of the paper following Section 4.

Remark 3.10. Since the distribution of a random matrix A is completely general in Lemma 3.9, by conditioning on $\mathcal{E}_K$ we can replace the conclusion (3.2) by

$$\mathbb{P}\Big\{ \inf_{x \in \mathrm{Incomp}(c_0,c_1)} \|Ax\|_2 \le \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\} \le \frac{1}{c_0 n} \sum_{k=1}^{n} \mathbb{P}\big\{ \mathrm{dist}(A_k, H_k) \le c_1^{-1}\varepsilon \wedge \mathcal{E}_K \big\}.$$

4 Invertibility for compressible vectors

In this section we establish a uniform lower bound for $\|Ax\|_2$ on the set of compressible vectors x. This solves the first part of the invertibility problem in (3.1).

4.1 Small ball probabilities for Ax

We shall first find a lower bound for $\|Ax\|_2$ for a fixed vector x. We start with a very general estimate. It will be improved later to a finer result, Proposition 6.11, which will take into account the additive structure of x.

Proposition 4.1 (Small ball probabilities for Ax)**.** *Let A be a random matrix which satisfies* **(A)**. *Then for every $x \in S^{n-1}$, one has*

$$\mathcal{L}(Ax, c_{4.1}\sqrt{n}) \le 2e^{-c_{4.1}n}.$$

Here $c_{4.1} > 0$ depends only on the parameter M_4 from assumptions (2.5).

Proof. Our goal is to prove that, for an arbitrary fixed vector $u \in \mathbb{R}^n$, one has

$$\mathbb{P}\big\{ \|Ax - u\|_2^2 \le c_{4.1}^2 n \big\} \le 2e^{-c_{4.1}n}.$$

Let us decompose the set of indices $[n]$ into two sets of roughly equal sizes, $\{1, \ldots, n_0\}$ and $\{n_0+1, \ldots, n\}$ where $n_0 = \lceil n/2 \rceil$. This induces the decomposition of the matrix A and both vectors in question, which we denote

$$A = \begin{pmatrix} D & G \\ G^* & E \end{pmatrix}, \quad x = \begin{pmatrix} y \\ z \end{pmatrix}, \quad u = \begin{pmatrix} v \\ w \end{pmatrix}.$$

This way, we express

$$\|Ax - u\|_2^2 = \|Dy + Gz - v\|_2^2 + \|G^*y + Ez - w\|_2^2. \tag{4.1}$$

We shall estimate the two terms separately, using that each of the matrices G and G^* has independent entries.

We condition on an arbitrary realization of D and E, and we express

$$\|Dy + Gz - v\|_2^2 = \sum_{j=1}^{n} \big(\langle G_j, z\rangle - d_j\big)^2$$

where G_j denote the rows of G and d_j denote the coordinates of the fixed vector $Dy - v$. For each j, we observe that $\langle G_j, z\rangle = \sum_{i=n_0+1}^{n} a_{ij} x_i$ is a sum of independent random variables, and $\sum_{i=n_0+1}^{n} x_i^2 = \|z\|_2^2$. Therefore Lemma 3.3 can be applied to control the small ball probabilities as

$$\mathcal{L}\Big(\big\langle G_j, \frac{z}{\|z\|_2}\big\rangle, \frac{1}{2}\Big) \le c_3 \in (0,1)$$

where c_3 depends only on the parameter M_4 from assumptions (2.5).

Further, we apply Tensorization Lemma 3.4 (part 2) for the vector $Gz/\|z\|_2$ with coordinates $\langle G_j, z/\|z\|_2\rangle$, $j = 1, \ldots, n_0$. It follows that there exist numbers $c_2 > 0$ and $c_3 \in (0,1)$ that depend only on M_4 and such that

$$\mathcal{L}(Gz, c_2\|z\|_2\sqrt{n_0}) = \mathcal{L}(Gz/\|z\|_2, c_2\sqrt{n_0}) \le c_3^{n_0}.$$

Since $Dy - v$ is a fixed vector, this implies

$$\mathbb{P}\big\{\|Dy + Gz - v\|_2^2 \le c_2^2\,\|z\|_2^2\, n_0\big\} \le c_3^{n_0}. \tag{4.2}$$

Since this holds conditionally on an arbitrary realization of D, E, it also holds unconditionally.

By a similar argument we obtain that

$$\mathbb{P}\big\{\|G^* y + Ez - w\|_2^2 \le c_2^2\,\|y\|_2^2\,(n - n_0)\big\} \le c_3^{n-n_0}. \tag{4.3}$$

Since $n_0 \ge n/2$ and $n - n_0 \ge n/3$ and $\|y\|_2^2 + \|z\|_2^2 = \|x\|_2^2 = 1$, we have $c_2^2\,\|z\|_2^2\, n_0 + c_2^2\,\|y\|_2^2\,(n - n_0) > \frac{1}{3}c_2^2 n$. Therefore, by (4.1), the inequality $\|Ax - u\|_2^2 \le \frac{1}{3}c_2^2 n$ implies that either the event in (4.2) holds, or the event in (4.3) holds, or both. By the union bound, we conclude that

$$\mathbb{P}\Big\{\|Ax - u\|_2^2 \le \frac{1}{3}c_2^2 n\Big\} \le c_3^{n_0} + c_3^{n-n_0} \le 2c_3^{n/3}.$$

This completes the proof. $\square$

4.2 Small ball probabilities for Ax uniformly over compressible x

An approximation argument allows us to extend Proposition 4.1 to a uniform invertibility bound on the set of compressible vectors x uniformly. The following result gives a satisfactory answer for the first part of the invertibility problem in (3.1), i.e. for the set of compressible vectors. We shall state a somewhat stronger result that is needed at this moment; the stronger form will be useful later in the proof of Lemma 7.2.

Proposition 4.2 (Small ball probabilities for compressible vectors)**.** *Let A be an $n \times n$ random matrix which satisfies* **(A)***, and let $K \geq 1$. There exist $c_0, c_1, c_{4.2} \in (0,1)$ that depend only on K and M_4 from assumptions* (2.3)*,* (2.5)*, and such that the following holds. For every $u \in \mathbb{R}^n$, one has*

$$\mathbb{P}\Big\{ \inf_{\frac{x}{\|x\|_2} \in \mathrm{Comp}(c_0,c_1)} \|Ax - u\|_2 / \|x\|_2 \leq c_{4.2}\sqrt{n} \wedge \mathcal{E}_K \Big\} \leq 2e^{-c_{4.2}n}. \tag{4.4}$$

Proof. Let us fix some small values of c_0, c_1 and $c_{4.2}$; the precise choice will be made shortly. According to Lemma 3.7, there exists a $(2c_1)$-net $\mathcal{N}$ of the set $\mathrm{Comp}(c_0, c_1)$ such that

$$|\mathcal{N}| \leq (9/c_0c_1)^{c_0 n}. \tag{4.5}$$

Let $\mathcal{E}$ denote the event in the left hand side of (4.4) whose probability we would like to bound. Assume that $\mathcal{E}$ holds. Then there exist vectors $x_0 := x/\|x\|_2 \in \mathrm{Comp}(c_0, c_1)$ and $u_0 := u/\|x\|_2 \in \mathrm{span}(u)$ such that

$$\|Ax_0 - u_0\|_2 \leq c_{4.2}\sqrt{n}. \tag{4.6}$$

By the definition of $\mathcal{N}$, there exists $y_0 \in \mathcal{N}$ such that

$$\|x_0 - y_0\|_2 \leq 2c_1. \tag{4.7}$$

On the one hand, by definition (2.3) of event $\mathcal{E}_K$, we have

$$\|Ay_0\|_2 \leq \|A\| \leq 3K\sqrt{n}. \tag{4.8}$$

On the other hand, it follows from (4.6) and (4.7) that

$$\|Ay_0 - u_0\|_2 \leq \|A\|\|x_0 - y_0\|_2 + \|Ax_0 - u_0\|_2 \leq 6c_1K\sqrt{n} + c_{4.2}\sqrt{n}. \tag{4.9}$$

This and (4.8) yield that

$$\|u_0\|_2 \leq 3K\sqrt{n} + 6c_1K\sqrt{n} + c_{4.2}\sqrt{n} \leq 10K\sqrt{n}.$$

So, we see that

$$u_0 \in \mathrm{span}(u) \cap 10K\sqrt{n}B_2^n =: E.$$

Let $\mathcal{M}$ be some fixed $(c_1K\sqrt{n})$-net of the interval E, such that

$$|\mathcal{M}| \leq \frac{20K\sqrt{n}}{c_1K\sqrt{n}} = \frac{20}{c_1}. \tag{4.10}$$

Let us choose a vector $v_0 \in \mathcal{M}$ such that $\|u_0 - v_0\|_2 \leq c_1K\sqrt{n}$. It follows from (4.9) that

$$\|Ay_0 - v_0\|_2 \leq 6c_1K\sqrt{n} + c_{4.2}\sqrt{n} + c_1K\sqrt{n} \leq (7c_1K + c_{4.2})\sqrt{n}.$$

Choose values of $c_1, c_{4.2} \in (0,1)$ so that $7c_1K + c_{4.2} \le c_{4.1}$, where $c_{4.1}$ is the constant from Proposition 4.1.

Summarizing, we have shown that the event $\mathcal{E}$ implies the existence of vectors $y_0 \in \mathcal{N}$ and $v_0 \in \mathcal{M}$ such that $\|Ay_0 - v_0\|_2 \le c_{4.1}\sqrt{n}$. Taking the union bound over $\mathcal{N}$ and $\mathcal{M}$, we conclude that

$$\mathbb{P}(\mathcal{E}) \le |\mathcal{N}| \cdot |\mathcal{M}| \max_{y_0 \in \mathcal{N},\, v_0 \in \mathcal{M}} \mathbb{P}\big\{ \|Ay_0 - v_0\|_2 \le c_{4.1}\sqrt{n} \big\}.$$

Applying Proposition 4.1 and using the estimates (4.5), (4.10) on the cardinalities of the nets, we obtain

$$\mathbb{P}(\mathcal{E}) \le \Big(\frac{9}{c_0 c_1}\Big)^{c_0 n} \cdot \frac{20}{c_1} \cdot 2e^{-c_{4.1}n}.$$

Choosing $c_0 > 0$ small enough depending on c_1 and $c_{4.1}$, we can ensure that

$$\mathbb{P}(\mathcal{E}) \le 2e^{-c_{4.1}n/2}$$

as required. This completes the proof. $\square$

As an immediate consequence of Proposition 4.2, we obtain a very good bound for the first half of the invertibility problem in (3.1). Indeed, since $\varepsilon n^{-1/2} \le c_{4.2}\sqrt{n}$, we have

$$\mathbb{P}\Big\{ \inf_{x \in \mathrm{Comp}(c_0,c_1)} \|Ax\|_2 \le \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\} \le 2e^{-c_{4.2}n}. \tag{4.11}$$

Remark 4.3 (Fixing c_0, c_1). At this point we fix some values $c_0 = c_0(K, M_4)$ and $c_1 = c_1(K, M_4)$ satisfying Proposition 4.2, for the rest of the argument.

5 Distance problem via small ball probabilities for quadratic forms

The second part of the invertibility problem in (3.1) – the one for for incompressible vectors – is more difficult. Recall that Lemma 3.9 reduces the invertibility problem to the distance problem, namely to an upper bound on the probability

$$\mathbb{P}\big\{ \mathrm{dist}(A_1, H_1) \le \varepsilon \big\}$$

where A_1 is the first column of A and H_1 is the span of the other columns. (By a permutation of the indices in $[n]$, the same bound would hold for all $\mathrm{dist}(A_k, H_k)$ as required in Lemma 3.9.)

The following proposition reduces the distance problem to the small ball probability for quadratic forms of random variables:

Proposition 5.1 (Distance problems via quadratic forms)**.** *Let $A = (a_{ij})$ be an arbitrary $n \times n$ matrix. Let A_1 denote the first column of A and H_1 denote the span of the other columns. Furthermore, let B denote the $(n-1) \times (n-1)$ minor of A obtained by removing the first row and the first column from A, and let $X \in \mathbb{R}^{n-1}$ denote the first column of A with the first entry removed. Then*

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

Proof. Let $h \in S^{n-1}$ denote a normal to the hyperplane H_1; choose the sign of the normal arbitrarily. We decompose

$$A = \begin{pmatrix} a_{11} & X^* \\ X & B \end{pmatrix}, \quad A_1 = \begin{pmatrix} a_{11} \\ X \end{pmatrix}, \quad h = \begin{pmatrix} h_1 \\ g \end{pmatrix},$$

where $h_1 \in \mathbb{R}$ and $g \in \mathbb{R}^{n-1}$. Then

$$\operatorname{dist}(A_1, H_1) = |\langle A_1, h\rangle| = |a_{11}h_1 + \langle X, g\rangle|. \tag{5.1}$$

Since h is orthogonal to the columns of the matrix $\binom{X^*}{B}$, we have

$$0 = \begin{pmatrix} X^* \\ B \end{pmatrix}^* h = h_1 X + Bg,$$

so

$$g = -h_1 B^{-1} X. \tag{5.2}$$

Furthermore,

$$1 = \|h\|_2^2 = h_1^2 + \|g\|_2^2 = h_1^2 + h_1^2 \|B^{-1}X\|_2^2.$$

Hence

$$h_1^2 = \frac{1}{1 + \|B^{-1}X\|_2^2}. \tag{5.3}$$

So, using (5.2) and (5.3), we can express the distance in (5.1) as

$$\operatorname{dist}(A_1, H_1) = \left|a_{11}h_1 - \langle h_1 B^{-1}X, X\rangle\right| = \frac{\left|\langle B^{-1}X, X\rangle - a_{11}\right|}{\sqrt{1 + \|B^{-1}X\|_2^2}}.$$

This completes the proof. □

Remark 5.2 (A versus B). Let us apply Proposition 5.1 to the $n \times n$ random matrix A which satisfies assumptions **(A)**. Recall that a_{11} is a fixed number, so the problem reduces to estimating the small ball probabilities for the quadratic form $\langle B^{-1}X, X\rangle$. Observe that X is a random vector that is independent of B, and whose entries satisfy the familiar moment assumptions (2.5).

The random matrix B has the same structure as A except it is $(n-1) \times (n-1)$ rather than $n \times n$. For this reason, it will be convenient to develop the theory in dimension n, that is for the quadratic forms $\langle A^{-1}X, X\rangle$, where X is an independent random vector. At the end, the theory will be applied in dimension $n-1$ for the matrix B.

6 Small ball probabilities for quadratic forms via additive structure

In order to produce good bounds (super-polynomial) for the small ball probabilities for the quadratic forms $\langle A^{-1}X, X\rangle$, we will have to take into account the additive structure of the vector $A^{-1}X$. Let us first review the corresponding theory for linear forms, which is sometimes called the *Littlewood-Offord theory*. We will later extend it (by decoupling) to quadratic forms.

6.1 Small ball probabilities via LCD

The linear Littlewood-Offord theory concerns the small ball probabilities for the sums of the form $\sum x_k \xi_k$ where ξ_k are identically distributed independent random variables, and $x = (x_1, \ldots, x_n) \in S^{n-1}$ is a given coefficient vector. Lemma 3.3 gives a general bound on the concentration function, $\mathcal{L}(S, \varepsilon) \le p$. But this bound is too weak – it produces a fixed probability p for all ε, even when ε approaches zero. Finer estimates are not possible for general sums; for example, the sum $S = \pm 1 \pm 1$ with random independent signs equals zero with fixed probability $1/2$. Nevertheless, one can break the barrier of fixed probability by taking into account the additive structure in the coefficient vector x.

The amount of additive structure in $x \in \mathbb{R}^n$ is captured by the *least common denominator* (LCD) of x. If the coordinates $x_k = p_k/q_k$ are rational numbers, then a suitable measure of additive structure in x is the least denominator $D(x)$ of these ratios, which is the common multiple of the integers q_k. Equivalently, $D(x)$ the smallest number $\theta > 0$ such that $\theta x \in \mathbb{Z}^n$. An extension of this concept for general vectors with real coefficients was developed in [13, 14], see also [15]; the particular form of this concept we shall use here is proposed by M. Rudelson (unpublished).

Definition 6.1 (LCD). *Let $L \ge 1$. We define the* least common denominator (LCD) *of $x \in S^{n-1}$ as*

$$D_L(x) = \inf\Big\{\theta > 0 :\ \mathrm{dist}(\theta x, \mathbb{Z}^n) < L\sqrt{\log_+(\theta/L)}\Big\}.$$

If the vector x is considered in $\mathbb{R}^I$ for some subset $I \subseteq [n]$, then in this definition we replace $\mathbb{Z}^n$ by $\mathbb{Z}^I$.

Clearly, one always has $D_L(x) > L$. A more sensitive but still quite simple bound is the following one:

Lemma 6.2. *For every $x \in S^{n-1}$ and every $L \ge 1$, one has*

$$D_L(x) \ge \frac{1}{2\|x\|_\infty}.$$

Proof. Let $\theta := D_L(x)$, and assume that $\theta < \frac{1}{2\|x\|_\infty}$. Then $\|\theta x\|_\infty < 1/2$. Therefore, by looking at the coordinates of the vector θx one sees that the vector $p \in \mathbb{Z}^n$ that minimizes $\|\theta x - p\|_2$ is $p = 0$. So

$$\operatorname{dist}(\theta x, \mathbb{Z}^n) = \|\theta x\|_2 = \theta.$$

On the other hand, by the definition of LCD, we have

$$\operatorname{dist}(\theta x, \mathbb{Z}^n) \leq L\sqrt{\log_+(\theta/L)}.$$

However, the inequality $\theta \leq L\sqrt{\log_+(\theta/L)}$ has no solutions in $\theta \geq 0$. This contradiction completes the proos. $\square$

The goal of our variant of Littlewood-Offord theory is to express the small ball probabilities of sums $\mathcal{L}(S, \varepsilon)$ in terms of $D(x)$. This is done in the following theorem, which is a version of results from [13, 14]; this particular simplified form is close to the form put forth by M. Rudelson (unpublished).

Theorem 6.3 (Small ball probabilities via LCD)**.** *Let $\xi_1, \ldots, \xi_n$ be independent and identically distributed random variables. Assume that there exist numbers $\varepsilon_0, p_0, M_1 > 0$ such that $\mathcal{L}(\xi_k, \varepsilon_0) \leq 1 - p_0$ and $\mathbb{E}|\xi_k| \leq M_1$ for all k. Then there exists $C_{6.3}$ which depends only on ε_0, p_0 and M_1, and such that the following holds. Let $x \in S^{n-1}$ and consider the sum $S = \sum_{k=1}^n x_k \xi_k$. Then for every $L \geq p_0^{-1/2}$ and $\varepsilon \geq 0$ one has*

$$\mathcal{L}(S, \varepsilon) \leq C_{6.3} L\Big(\varepsilon + \frac{1}{D_L(x)}\Big).$$

The proof of Theorem 6.3 is based on Esseen's Lemma, see e.g. [17], p. 290.

Lemma 6.4 (C.-G. Esseen)**.** *Let Y be a random variable. Then*

$$\mathcal{L}(Y, 1) \leq C_{6.4} \int_{-1}^{1} |\phi_Y(\theta)| \, d\theta$$

where $\phi_Y(\theta) = \mathbb{E}\exp(2\pi i \theta Y)$ is the characteristic function of Y, and $C_{6.4}$ is an absolute constant.

Proof of Theorem 6.3. By replacing ξ_k with ξ_k/ε_0, we can assume without loss of generality that $\varepsilon_0 = 1$. We apply Esseen's Lemma 6.4 for $Y = S/\varepsilon$. Using independence of ξ_k, we obtain

$$\mathcal{L}(S, \varepsilon) \leq C_{6.4} \int_{-1}^{1} \prod_{k=1}^{n} \Big|\phi\Big(\frac{\theta x_k}{\varepsilon}\Big)\Big| \, d\theta, \tag{6.1}$$

where $\phi(t) = \mathbb{E}\exp(2\pi i t\xi)$ is the characteristic function of $\xi := \xi_1$.

We proceed with a conditioning argument similar to the ones used in [12, 13, 14]. Let ξ' denote an independent copy of ξ, and let $\bar{\xi} = \xi - \xi'$; then $\bar{\xi}$ is a symmetric random variable. By symmetry, we have

$$|\phi(t)|^2 = \mathbb{E}\exp(2\pi i t\bar{\xi}) = \mathbb{E}\cos(2\pi t\bar{\xi}).$$

Using the inequality $|x| \le \exp\big[-\frac{1}{2}(1-x^2)\big]$ which is valid for all $x \in \mathbb{R}$, we obtain

$$|\phi(t)| \le \exp\Big[-\frac{1}{2}\big(1 - \mathbb{E}\cos(2\pi t\bar{\xi})\big)\Big]. \tag{6.2}$$

By assumption, we have $\mathcal{L}(\xi, 1) \le 1 - p_0$. Conditioning on $\bar{\xi}$ we see that $\mathbb{P}\{|\bar{\xi}| \ge 1\} \ge p_0$. Furthermore, another assumption of the theorem implies that $\mathbb{E}|\bar{\xi}| \le 2\mathbb{E}|\xi| \le 2M_1$. Using Markov's inequality, we conclude that $\mathbb{P}\{|\bar{\xi}| \ge 4M_1/p_0\} \le p_0/2$. Combining the two probability bounds, we see that the event

$$\mathcal{E} := \big\{1 \le |\bar{\xi}| \le C_0\big\} \quad \text{satisfies} \quad \mathbb{P}\{\mathcal{E}\} \ge p_0/2, \quad \text{where} \quad C_0 := \frac{4M_1}{p_0}.$$

We then estimate the expectation appearing in (6.2) by conditioning on $\mathcal{E}$:

$$\begin{aligned} 1 - \mathbb{E}\cos(2\pi t\bar{\xi}) &\ge \mathbb{P}\{\mathcal{E}\} \cdot \mathbb{E}\big[1 - \cos(2\pi t\bar{\xi}) \,\big|\, \mathcal{E}\big] \\ &\ge \frac{p_0}{2} \cdot \mathbb{E}\Big[\frac{4}{\pi^2}\min_{q\in\mathbb{Z}} |2\pi t\bar{\xi} - 2\pi q|^2 \,\Big|\, \mathcal{E}\Big] \\ &= 8p_0\, \mathbb{E}\Big[\min_{q\in\mathbb{Z}} |t\bar{\xi} - q|^2 \,\Big|\, \mathcal{E}\Big]. \end{aligned}$$

Substituting this into (6.2) and then into (6.1), and using Jensen's inequality, we obtain

$$\begin{aligned} \mathcal{L}(S,\varepsilon) &\le C_{6.4} \int_{-1}^{1} \exp\Big(-4p_0 \mathbb{E}\Big[\min_{q_k\in\mathbb{Z}} \sum_{k=1}^n \Big|\frac{\bar{\xi}\theta}{\varepsilon}x_k - q_k\Big|^2 \,\Big|\, \mathcal{E}\Big]\Big)\, d\theta \\ &\le C_{6.4}\, \mathbb{E}\Big[\int_{-1}^{1} \exp\Big(-4p_0 \operatorname{dist}\Big(\frac{\bar{\xi}\theta}{\varepsilon}x, \mathbb{Z}^n\Big)^2\Big)\, d\theta \,\Big|\, \mathcal{E}\Big]. \end{aligned}$$

Since the integrand is an even function of θ, we can integrate over $[0,1]$ instead of $[-1,1]$ at the cost of an extra factor of 2. Also, replacing the expectation by the maximum and using the definition of the event $\mathcal{E}$, we obtain

$$\mathcal{L}(S,\varepsilon) \le 2C_{6.4} \sup_{1\le z\le C_0} \int_0^1 \exp\big(-4p_0 f_z^2(\theta)\big)\, d\theta \tag{6.3}$$

where

$$f_z(\theta) = \operatorname{dist}\Big(\frac{z\theta}{\varepsilon}x, \mathbb{Z}^n\Big).$$

Suppose that

$$\varepsilon > \varepsilon_0 := \frac{C_0}{D_L(x)}.$$

Then, for every $1 \le z \le C_0$ and every $\theta \in [0,1]$, we have $\frac{z\theta}{\varepsilon} < D_L(x)$. By the definition of $D_L(x)$, this means that

$$f_z(\theta) = \operatorname{dist}\Big(\frac{z\theta}{\varepsilon}x, \mathbb{Z}^n\Big) \ge L\sqrt{\log_+\Big(\frac{z\theta}{\varepsilon L}\Big)}.$$

Putting this estimate back into (6.3), we obtain

$$\mathcal{L}(S,\varepsilon) \le 2C_{6.4} \sup_{z\ge 1} \int_0^1 \exp\Big(-4p_0 L^2 \log_+\Big(\frac{z\theta}{\varepsilon L}\Big)\Big)\, d\theta.$$

After change of variable $t = \frac{z\theta}{\varepsilon L}$ and using that $z \ge 1$ we have

$$\mathcal{L}(S,\varepsilon) \le 2C_{6.4}L\varepsilon \int_0^\infty \exp\big(-4p_0L^2\log_+ t\big)\, dt = 2C_{6.4}L\varepsilon\Big(1 + \int_1^\infty t^{-4p_0L^2}\, dt\Big).$$

Since $p_0L^2 \ge 1$ by assumption, the integral in the right hand side is bounded by an absolute constant, so

$$\mathcal{L}(S,\varepsilon) \le C_1 L\varepsilon$$

where C_1 is an absolute constant.

Finally, suppose that $\varepsilon \le \varepsilon_0$. Applying the previous part for $2\varepsilon_0$, we get

$$\mathcal{L}(S,\varepsilon) \le \mathcal{L}(S,2\varepsilon_0) \le 2C_1L\varepsilon_0 = \frac{2C_1C_0L}{D_L(x)}.$$

This completes the proof of Theorem 6.3. □

Remark 6.5. For a general, not necessarily unit vector $x \in \mathbb{R}^n$, the conclusion of Theorem 6.3 reads as

$$\mathcal{L}(S,\varepsilon) = \mathcal{L}\Big(\frac{S}{\|x\|_2}, \frac{\varepsilon}{\|x\|_2}\Big) \le C_{6.3}L\Big(\frac{\varepsilon}{\|x\|_2} + \frac{1}{D_L(x/\|x\|_2)}\Big).$$

6.2 Regularized LCD

As we saw in Proposition 5.1, the distance problem reduces to a quadratic Littlewood-Offord problem, for quadratic forms of the type $\sum_{ij} x_{ij}\xi_i\xi_j$. We will seek to reduce the quadratic problem to a linear one by decoupling and conditioning arguments. This process requires a more robust version of the concept of the LCD, which we develop now.

Let $x \in \mathrm{Incomp}(c_0, c_1)$; recall that we have fixed the values $c_0 = c_0(K, M_4)$, $c_1 = c_1(K, M_4)$ in Remark 4.3. By Lemma 3.8, at least $\frac{1}{2}c_0c_1^2n$ coordinates x_k of x satisfy

$$\frac{c_1}{\sqrt{2n}} \le |x_k| \le \frac{1}{\sqrt{c_0 n}}. \tag{6.4}$$

Let us fix some constant c_{oo} such that

$$\frac{1}{4}c_0c_1^2 \le c_{oo} \le \frac{1}{4};$$

we can make the value of c_{oo} depend only on c_0 and c_1 (hence only on parameters K and M_4). Then for every vector $x \in \mathrm{Incomp}(c_0, c_1)$ we can assign a subset called $\mathrm{spread}(x) \subseteq [n]$ so that

$$|\,\mathrm{spread}(x)| = \lceil c_{oo} n \rceil$$

and so that (6.4) holds for all $k \in \mathrm{spread}(x)$.

The point here is that not all of the coordinates x_k satisfying (6.4) will be good in the future; the set $\mathrm{spread}(x)$ will allow us to include only the good ones. At this point, we consider an arbitrary valid assignment of $\mathrm{spread}(x)$ to x; the particular choice of the assignment will be determined later.

Our new version of LCD is designed to capture the amount of structure in the *least structured* part of the coefficients of x.

Definition 6.6 (Regularized LCD)**.** *Let $\lambda \in (0, c_{oo})$ and $L \ge 1$. We define the* regularized LCD *of a vector $x \in \mathrm{Incomp}(c_0, c_1)$ as*

$$\widehat{D}_L(x, \lambda) = \max\Big\{D_L\big(x_I/\|x_I\|_2\big) \,:\, I \subseteq \mathrm{spread}(x),\ |I| = \lceil \lambda n \rceil\Big\}.$$

Denote by $I(x)$ the maximizing set I in this definition.

Remark 6.7. Since the sets I in this definition are subsets of $\mathrm{spread}(x)$, inequalities (6.4) imply that

$$c_{6.7}\sqrt{\lambda} \le \|x_I\|_2 \le C_{6.7}\sqrt{\lambda}$$

where $c_{6.7} = c_1/\sqrt{2}$ and $C_{6.7} = 1/\sqrt{c_0}$.

Lemma 6.8. *For every $x \in \mathrm{Incomp}(c_0, c_1)$ and every $\lambda \in (0, c_{oo})$ and $L \ge 1$, one has*

$$\widehat{D}_L(x, \lambda) \ge c_{6.8}\sqrt{\lambda n}.$$

Here $c_{6.8} \in (0, 1)$ depends only on c_0 and c_1.

Proof. Consider a subset I as in the definition of $\widehat{D}_L(x,\lambda)$. Denote $z_I := x_I/\|x_I\|_2$. By (6.4) and Remark 6.7, we have $\|z_I\|_\infty \le C/\sqrt{\lambda n}$ where $C \in (0,1)$ depends only on c_0 and c_1. Then Lemma 6.2 implies that

$$D_L(z_I) \ge \frac{1}{2C}\sqrt{\lambda n}.$$

By the definition of $\widehat{D}_L(x,\lambda)$, the proof is complete. □

Now we state a version of Theorem 6.3 for the regularized LCD.

Proposition 6.9 (Small ball probabilities via regularized LCD)**.** *Let $\xi_1,\dots,\xi_n$ be independent and identically distributed random variables. Assume that there exist numbers $\varepsilon_0, p_0 > 0$ such that $\mathcal{L}(\xi_k,\varepsilon_0) \le 1 - p_0$ and $\mathbb{E}|\xi_k| \le M_1$ for all k. Then there exist $C_{6.9}$ which depends only on ε_0, p_0, and M_1, and such that the following holds.*

Consider a vector $x \in \mathrm{Incomp}(c_0,c_1)$ and a subset $J \subseteq [n]$ such that $J \supseteq I(x)$. Consider also the sum $S_J = \sum_{k\in J} x_k\xi_k$. Then for every $\lambda \in (0, c_{oo})$, $L \ge p_0^{-1/2}$ and $\varepsilon \ge 0$, one has

$$\mathcal{L}(S_J,\varepsilon) \le C_{6.9} L\Big(\frac{\varepsilon}{\sqrt{\lambda}} + \frac{1}{\widehat{D}_L(x,\lambda)}\Big).$$

Proof. Note that for every two sets $I \subseteq J \subseteq [n]$, the corresponding sums satisfy $\mathcal{L}(S_J,\varepsilon) \le \mathcal{L}(S_I,\varepsilon)$; this follows by conditioning on the random variables ξ_k with $k \in J \setminus I$. Applying this relation for $I := I(x) \subseteq J$, we obtain

$$\begin{aligned}\mathcal{L}(S_J,\varepsilon) \le \mathcal{L}(S_I,\varepsilon) &\le C_{6.3} L\Big(\frac{\varepsilon}{\|x_I\|_2} + \frac{1}{D_L(x_I/\|x_I\|_2)}\Big) \qquad \text{(by Remark 6.5)}\\ &\le C_{6.3} L\Big(\frac{\varepsilon}{c_{6.7}\sqrt{\lambda}} + \frac{1}{\widehat{D}_L(x,\lambda)}\Big) \qquad \text{(by Remark 6.7).}\end{aligned}$$

This completes the proof. □

Remark 6.10. By Lemma 3.2, both Theorem 6.3 and Proposition 6.9 can be applied for arbitrary independent and identically distributed random variables $\xi_1,\dots,\xi_n$ that have unit variance and finite fourth moment. In particular, Theorem 6.3 and Proposition 6.9 apply if ξ_k satisfy the same moment assumptions (2.5) as the entries a_{ij} of A. The constants $C_{6.3}$ and $C_{6.9}$ in this case depends only on the fourth moment parameter M_4 from the assumptions **(A)** on the random matrix A.

6.3 Small ball probabilities for Ax via regularized LCD

We will now develop a refinement of Proposition 4.1 that is sensitive to the additive structure of the vector x.

Proposition 6.11 (Small ball probabilities for Ax via regularized LCD)**.** *Let A be a random matrix which satisfies* **(A)***. Let $x \in \mathrm{Incomp}(c_0, c_1)$ and $\lambda \in (0, c_{oo})$. Then for every $L \geq L_0$ and $\varepsilon \geq 0$, one has*

$$\mathcal{L}(Ax, \varepsilon\sqrt{n}) \leq \left[\frac{C_{6.11} L\varepsilon}{\sqrt{\lambda}} + \frac{C_{6.11} L}{\widehat{D}_L(x, \lambda)} \right]^{n - \lceil \lambda n \rceil}.$$

Here $C_{6.11}$ and L_0 depend only on the parameters K and M_4 from assumptions (2.3)*,* (2.5)*.*

Proof. Our goal is to bound above the probability

$$\mathbb{P}\big\{ \|Ax - u\|_2 \leq \varepsilon\sqrt{n} \big\}$$

for an arbitrary fixed vector $u \in \mathbb{R}^n$.

Let $I = I(x)$ be the maximizing set from the definition of $\widehat{D}_L(x, \lambda)$. We decompose the set of indices $[n]$ into sets $I \cup I^c$ similarly to how we did it in the proof of Proposition 4.1. This induces the decomposition of the matrix A and both vectors in question, which we denote

$$A = \begin{pmatrix} D & G \\ G^* & E \end{pmatrix}, \quad x = \begin{pmatrix} y \\ z \end{pmatrix}, \quad u = \begin{pmatrix} v \\ w \end{pmatrix},$$

where D is a $I^c \times I^c$ matrix, G is a $I^c \times I$ matrix, $y, v \in \mathbb{R}^{I^c}$ and $z, w \in \mathbb{R}^I$. This way, we express

$$\|Ax - u\|_2^2 = \|Dy + Gz - v\|_2^2 + \|G^* y + Ez - w\|_2^2.$$

Let us condition on an arbitrary realization of the minors D and E. Denoting $u_0 := v - Dy$, we have

$$\|Ax - u\|_2 \geq \|Gz - u_0\|_2.$$

We will use the crucial facts that G is a $I^c \times I$ matrix with independent entries, and u_0 is a fixed vector in $\mathbb{R}^{I^c}$. The i-th coordinate of the vector $Gz \in \mathbb{R}^{I^c}$ is

$$(Gz)_i = \sum_{j \in I} a_{ij} x_j, \quad i \in I^c.$$

All random variables a_{ij} here are independent. So we can apply Proposition 6.9 with $J = I = I(x)$ (see Remark 6.10), and we obtain

$$\mathcal{L}\big((Gz)_i, \varepsilon\big) \leq C_{6.9} L \Big(\frac{\varepsilon}{\sqrt{\lambda}} + \frac{1}{\widehat{D}_L(x, \lambda)} \Big), \quad i \in I^c.$$

Since the coordinates $(Gz)_i$ of the random vector Gz are independent, Tensorization Lemma 3.4 (see Remark 3.5) implies that

$$\mathcal{L}\big(Gz, \varepsilon\sqrt{|I^c|}\big) \le \left[\frac{CL\varepsilon}{\sqrt{\lambda}} + \frac{CL}{\widehat{D}_L(x,\lambda)}\right]^{|I^c|},$$

where C depends on $C_{6.9}$ only. This concludes the proof since $|I^c| = n - \lceil \lambda n \rceil \ge n/2$. $\square$

7 Estimating additive structure

Recall that our goal is to estimate the small ball probabilities for the quadratic forms of the type $\langle A^{-1}X, X\rangle$. In accordance with the spirit of Littlewood-Offord theory, we will first need to estimate the amount of additive structure in the random vector $A^{-1}X$. In this section, we indeed show that the regularized LCD of $A^{-1}X$ is large for every fixed X. This will be used later along with a decoupling argument to bound the small ball probabilities for $\langle A^{-1}X, X\rangle$.

Recall that the values of constants c_0, c_1, c_{oo} are already chosen in Remark 4.3; they depend only on parameters K, M_4.

Theorem 7.1 (Structure theorem)**.** *Let A be a random matrix which satisfies* **(A)***. There exist $c_{7.1} > 0$ and $L_0 \ge 1$ that depend only on the parameters K and M_4 from assumptions* (2.3), (2.5)*, and such that the following holds. Let $u \in \mathbb{R}^n$ be an arbitrary fixed vector, and consider $x_0 := A^{-1}u/\|A^{-1}u\|_2$. Let $L \ge L_0$ and $n^{-c_{7.1}} \le \lambda \le c_{oo}/3$. Consider the event*

$$\mathcal{E} = \Big\{x_0 \in \mathrm{Incomp}(c_0, c_1) \text{ and } \widehat{D}_L(x_0, \lambda) \ge L^{-2} n^{c_{7.1}/\lambda}\Big\}.$$

Then

$$\mathbb{P}(\mathcal{E}^c \cap \mathcal{E}_K) \le 2e^{-c_{7.1}n}.$$

We shall first prove the easier fact that $x_0 \in \mathrm{Incomp}(c_0, c_1)$. The more difficult part of the theorem is the estimate on the LCD. Its proof will be based on the probability bound of Proposition 6.11 and nontrivial covering estimates for the sets of vectors with given LCD, which we shall develop in Section 7.1.

Lemma 7.2 ($A^{-1}u$ is incompressible)**.** *In the setting of Theorem 7.1, consider the event*

$$\mathcal{E}_1 = \big\{x_0 \in \mathrm{Incomp}(c_0, c_1)\big\}.$$

Then

$$\mathbb{P}(\mathcal{E}_1^c \cap \mathcal{E}_K) \le 2e^{-c_{7.2}n}.$$

Here $c_{7.2} > 0$ depends only on the parameters K and M_4 from assumptions (2.3), (2.5)*.*

Proof. Denote $x = A^{-1}u$; then $Ax = u$. Therefore

$$\mathcal{E}_1^c \subseteq \Big\{\exists x \in \mathbb{R}^n : \frac{x}{\|x\|_2} \in \mathrm{Comp}(c_0, c_1) \wedge Ax = u\Big\}.$$

By Proposition 4.2, $\mathbb{P}(\mathcal{E}_1^c \cap \mathcal{E}_K) \leq 2e^{-c_{4.2}n}$ as claimed. □

7.1 Covering sets of vectors with small LCD

Definition 7.3 (Sublevel sets of LCD)**.** *Let us fix $\lambda \in (0, c_{oo})$. For every value $D \geq 1$, we define the set*

$$S_D = \big\{x \in \mathrm{Incomp}(c_0, c_1) : \widehat{D}_L(x, \lambda) \leq D\big\}.$$

Our present goal is to bound the covering numbers of S_D.

Proposition 7.4 (Covering sublevel sets of regularized LCD)**.** *There exist $C_{7.4}, c_{7.4} > 0$ which depend only on c_0, c_1, and such that the following holds. Let $\lambda \in (C_{7.4}/n, c_{oo}/3)$ and $L \geq 1$. For every $D \geq 1$, the sublevel set S_D has a β-net $\mathcal{N}$ such that*

$$\beta = \frac{L\sqrt{\log D}}{\sqrt{\lambda}D}, \quad |\mathcal{N}| \leq \left[\frac{C_{7.4}D}{(\lambda n)^{c_{7.4}}}\right]^n D^{1/\lambda}.$$

The main point of this result is the presence of the term $(\lambda n)^{c_{7.4}} \gg 1$ in the estimate of the cardinality of $\mathcal{N}$. This makes $|\mathcal{N}|$ substantially smaller than $(3/\beta)^n$, which is a trivial estimate on the β-net for the whole sphere S^{n-1}, see Lemma 2.2.

The proof of Proposition 7.4 relies on a series of lemmas of increasing generality. We begin by covering the level sets of the usual (not regularized) LCD. We shall work in a lower dimension m for the time being; the definition of LCD is thus considered in $\mathbb{R}^m$.

Lemma 7.5. *Let $c \in (0, 1)$, $D_0 \geq c\sqrt{m} \geq 1$ and $L \geq 1$. Then the set*

$$\big\{x \in S^{m-1} : D_L(x) \in (D_0, 2D_0]\big\}$$

has a β-net $\mathcal{N}$ such that

$$\beta = \frac{2L\sqrt{\log(2D_0)}}{D_0}, \quad |\mathcal{N}| \leq \left(\frac{CD_0}{\sqrt{m}}\right)^m.$$

Here C depends only on c.

Proof. Let x be a vector from the set in question. By the definition of LCD, there exists $p \in \mathbb{Z}^m$ such that

$$\|D_L(x)x - p\|_2 \leq L\sqrt{\log_+(2D_0/L)}. \tag{7.1}$$

Dividing both sides by $D_L(x)$ and using trivial estimates in the right hand side, we get

$$\Big\| x - \frac{p}{D_L(x)} \Big\|_2 \leq \frac{L\sqrt{\log(2D_0)}}{D_0}.$$

Since $\|x\|_2 = 1$, the last two inequalities imply that

$$\Big\| x - \frac{p}{\|p\|_2} \Big\|_2 \leq \frac{2L\sqrt{\log(2D_0)}}{D_0}.$$

Moreover, since $\|x\|_2 = 1$, we have

$$\|p\|_2 \leq \|D_L(x)x - p\|_2 + \|D_L(x)x\|_2 \leq L\sqrt{\log_+(2D_0/L)} + 2D_0 \leq 4D_0.$$

This shows that the set

$$\mathcal{N} := \Big\{ \frac{p}{\|p\|_2} : p \in \mathbb{Z}^n \cap 4D_0 B_2^m \Big\}$$

is indeed an β-net of the set in question. Counting the number of integer points in a ball by a standard volume argument, we estimate

$$|\mathcal{N}| \leq \Big(1 + \frac{12D_0}{\sqrt{m}} \Big)^m \leq \Big(\frac{CD_0}{\sqrt{m}} \Big)^m.$$

This completes the proof. □

The next step is toward removing the lower bound for $D_L(x)$ in Lemma 7.5.

Lemma 7.6. *Let $c \in (0,1)$, $D \geq D_0 \geq c\sqrt{m} \geq 1$ and $L \geq 1$. Then the set*

$$\big\{ x \in S^{m-1} : D_L(x) \in (D_0, 2D_0] \big\}$$

has a β-net $\mathcal{N}$ such that

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

Here C depends only on c.

Proof. By Lemma 7.5, we can cover the set in question with $\big(\frac{C_0 D_0}{\sqrt{m}}\big)^m$ Euclidean balls of radius $\beta_0 = \frac{2L\sqrt{\log(2D_0)}}{D_0}$ centered in the set, where C_0 depends only on c. If $\beta_0 \leq \beta$ then the lemma is proved. Assume that $\beta_0 \geq \beta$. We can further cover every ball of radius β_0 by balls of the smaller radius $\beta/2$. According to Lemma 2.2, the number of smaller balls per larger ball is at most

$$\Big(1 + \frac{4\beta_0}{\beta} \Big)^m \leq \Big(\frac{5\beta_0}{\beta} \Big)^m \leq \Big(\frac{3D}{D_0} \Big)^m.$$

The total number of smaller balls is then at most

$$\left(\frac{C_0 D_0}{\sqrt{m}}\right)^m \cdot \left(\frac{3D}{D_0}\right)^m \le \left(\frac{3C_0 D}{\sqrt{m}}\right)^m.$$

By enlarging the radius of the balls from $\beta/2$ to β as in Remark 2.1, one can assume that they are centered in the set in question. This completes the proof. $\square$

Now we can remove the flexible lower bound on $D_L(x)$ in Lemma 7.5.

Lemma 7.7. *Let $c \in (0,1)$ such that $D > c\sqrt{m} \ge 2$ and $L \ge 1$. Then the set*

$$\big\{x \in S^{m-1} : c\sqrt{m} < D_L(x) \le D\big\}$$

has a β-net $\mathcal{N}$ such that

$$\beta = \frac{4L\sqrt{\log(2D)}}{D}, \quad |\mathcal{N}| \le \left(\frac{CD}{\sqrt{m}}\right)^m \log_2 D.$$

Here C depends only on c.

Proof. We decompose the set

$$\big\{x \in S^{m-1} : D_L(x) \le D\big\} \subseteq \bigcup_k \big\{x \in S^{m-1} : D_L(x) \in (2^{-k}D, 2^{-k+1}D]\big\},$$

where the union is over the integers k such that the interval $(2^{-k}D, 2^{-k+1}D]$ has a nonempty intersection with the interval $(c\sqrt{m}, D]$. The assumptions imply that all such k are nonnegative and $2^{-k}D \ge c\sqrt{m}/2 \ge 1$. So there are at most $\log_2 D$ terms in this union, and for each term one can construct an β-net using Lemma 7.6. The union of these nets forms a required net $\mathcal{N}$. $\square$

Further, we remove the normalization requirement from the set to be covered.

Lemma 7.8. *Let $c \in (0,1)$ such that $D > c\sqrt{m} \ge 2$ and $L \ge 1$. Then the set*

$$\big\{x \in B_2^m : c\sqrt{m} < D_L(x/\|x\|_2) \le D\big\} \tag{7.2}$$

has a β-net $\mathcal{N}$ such that

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

Here C depends only on c.

Proof. Let $\mathcal{N}_0$ be a β-net of the set $\{x \in S^{m-1} : c\sqrt{m} < D_L(x) \le D\}$ as in Lemma 7.7. For each $x \in \mathcal{N}_0$, let $\mathcal{M}_x$ denote a $\beta/2$-net of the interval $\mathrm{span}(x) \cap B_2^m$ such that $|\mathcal{M}_x| \le 4/\beta$. Then $\mathcal{N} := \cup_{x \in \mathcal{N}_0} \mathcal{M}_x$ clearly forms a β-net of the set in (7.2), and

$$|\mathcal{N}| \le |\mathcal{N}_0| \cdot \frac{4}{\beta} \le \left(\frac{CD}{\sqrt{m}}\right)^m \log_2 D \cdot \frac{D}{L\sqrt{\log(2D)}}.$$

A trivial estimate of the right hand side completes the proof. $\square$

Proof of Proposition 7.4.
Step 1: decomposition. Consider a vector $x \in S_D$. Recall from Section 6.2 that $\mathrm{spread}(x) \subseteq [n]$ and $|\,\mathrm{spread}(x)| = \lceil c_{oo} n \rceil$. Let us decompose $\mathrm{spread}(x)$ into disjoint sets

$$\mathrm{spread}(x) = I_1 \cup \cdots \cup I_{k_0} \cup J$$

for some k_0 such that

$$|I_k| = \lceil \lambda n \rceil \text{ for } k \le k_0, \quad |J| < \lceil \lambda n \rceil,$$

and so that the sets fill $\mathrm{spread}(x)$ from left to right, i.e. $\sup I_k < \inf I_{k+1}$ and $\sup I_k < \inf J$ for all k. Since $\lambda \le c_{oo}$, we have $k_0 \ge 1$. Moreover, let

$$I_0 = [n] \setminus (I_1 \cup \cdots \cup I_{k_0}).$$

This produces a decomposition of $[n]$ into disjoint sets

$$[n] = I_0 \cup I_1 \cup \cdots I_{k_0}. \tag{7.3}$$

This decomposition is obviously uniquely determined by the subset $\mathrm{spread}(x)$, and it does not otherwise depend on x.

We notice two useful bounds that will help us later. Since $I_1 \cup \cdots \cup I_{k_0} = \mathrm{spread}(x) \setminus J$, we have

$$|I_1 \cup \cdots \cup I_{k_0}| \ge \lceil c_{oo} n \rceil - \lceil \lambda n \rceil \ge c_{oo} n/2 \tag{7.4}$$

and

$$k_0 \le \frac{\lceil c_{oo} n \rceil}{\lceil \lambda n \rceil} \le \frac{2c_{oo}}{\lambda}. \tag{7.5}$$

Step 2: constructing nets for each component. Let consider a fixed decomposition (7.3), and decompose the vector x accordingly:

$$x = (x_{I_0}, x_{I_1}, \ldots, x_{I_{k_0}}).$$

We are going to construct separate β-nets for each component x_{I_k}, and combine them in to one net for S_D.

A net $\mathcal{N}_0$ for the first component of x is chosen trivially. Note that $x_{I_0} \in B_2^{I_0}$. By Lemma 2.2, we can choose a $(1/D)$-net $\mathcal{N}_0$ of $B_2^{I_0}$ with

$$|\mathcal{N}_0| \leq (3D)^{|I_0|}.$$

For the other components of x, we will choose β_0-nets non-trivially, where

$$\beta_0 = \frac{4L\sqrt{\log(2D)}}{D}. \tag{7.6}$$

To this end, let us fix $k \leq k_0$. Since $x \in S_D$, the definition of the regularized LCD yields that

$$D_L\big(x_{I_k}/\|x_{I_k}\|_2\big) \leq \widehat{D}_L(x, \lambda) \leq D.$$

On the other hand, the argument in Lemma 6.8 yields

$$D_L\big(x_{I_k}/\|x_{I_k}\|_2\big) \geq c_{6.8}\sqrt{\lambda n}.$$

By the assumptions,

$$c_{6.8}\sqrt{\lambda n} \geq \frac{c_{6.8}}{2}\sqrt{|I_k|} \geq \frac{c_{6.8}}{2}\sqrt{\lambda n} \geq 2.$$

(We can choose a value of $C_{7.4}$ large enough so that this holds). Thus

$$D_L\big(x_{I_k}/\|x_{I_k}\|_2\big) \geq \frac{c_{6.8}}{2}\sqrt{|I_k|} \geq 2.$$

We have shown that x_{I_k} belongs to the set

$$V_k := \Big\{y \in B_2^{I_k} : \frac{c_{6.8}}{2}\sqrt{|I_k|} < D_L(y/\|y\|_2) \leq D\Big\}.$$

By Lemma 7.8, there exists a β_0-net $\mathcal{N}_k$ of V_k with

$$\Big(\frac{CD}{\sqrt{|I_k|}}\Big)^{|I_k|} D^2.$$

Step 3: combining the nets. We are going to combine the nets $\mathcal{N}_k$ into one net for S_D. So far we have shown that for every $x \in S_D$, there exist a decomposition (7.3) and nets $\mathcal{N}_0, \mathcal{N}_1, \ldots, \mathcal{N}_{k_0}$ which are uniquely determined by the index set spread(x), and there exist vectors $y_k \in \mathcal{N}_k$ such that

$$\|x_{I_k} - y_{I_k}\|_2 \leq \beta_0, \quad k = 0, 1, \ldots, k_0.$$

Consider the vector

$$y = (y_{I_0}, y_{I_1}, \ldots, y_{I_{k_0}}). \tag{7.7}$$

It follows that

$$\|x-y\|_2 = \Big(\sum_{k=0}^{k_0} \|x_{I_k} - y_{I_k}\|_2^2\Big)^{1/2} \le \beta_0\sqrt{k_0+1}.$$

By (7.5) and since $\lambda \le c_{oo}$, we have $k_0+1 \le 3c_{oo}/\lambda$. Recalling the definition (7.6) of β_0, we conclude that

$$\|x-y\|_2 \le \frac{7\sqrt{c_{oo}}L\sqrt{\log L}}{\sqrt{\lambda}D} \le \frac{L\sqrt{\log(2D)}}{\sqrt{\lambda}D} = \beta,$$

where we used that the value of c_{oo} can be chosen small enough (smaller than 1/49 in this case).

Consider the set $\mathcal{N}$ of vectors y that can arise in (7.7). We showed that $\mathcal{N}$ is a β-net of S_D. Moreover, since the index set spread(x) can be chosen in at most 2^n ways, it follows that

$$|\mathcal{N}| \le 2^n|\mathcal{N}_0||\mathcal{N}_1|\cdots|\mathcal{N}_{k_0}| \le 2^n \cdot (3D)^{|I_0|} \cdot \prod_{k=1}^{k_0}\Big(\frac{CD}{\sqrt{|I_k|}}\Big)^{|I_k|}D^2.$$

To simplify this bound, note that $\sum_{k=1}^{k_0}|I_k| \ge c_{oo}n/2$ by (7.4) and that $\sum_{k=0}^{k_0}|I_k| = n$ and $|I_k| \ge \lambda n \ge 1$ by construction. It follows that

$$|\mathcal{N}| \le \frac{(6CD)^n}{(\sqrt{\lambda n})^{c_{oo}n/2}}D^{2k_0}.$$

Estimate (7.5) on k_0 implies that $2k_0 \le 1/\lambda$, which completes the proof of Proposition 7.4. □

7.2 Proof of Structure Theorem 7.1.

In Proposition 6.11 we estimated the small ball probabilities for the random vector Ax for a fixed vector x. Now we combine this with the covering results of the previous section to obtain a bound that is uniform over all x with small regularized LCD. Recall that S_D denotes the sub-lebel set of regularized LCD according to Definition 7.3.

Lemma 7.9 (Small ball probabilities on a sublevel set of LCD)**.** *There exist $c, c' > 0$ and $L_0 \ge 1$ that depend only on the parameters K and M_4 from the assumptions* (2.3), (2.5), *and such that the following holds. Let $L \ge L_0$, $n^{-c} \le \lambda \le c_{oo}/3$ and $1 \le D \le L^{-2}n^{c/\lambda}$. Then*

$$\mathbb{P}\big\{\exists x \in S_D : \|Ax-u\|_2 \le K\beta\sqrt{n} \wedge \mathcal{E}_K\big\} \le n^{-c'n}, \tag{7.8}$$

where

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

Proof. We will first compute the probability for $S_D \setminus S_{D/2}$ instead of S_D in (7.8). Proposition 6.11 implies that for every $x \in S_D \setminus S_{D/2}$, one has

$$\mathbb{P}\big\{\|Ax-u\|_2 \le \varepsilon\sqrt{n}\big\} \le \left[\frac{C_{6.11}L\varepsilon}{\sqrt{\lambda}} + \frac{C_{6.11}L}{D}\right]^{n-\lceil\lambda n\rceil}, \quad \varepsilon \ge 0.$$

Let us use this inequality for $\varepsilon = 4K\beta$. Clearly, the term $\frac{\varepsilon}{\sqrt{\lambda}}$ dominates the term $\frac{1}{4D}$. So we obtain

$$\mathbb{P}\big\{\|Ax-u\|_2 \le 4K\beta\sqrt{n}\big\} \le \left[\frac{C'L^2\sqrt{\log(2D)}}{\lambda D}\right]^{n-\lceil\lambda n\rceil} =: p_0.$$

(Here the constant $C' = C'(K, M_4)$ absorbs the factor K.)

Let us choose a β-net $\mathcal{N}$ of $S_D \setminus S_{D/2}$ according to Proposition 7.4. The union bound yields

$$\begin{aligned}\mathbb{P}\big\{\exists x \in \mathcal{N} :\, \|Ax-u\|_2 \le 4K\beta\sqrt{n}\big\} &\le |\mathcal{N}| \cdot p_0 \\ &\le \left[\frac{C_{7.4}D}{(\lambda n)^{c_{7.4}}}\right]^n D^{1/\lambda} \cdot \left[\frac{C'L^2\sqrt{\log(2D)}}{\lambda D}\right]^{n-\lceil\lambda n\rceil} =: p_1.\end{aligned}$$

One can estimate p_1 using the assumptions that n is sufficiently large, $n^{-c} \le \lambda \le c_{oo}/3$ and $1 \le D \le L^{-2}n^{c/\lambda}$. Choosing the constant $c > 0$ sufficiently small and making simplifications, we obtain

$$p_1 \le n^{-c''n}.$$

Suppose event $\mathcal{E}_K$ occurs, and suppose there exists $x \in S_D \setminus S_{D/2}$ such that $\|Ax-u\|_2 \le K\beta\sqrt{n}$. There exists $x_0 \in \mathcal{N}$ such that $\|x-x_0\|_2 \le \beta$. Then

$$\begin{aligned}\|Ax_0-u\|_2 &\le \|Ax-u\|_2 + \|A(x-x_0)\|_2 \le \|Ax-u\|_2 + \|A\|\|x-x_0\|_2 \\ &\le K\beta\sqrt{n} + 3K\sqrt{n}\cdot\beta = 4K\beta\sqrt{n}.\end{aligned}$$

As we know, the probability of the latter event is at most $p_1 \le n^{-c''n}$. So we have shown that

$$\mathbb{P}\big\{\exists x \in S_D \setminus S_{D/2} :\, \|Ax-u\|_2 \le K\beta\sqrt{n} \wedge \mathcal{E}_K\big\} \le n^{-c''n}.$$

Finally, we get rid of $S_{D/2}$ in this bound. Since β decreases in D, as long as $D/2 \ge 1$ the previous result can be applied for $D/2$ instead of D, and we get

$$\mathbb{P}\big\{\exists x \in S_{D/2} \setminus S_{D/4} :\, \|Ax-u\|_2 \le K\beta\sqrt{n} \wedge \mathcal{E}_K\big\} \le n^{-c''n}.$$

We can continue this way for $S_{D/4} \setminus S_{D/8}$, etc. So we decompose $S = \bigcup_{k=0}^{k_0}(S_{2^{-k}D} \setminus S_{2^{-k-1}D})$, where k_0 is the largest integer such that $2^{-k_0}D \ge c_{6.8}\sqrt{\lambda n}$. (Recall that by Proposition 6.8, the set S_{D_0} is empty if $D_0 < c_{6.8}\sqrt{\lambda n}$. Since $c_{6.8}\sqrt{\lambda n} \geq 1$, we have $k_0 \leq \log_2 D$. The union bound then gives

$$\mathbb{P}\big\{\exists x \in S_D : \|Ax - u\|_2 \leq K\beta\sqrt{n} \wedge \mathcal{E}_K\big\} \leq k_0 \cdot n^{-c''n} \leq \log_2(D) n^{-c''n} \leq n^{-c'n}$$

if the constant $c' > 0$ is chosen appropriately small. This completes the proof. $\square$

Proof of Structure Theorem 7.1. We fix constants c, c', L_0 given by Lemma 7.9. Consider the following two events:

$$\mathcal{E}_0 = \big\{\widehat{D}_L(x_0, \lambda) > L^{-2} n^{c/\lambda} =: D_0 \text{ or } \widehat{D}_L(x_0, \lambda) \text{ is undefined}\big\},$$
$$\mathcal{E}_1 = \big\{x_0 \in \mathrm{Incomp}(c_0, c_1)\big\}.$$

Recall that if $\mathcal{E}_1$ holds then $\widehat{D}_L(x_0, \lambda)$ is defined. So our desired event $\mathcal{E}$ can be written as

$$\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_0.$$

Then $\mathcal{E}^c = \mathcal{E}_1^c \cup (\mathcal{E}_1 \cap \mathcal{E}^c) = \mathcal{E}_1^c \cup (\mathcal{E}_1 \cap \mathcal{E}_0^c)$. Finally, the event whose probability we need to estimate is $\mathcal{E}^c \cap \mathcal{E}_K \subseteq (\mathcal{E}_1^c \cap \mathcal{E}_K) \cup (\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K)$. Hence

$$\mathbb{P}(\mathcal{E}^c \cap \mathcal{E}_K) \leq \mathbb{P}(\mathcal{E}_1^c \cap \mathcal{E}_K) + \mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K).$$

The first term was estimated in Lemma 7.2 as

$$\mathbb{P}(\mathcal{E}_1^c \cap \mathcal{E}_K) \leq 2e^{-c_{7.2}n}.$$

It remains to obtain a similar estimate on the second term $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K)$. We can express

$$\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K = \big\{x_0 := A^{-1}u/\|A^{-1}u\|_2 \in S_{D_0} \wedge \mathcal{E}_K\big\}.$$

Let $u_0 := Ax_0 = u/\|A^{-1}u\|_2$. Event $\mathcal{E}_K$ implies that $\|u_0\|_2 = \|Ax_0\|_2 \leq \|A\| \leq 3K\sqrt{n}$. Therefore u_0 lies on a one-dimensional interval:

$$u_0 \in \mathrm{span}(u) \cap 3K\sqrt{n}B_2^n =: E.$$

So

$$\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K \subseteq \big\{\exists x_0 \in S_{D_0}, \exists u_0 \in E : Ax_0 = u_0 \wedge \mathcal{E}_K\big\}.$$

In view of an application of Lemma 7.9, let us choose

$$\beta_0 = \frac{L\sqrt{\log(2D_0)}}{D_0}.$$

Let $\mathcal{M}$ denote some fixed $(K\beta_0\sqrt{n})$-net of the interval E, such that

$$|\mathcal{M}| \leq \frac{3K\sqrt{n}}{6K\beta_0\sqrt{n}} = \frac{6}{\beta_0} \leq 6D_0.$$

So, for every $u_0 \in E$ we can choose $v_0 \in \mathcal{M}$ such that $\|u_0 - v_0\|_2 \le K\beta_0\sqrt{n}$. Since $Ax_0 = u_0$, it follows that $\|Ax_0 - v_0\|_2 \le K\beta_0\sqrt{n}$. We have shown that

$$\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K \subseteq \big\{\exists x_0 \in S_{D_0},\, \exists v_0 \in \mathcal{M} :\, \|Ax_0 - v_0\|_2 \le K\beta_0\sqrt{n} \wedge \mathcal{E}_K\big\}.$$

An application of Lemma 7.9 and a union bound over $v_0 \in \mathcal{M}$ give

$$\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K) \le |\mathcal{M}| \cdot n^{-c'n} \le 6D_0 \cdot n^{-c'n} \le n^{-c'n/2}$$

where we used that $D_0 \le n^{c/\lambda}$, and since we can assume that constant $c > 0$ appropriately small. The proof of Structure Theorem 7.1 is complete. $\square$

8 Small ball probabilities for quadratic forms

Now that we developed a machinery for estimating small ball probabilities, we can come back to our main task, estimating the small ball probability for quadratic forms. Recall that by Proposition 5.1 and Remark 5.2, the distance problem reduces to estimating Lévy concentration function for the self-normalized quadratic forms:

$$\mathcal{L}\Big(\frac{\langle A^{-1}X, X\rangle}{\sqrt{1+\|A^{-1}X\|_2^2}}, \varepsilon\Big) \le ? \tag{8.1}$$

Here and throughout this section, A denotes the $n \times n$ symmetric random matrix satisfying assumptions (2.5). X denotes a random vector whose entries are independent of A and of each other, identically distributed, and satisfy the same moment assumptions (2.5) as those of A, namely they have zero mean, unit variance, and fourth moment bounded by M_4^4.

The goal of this section is to prove the following estimate.

Theorem 8.1 (Small ball probabilities for quadratic forms)**.** *Let A be an $n \times n$ random matrix which satisfies* **(A)***, and let X be a random vector in $\mathbb{R}^n$ whose entries are independent of each other and of A, identically distributed, and satisfy the same moment assumptions* (2.5) *as those of A, namely they have zero mean, unit variance, and fourth moment bounded by M_4^4. There exist constants $C_{8.1}, c_{8.1} > 0$ that depend only on the parameters K and M_4 from the assumptions* (2.3), (2.5)*, and such that the following holds. For every $\varepsilon \ge 0$ and every $u \in \mathbb{R}$, one has*

$$\mathbb{P}\Big\{\frac{|\langle A^{-1}X, X\rangle - u|}{\sqrt{1+\|A^{-1}X\|_2^2}} \le \varepsilon \wedge \mathcal{E}_K\Big\} \le C_{8.1}\varepsilon^{1/9} + 2\exp(-n^{c_{8.1}}). \tag{8.2}$$

In particular, we have a desired bound for Lévy concentration function in (8.1), namely $C_{8.1}\varepsilon^{1/9} + 2\exp(-n^{c_{8.1}}) + \mathbb{P}(\mathcal{E}_K^c)$.

To prove Theorem 8.1, we will first decouple the enumerator $\langle A^{-1}X, X\rangle$ from the denominator $\sqrt{1+\|A^{-1}X\|_2^2}$ by showing that $\|A^{-1}X\|_2 \sim \|A^{-1}\|_{\mathrm{HS}}$ with high probability. This is done in Section 8.1. Then we decouple the quadratic form $\langle A^{-1}X, X\rangle$. An ideal decoupling argument would replace $\langle A^{-1}X, X\rangle$ by $\langle A^{-1}X, X'\rangle$ where X' is independent random vector; our argument will be of similar nature. Then by conditioning on X we obtain a linear form, and we can estimate its small ball probabilities using the Littlewood-Offord theory (specifically, using Proposition 6.9 and Structure Theorem 7.1). This will be done in Section 8.3.

8.1 Size of $A^{-1}X$

The following result compares the size of the denominator $\sqrt{1+\|A^{-1}X\|_2^2}$ appearing in (8.2) to $\|A^{-1}\|_{\mathrm{HS}}$.

Proposition 8.2 (Size of $A^{-1}X$)**.** *Let A be a random matrix which satisfies* **(A)**. *There exist constants $c, C_{8.2}, c_{8.2} > 0$ that depend only on the parameters K and M_4 from the assumptions* (2.3), (2.5), *and such that the following holds. Let $n^{-c} \le \lambda \le c$. The random matrix A has the following property with probability at least $1 - e^{-cn}$. If $\mathcal{E}_K$ holds, then for every $\varepsilon > 0$ one has:*

(i) with probability at least $1 - e^{-c_{8.2}n}$ in X, we have

$$\|A^{-1}X\|_2 \ge C_{8.2};$$

(ii) with probability at least $1 - \varepsilon$ in X, we have

$$\|A^{-1}X\|_2 \le \varepsilon^{-1/2}\|A^{-1}\|_{\mathrm{HS}};$$

(iii) with probability at least $1 - C_{8.2}\varepsilon/\sqrt{\lambda} - n^{-c_{8.2}/\lambda}$ in X, we have

$$\|A^{-1}X\|_2 \ge \varepsilon\|A^{-1}\|_{\mathrm{HS}}.$$

The proof of this result uses the following elementary lemma.

Lemma 8.3 (Sums of dependent random variables)**.** *Let $Z_1, \ldots, Z_n$ be arbitrary non-negative random variables (not necessarily independent), and $p_1, \ldots, p_n$ be non-negative numbers such that*

$$\sum_{k=1}^{n} p_k = 1.$$

Then for every $\varepsilon \in \mathbb{R}$ one has

$$\mathbb{P}\Big\{\sum_{k=1}^{n} p_k Z_k \le \varepsilon\Big\} < 2\sum_{k=1}^{n} p_k\, \mathbb{P}\{Z_k \le 2\varepsilon\}.$$

Proof. By Markov's inequality, the event $\sum_{k=1}^n p_k Z_k \le \varepsilon$ implies $\sum_k p_k \mathbf{1}_{\{Z_k > 2\varepsilon\}} < 1/2$ and, consequently, $\sum_k p_k \mathbf{1}_{\{Z_k \le 2\varepsilon\}} > 1/2$. Therefore,

$$\begin{aligned}\mathbb{P}\Big\{ \sum_{k=1}^n p_k Z_k \le \varepsilon \Big\} &\le \mathbb{P}\Big\{ \sum_k p_k \mathbf{1}_{\{Z_k \le 2\varepsilon\}} > 1/2 \Big\} \\ &< 2\mathbb{E} \sum_k p_k \mathbf{1}_{\{Z_k \le 2\varepsilon\}} \quad \text{(again by Markov's inequality)} \\ &= 2 \sum_{k=1}^n p_k \, \mathbb{P}\{Z_k \le 2\varepsilon\}.\end{aligned}$$

The proof is complete. □

Proof of Proposition 8.2. Let $e_1, \ldots, e_n$ denotes the canonical basis of $\mathbb{R}^n$, and let

$$x_k := \frac{A^{-1} e_k}{\|A^{-1} e_k\|_2}, \quad k = 1, \ldots, n.$$

Let us apply Structure Theorem 7.1 combined with the union bound over $k = 1, \ldots, n$. We do this with $L = L_0$ a suitably large constant depending on parameters K and M only (chosen so that Proposition 6.9 can be applied below). We see that the random matrix A has the following property with probability at least $1 - n \cdot 2e^{-c_{7.1} n} \ge 1 - 2e^{-c_{7.1} n/2}$: if $\mathcal{E}_K$ holds then

$$x_k \in \mathrm{Incomp}(c_0, c_1),\ \widehat{D}_L(x_k, \lambda) \ge L^{-2} n^{c_{7.1}/\lambda} \quad k = 1, \ldots, n. \tag{8.3}$$

Let us fix a realization of A with this property. We shall deduce properties (i), (ii), (iii) from it. Without loss of generality we may assume that $\mathcal{E}_K$ holds.

(i) We have

$$\|X\|_2 \le \|A\| \|A^{-1} X\|_2.$$

By $\mathcal{E}_K$, we have $\|A\| \le 3K\sqrt{n}$. Moreover, Lemma 3.2 and Tensorization Lemma 3.4 imply that the random vector X satisfies $\|X\|_2 \ge c' \sqrt{n}$ with probability at least $1 - e^{-c'n}$, for some constant $c' = c'(K, M) > 0$. It follows that $\|A^{-1} X\|_2 \ge c'/3K$ with the same probability, so part (i) of the proposition is proved.

(ii) Using that A is a symmetric matrix, we express

$$\|A^{-1} X\|_2^2 = \sum_{k=1}^n \langle A^{-1} X, e_k \rangle^2 = \sum_{k=1}^n \langle A^{-1} e_k, X \rangle^2 = \sum_{k=1}^n \|A^{-1} e_k\|_2^2 \, \langle x_k, X \rangle^2. \tag{8.4}$$

Recall that the coordinates of X are independent random variables with zero mean and unit variance. Therefore $\mathbb{E}_X \langle x_k, X \rangle^2 = 1$ for all k, so

$$\mathbb{E}_X \|A^{-1} X\|_2^2 = \sum_{k=1}^n \|A^{-1} e_k\|_2^2 = \|A^{-1}\|_{\mathrm{HS}}^2.$$

An application of Markov's inequality yields part (ii) of the proposition.

(iii) Fix $k \le n$. Then $\langle x_k, X\rangle$ can be expressed as a sum of independent random variables $\sum_{i=1}^n x_{ki} X_i$, where x_{ki} and X_i denote the coordinates of x_k and of X respectively. This sum can be estimated using Proposition 6.9 (with $J = [n]$) combined with the estimate (8.3) on the regularized LCD of x_k. This gives

$$\mathbb{P}_X\big\{|\langle x_k, X\rangle| \le \sqrt{2}\,\varepsilon\big\} \le C_{6.9} L\Big(\frac{\varepsilon}{\sqrt{\lambda}} + L^2 n^{-c_{7.1}/\lambda}\Big). \tag{8.5}$$

Now we combine these estimates for all k using (8.4) and Lemma 8.3 with $p_k = \|A^{-1}e_k\|_2^2/\|A^{-1}\|_{\mathrm{HS}}^2$; note that $\sum_{k=1}^n p_k = 1$. We obtain

$$\begin{aligned}
\mathbb{P}_X\big\{\|A^{-1}X\|_2 \le \varepsilon\|A^{-1}\|_{\mathrm{HS}}\big\} &= \mathbb{P}\Big\{\sum_{k=1}^n p_k \langle x_k, X\rangle^2 \le \varepsilon^2\Big\} \\
&\le 2\sum_{k=1}^n p_k\, \mathbb{P}\big\{\langle x_k, X\rangle^2 \le 2\varepsilon^2\big\} \quad \text{(by Lemma 8.3)} \\
&\le 2C_{6.9} L\Big(\frac{\varepsilon}{\sqrt{\lambda}} + L^2 n^{-c_{7.1}/\lambda}\Big) \quad \text{(by (8.5))}.
\end{aligned}$$

This proves part (iii), and completes the proof of Proposition 8.2. □

8.2 Decoupling quadratic forms

Decoupling the quadratic form $\langle A^{-1}X, X\rangle$ is based on the following general result. Similar decoupling techniques for quadratic forms were first applied by Götze [6] and used in literature many times since then; in particular such a decoupling argument was used in [4, 3] in a context similar to ours.

Lemma 8.4 (Decoupling quadratic forms)**.** *Let G be an arbitrary symmetric $n \times n$ matrix, and let X be a random vector in $\mathbb{R}^n$ with independent coordinates. Let X' denote an independent copy of X. Consider a subset $J \subseteq [n]$. Then for every $\varepsilon \ge 0$ one has*

$$\begin{aligned}
\mathcal{L}\big(\langle GX, X\rangle, \varepsilon\big)^2 &= \sup_{u\in\mathbb{R}} \mathbb{P}\big\{|\langle GX, X\rangle - u| \le \varepsilon\big\}^2 \\
&\le \mathbb{P}_{X,X'}\Big\{\big|\langle G(P_{J^c}(X - X')), P_J X\rangle - v\big| \le \varepsilon\Big\}
\end{aligned}$$

where v is some random variable whose value is determined by the $J^c \times J^c$ minor of G and the random vectors $P_{J^c}X$, $P_{J^c}X'$.

The point of this result is that, upon conditioning on the coordinates of X and X' in J^c, the vectors $x_0 := G(P_J^c(X - X'))$ and v become fixed. So the Lévy concentration function of the quadratic form $\langle GX, X\rangle$ gets bounded by the Lévy concentration function of the linear form $\langle x_0, P_J X\rangle$. The latter, as we know, can be estimated using the Littlewood-Offord theory.

The proof of Lemma 8.4 is based on the general decoupling lemma from [16], which was already used for a purpose similar to ours in [3].

Lemma 8.5. *Let Y and Z be independent random variables or vectors, and let Z' be an independent copy of Z. Let $\mathcal{E}(Y,Z)$ be an event which is determined by the values of Y and Z. Then*

$$\mathbb{P}\big\{\mathcal{E}(Y,Z)\big\}^2 \le \mathbb{P}\big\{\mathcal{E}(Y,Z) \cap \mathcal{E}(Y,Z')\big\}. \qquad \square$$

Proof of Decoupling Lemma 8.4. By permuting the coordinates, without loss of generality we can assume that J and J^c are intervals of coordinates with $\sup J \le \inf J^c$. The decomposition $[n] = J \cup J^c$ induces the decomposition of the matrix A and all the vectors in question,

$$G = \begin{pmatrix} E & F \\ F^* & H \end{pmatrix}, \quad X = \begin{pmatrix} Y \\ Z \end{pmatrix}, \quad X' = \begin{pmatrix} Y' \\ Z' \end{pmatrix}; \quad \text{let } \widetilde{X} = \begin{pmatrix} Y \\ Z' \end{pmatrix}.$$

Here E is a $J \times J$ minor of G, H is a $J \times J^c$ minor, etc., and similarly $Y \in \mathbb{R}^J$, $Z \in \mathbb{R}^{J^c}$, etc. Let us fix a $u \in \mathbb{R}$ and apply Lemma 8.5; this gives

$$p^2 := \mathbb{P}\big\{|\langle GX, X\rangle - u| \le \varepsilon\big\}^2 \le \mathbb{P}_{X,\widetilde{X}}\big\{|\langle GX, X\rangle - u| \le \varepsilon \wedge |\langle G\widetilde{X}, \widetilde{X}\rangle - u| \le \varepsilon\big\}. \tag{8.6}$$

By the triangle inequality,

$$p^2 \le \mathbb{P}_{X,\widetilde{X}}\big\{|\langle GX, X\rangle - \langle G\widetilde{X}, \widetilde{X}\rangle| \le 2\varepsilon\big\}.$$

By our decomposition, we have

$$\langle GX, X\rangle = \langle EY, Y\rangle + 2\langle FZ, Y\rangle + \langle HZ, Z\rangle,$$
$$\langle G\widetilde{X}, \widetilde{X}\rangle = \langle EY, Y\rangle + 2\langle FZ', Y\rangle + \langle HZ', Z'\rangle.$$

Hence

$$\langle GX, X\rangle - \langle G\widetilde{X}, \widetilde{X}\rangle = 2\langle F(Z - Z'), Y\rangle + \langle HZ, Z\rangle - \langle HZ', Z'\rangle.$$

Recall that F is the restriction of the matrix G onto the pairs of coordinates in $J \times J^c$, that $Z - Z'$ is the restriction of the vector $X - X'$ onto the coordinates in J^c, and that Y is the restriction of X onto the coordinates in J. So

$$\langle F(Z - Z'), Y\rangle = \langle G(P_{J^c}(X - X')), P_J X\rangle.$$

Similarly we can see that the value of $\langle HZ, Z\rangle - \langle HZ', Z'\rangle$ depends on the $J^c \times J^c$ minor H and on the restrictions of X and X' onto the coordinates in J^c. So setting $v = 2\langle HZ, Z\rangle - 2\langle HZ', Z'\rangle$, we express

$$\langle GX, X\rangle - \langle G\widetilde{X}, \widetilde{X}\rangle = 2\langle G(P_{J^c}(X - X')), P_J X\rangle + v.$$

This and (8.6) completes the proof of Decoupling Lemma 8.4. $\square$

8.3 Proof of Theorem 8.1

Our argument will be based on decoupling the quadratic form $\langle AX, X\rangle$, and treating the resulting linear form using the Littlewood-Offord theory developed earlier in this paper.

Step 1: Constructing a random subset J and assignment $\mathrm{spread}(x)$**.** The decoupling starts by decomposing $[n]$ into two random sets J and J^c. To this end, we consider independent $\{0,1\}$-valued random variables $\delta_1, \ldots, \delta_n$ ("selectors") with $\mathbb{E}\delta_i = c_{oo}/2$. (Recall that the constant c_{oo}, which depends on K and M_4 only, was fixed in the definition of the regularized LCD in Section 6.2.) We then define

$$J := \{i \in [n] : \delta_i = 0\}.$$

Then $\mathbb{E}|J^c| = c_{oo}n/2$. By a basic result in large deviations (see e.g. [1] Theorem A.1.4), the bound

$$|J^c| \le c_{oo}n \tag{8.7}$$

holds with high probability:

$$\mathbb{P}_J\{(8.7)\text{ holds}\} \ge 1 - 2e^{-c'_{oo}n}$$

where $c'_{oo} = c_{oo}^2/2$.

Consider a fixed realization of J that satisfies (8.7). As we know from Section 6.2, at least $2c_{oo}n$ coordinates of a vector $x \in \mathrm{Incomp}(c_0, c_1)$ satisfy the regularity condition (6.4). It follows that for each vector $x \in \mathrm{Incomp}(c_0, c_1)$ we can assign a subset

$$\mathrm{spread}(x) \subseteq J, \quad |\,\mathrm{spread}(x)| = \lceil c_{oo}n \rceil \tag{8.8}$$

and so that the regularity condition (6.4) holds for all $k \in \mathrm{spread}(x)$. If there is more than one way to assign $\mathrm{spread}(x)$ to x, we choose one fixed way to do so. This results in a valid assignment (per Section 6.2) that depends only on the choice of the random set J. We shall use this assignment in applications of Definition 6.6 of the regularized LCD of x.

Step 2: Estimating the denominator $\sqrt{1 + \|A^{-1}X\|_2^2}$ and LCD of the inverse. Lemma 8.2 will allow us to replace in (8.2) the denominator $\sqrt{1 + \|A^{-1}X\|_2^2}$ by $\|A^{-1}\|_2^2$. However, we have to do this carefully in order to withstand losses that will occur at the decoupling step. So, let $\varepsilon_0 \in (0,1)$ and let X' denote an independent copy of the random vector X. We consider the following event that is determined by the random matrix A, random vectors X, X' and the random set J:

$$\sqrt{\varepsilon_0}\sqrt{1 + \|A^{-1}X\|_2^2} \le \|A^{-1}\|_{\mathrm{HS}} \le \frac{1}{\varepsilon_0}\|A^{-1}(P_{J^c}(X - X'))\|_2. \tag{8.9}$$

Recall that the coordinates X_i of X are independent random variables with zero mean, unit variance, and $\mathbb{E}X_i^4 \le M_4^4$. It follows that the coordinates $Y_i = \delta_i(X_i - X_i')$ of the random vector $Y := P_{J^c}(X - X')$ are again independent random variables with

$$\mathbb{E}Y_i = 0, \quad \mathbb{E}Y_i^2 = c_{oo}, \quad \mathbb{E}Y_i^4 \le 8c_{oo}M_4^4.$$

We see that Proposition 8.2 applies for X, and also for X replaced by $c_{oo}^{-1/2}X$ (in the latter case with M_4 replaced by $2c_{oo}^{-1/4}M_4$). It follows that

$$\mathbb{P}_{A,X,X',J}\{(8.7) \text{ holds} \vee \mathcal{E}_K^c\} \ge 1 - \frac{C'\varepsilon_0}{\sqrt{\lambda}} - n^{-c'/\lambda} - 2e^{-c'n}$$

where $C', c' > 0$ depend only on K and M_4.

Consider the random vector

$$x_0 := \frac{A^{-1}(P_{J^c}(X - X'))}{\|A^{-1}(P_{J^c}(X - X'))\|_2}. \tag{8.10}$$

(If the denominator equals zero, assign to x_0 an arbitrary fixed vector in S^{n-1}.) Let us condition on an arbitrary realization of random vectors X, X' and on a realization of J which satisfies (8.7). Fix some value of the parameter λ satisfying $n^{-c_{7.1}} \le \lambda \le c_{oo}/3$ as required in Structure Theorem 7.1, and consider the event

$$x_0 \in \mathrm{Incomp}(c_0, c_1) \quad \text{and} \quad \widehat{D}_{L_0}(x_0, \lambda) \ge C'' n^{c''/\lambda}, \tag{8.11}$$

which depends on the random matrix A. By Structure Theorem 7.1, the conditional probability is

$$\mathbb{P}_A\big\{(8.11) \text{ holds} \vee \mathcal{E}_K^c \,|\, X, X', J \text{ satisfies } (8.7)\big\} \ge 1 - 2e^{-c''n}.$$

Here $L_0, C'', c'' > 0$ depend only on K and M_4.

Combining the three probabilities, we obtain

$$\begin{aligned}
&\mathbb{P}_{A,X,X',J}\big\{((8.7),\ (8.9),\ (8.11) \text{ hold}) \vee \mathcal{E}_K^c\big\} \\
&\qquad \ge 1 - 2e^{-c'_{oo}n} - \Big(\frac{C'\varepsilon_0}{\sqrt{\lambda}} + n^{-c'/\lambda} + 2e^{-c'n}\Big) - 2e^{-c''n} \\
&\qquad =: 1 - p_0. \qquad (8.12)
\end{aligned}$$

It follows that there exists a realization of J that satisfies (8.7) and such that

$$\mathbb{P}_{A,X,X'}\big\{((8.9),\ (8.11) \text{ hold}) \vee \mathcal{E}_K^c\big\} \ge 1 - p_0.$$

Let us fix such a realization of J for the rest of the proof. An application of Fubini's theorem shows that the random matrix A has the following property with probability at least $1 - \sqrt{p_0}$:

$$\mathbb{P}_{X,X'}\big\{((8.9),\ (8.11) \text{ hold}) \vee \mathcal{E}_K^c \,|\, A\big\} \ge 1 - \sqrt{p_0}.$$

But the event $\mathcal{E}_K^c$ depends on A only and not on X or X'. Therefore, the random matrix A has the following property with probability at least $1-\sqrt{p_0}$. Either $\mathcal{E}_K^c$ holds, or:

$$\mathcal{E}_K \text{ holds and } \mathbb{P}_{X,X'}\big\{(8.9),\ (8.11) \text{ hold} \,|\, A\big\} \geq 1-\sqrt{p_0}. \tag{8.13}$$

Step 3: decoupling. The event we are interested in is

$$\mathcal{E} := \left\{ \frac{|\langle A^{-1}X, X\rangle - u|}{\sqrt{1+\|A^{-1}X\|_2^2}} \leq \varepsilon \right\}.$$

We need to estimate the probability

$$\mathbb{P}_{A,X}(\mathcal{E}\cap\mathcal{E}_K) \leq \mathbb{P}_{A,X}\{\mathcal{E}\wedge (8.13) \text{ holds}\} + \mathbb{P}_{A,X}\{\mathcal{E}_K \wedge (8.13) \text{ fails}\}.$$

By the previous step in the proof, the second term here is bounded by $\sqrt{p_0}$. Therefore

$$\mathbb{P}_{A,X}(\mathcal{E}\cap\mathcal{E}_K) \leq \sup_{A \text{ satisfies } (8.13)} \mathbb{P}_X(\mathcal{E}\,|\,A) + \sqrt{p_0}.$$

Computing the same probability in the larger space determined by the random vectors X, X', and using property (8.13), we write

$$\mathbb{P}_{A,X}(\mathcal{E}\cap\mathcal{E}_K) \leq \sup_{A \text{ satisfies } (8.13)} \mathbb{P}_{X,X'}\big\{\mathcal{E}\wedge (8.9) \text{ holds} \,|\, A\big\} + 2\sqrt{p_0}. \tag{8.14}$$

Let us fix a realization of a random matrix A satisfying (8.13) for the rest of the proof. So our goal is to bound the probability

$$p_1 := \mathbb{P}_{X,X'}\big\{\mathcal{E}\wedge (8.9) \text{ holds}\big\}.$$

Using definition of $\mathcal{E}$ and the first inequality in property (8.9), we have

$$p_1 \leq P_{X,X'}\Big\{|\langle A^{-1}X, X\rangle - u| \leq \frac{\varepsilon}{\sqrt{\varepsilon_0}}\|A^{-1}\|_{\mathrm{HS}}\Big\}.$$

We apply Decoupling Lemma 8.4, and obtain

$$p_1^2 \leq \mathbb{P}_{X,X'}\{\mathcal{E}_0\}$$

where

$$\mathcal{E}_0 = \Big\{\big|\langle A^{-1}(P_{J^c}(X-X')), P_J X\rangle - v\big| \leq \frac{\varepsilon}{\sqrt{\varepsilon_0}}\|A^{-1}\|_{\mathrm{HS}}\Big\}$$

and where $v = v(A^{-1}, P_{J^c}X, P_{J^c}X')$ denotes a number that depends on A^{-1}, $P_{J^c}X$, $P_{J^c}X'$ only. Further, using property (8.13) (in which conditioning on A is no longer needed as we are treating A as a fixed matrix), we get

$$p_1^2 \leq \mathbb{P}_{X,X'}\{\mathcal{E}_0\} \leq \mathbb{P}_{X,X'}\big\{\mathcal{E}_0 \wedge (8.9),\ (8.11) \text{ hold}\big\} + \sqrt{p_0}.$$

Let us divide both sides in the inequality defining the event $\mathcal{E}_0$ by $\|A^{-1}(P_{J^c}(X - X'))\|_2$. Using definition (8.10) of x_0 and the second inequality in (8.9), we obtain

$$p_1^2 \le \mathbb{P}_{X,X'}\Big\{\big|\langle x_0, P_J X\rangle - w\big| \le \varepsilon_0^{-3/2}\varepsilon \wedge (8.11) \text{ holds}\Big\} + \sqrt{p_0} \tag{8.15}$$

where $w = w(A^{-1}, P_{J^c}X, P_{J^c}X')$ is an appropriate number.

Step 4: The small ball probabilities of a linear form. By definition, the random vector x_0 is determined by the random vector $P_{J^c}(X - X')$, which is independent of the random vector $P_J X$. So if we fix an arbitrary realization of the random vectors $P_{J^c}X$ and $P_{J^c}X'$, this will fix the vector x_0 and the number w in (8.15). Since moreover (8.11) is a property of x_0, we conclude that

$$p_1^2 \le \sup_{\substack{x_0 \text{ satisfies } (8.11)\\ w\in\mathbb{R}}} \mathbb{P}_{P_J X}\Big\{\big|\langle x_0, P_J X\rangle - w\big| \le \varepsilon_0^{-3/2}\varepsilon\Big\} + \sqrt{p_0}.$$

So let us fix a vector $x_0 = (x_{01}, \ldots, x_{0n}) \in S^{n-1}$ that satisfies (8.11) and a number $w \in \mathbb{R}$. We have reduced the problem to estimating the small ball probabilities for the sum of independent random variables

$$\langle x_0, P_J X\rangle = \sum_{k\in J} x_{0k}\xi_k$$

where we denote $X = (\xi_1, \ldots, \xi_n)$.

We can apply Proposition 6.9 for this sum, noting that by (8.8) we have $J \supseteq \operatorname{spread}(x_0) \supseteq I(x)$ as required there. (The last inclusion follows by the definition of the maximizing set $I(x)$, recall Definition 6.6.) It follows that

$$P_{P_J X}\Big\{\big|\langle x_0, P_J X\rangle - w\big| \le \varepsilon_0^{-3/2}\varepsilon\Big\} \le \frac{C_1\varepsilon_0^{-3/2}\varepsilon}{\sqrt{\lambda}} + \frac{C_1}{\widehat{D}_{L_0}(x_0, \lambda)},$$

for some $C_1 = C_1(K, M_4)$. Using property (8.11) to bound the second term in the right hand side, we obtain

$$p_1^2 \le \frac{C_1\varepsilon_0^{-3/2}\varepsilon}{\sqrt{\lambda}} + C_1' n^{-c''/\lambda} + \sqrt{p_0},$$

Now we estimate the probability of the desired event in (8.14) as

$$\begin{aligned}\mathbb{P}_{A,X}(\mathcal{E} \cap \mathcal{E}_K) &\le p_1 + 2\sqrt{p_0}\\ &\le \Big(\frac{C_1\varepsilon_0^{-3/2}\varepsilon}{\sqrt{\lambda}}\Big)^{1/2} + \big(C_1' n^{-c''/\lambda}\big)^{1/2} + p_0^{1/4} + 2\sqrt{p_0}.\end{aligned}$$

Recalling the definition (8.12) of p_0 and simplifying, we obtain

$$\mathbb{P}_{A,X}(\mathcal{E} \cap \mathcal{E}_K) \le \Big(\frac{C_1\varepsilon_0^{-3/2}\varepsilon}{\sqrt{\lambda}}\Big)^{1/2} + \Big(\frac{C'\varepsilon_0}{\sqrt{\lambda}}\Big)^{1/4} + C_1' n^{-c_1'/\lambda} + C_1' e^{-c_1' n}.$$

Step 5: Optimizing the parameters. This inequality holds for all $\varepsilon_0 > 0$, so we can optimize in ε_0. Setting $\varepsilon_0 = \varepsilon^{1/2}/\lambda^{1/8}$, we obtain after some simplification that

$$\mathbb{P}_{A,X}(\mathcal{E} \cap \mathcal{E}_K) \le \frac{C_2 \varepsilon^{1/8}}{\lambda^{5/32}} + C_1' n^{-c_1'/\lambda} + C_1' e^{-c_1' n}.$$

By assumption, $\lambda \ge n^{-c_{7.1}}$ where $c_{7.1} > 0$ is a small constant. So, for appropriately chosen constants, the term $n^{-c_1'/\lambda}$ dominates the term $e^{-c_1' n}$. We obtain

$$\mathbb{P}_{A,X}(\mathcal{E} \cap \mathcal{E}_K) \le \frac{C_2 \varepsilon^{1/8}}{\lambda^{5/32}} + 2C_1' n^{-c_1'/\lambda}.$$

Recall that this inequality holds for all $\varepsilon \ge 0$ and $n^{-c_{7.1}} \le \lambda \le c_{oo}/3$, so we can also optimize in λ. For convenience, we isolate this step as a separate elementary observation.

Fact 8.6 (Optimization)**.** *Let $C \ge 1$, $a, b, c', c > 0$. There exists numbers C_0 and n_0 that depend only on a, b, c', C, c and such that the following holds. Let $n \ge n_0$. Consider a function $p(\varepsilon) : [0,1] \to \mathbb{R}_+$ which satisfies*

$$p(\varepsilon) \le M^a \varepsilon^b + n^{-c'M} \quad \text{for all } \varepsilon \in [0,1] \text{ and } C \le M \le n^c.$$

Then

$$p(\varepsilon) \le C_0 \varepsilon^{b-0.01} + n^{-c'n^c} \quad \text{for all } \varepsilon \in [0,1].$$

Proof of Fact 8.6. Choose some number $C \le M_0 \le n^c$ whose value will be determined later. By the assumption, the inequality

$$p \le M_0^a \varepsilon^b + n^{-c'M_0} \le (M_0^a + 1)\varepsilon^b \tag{8.16}$$

holds for all $\varepsilon \ge n^{-c'M_0/b}$. On the other hand, using the assumption with $M = n^c$, we see that the inequality

$$p \le n^{ac} \varepsilon^b + n^{-c'n^c} \le \varepsilon^{b-0.01} + n^{-c'n^c}$$

holds for all $\varepsilon \le n^{-100ac}$. Let us choose M_0 as the minimal number such that $M_0 \ge C$ and $c'M_0/b \ge 100ac$. Note that we have $C \le M \le n^c$ as required, for sufficiently large n_0. Therefore, every ε belongs to the range where inequality (8.16) holds or (8.17) holds, or both. So at least one of these inequalities holds for all $\varepsilon \ge 0$. This completes the proof with $C_0 = M_0^a + 1$. □

Applying Fact 8.6 with $M = 1/\lambda$, $a = 5/32$ and $b = 1/8$, we conclude that

$$\mathbb{P}_{A,X}(\mathcal{E} \cap \mathcal{E}_K) \le C_0 \varepsilon^{1/9} + n^{-c'n^c} \tag{8.17}$$

holds for all $\varepsilon \in [0,1]$, where $c = c_{7.1}$. Since we can choose $C_0 \ge 1$, the same inequality trivially holds for $\varepsilon > 1$ as the right hand side becomes larger than 1. The proof of Theorem 8.1 is complete. □

9 Consequences: the distance problem and invertibility of random matrices

9.1 The distance theorem

An application of Theorem 8.1 together with Proposition 5.1 produces a satisfactory solution to the distance problem posed in the beginning of Section 5.

Corollary 9.1 (Distance between random vectors and subspaces)**.** *Let A be a random matrix satisfying* **(A)***. There exist constants $C, c > 0$ that depend only on the parameters K and M_4 from* (2.3), (2.5)*, and such that the following holds. Let A_k denote the k-th column of A and H_k denote the span of the other columns. For every $\varepsilon \ge 0$, one has*

$$\mathbb{P}\Big\{ \operatorname{dist}(A_k, H_k) \le \varepsilon \wedge \mathcal{E}_K \Big\} \le C_{9.1} \varepsilon^{1/9} + 2\exp(-n^{c_{9.1}}).$$

Proof. By permuting the coordinates, we can assume without loss of generality that $k = 1$. Proposition 5.1 states that

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

where B denotes the $(n-1)\times(n-1)$ minor of A obtained by removing the first row and the first column from A and $X \in \mathbb{R}^{n-1}$ denotes the first column of A with the first entry removed. By assumptions, B is a random matrix which satisfies the same assumptions **(A)** as A (except the dimension is one less), and X is an independent random vector whose entries also satisfy the same assumptions (2.5). So we can apply Theorem 8.1 for B and X. Conditioning on the independent entry $a_{11} = u$, we obtain that

$$\mathbb{P}\Big\{ \frac{\big|\langle B^{-1}X, X\rangle - a_{11}\big|}{\sqrt{1 + \|B^{-1}X\|_2^2}} \le \varepsilon \wedge \mathcal{E}_K \Big\} \le C_{8.1} \varepsilon^{1/9} + 2\exp(-(n-1)^{c_{8.1}}).$$

This completes the proof. $\square$

9.2 Invertibility of random matrices: proof of Theorem 1.1.

We can now derive the main result of the paper, Theorem 1.1. In Section 2.3, we reduced the problem to proving the invertibility bound (2.4). We shall now establish this bound, which immediately implies Theorem 1.1.

Theorem 9.2 (Invertibility of symmetric random matrices)**.** *Let A be a random matrix which satisfies* **(A)***. Consider a number $K > 0$. Then, for all $\varepsilon \ge 0$, one has*

$$\mathbb{P}\Big\{ \min_k |\lambda_k(A)| \le \varepsilon n^{-1/2} \wedge \|A\| \le 3K \Big\} \le C\varepsilon^{1/9} + 2\exp(-n^c),$$

where $C, c > 0$ depend only on the fourth moment bound M_4 from (2.5) *and on* K.

Proof. Denote by p the probability in question. As we observed in Section 2.3,

$$p = \mathbb{P}\Big\{ \min_{x \in S^{n-1}} \|Ax\|_2 \le \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\}.$$

In (3.1), we split the invertibility problem into two, for compressible and incompressible vectors:

$$p \le \mathbb{P}\Big\{ \inf_{x \in \mathrm{Comp}(c_0,c_1)} \|Ax\|_2 \le \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\} \\ + \mathbb{P}\Big\{ \inf_{x \in \mathrm{Incomp}(c_0,c_1)} \|Ax\|_2 \le \varepsilon n^{-1/2} \wedge \mathcal{E}_K \Big\}.$$

The values of c_0, c_1 were then fixed in Remark 4.3. The probability for the compressible vectors is bounded by $2e^{-c_{4.2}n}$ by (4.11). The probability for the incompressible vectors is estimated via distances in Lemma 3.9, see Remark 3.10. This gives

$$p \le 2e^{-c_{4.2}n} + \frac{1}{c_0 n} \sum_{k=1}^{n} \mathbb{P}\big\{ \mathrm{dist}(A_k, H_k) \le c_1^{-1}\varepsilon \wedge \mathcal{E}_K \big\}.$$

Finally, the distances are estimated in Corollary 9.1, which gives

$$p \le 2e^{-c_{4.2}n} + C_{9.1}\varepsilon^{1/9} + 2\exp(-n^{c_{9.1}}).$$

Choosing the values of the constant $c > 0$ sufficiently small, we complete the proof of Theorem 9.2. $\square$

References

[1] N. Alon, J. Spencer, *The probabilistic method*, Second edition. Wiley-Interscience, New York, 2000.

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

[3] K. Costello, *Bilinear and quadratic variants on the Littlewood-Offord problem*, submitted (2009).

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

[5] L. Erdös, B. Schlein, H.-T. Yau, *Wegner estimate and level repulsion for Wigner random matrices*, Int. Math. Res. Not. 3 (2010), 436–479.

[6] F. Götze, *Asymptotic expansions for bivariate von Mises functionals*, Z. Wahrsch. Verw. Gebiete 50 (1979), 333–355.

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

[8] R. Latala, *Some estimates of norms of random matrices*, Proc. Amer. Math. Soc. 133 (2005), 1273–1282.

[9] M. Ledoux, M. Talagrand, *Probability in Banach spaces*, Springer, 1991.

[10] H. Nguyen, *Inverse Littlewood-Offord problems and the singularity of random symmetric matrices*, Duke Math. J., to appear.

[11] H. Nguyen, *On the least singular value of random symmetric matrices*, submitted.

[12] M. Rudelson, *Invertibility of random matrices: norm of the inverse*, Annals of Mathematics 168 (2008), 575–600.

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

[14] M. Rudelson, R. Vershynin, *Smallest singular value of a random rectangular matrix*, Communications on Pure and Applied Mathematics 62 (2009), 1707–1739.

[15] M. Rudelson, R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*, Proceedings of the International Congress of Mathematicians, Hyderabad, India, 2010.

[16] A. Sidorenko, *A correlation inequality for bipartite graphs*, Graphs Combin. 9 (1993), 201–204

[17] T. Tao, V. Vu, *Additive combinatorics.* Cambridge Studies in Advanced Mathematics, 105. Cambridge University Press, Cambridge, 2006.

[18] T. Tao, V. Vu, *From the Littlewood-Offord problem to the Circular Law: universality of the spectral distribution of random matrices*, Bull. Amer. Math. Soc. 46 (2009), 377–396.

[19] T. Tao, V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2) 169 (2009), 595–632.

[20] T. Tao, V. Vu, *Random matrices: the distribution of the smallest singular values*, Geom. Funct. Anal. 20 (2010), 260–297.

[21] T. Tao, V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math. 206 (2011), 127–204.

[22] T. Tao, V. Vu, *Random matrices: Universality of local eigenvalue statistics up to the edge*, Comm. Math. Phys. 298 (2010), 549–572.

[23] T. Tao, V. Vu, *Random matrices: Localization of the eigenvalues and the necessity of four moments*, Acta Mathematica Vietnamica 36 (2011), 431–449.

[24] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, in: Compressed sensing: theory and applications, eds. Y. Eldar and G. Kutyniok, Cambridge University Press, to appear.