Available online at www.sciencedirect.com
ScienceDirect
Advances in Mathematics 218 (2008) 600–633
ADVANCES IN Mathematics
www.elsevier.com/locate/aim

The Littlewood–Offord problem and invertibility of random matrices

Mark Rudelson [a,1], Roman Vershynin [b,*,2]

[a] *Department of Mathematics, University of Missouri, Columbia, MO 65211, USA*
[b] *Department of Mathematics, University of California, Davis, CA 95616, USA*

Received 27 August 2007; accepted 21 January 2008

Available online 7 March 2008

Communicated by Michael J. Hopkins

Abstract

We prove two basic conjectures on the distribution of the smallest singular value of random $n \times n$ matrices with independent entries. Under minimal moment assumptions, we show that the smallest singular value is of order $n^{-1/2}$, which is optimal for Gaussian matrices. Moreover, we give a optimal estimate on the tail probability. This comes as a consequence of a new and essentially sharp estimate in the Littlewood–Offord problem: for i.i.d. random variables X_k and real numbers a_k, determine the probability p that the sum $\sum_k a_k X_k$ lies near some number v. For arbitrary coefficients a_k of the same order of magnitude, we show that they essentially lie in an arithmetic progression of length $1/p$.
Published by Elsevier Inc.

Keywords: Random matrices; Singular values; Condition number; Small ball probability; Littlewood–Offord problem

1. Introduction

1.1. Invertibility of random matrices

In this paper we solve two open problems on the distribution of the smallest singular value of random matrices.

* Corresponding author.
E-mail addresses: rudelson@math.missouri.edu (M. Rudelson), vershynin@math.ucdavis.edu (R. Vershynin).

[1] Supported by NSF DMS grants 0556151 and 0652684.
[2] Supported by the Alfred P. Sloan Foundation and by NSF DMS grants 0401032 and 0652617.

0001-8708/$ – see front matter Published by Elsevier Inc.
doi:10.1016/j.aim.2008.01.010

Let A be an $n \times n$ matrix with real or complex entries. The *singular values* $s_k(A)$ of A are the eigenvalues of $|A| = \sqrt{A^*A}$ arranged in the non-increasing order. Of particular significance are the largest and the smallest singular values

$$s_1(A) = \sup_{x:\ \|x\|_2=1} \|Ax\|_2, \qquad s_n(A) = \inf_{x:\ \|x\|_2=1} \|Ax\|_2.$$

These quantities can obviously be expressed in terms of the spectral norm—the operator norm of A considered as an operator on ℓ_2^n. Indeed, $s_1(A) = \|A\|$, and if the matrix A is non-singular then $s_n(A) = 1/\|A^{-1}\|$. The smallest singular value thus equals the distance from A to the set of singular matrices in the spectral norm.

The behavior of the largest singular value of random matrices A with i.i.d. entries is well studied. The weakest assumption for its regular behavior is boundedness of the *fourth moment* of the entries; then

$$s_1(A) \sim n^{1/2} \quad \text{with high probability.} \tag{1.1}$$

Indeed, by [33,1] the finite fourth moment is *necessary and sufficient* for $s_1(A)/n^{1/2}$ to have an almost sure limit as $n \to \infty$, and this limit equals 2. Latala [15] showed that (1.1) holds under the forth moment assumption even if entries are not identically distributed.

Much less has been known about the behavior of the smallest singular value. In the classic work on numerical inversion of large matrices, von Neumann and his associates used random matrices to test their algorithms, and they speculated that

$$s_n(A) \sim n^{-1/2} \quad \text{with high probability} \tag{1.2}$$

(see [32, pp. 14, 477, 555]). In a more precise form, this estimate was conjectured by Smale [24] and proved by Edelman [6] and Szarek [28] for *random Gaussian matrices* A, those with i.i.d. standard normal entries. Edelman's theorem states that for every $\varepsilon \geqslant 0$

$$\mathbb{P}\big(s_n(A) \leqslant \varepsilon n^{-1/2}\big) \sim \varepsilon. \tag{1.3}$$

Prediction (1.2) for general random matrices has been an open problem, unknown even for the *random sign matrices* A, those whose entries are ± 1 symmetric random variables. In this paper we prove the prediction (1.2) in full generality under the aforementioned fourth moment assumption.

Theorem 1.1 *(Invertibility: fourth moment). Let A be an $n \times n$ matrix whose entries are independent real random variables with variances at least 1 and fourth moments bounded by B. Then, for every $\delta > 0$ there exist $\varepsilon > 0$ and n_0 which depend (polynomially) only on δ and B, and such that*

$$\mathbb{P}\big(s_n(A) \leqslant \varepsilon n^{-1/2}\big) \leqslant \delta \quad \textit{for all } n \geqslant n_0.$$

This shows in particular that the median of $s_n(A)$ is of order $n^{-1/2}$.

Under stronger moment assumptions, more is known about the distribution of the largest singular value, and similarly one hopes to know more about the smallest singular value.

Indeed, Soshnikov [25] proved that the limiting distribution of $s_1(A)$ is precisely the Tracy–Widom law for all matrices with i.i.d. subgaussian entries. Recall that a random variable ξ is called *subgaussian* if its tail is dominated by that of the standard normal random variable: there exists $B > 0$ such that

$$\mathbb{P}\big(|\xi| > t\big) \leqslant 2\exp\big(-t^2/B^2\big) \quad \text{for all } t > 0. \tag{1.4}$$

The minimal B here is called the *subgaussian moment*[3] of ξ. Inequality (1.4) is often equivalently formulated as a moment condition

$$\big(\mathbb{E}|\xi|^p\big)^{1/p} \leqslant CB\sqrt{p} \quad \text{for all } p \geqslant 1, \tag{1.5}$$

where C is an absolute constant. The class of subgaussian random variables includes many random variables that arise naturally in applications, such as normal, symmetric ± 1, and in general all bounded random variables.

One might then expect that the estimate (1.3) for the distribution of the smallest singular value of Gaussian matrices should hold for all subgaussian matrices. Note however that (1.3) fails for the random sign matrices, since they are singular with positive probability. Estimating the singularity probability for random sign matrices is a longstanding open problem. Even proving that it converges to 0 as $n \to \infty$ is a nontrivial result due to Komlós [14]. Later Kahn, Komlós and Szemerédi [13] showed that it is exponentially small

$$\mathbb{P}(\text{random sign matrix } A \text{ is singular}) < c^n \tag{1.6}$$

for some universal constant $c \in (0, 1)$. The often conjectured optimal value of c is $1/2 + o(1)$ [20,13], and the best known value $3/4 + o(1)$ is due to Tao and Vu [29,30].

Spielman and Teng [26] conjectured that (1.3) should hold for the random sign matrices up to an exponentially small term that accounts for their singularity probability

$$\mathbb{P}\big(s_n(A) \leqslant \varepsilon n^{-1/2}\big) \leqslant \varepsilon + c^n.$$

In this paper, we prove Spielman–Teng's conjecture for all matrices with subgaussian i.i.d. entries, and up to a constant factor which depends only on the subgaussian moment.

Theorem 1.2 *(Invertibility: subgaussian). Let $\xi_1, \dots, \xi_n$ be independent centered real random variables with variances at least 1 and subgaussian moments bounded by B. Let A be an $n \times n$ matrix whose rows are independent copies of the random vector $(\xi_1, \dots, \xi_n)$. Then for every $\varepsilon \geqslant 0$ one has*

$$\mathbb{P}\big(s_n(A) \leqslant \varepsilon n^{-1/2}\big) \leqslant C\varepsilon + c^n, \tag{1.7}$$

where $C > 0$ and $c \in (0, 1)$ depend (polynomially) only on B.

[3] In the literature in geometric functional analysis, the subgaussian moment is often called the ψ_2-norm.

Remarks. 1. For $\varepsilon = 0$, Theorem 1.2 yields an exponential bound for the singularity probability

$$\mathbb{P}(\text{random subgaussian matrix } A \text{ is singular}) < c^n.$$

Thus Kahn–Komlós–Szemerédi's bound (1.6) holds for all subgaussian matrices. Moreover, while (1.6) estimates the probability that a random matrix *belongs* to the set of singular matrices, Theorem 1.2 estimates the *distance* to that set.

2. The bounds in Theorem 1.2 are precise. Edelman's bound (1.3) shows that the term $\varepsilon n^{-1/2}$ is optimal for the Gaussian matrix, while the term c^n is optimal for a random sign matrix.

3. For simplicity, we state and prove all our results over the real field. However, our arguments easily generalize to the complex field; see e.g. [21].

4. A weaker result was recently proved by the first author [22] who showed that $\mathbb{P}(s_n(A) \leqslant \varepsilon n^{-3/2}) \leqslant C\varepsilon + Cn^{-1/2}$. He later improved the term $n^{-1/2}$ to c^n. Shortly after that, both authors of this paper independently discovered how to reduce the term $n^{-3/2}$ to the sharp order $n^{-1/2}$. In December 2006, the second author found a new way to prove the sharp invertibility estimate by obtaining an essentially optimal result for the Littlewood–Offord problem as stated in Theorem 1.5. We thus decided to publish jointly, and some of the arguments were improved during the final stage of our work.

5. Another weaker result was recently proved by Tao and Vu [31] for random sign matrices. They showed that for every $A > 0$ there exists $B > 0$ such that $s_n(A) \geqslant n^{-B}$ holds with probability $1 - O_A(n^{-A})$.

1.2. The Littlewood–Offord problem

Our results on random matrices come as a consequence of a new and essentially sharp estimate in the Littlewood–Offord problem [2,10]. A classical theme in Probability Theory is the study of the random sums

$$S := \sum_{k=1}^{n} a_k \xi_k, \tag{1.8}$$

where $\xi_1, \dots, \xi_n$ are independent identically distributed random variables and $a = (a_1, \dots, a_n)$ is a vector of real coefficients.

The large deviation theory demonstrates that S nicely concentrates around its mean. On the other hand, by the central limit theorem, one cannot expect tighter concentration than that of the appropriately scaled Gaussian random variable. However, rigorous anti-concentration estimates are hard to prove (see [17]), especially for discrete random variables ξ_k. The Littlewood–Offord problem thus asks to estimate the *small ball probability*

$$p_\varepsilon(a) := \sup_{v \in \mathbb{R}} \mathbb{P}\big(|S - v| \leqslant \varepsilon\big).$$

A small value of $p_\varepsilon(a)$ would mean that the random sums S are well spread.

For the *random Gaussian sums*, i.e. for ξ_k being standard normal random variables, the small ball probability for each ε depends only on the Euclidean norm of the coefficient vector a and not on its direction, and one has $p_\varepsilon(a) \sim \varepsilon / \|a\|_2$.

For most other distributions, $p_\varepsilon(a)$ depends on the direction of a, and determining the asymptotics is hard. A remarkable and extensively studied case is for the *random sign-sums* $\sum \pm a_k$,

i.e. for symmetric ± 1 random variables ξ_k. The small ball probability strongly depends on the direction of the coefficient vector: for example, $p_0(a) = 1/2$ for $a = (1, 1, 0, \dots, 0)$ while $p_0(a) \sim n^{-1/2}$ for $a = (1, 1, \dots, 1)$.

The coefficient vectors with few nonzero coordinates turn out to be the only obstacle for nontrivial estimates on the small ball probability. The classical result of Littlewood and Offord strengthened by Erdös [7] states that if all $|a_k| \geqslant 1$ then for the random sign-sums one has

$$p_1(a) \lesssim n^{-1/2}. \tag{1.9}$$

This is sharp for $a_k = 1$: there are lots of cancelations in most of the sign-sums $\sum \pm 1$. However, if $|a_j - a_k| \geqslant 1$ for $k \neq j$, then the small ball probability is even smaller:

$$p_1(a) \lesssim n^{-3/2}. \tag{1.10}$$

This was proved by Erdös and Moser [8] for $p_0(a)$ and with an extra $\log n$ factor, which was removed by Sárközi and Szemerédi [23]. Hálasz [12] proved this estimate for $p_1(a)$ and generalized it to higher dimensions. Estimate (1.10) is sharp for $a_k = k$: there are still many cancelations in most of the sign-sums $\sum \pm k$.

Tao and Vu [31] recently proposed a method to reduce the small ball probability to an arbitrary polynomial order. They suggested to look at the inverse problem and to study the following phenomenon:

If the small ball probability $p_0(a)$ is large then the coefficient vector a has a rich additive structure.

Thus, the only reason for many cancelations in the sign-sums $\sum \pm a_k$ is that most coefficients a_k are arithmetically well comparable. By removing this obstacle one can force the small ball probability down to an arbitrary polynomial order:

Theorem 1.3. *(See Tao and Vu [31].) Let $a_1, \dots, a_n$ be integers, and let $A \geqslant 1$, $\varepsilon \in (0, 1)$. Suppose for the random sign-sums one has*

$$p_0(a) \geqslant n^{-A}.$$

Then all except $O_{A,\varepsilon}(n^{\varepsilon})$ coefficients a_k are contained in the Minkowski sum of $O(A/\varepsilon)$ arithmetic progressions of lengths $n^{O_{A,\varepsilon}(1)}$.

(Recall that the Minkowski sum of sets is defined as $U + V = \{u + v\colon\ u \in U,\ v \in V\}$.)

In this paper we demonstrate that a similar, and even simpler, phenomenon holds for real rather than integer numbers a_k, for the small ball probabilities $p_\varepsilon(a)$ rather than the probability $p_0(a)$ of exact values, and for general random sums (1.8) rather than the random sign-sums.

We thus provide an essentially sharp solution to the Littlewood–Offord problem for coefficients a_k of equal order of magnitude. We show that one can force the small ball probability $p_\varepsilon(a)$ down to an *arbitrary function* of n, up to an exponentially small order, which is best possible. We prove that:

The coefficients of a are essentially contained in one arithmetic progression of length $\lesssim p_\varepsilon(a)^{-1}$.

By "essentially" we mean that for arbitrary $\alpha \in (0, 1)$ and $\kappa > c(\alpha)$ we can guarantee that all but κ coefficients a_k are within αd from the elements of some arithmetic progression, where d is the gap between its elements. It is convenient to state this result in terms of the essential least common denominator of real numbers:

Definition 1.4 *(Essential LCD).* Let $\alpha \in (0, 1)$ and $\kappa \geqslant 0$. The *essential least common denominator* $D(a) = D_{\alpha,\kappa}(a)$ of a vector $a \in \mathbb{R}^n$ is defined as the infimum of $t > 0$ such that all except κ coordinates of the vector ta are of distance at most α from nonzero integers.

For numbers $a_k = O(1)$, the essential LCD has an obvious interpretation in terms of arithmetic progressions: all except κ coefficients a_k are within distance $\alpha/D(a) = O(\alpha)$ from the elements of an arithmetic progression of length $O(D(a))$.

Theorem 1.5 *(Small Ball Probability). Let $\xi_1, \dots, \xi_n$ be independent identically distributed centered random variables with variances at least 1 and third moments bounded by B. Let $a = (a_1, \dots, a_n)$ be a vector of real coefficients such that, for some $K_1, K_2 > 0$ one has*

$$K_1 \leqslant |a_k| \leqslant K_2 \quad \text{for all } k. \tag{1.11}$$

Let $\alpha \in (0, 1)$ and $\kappa \in (0, n)$. Then for every $\varepsilon \geqslant 0$ one has

$$p_\varepsilon(a) \leqslant \frac{C}{\sqrt{\kappa}}\left(\varepsilon + \frac{1}{D_{\alpha,\kappa}(a)}\right) + Ce^{-c\alpha^2\kappa},$$

where $C, c > 0$ depend (polynomially) only on B, K_1, K_2.

A more precise version of this result is Theorem 4.1 below.

Remarks. 1. By the definition, one always has $D_{\alpha,\kappa}(a) \gtrsim 1/K_2$ (e.g. with $\alpha = 1/3$ and $\kappa = n/4$). Theorem 1.5 thus yields $p_1(a) \lesssim n^{-1/2}$, which agrees with Littlewood–Offord and Erdös inequality (1.9).

2. Suppose the components of a are uniformly spread between two comparable values; say $a = (n, n + 1, n + 2, \dots, 2n)$. Obviously, $D_{\alpha,\kappa}(a/n) \sim n$ (e.g. with $\alpha = 1/3$ and $\kappa = n/4$). Theorem 1.5 thus yields $p_1(a) = p_{1/n}(a/n) \lesssim n^{-3/2}$. This agrees with Erdös–Moser inequality (1.10).

3. By making coefficients of a more arithmetically incomparable, such as by considering polynomial progressions, one can force the small ball probability $p_\varepsilon(a)$ down to an arbitrarily small value, up to an exponentially small order.

One can restate Theorem 1.5 as an inverse Littlewood–Offord theorem:

Corollary 1.6 *(Inverse Littlewood–Offord Theorem). Let $a_1, \dots, a_n$ be real numbers satisfying* (1.11) *and $\xi_1, \dots, \xi_n$ be random variables as in Theorem* 1.5. *Let $A \geqslant 1/2$, $\kappa \in (0, n)$ and $\varepsilon > 0$. Suppose for the random sums* (1.8) *one has*

$$p_\varepsilon(a) \geqslant n^{-A}.$$

Then there exists an arithmetic progression of length $L = O(n^A \kappa^{-1/2})$ *and with gap between its elements* $d \leqslant 1$*, and such that all except* κ *coefficients* a_k *are within distance* $O(A \log(n)/\kappa)^{1/2} \cdot d$ *from the elements of the progression, provided that* $\varepsilon \leqslant 1/L$.

By Remark 1 above, the assumption $A \geqslant 1/2$ is optimal.

In contrast with Theorem 1.3, Corollary 1.6 guarantees an approximate, rather than exact, embedding of the coefficients $a_1, \dots, a_n$ into an arithmetic progression. On the other hand, Corollary 1.6: (a) applies for real rather integer coefficients; (b) embeds into one arithmetic progression rather than a Minkowski sum of several progressions; (c) provides a significantly sharper bound on the length of the progression; (d) characterizes general small ball probabilities $p_\varepsilon(a)$ rather than the probability of exact values $p_0(a)$; (e) holds for general sums of i.i.d. random variables rather than the random sign-sums.

1.3. Outline of the argument

We develop a general approach to the invertibility of random matrices. Our main result, the Strong Invertibility Theorem 5.1, *reduces estimating the smallest singular value of random matrices to estimating the largest singular value.* Because the largest singular value is much more studied, this immediately implies both our invertibility results stated above, Theorems 1.1 and 1.2.

The general approach to invertibility is developed in two stages. In Section 3 we present a "soft" and rather short argument that leads to a weaker result. It yields the Fourth Moment Theorem 1.1 and also a weaker version of the Subgaussian Theorem 1.2 with $Cn^{-1/2}$ instead of the exponential term c^n in (1.7).

Our soft argument does not use any new estimates of the small ball probability. To bound $\|Ax\|_2$ below for all vectors x in the unit sphere, we give two separate arguments for *compressible* vectors x, whose norm is concentrated in a small number of coordinates, and for *incompressible* vectors comprising the rest of the sphere.

For a compressible vector, the main contribution in the quantity $\|Ax\|_2$ comes from the few (say, $n/10$) columns of A corresponding to the biggest coordinates of x. This allows us to replace A by its $n \times n/10$ submatrix with the chosen columns. Such rectangular random matrices are known to have big smallest singular value (see e.g. [18]), which establishes a nice lower bound on $\|Ax\|_2$ for all compressible vectors.

For the incompressible vectors, we show the invertibility differently. Clearly, $s_n(A)$ is bounded above by the distance from its nth row vector X_n to the span H_n of the others. We use a careful *average union* argument (Lemma 3.5) to show a reverse inequality for A restricted to the set of incompressible vectors.

Next, this distance can be bounded below as $\operatorname{dist}(X_n, H_n) \geqslant |\langle X^*, X_n \rangle|$, where X^* is a unit normal of H_n. Since X^* and H_n are independent, the inner product $\langle X^*, X_n \rangle$ can be written as a sum of independent random variables of the form (1.8). This reduces the invertibility problem to the Littlewood–Offord problem.

A useful small ball probability bound can be deduced from the central limit theorem, by approximating the random sum (1.8) with a Gaussian random variable for which the small ball probability is easy to compute. With such bound, the argument above yields a weaker version of the invertibility estimate (1.7) with $Cn^{-1/2}$ instead of c^n.

This weaker estimate is a limitation of using the central limit theorem. To prove the Strong Invertibility Theorem 5.1, and thus deduce the Subgaussian Theorem 1.2, we will use the full

strength of the Small Ball Probability Theorem 1.5 instead. This argument is presented in Section 5.

Our proof of Theorem 1.5 starts with the method developed by Halász [11,12]. It allows us to bound the small ball probability $p_\varepsilon(a)$ by a quantity of ergodic nature—the measure of the *recurrence set* of a. It indicates how often a particle in $\mathbb{R}^n$ moving in the direction a with unit speed gets close to the points of the integer lattice. If this happens often, then a density argument shows that the particle must get close to two distinct lattice points over a short period of time, say at times t_1 and t_2. It then follows that $(t_2 - t_1)a$ is close to an integer, which implies that the essential LCD of a is small. This argument is given in Section 4.

2. Preliminaries

In the sequel n denotes a sufficiently large integer, i.e. an integer bigger than a suitable absolute constant. The standard inner product on $\mathbb{R}^n$ is denoted by $\langle x, y\rangle$. The ℓ_p norm on $\mathbb{R}^n$ is defined as $\|x\|_p = (\sum_{k=1}^n |x_k|^p)^{1/p}$ for $0 < p < \infty$, and $\|x\|_\infty = \max_k |x_k|$. The unit Euclidean ball and the sphere in $\mathbb{R}^n$ are denoted by B_2^n and S^{n-1} respectively. For a subset $\sigma \subseteq \{1, \dots, n\}$, the orthogonal projection onto $\mathbb{R}^\sigma$ in $\mathbb{R}^n$ is denoted by P_σ.

The following observation will allow us to select a nice subset of the coefficients a_k when computing the small ball probability.

Lemma 2.1 *(Restriction). For any $a \in \mathbb{R}^n$, any $\sigma \subseteq \{1, \dots, n\}$ and any $\varepsilon \geqslant 0$, we have*

$$p_\varepsilon(a) \leqslant p_\varepsilon(P_\sigma a).$$

Proof. For fixed $v \in \mathbb{R}$ and for the random sum (1.8), we write $S - v = S_\sigma - v_\sigma$, where $S_\sigma := \sum_{k\in\sigma} a_k\xi_k$ and $v_\sigma := v - \sum_{k\in\sigma^c} a_k\xi_k$. We condition on a realization of $(\xi_k)_{k\in\sigma^c}$, and denote by $\mathbb{P}_\sigma$ the probability with respect to $(\xi_k)_{k\in\sigma}$. Then a realization of v_σ is fixed, so

$$\mathbb{P}_\sigma\big(|S - v| \leqslant \varepsilon\big) = \mathbb{P}_\sigma\big(|S_\sigma - v_\sigma| \leqslant \varepsilon\big) \leqslant p_\varepsilon(P_\sigma a).$$

Taking the expectation of both sides with respect to $(\xi_k)_{k\in\sigma^c}$ completes the proof. □

The following tensorization lemma transfers one-dimensional small ball probability estimates to the multidimensional case. It is a minor variant of Lemma 4.4 of [22].

Lemma 2.2 *(Tensorization). Let $\zeta_1, \dots, \zeta_n$ be independent non-negative random variables, and let $K, \varepsilon_0 \geqslant 0$.*

(1) *Assume that for each k*

$$\mathbb{P}(\zeta_k < \varepsilon) \leqslant K\varepsilon \quad \text{for all } \varepsilon \geqslant \varepsilon_0.$$

Then

$$\mathbb{P}\left(\sum_{k=1}^n \zeta_k^2 < \varepsilon^2 n\right) \leqslant (CK\varepsilon)^n \quad \text{for all } \varepsilon \geqslant \varepsilon_0,$$

where C is an absolute constant.

(2) *Assume that there exist* $\lambda > 0$ *and* $\mu \in (0,1)$ *such that for each* k

$$\mathbb{P}(\zeta_k < \lambda) \leqslant \mu.$$

Then there exist $\lambda_1 > 0$ *and* $\mu_1 \in (0,1)$ *that depend on* λ *and* μ *only and such that*

$$\mathbb{P}\left(\sum_{k=1}^{n} \zeta_k^2 < \lambda_1 n\right) \leqslant \mu_1^n.$$

We give a proof of the first part for completeness. The second part is similar, cf. [18] proof of Proposition 3.4.

Proof. Let $\varepsilon \geqslant \varepsilon_0$. By Chebychev's inequality,

$$\begin{aligned}\mathbb{P}\left(\sum_{k=1}^{n} \zeta_k^2 < \varepsilon^2 n\right) &= \mathbb{P}\left(n - \frac{1}{\varepsilon^2}\sum_{k=1}^{n} \zeta_k^2 > 0\right) \leqslant \mathbb{E}\exp\left(n - \frac{1}{\varepsilon^2}\sum_{k=1}^{n} \zeta_k^2\right) \\ &= e^n \prod_{k=1}^{n} \mathbb{E}\exp\left(-\zeta_k^2/\varepsilon^2\right). \end{aligned} \tag{2.1}$$

By the distribution integral formula,

$$\mathbb{E}\exp\left(-\zeta_k^2/\varepsilon^2\right) = \int_0^1 \mathbb{P}\left(\exp\left(-\zeta_k^2/\varepsilon^2\right) > s\right) ds = \int_0^\infty 2ue^{-u^2}\mathbb{P}(\zeta_k < \varepsilon u)\, du.$$

For $u \in (0,1)$, we have $\mathbb{P}(\zeta_k < \varepsilon u) \leqslant \mathbb{P}(\zeta_k < \varepsilon) \leqslant K\varepsilon$. This and the assumption of the lemma yields

$$\mathbb{E}\exp\left(-\zeta_k^2/\varepsilon^2\right) \leqslant \int_0^1 2ue^{-u^2} K\varepsilon\, du + \int_1^\infty 2ue^{-u^2} K\varepsilon u\, du \leqslant CK\varepsilon.$$

Putting this into (2.1) yields

$$\mathbb{P}\left(\sum_{k=1}^{n} \zeta_k^2 < \varepsilon^2 n\right) \leqslant e^n (CK\varepsilon)^n.$$

This completes the proof. □

2.1. Largest singular value

We recall some known bounds on the largest singular value of random matrices under the fourth moment assumption and the subgaussian moment assumption. The following result is a partial case of a recent result of Latala.

Theorem 2.3 *(Largest singular value: fourth moment [15]). Let A be an $n \times n$ matrix whose entries are independent centered random variables with variances at least* 1 *and fourth moments bounded by B. Then*

$$\mathbb{E}\|A\| \leqslant C_1 n^{1/2}$$

where $C_1 = CB^{1/4}$, and where C is an absolute constant.

Under the stronger subgaussian moment assumption, a standard observation shows that $\|A\| \sim n^{1/2}$ with exponentially large probability (see e.g. [4] and [18, Fact 2.4]):

Lemma 2.4 *(Largest singular value: subgaussian). Let A be an $n \times n$ matrix whose entries are independent centered random variables with variances at least* 1 *and subgaussian moments bounded by B. Then*

$$\mathbb{P}\big(\|A\| > C_1 n^{1/2}\big) \leqslant 2e^{-n},$$

where C_1 depends only on B.

2.2. Smallest singular value of rectangular matrices

Estimates on the smallest singular value are known for *rectangular* random matrices [18].

Proposition 2.5 *(Smallest singular value of rectangular matrices). Let G be an $n \times k$ matrix whose entries are independent centered random variables with variances at least* 1 *and fourth moments bounded by B. Let $K \geqslant 1$. Then there exist $c_1, c_2 > 0$ and $\delta_0 \in (0,1)$ that depend only on B and K such that if $k < \delta_0 n$ then*

$$\mathbb{P}\Big(\inf_{x \in S^{k-1}} \|Gx\|_2 \leqslant c_1 n^{1/2} \text{ and } \|G\| \leqslant K n^{1/2}\Big) \leqslant e^{-c_2 n}. \tag{2.2}$$

Under the stronger subgaussian assumption, the condition $\|G\| \leqslant K n^{1/2}$ can clearly be removed from (2.2) by Lemma 2.4. This is not so under the fourth moment assumption. So here and later in the paper, this condition will often appear in order to deduce the Fourth Moment Theorem 1.1. The reader interested only in the Subgaussian Theorem 1.2 can disregard this condition.

A result stronger than Proposition 2.5, for the aspect ratio δ_0 arbitrarily close to 1, follows by modifying the argument of [18]. For completeness, we shall prove Proposition 2.5. We start with the most general (but weakest possible) estimate on the small ball probability.

Lemma 2.6. *Let $\xi_1, \dots, \xi_n$ be independent centered random variables with variances at least* 1 *and fourth moments bounded by B. Then there exists $\mu \in (0,1)$ depending only on B, such that for every coefficient vector $a = (a_1, \dots, a_n) \in S^{n-1}$ the random sum $S = \sum_{k=1}^n a_k \xi_k$ satisfies*

$$\mathbb{P}\big(|S| < 1/2\big) \leqslant \mu.$$

Proof. Let $\varepsilon_1, \dots, \varepsilon_n$ be independent symmetric ± 1 random variables, which are independent of $\xi_1, \dots, \xi_n$. By the standard symmetrization inequality (see [16] Lemma 6.3),

$$\mathbb{E}S^4 \leqslant 16\mathbb{E}\left(\sum_{k=1}^n \varepsilon_k \xi_k a_k\right)^4.$$

We first condition on $\xi_1, \dots, \xi_n$ and take the expectation with respect to $\varepsilon_1, \dots, \varepsilon_n$. Khinchine's inequality (see e.g. [16, Lemma 4.1]) and our assumptions on ξ_k then yield

$$\begin{aligned}\mathbb{E}S^4 &\leqslant C\mathbb{E}\left(\sum_{k=1}^n \xi_k^2 a_k^2\right)^2 = C\mathbb{E}\sum_{k,j=1}^n \xi_k^2 \xi_j^2 a_k^2 a_j^2 \\ &\leqslant C\sum_{k,j=1}^n \left(\mathbb{E}\xi_k^4\right)^{1/2}\left(\mathbb{E}\xi_j^4\right)^{1/2} a_k^2 a_j^2 \leqslant CB\left(\sum_{k=1}^n a_k^2\right)^2 = CB.\end{aligned}$$

The Paley–Zygmund inequality (see e.g. [18, Lemma 3.5]) implies that for any $\lambda > 0$

$$\mathbb{P}\big(|S| > \lambda\big) \geqslant \frac{(\mathbb{E}S^2 - \lambda^2)^2}{\mathbb{E}S^4} \geqslant \frac{(1-\lambda^2)^2}{CB}.$$

To finish the proof, set $\lambda = 1/2$. □

Combining Lemma 2.6 with the tensorization Lemma 2.2, we obtain the following invertibility estimate for a fixed vector.

Corollary 2.7. *Let G be a matrix as in Proposition* 2.5. *Then there exist constants $\eta, \nu \in (0,1)$ depending only on B, such that for every $x \in S^{k-1}$*

$$\mathbb{P}\big(\|Gx\|_2 < \eta n^{1/2}\big) \leqslant \nu^n.$$

Proof of Proposition 2.5. Let $\varepsilon > 0$ to be chosen later. There exists an ε-net $\mathcal{N}$ in S^{k-1} (in the Euclidean norm) of cardinality $|\mathcal{N}| \leqslant (3/\varepsilon)^k$ (see e.g. [19]). Let η and ν be the numbers in Corollary 2.7. Then by the union bound,

$$\mathbb{P}\big(\exists x \in \mathcal{N}\colon\ \|Gx\|_2 < \eta n^{1/2}\big) \leqslant (3/\varepsilon)^k \cdot \nu^n. \tag{2.3}$$

Let V be the event that $\|G\| \leqslant Kn^{1/2}$ and $\|Gy\|_2 \leqslant \frac{1}{2}\eta n^{1/2}$ for some point $y \in S^{k-1}$. Assume that V occurs, and choose a point $x \in \mathcal{N}$ such that $\|y - x\|_2 < \varepsilon$. Then

$$\|Gx\|_2 \leqslant \|Gy\|_2 + \|G\| \cdot \|x - y\|_2 \leqslant \frac{1}{2}\eta n^{1/2} + Kn^{1/2} \cdot \varepsilon = \eta n^{1/2},$$

if we set $\varepsilon = \eta/2K$. Hence, by (2.3),

$$\mathbb{P}(V) \leqslant \big(\nu \cdot (3/\varepsilon)^{k/n}\big)^n \leqslant e^{-c_2 n},$$

if we assume that $k/n \leqslant \delta_0$ for an appropriately chosen $\delta_0 < 1$. This completes the proof. □

2.3. The small ball probability via the central limit theorem

The central limit theorem can be used to estimate the small ball probability, as observed in [18]. Specifically, one can use the Berry–Esseen version of the central limit theorem (see [27, Section 2.1]):

Theorem 2.8 *(Berry–Esseen CLT). Let $\zeta_1, \dots, \zeta_n$ be independent centered random variables with finite third moments, and let $\sigma^2 := \sum_{k=1}^n \mathbb{E}|\zeta_k|^2$. Consider a standard normal random variable g. Then for every $t > 0$*

$$\left| \mathbb{P}\left(\frac{1}{\sigma} \sum_{k=1}^n \zeta_k \leqslant t \right) - \mathbb{P}(g \leqslant t) \right| \leqslant C\sigma^{-3} \sum_{k=1}^n \mathbb{E}|\zeta_k|^3, \tag{2.4}$$

where C is an absolute constant.

The following corollary is essentially given in [18]. We shall include a proof for the reader's convenience.

Corollary 2.9 *(Small ball probability via CLT). Let $\xi_1, \dots, \xi_n$ be independent centered random variables with variances at least 1 and third moments bounded by B. Then for every $a \in \mathbb{R}^n$ and every $\varepsilon \geqslant 0$, one has*

$$p_\varepsilon(a) \leqslant \sqrt{\frac{2}{\pi}} \frac{\varepsilon}{\|a\|_2} + C_1 B \left(\frac{\|a\|_3}{\|a\|_2} \right)^3,$$

where C_1 is an absolute constant.

Proof. We shall use Theorem 2.8 for $\zeta_k = a_k \xi_k$. There, $\sigma \geqslant \|a\|_2$ and $\sum_{k=1}^n \mathbb{E}|\zeta_k|^3 \leqslant B\|a\|_3^3$. Thus for every $u \in \mathbb{R}$ we have

$$\mathbb{P}\left(\left| \frac{1}{\|a\|_2} \sum_{k=1}^n a_k \xi_k - u \right| \leqslant t \right) \leqslant \mathbb{P}\big(|g - u| \leqslant t\big) + 2CB \left(\frac{\|a\|_3}{\|a\|_2} \right)^3. \tag{2.5}$$

Since the density of the standard normal random variable g is uniformly bounded by $1/\sqrt{2\pi}$, we have

$$\mathbb{P}\big(|g - u| \leqslant t\big) \leqslant \frac{2t}{\sqrt{2\pi}} = \sqrt{\frac{2}{\pi}} t.$$

With $u = \frac{v}{\|a\|_2}$ and $t = \frac{\varepsilon}{\|a\|_2}$, the left-hand side of (2.5) equals $\mathbb{P}(|S - v| \leqslant \varepsilon)$, which completes the proof with $C_1 = 2C$. □

As an immediate corollary, we get:

Corollary 2.10 *(Small ball probability for big ε). Let $\xi_1, \dots, \xi_n$ be independent centered random variables with variances at least* 1 *and third moments bounded by B. Assume that a coefficient vector a satisfies* (1.11). *Then for every $\varepsilon \geqslant 0$ one has*

$$p_\varepsilon(a) \leqslant \frac{C_2}{\sqrt{n}}\big(\varepsilon/K_1 + B(K_2/K_1)^3\big),$$

where C_2 is an absolute constant.

3. Invertibility of random matrices: Soft approach

In this section, we develop a soft approach to the invertibility of random matrices. Instead of using the new estimates on the small ball probability, we will rely on the central limit theorem (Corollary 2.10). This approach will yield a weaker bound, with polynomial rather than exponential term for the singularity probability. In Section 5 we shall improve upon the weak point of this argument, so the Small Ball Probability Theorem 1.5 will be used instead.

Theorem 3.1 *(Weak invertibility). Let A be an $n \times n$ matrix whose entries are independent random variables with variances at least* 1 *and fourth moments bounded by B. Let $K \geqslant 1$. Then for every $\varepsilon \geqslant 0$ one has*

$$\mathbb{P}\big(s_n(A) \leqslant \varepsilon n^{-1/2}\big) \leqslant C\varepsilon + Cn^{-1/2} + \mathbb{P}\big(\|A\| > Kn^{1/2}\big), \tag{3.1}$$

where C depends (polynomially) only on B and K.

To make this bound useful, we recall that the last term in (3.1) can be bounded using Theorem 2.3 under the fourth moment assumption and by Lemma 2.4 under the subgaussian assumption. In particular, this proves Fourth Moment Theorem 1.1:

Proof of the Fourth Moment Theorem 1.1. Let $\delta > 0$. By Theorem 2.3 and using Chebychev's inequality, we have

$$\mathbb{P}\left(\|A\| > \frac{3C_1}{\delta} n^{1/2}\right) < \delta/3.$$

Then setting $K = 3C_1/\delta$, $\varepsilon = \delta/3C$ and $n_0 = (3C/\delta)^2$, we make each of the three terms in the right-hand side of (3.1) bounded by $\delta/3$. This completes the proof. □

Remark. Theorem 3.1 in combination with Lemma 2.4 yields a weaker version of the Subgaussian Theorem 1.2, with $Cn^{-1/2}$ instead of c^n.

3.1. Decomposition of the sphere

To prove Theorem 3.1, we shall partition the unit sphere S^{n-1} into the two sets of *compressible* and *incompressible* vectors, and will show the invertibility of A on each set separately.

Definition 3.2 *(Compressible and incompressible vectors).* Let $\delta, \rho \in (0,1)$. A vector $x \in \mathbb{R}^n$ is called *sparse* if $|\mathrm{supp}(x)| \leqslant \delta n$. A vector $x \in S^{n-1}$ is called *compressible* if x is within Euclidean distance ρ 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 sparse, compressible and incompressible vectors will be denoted by $\mathit{Sparse} = \mathit{Sparse}(\delta)$, $\mathit{Comp} = \mathit{Comp}(\delta, \rho)$ and $\mathit{Incomp} = \mathit{Incomp}(\delta, \rho)$ respectively.

Remarks. 1. Here we borrow the terminology from the signal processing and the sparse approximation theory. Efficient compression of many real-life signals, such as images and sound, relies on the assumption that their coefficients (Fourier, wavelet, frame, etc.) decay in a fast way. Essential information about the signal is thus contained in few most significant coefficients, which can be stored in small space (see [5,3]). Such coefficient vector is close to a sparse vector, and is thus compressible in the sense of our definition.

2. Sets similar to those of compressible and incompressible vectors were previously used for the invertibility problem in [18,22].

3. In our argument, the parameters δ, ρ will be chosen as small constants that depend only on B and K.

Using the decomposition of the sphere $S^{n-1} = \mathit{Comp} \cup \mathit{Incomp}$, we break the invertibility problem into two subproblems, for compressible and incompressible vectors:

$$\begin{aligned}
&\mathbb{P}\big(s_n(A) \leqslant \varepsilon n^{-1/2} \text{ and } \|A\| \leqslant K n^{1/2}\big) \\
&\quad \leqslant \mathbb{P}\Big(\inf_{x \in \mathit{Comp}(\delta,\rho)} \|Ax\|_2 \leqslant \varepsilon n^{-1/2} \text{ and } \|A\| \leqslant K n^{1/2}\Big) \\
&\qquad + \mathbb{P}\Big(\inf_{x \in \mathit{Incomp}(\delta,\rho)} \|Ax\|_2 \leqslant \varepsilon n^{-1/2} \text{ and } \|A\| \leqslant K n^{1/2}\Big). \qquad (3.2)
\end{aligned}$$

The compressible vectors are close to a coordinate subspace of a small dimension δn. The restriction of our random matrix A onto such a subspace is a random *rectangular* $n \times \delta n$ matrix. Such matrices are well invertible with exponentially high probability (see Proposition 2.5). By taking the union bound over all coordinate subspaces, we will deduce the invertibility of the random matrix on the set of compressible vectors.

Showing the invertibility on the set of incompressible vectors is generally harder, for this set is bigger in some sense. By a careful *average union* argument, we shall reduce the problem to a small ball probability estimate.

3.2. Invertibility for the compressible vectors

On the set of compressible vectors, a much stronger invertibility holds than we need in (3.2):

Lemma 3.3 *(Invertibility for compressible vectors). Let A be a random matrix as in Theorem* 3.1*, and let $K \geqslant 1$. Then there exist $\delta, \rho, c_3, c_4 > 0$ that depend only on B and K, and such that*

$$\mathbb{P}\Big(\inf_{x \in \mathit{Comp}(\delta,\rho)} \|Ax\|_2 \leqslant c_3 n^{1/2} \text{ and } \|A\| \leqslant K n^{1/2}\Big) \leqslant e^{-c_4 n}.$$

Remark. The bound in Lemma 3.3 is much stronger than we need in (3.2). Indeed, by choosing the constant C in Theorem 3.1 large enough, we can assume that $n > 1/c_3$ and $\varepsilon < 1$. Then the value $c_3 n^{1/2}$ in Lemma 3.3 is bigger than $\varepsilon n^{-1/2}$ in (3.2).

Proof. We first prove a similar invertibility estimate for the sparse vectors. To this end, we can assume that $\delta_0 < 1/2$ in Proposition 2.5. We use this result with $k = \delta n$ and take the union bound over all $\lceil \delta n \rceil$-element subsets σ of $\{1, \dots, n\}$:

$$\begin{aligned}
&\mathbb{P}\Big(\inf_{x \in \mathit{Sparse}(\delta),\, \|x\|_2 = 1} \|Ax\|_2 \leqslant c_1 n^{1/2} \text{ and } \|A\| \leqslant K n^{1/2} \Big) \\
&\quad = \mathbb{P}\Big(\exists \sigma,\ |\sigma| = \lceil \delta n \rceil \colon \inf_{x \in \mathbb{R}^\sigma,\, \|x\|_2 = 1} \|Ax\|_2 \leqslant c_1 n^{1/2} \text{ and } \|A\| \leqslant K n^{1/2} \Big) \\
&\quad \leqslant \binom{n}{\lceil \delta n \rceil} e^{-c_2 n} \leqslant \exp\big(4e\delta \log(e/\delta) n - c_2 n \big) \leqslant e^{-c_2 n/2}
\end{aligned} \tag{3.3}$$

with an appropriate choice of $\delta < \delta_0$, which depends only on c_2 (which in turn depends only on B and K).

Now we deduce the invertibility estimate for the compressible vectors. Let $c_3 > 0$ and $\rho \in (0, 1/2)$ to be chosen later. We need to bound the event V that $\|Ax\|_2 \leqslant c_3 n^{1/2}$ for some vector $x \in \mathit{Comp}(\delta, \rho)$ and $\|A\| \leqslant K n^{1/2}$. Assume V occurs. Every such vector x can be written as a sum $x = y + z$, where $y \in \mathit{Sparse}(\delta)$ and $\|z\|_2 \leqslant \rho$. Thus $\|y\|_2 \geqslant 1 - \rho \geqslant 1/2$, and

$$\|Ay\|_2 \leqslant \|Ax\|_2 + \|A\|\|z\|_2 \leqslant c_3 n^{1/2} + \rho K n^{1/2}.$$

We choose $c_3 := c_1/4$ and $\rho := c_1/4K$ so that $\|Ay\|_2 \leqslant \frac{1}{2} c_1 n^{1/2}$. Since $\|y\|_2 \geqslant 1/2$, we have found a unit vector $u \in \mathit{Sparse}(\delta)$ such that $\|Au\|_2 \leqslant c_1 n^{1/2}$ (choose $u = y/\|y\|_2$). This shows that the event V implies the event in (3.3), so we have $\mathbb{P}(V) \leqslant e^{-c_2 n/2}$. This completes the proof. □

3.3. Invertibility for the incompressible vectors via distance

For the incompressible vectors, we shall reduce the invertibility problem to a lower bound on the distance between a random vector and a random hyperplane.

We first show that incompressible vectors are well spread in the sense that they have many coordinates of the order $n^{-1/2}$.

Lemma 3.4 *(Incompressible vectors are spread). Let $x \in \mathit{Incomp}(\delta, \rho)$. Then there exists a set $\sigma \subseteq \{1, \dots, n\}$ of cardinality $|\sigma| \geqslant \frac{1}{2}\rho^2 \delta n$ and such that*

$$\frac{\rho}{\sqrt{2n}} \leqslant |x_k| \leqslant \frac{1}{\sqrt{\delta n}} \quad \textit{for all } k \in \sigma.$$

Proof. Consider the subsets of $\{1, \dots, n\}$ defined as

$$\sigma_1 := \left\{ k \colon |x_k| \leqslant \frac{1}{\sqrt{\delta n}} \right\}, \qquad \sigma_2 := \left\{ k \colon |x_k| \geqslant \frac{\rho}{\sqrt{2n}} \right\},$$

and put $\sigma := \sigma_1 \cap \sigma_2$.

By Chebychev's inequality, $|\sigma_1^c| \leqslant \delta n$. Then $y := P_{\sigma_1^c} x \in \mathit{Sparse}(\delta n)$, so the incompressibility of x implies that $\|P_{\sigma_1} x\|_2 = \|x - y\|_2 > \rho$. By the definition of σ_2, we have $\|P_{\sigma_2^c} x\|_2^2 \leqslant n \cdot \frac{\rho^2}{2n} = \rho^2/2$. Hence

$$\|P_\sigma x\|_2^2 \geqslant \|P_{\sigma_1} x\|_2^2 - \|P_{\sigma_2^c} x\|_2^2 \geqslant \rho^2/2. \tag{3.4}$$

On the other hand, by the definition of $\sigma_1 \supseteq \sigma$,

$$\|P_\sigma x\|_2^2 \leqslant \|P_\sigma x\|_\infty^2 \cdot |\sigma| \leqslant \frac{1}{\delta n} \cdot |\sigma|. \tag{3.5}$$

It follows from (3.4) and (3.5) that $|\sigma| \geqslant \frac{1}{2}\rho^2 \delta n$. □

Lemma 3.5 *(Invertibility via distance). Let A be any random matrix. Let $X_1, \dots, X_n$ denote the column vectors of A, and let H_k denote the span of all column vectors except the kth. Then for every $\delta, \rho \in (0,1)$ and every $\varepsilon > 0$, one has*

$$\mathbb{P}\Big(\inf_{x \in \mathit{Incomp}(\delta,\rho)} \|Ax\|_2 < \varepsilon \rho n^{-1/2} \Big) \leqslant \frac{1}{\delta n} \sum_{k=1}^n \mathbb{P}\big(\operatorname{dist}(X_k, H_k) < \varepsilon\big). \tag{3.6}$$

Remark. The main point of this bound is the *average*, rather than the maximum, of the distances in the right-hand side of (3.6). This will allow us to avoid estimating the union of n events and thus bypass a loss of the n factor in the invertibility theorem.

Proof. Let $x \in \mathit{Incomp}(\delta, \rho)$. Writing $Ax = \sum_{k=1}^n x_k X_k$, we have

$$\begin{aligned} \|Ax\|_2 &\geqslant \max_{k=1,\dots,n} \operatorname{dist}(Ax, H_k) \\ &= \max_{k=1,\dots,n} \operatorname{dist}(x_k X_k, H_k) = \max_{k=1,\dots,n} |x_k| \operatorname{dist}(X_k, H_k). \end{aligned} \tag{3.7}$$

Denote

$$p_k := \mathbb{P}\big(\operatorname{dist}(X_k, H_k) < \varepsilon\big).$$

Then

$$\mathbb{E}\big|\big\{k\colon \operatorname{dist}(X_k, H_k) < \varepsilon\big\}\big| = \sum_{k=1}^n p_k.$$

Denote by U the event that the set $\sigma_1 := \{k\colon \operatorname{dist}(X_k, H_k) \geqslant \varepsilon\}$ contains more than $(1-\delta)n$ elements. Then by Chebychev's inequality,

$$\mathbb{P}\big(U^c\big) \leqslant \frac{1}{\delta n} \sum_{k=1}^n p_k.$$

On the other hand, for every incompressible vector x, the set $\sigma_2(x) := \{k\colon |x_k| \geqslant \rho n^{-1/2}\}$ contains at least δn elements. (Otherwise, since $\|P_{\sigma_2(x)^c} x\|_2 \leqslant \rho$, we would have $\|x - y\|_2 \leqslant \rho$ for the sparse vector $y := P_{\sigma_2(x)} x$, which would contradict the incompressibility of x.)

Assume that the event U occurs. Fix any incompressible vector x. Then $|\sigma_1| + |\sigma_2(x)| > (1-\delta)n + \delta n > n$, so the sets σ_1 and $\sigma_2(x)$ have nonempty intersection. Let $k \in \sigma_1 \cap \sigma_2(x)$. Then by (3.7) and by the definitions of the sets σ_1 and $\sigma_2(x)$, we have

$$\|Ax\|_2 \geqslant |x_k| \operatorname{dist}(X_k, H_k) \geqslant \rho n^{-1/2} \cdot \varepsilon.$$

Summarizing, we have shown that

$$\mathbb{P}\Big(\inf_{x \in Incomp(\delta,\rho)} \|Ax\|_2 < \varepsilon \rho n^{-1/2} \Big) \leqslant \mathbb{P}\big(U^c\big) \leqslant \frac{1}{\delta n} \sum_{k=1}^{n} p_k.$$

This completes the proof. □

3.4. Distance via the small ball probability

Lemma 3.5 reduces the invertibility problem to a lower bound on the distance between a random vector and a random hyperplane. Now we reduce bounding the distance to a small ball probability estimate.

Let $X_1, \dots, X_n$ be the column vectors of A. These are independent random vectors in $\mathbb{R}^n$. Consider the subspace $H_n = \operatorname{span}(X_1, \dots, X_{n-1})$. Our goal is to bound the distance between the random vector X_n and the random subspace H_n.

To this end, let X^* be any unit vector orthogonal to $X_1, \dots, X_{n-1}$. We call it a *random normal*. We can choose X^* so that it is a random vector that depends only on $X_1, \dots, X_{n-1}$ and is independent of X_n.

We clearly have

$$\operatorname{dist}(X_n, H_n) \geqslant \big|\big\langle X^*, X_n \big\rangle\big|. \tag{3.8}$$

Since the vectors $X^* =: (a_1, \dots, a_n)$ and $X_n =: (\xi_1, \dots, \xi_n)$ are independent, we should be able to use the small ball probability estimates, such as Corollary 2.10, to deduce a lower bound on the magnitude of

$$\big\langle X^*, X_n \big\rangle = \sum_{k=1}^{n} a_k \xi_k.$$

To this end, we first need to check that the coefficients of the vector X^* are well spread.

Lemma 3.6 *(Random normal is incompressible). Let $\delta, \rho, c_4 > 0$ be as in Lemma* 3.3. *Then*

$$\mathbb{P}\big(X^* \in Comp(\delta, \rho) \text{ and } \|A\| \leqslant K n^{1/2}\big) \leqslant e^{-c_4 n}.$$

Proof. Let A' be the $(n-1) \times n$ random matrix with rows $X_1, \dots, X_{n-1}$, i.e. the submatrix of A^T obtained by removing the last row. By the definition of the random normal,

$$A'X^* = 0. \tag{3.9}$$

Therefore, if $X^* \in Comp(\delta, \rho)$ then $\inf_{x \in Comp(\delta,\rho)} \|A'x\|_2 = 0$. By replacing n with $n-1$, one can easilty check that the proof Lemma 3.3 remains valid for A' as well as for A; note also that $\|A'\| \leqslant \|A\|$. This completes the proof. □

Now we recall our small ball probability estimate, Corollary 2.10, in a form useful for the incompressible vectors:

Lemma 3.7 *(Small ball probability for incompressible vectors). Let $\xi_1, \dots, \xi_n$ be random variables as in Corollary* 2.10. *Let $\delta, \rho \in (0,1)$, and consider a coefficient vector $a \in Incomp(\delta, \rho)$. Then for every $\varepsilon \geqslant 0$ one has*

$$p_\varepsilon(a) \leqslant C_5\big(\varepsilon + Bn^{-1/2}\big),$$

where C_5 depends (polynomially) only on δ and ρ.

Proof. Let σ denote the set of the spread coefficients of a constructed in Lemma 3.4. Then $|\sigma| \geqslant \frac{1}{2}\rho^2\delta n$, and the vector $b := n^{1/2} P_\sigma a$ satisfies $K_1 \leqslant |b_k| \leqslant K_2$ for all $k \in \sigma$, where $K_1 = \rho/\sqrt{2}$ and $K_2 = 1/\sqrt{\delta}$. By Restriction Lemma 2.1 and Corollary 2.10, we have

$$p_\varepsilon(a) = p_{n^{1/2}\varepsilon}\big(n^{1/2}a\big) \leqslant p_{n^{1/2}\varepsilon}(b) \leqslant C_5\big(\varepsilon + Bn^{-1/2}\big).$$

This completes the proof. □

Lemmas 3.7 and 3.6 imply the desired distance bound:

Lemma 3.8 *(Weak Distance Bound). Let A be a random matrix as in Theorem* 3.1. *Let $X_1, \dots, X_n$ denote its column vectors, and consider the subspace $H_n = \operatorname{span}(X_1, \dots, X_{n-1})$. Let $K \geqslant 1$. Then for every $\varepsilon \geqslant 0$, one has*

$$\mathbb{P}\big(\operatorname{dist}(X_n, H_n) < \varepsilon \text{ and } \|A\| \leqslant Kn^{1/2}\big) \leqslant C_6\big(\varepsilon + n^{-1/2}\big),$$

where C_6 depends only on B and K.

Remark. In Theorem 5.2 below, we shall improve this distance bound by reducing the polynomial term $n^{-1/2}$ by the exponential term e^{-cn}.

Proof. We condition upon a realization of the random vectors $X_1, \dots, X_{n-1}$. This fixes realizations of the subspace H_n and the random normal X^*. Recall that X_n is independent of X^*. We denote the probability with respect to X_n by $\mathbb{P}_n$, and the expectation with respect to $X_1, \dots, X_{n-1}$ by $\mathbb{E}_{1,\dots,n-1}$. Then

$$\begin{aligned}&\mathbb{P}\big(\big|\langle X^*, X_n\rangle\big| < \varepsilon \text{ and } \|A\| \leqslant Kn^{1/2}\big)\\ &\quad\leqslant \mathbb{E}_{1,\dots,n-1}\mathbb{P}_n\big(\big|\langle X^*, X_n\rangle\big| < \varepsilon \text{ and } X^* \in \mathit{Incomp}(\delta,\rho)\big)\\ &\qquad + \mathbb{P}\big(X^* \in \mathit{Comp}(\delta,\rho) \text{ and } \|A\| \leqslant Kn^{1/2}\big). \end{aligned} \tag{3.10}$$

Fix $\delta, \rho > 0$ so that the conclusion of Lemma 3.6 holds. This bounds the last term in the right-hand side of (3.10) by $e^{-c_4 n}$. Furthermore, by Lemma 3.7, for any fixed realization of $X_1, \dots, X_n$ such that $X^* \in \mathit{Incomp}(\delta, \rho)$ we have

$$\mathbb{P}_n\big(\big|\langle X^*, X_n\rangle\big| < \varepsilon\big) \leqslant C_5'\big(\varepsilon + n^{-1/2}\big),$$

where C_5' depends only on B and K. It follows that

$$\mathbb{P}\big(\big|\langle X^*, X_n\rangle\big| < \varepsilon \text{ and } \|A\| \leqslant Kn^{1/2}\big) \leqslant C_5'\big(\varepsilon + n^{-1/2}\big) + e^{-c_4 n}.$$

By (3.8), the proof is complete. □

Combining Lemmas 3.5 and 3.8, we have shown the invertibility of a random matrix on the set of incompressible vectors:

Lemma 3.9 *(Invertibility for incompressible vectors). Let A be a random matrix as in Theorem* 3.1. *Let $K \geqslant 1$ and $\delta, \rho \in (0,1)$. Then for every $\varepsilon \geqslant 0$, one has*

$$\mathbb{P}\Big(\inf_{x \in \mathit{Incomp}(\delta,\rho)} \|Ax\|_2 \leqslant \varepsilon\rho n^{-1/2}\Big) \leqslant \frac{C_7}{\delta}\big(\varepsilon + n^{-1/2}\big) + \mathbb{P}\big(\|A\| > Kn^{1/2}\big),$$

where C_7 depends only on B and K.

3.5. Invertibility on the whole sphere

The Weak Invertibility Theorem 3.1 now follows from the decomposition of the sphere (3.2) into compressible and incompressible vectors, and from the invertibility on each of the two parts established in Lemma 3.3 (see the remark below it) and Lemma 3.9 (used for δ, ρ as in Lemma 3.3 and for ε/ρ rather than ε).

4. Small ball probability

In this section, we prove the following more precise version of Theorem 1.5.

Theorem 4.1 *(Small Ball Probability). Let ξ be a centered random variable with variance at least* 1 *and with the third moment bounded by B. Consider independent copies $\xi_1, \dots, \xi_n$ of ξ. Let $a = (a_1, \dots, a_n)$ be a coefficient vector and let $K \geqslant 1$ be such that*

$$1 \leqslant |a_k| \leqslant K \quad \textit{for all } k. \tag{4.1}$$

Let $0 < \alpha < 1/6K$ and $0 < \kappa < n$. Then for every $\varepsilon \geqslant 0$ one has

$$p_\varepsilon(a) \leqslant \frac{CBK^3}{\sqrt{\kappa}}\left(\varepsilon + \frac{1}{D_{2\alpha,2\kappa}(a)}\right) + C\exp\left(-\frac{c\alpha^2\kappa}{B^2}\right),$$

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

Remarks. 1. This result clearly implies Theorem 1.5. (Indeed, in Theorem 1.5 one can assume that $K_1 = 1$ by rescaling the coefficients a_k, and that $\alpha < 1/6K_2$ by considering $\alpha/6K_2$ instead of α.)

2. Since the definition of $p_\varepsilon(a)$ includes shifts, Theorem 4.1 holds also for the shifted random variables $\xi'_j = \xi_j + t_j$ for any real numbers $t_1, \dots, t_n$.

The approach based on the central limit theorem establishes Theorem 4.1 for the values of ε of constant order and above. Indeed, for $\varepsilon > \varepsilon_0 > 0$, Corollary 2.10 yields

$$p_\varepsilon(a) \leqslant \frac{C'_2 B K^3}{\sqrt{n}} \varepsilon$$

where C'_2 depends only on ε_0.

For ε below the constant order, this bound cannot hold without any additional information about the coefficient vector a. Indeed, if all $a_k = 1$ then random sign-sums satisfy $p_0(a) \geqslant \mathbb{P}(S = 0) \sim n^{-1/2}$.

We thus need to develop a tool sharper that the central limit theorem to handle smaller ε. Our new method uses the approach of Halász [11,12], which was also used in [22].

4.1. Initial reductions, symmetrization, truncation

Throughout the proof, absolute constants will be denoted by $C, c, c_1, \dots$. The particular value of each constant can be different in different instances.

As explained above, we can assume in the sequel that ε is below a constant, such as

$$\varepsilon < \pi/4. \tag{4.2}$$

We can also assume that $\kappa < n/2$ and that $a_k \geqslant 1$ by replacing, if necessary, ξ_k by $-\xi_k$.

We shall symmetrize the random variables ξ_k and remove any small values they can possibly take. For many random variables, such as random ± 1, this step is not needed.

Let ξ' be an independent copy of ξ and define the random variable $\zeta := |\xi - \xi'|$. Then

$$\mathbb{E}\zeta^2 = 2\mathbb{E}|\xi|^2 \geqslant 2 \quad \text{and} \quad \mathbb{E}\zeta^3 \leqslant 8\mathbb{E}|\xi|^3 \leqslant 8B.$$

The Paley–Zygmund inequality (see e.g. [18, Lemma 3.5]) implies that

$$\mathbb{P}(\zeta > 1) \geqslant \frac{(\mathbb{E}\zeta^2 - 1)^3}{(\mathbb{E}\zeta^3)^2} \geqslant \frac{1}{64B^2} =: \beta. \tag{4.3}$$

Denote by $\bar{\zeta}$ the random variable ζ conditioned on $\zeta > 1$. Formally, $\bar{\zeta}$ is a random variable such that for every measurable function f one has

$$\mathbb{E}f(\bar{\zeta}) = \frac{1}{\mathbb{P}(\zeta > 1)} \mathbb{E}f(\zeta)\mathbf{1}_{\{\zeta > 1\}}.$$

It then follows by (4.3) that for every measurable non-negative function f, one has

$$\mathbb{E}f(\zeta) \geqslant \beta \mathbb{E}f(\bar{\zeta}). \tag{4.4}$$

4.2. Small ball probability via characteristic functions

An inequality of Esseen ([9], see also [12]), bounds the small ball probability of a random variable S by the L_1 norm of its characteristic function

$$\phi(t) = \phi_S(t) = \mathbb{E}\exp(iSt).$$

Lemma 4.2 *(Esseen's Inequality). For every random variable S and for every $\varepsilon > 0$, one has*

$$\sup_{v\in\mathbb{R}} \mathbb{P}\big(|S - v| \leqslant \varepsilon\big) \leqslant C \int_{-\pi/2}^{\pi/2} \big|\phi(t/\varepsilon)\big|\, dt,$$

where C is an absolute constant.

We want to use Esseen's Inequality for the random sum $S = \sum_{k=1}^n a_k\xi_k$. The characteristic function of $a_k\xi_k$ is

$$\phi_k(t) := \mathbb{E}\exp(ia_k\xi_k t) = \mathbb{E}\exp(ia_k\xi t),$$

so the characteristic function of S is then

$$\phi(t) = \prod_{k=1}^n \phi_k(t).$$

To estimate the integral in Esseen's Lemma 4.2, we first observe that

$$\big|\phi_k(t)\big|^2 = \mathbb{E}\cos(a_k\zeta t).$$

Using the inequality $|x| \leqslant \exp(-\frac{1}{2}(1 - x^2))$ valid for all x, we then obtain

$$\begin{aligned}\big|\phi(t)\big| &\leqslant \prod_{k=1}^n \exp\left(-\frac{1}{2}\big(1 - \big|\phi_k(t)\big|^2\big)\right)\\ &= \exp\left(-\mathbb{E}\sum_{k=1}^n \frac{1}{2}\big(1 - \cos(a_k\zeta t)\big)\right) = \exp\big(-\mathbb{E}f(\zeta t)\big),\end{aligned}$$

where

$$f(t) := \sum_{k=1}^n \sin^2\left(\frac{1}{2}a_k t\right).$$

Hence by (4.4), we have

$$\big|\phi(t)\big| \leqslant \exp\big(-\beta\mathbb{E}f(\bar{\zeta}t)\big).$$

Then by Esseen's Lemma 4.2 and using Jensen's inequality, we estimate the small ball probability as

$$p_\varepsilon(a) \leqslant C \int_{-\pi/2}^{\pi/2} \bigl|\phi(t/\varepsilon)\bigr|\,dt \leqslant C \int_{-\pi/2}^{\pi/2} \exp\bigl(-\beta \mathbb{E} f(\bar{\zeta} t/\varepsilon)\bigr)\,dt$$
$$\leqslant C\mathbb{E} \int_{-\pi/2}^{\pi/2} \exp\bigl(-\beta f(\bar{\zeta} t/\varepsilon)\bigr)\,dt \leqslant C \sup_{z\geqslant 1} \int_{-\pi/2}^{\pi/2} \exp\bigl(-\beta f(zt/\varepsilon)\bigr)\,dt. \tag{4.5}$$

Fix $z \geqslant 1$. First we estimate the maximum

$$M := \max_{|t|\leqslant \pi/2} f(zt/\varepsilon) = \max_{|t|\leqslant \pi/2} \sum_{k=1}^{n} \sin^2(a_k zt/2\varepsilon).$$

Lemma 4.3. *We have*

$$\frac{n}{4} \leqslant M \leqslant n.$$

Proof. The upper bound is trivial. For the lower bound, we estimate the maximum by the average:

$$M \geqslant \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} f(zt/\varepsilon)\,dt = \frac{1}{2} \sum_{k=1}^{n} \left(1 - \frac{\sin(\pi a_k z/2\varepsilon)}{\pi a_k z/2\varepsilon}\right).$$

By our assumptions, $a_k \geqslant 1$, $z \geqslant 1$ and $\varepsilon < \pi/4$. Hence $\pi a_k z/2\varepsilon \geqslant 2$, so

$$M \geqslant \frac{n}{2} \inf_{t\geqslant 2} \left(1 - \frac{\sin t}{t}\right) \geqslant \frac{n}{4}.$$

This completes the proof. □

Now we consider the level sets of f, defined for $m, r \geqslant 0$ as

$$T(m,r) := \bigl\{t\colon\ |t| \leqslant r,\ \ f(zt/\varepsilon) \leqslant m\bigr\}.$$

By a crucial lemma of Halász, the Lebesgue measure of the level sets $|T(m,r)|$ behaves in a regular way ([12], see [22, Lemma 3.2]):

Lemma 4.4 *(Regularity). Let $l \in \mathbb{N}$ be such that $l^2 m \leqslant M$. Then*

$$\left|T\left(m, \frac{\pi}{2}\right)\right| \leqslant \frac{2}{l} \cdot \bigl|T\bigl(l^2 m, \pi\bigr)\bigr|.$$

Hence, for every $\eta \in (0, 1)$ such that $m \leqslant \eta M$, one has:

$$\left|T\left(m, \frac{\pi}{2}\right)\right| \leqslant 4\sqrt{\frac{m}{\eta M}} \cdot \left|T(\eta M, \pi)\right|. \tag{4.6}$$

(Apply Lemma 4.4 with $l = \lfloor\sqrt{\frac{\eta M}{m}}\rfloor$.)

Now we can estimate the integral in (4.5) by the integral distribution formula. Using (4.6) for small m and the trivial bound $|T(m, \pi/2)| \leqslant \pi$ for large m, we get

$$\begin{aligned}
p_\varepsilon(a) &\leqslant C \sup_{z \geqslant 1} \int_{-\pi/2}^{\pi/2} \exp\bigl(-\beta f(zt/\varepsilon)\bigr)\, dt \\
&\leqslant C \int_0^\infty \left|T\left(m, \frac{\pi}{2}\right)\right| \beta e^{-\beta m}\, dm \\
&\leqslant C \int_0^{\eta M} 4\sqrt{\frac{m}{\eta M}} \cdot \left|T(\eta M, \pi)\right| \beta e^{-\beta m}\, dm + C \int_{\eta M}^\infty \pi\ \beta e^{-\beta m}\, dm \\
&\leqslant \frac{C_1}{\sqrt{\beta \eta M}} \cdot \left|T(\eta M, \pi)\right| + C\pi e^{-\beta \eta M} \\
&\leqslant \frac{C_2 B}{\sqrt{\eta n}} \cdot \left|T(\eta n, \pi)\right| + C\pi e^{-c_2 \eta n / B^2}.
\end{aligned} \tag{4.7}$$

In the last line, we used Lemma 4.3 and the definition (4.3) of β.

4.3. Recurrence set

We shall now bound the measure of the level set $|T(\eta n, \pi)|$ by a quantity of ergodic nature, the density of the *recurrence set* of a.

Consider any $t \in T(\eta n, \pi)$ and set $y := z/2\varepsilon$. Then $y \geqslant 1/2\varepsilon$, and

$$f(zt/\varepsilon) = \sum_{k=1}^n \sin^2(a_k y t) \leqslant \eta n. \tag{4.8}$$

Let us fix

$$\eta := \frac{\alpha^2 \kappa}{4n}. \tag{4.9}$$

Then at least $n - \kappa$ terms in the sum in (4.8) satisfy

$$\sin^2(a_k y t) \leqslant \frac{\eta n}{\kappa} = \frac{\alpha^2}{4} < \frac{1}{144},$$

which implies for those terms that $\operatorname{dist}(a_k y t, \pi\mathbb{Z}) \leqslant \alpha$. Thus yt/π belongs to the recurrence set of a, which we define as follows:

Definition 4.5 *(Recurrence set).* Let $\alpha \in (0,1)$ and $\kappa \geqslant 0$. The *recurrence set* $I(a) = I_{\alpha,\kappa}(a)$ of a vector $a \in \mathbb{R}^n$ is defined as the set of all $t \in \mathbb{R}$ such that all except κ coordinates of the vector ta are of distance at most α from $\mathbb{Z}$.

Regarding t as time, we can think of the recurrence set as the moments when most of the particles moving along the unit torus with speeds $a_1, \dots, a_n$ return close to their initial positions.

Our argument thus shows that $T(\eta n, \pi) \subseteq \frac{\pi}{y} I_{\alpha,\kappa}(a)$. Thus

$$\bigl|T(\eta n, \pi)\bigr| \leqslant \left|\frac{\pi}{y} I_{\alpha,\kappa}(a) \cap [-\pi, \pi]\right| = \frac{\pi}{y} \cdot \bigl|I_{\alpha,\kappa}(a) \cap [-y, y]\bigr|.$$

The quantity

$$\operatorname{dens}(I, y) := \frac{1}{2y} \cdot \bigl|I \cap [-y, y]\bigr|$$

can be interpreted as the *density* of the set I. We have thus shown that

$$\bigl|T(\eta n, \pi)\bigr| \leqslant 2\pi \operatorname{dens}\bigl(I_{\alpha,\kappa}(a), y\bigr).$$

Using this bound and our choice (4.9) of η in (4.7), we conclude that

$$p_\varepsilon(a) \leqslant \frac{C_3 B}{\alpha\sqrt{\kappa}} \cdot \sup_{y \geqslant 1/2\varepsilon} \operatorname{dens}\bigl(I_{\alpha,\kappa}(a), y\bigr) + C\pi e^{-c_3 \alpha^2 \kappa / B^2}. \tag{4.10}$$

4.4. Density of the recurrence set

It remains to bound the density of the recurrence set $I(a)$ by the reciprocal of the essential LCD $D(a)$. We will derive this from the following structural lemma, which shows that: (1) the recurrence set has lots of gaps; (2) each gap bounds below the essential LCD of a.

For $t \in \mathbb{R}$, by $[t]$ we denote an integer nearest to t.

Lemma 4.6 *(Gaps in the recurrence set). Under the assumptions of Theorem* 4.1*, let* $t_0 \in I_{\alpha,\kappa}(a)$*. Then:*

(1) $t_0 + 3\alpha \notin I_{\alpha,\kappa}(a)$.
(2) *Let* $t_1 \in I_{\alpha,\kappa}(a)$ *be such that* $t_1 > t_0 + 3\alpha$*. Then* $t_1 - t_0 \geqslant D_{2\alpha,2\kappa}(a)$.

Since $D_{2\alpha,2\kappa}(a) \geqslant (1 - 2\alpha)/K > 4\alpha$, this lemma implies that the recurrence set I has gaps of size at least $D_{2\alpha,2\kappa}(a) - 4\alpha$.

Proof. *Part* 1. Since $t_0 \in I_{\alpha,\kappa}(a)$, there exists a set $\sigma_0 \subseteq \{1, \dots, n\}$ of cardinality $|\sigma_0| \geqslant n - \kappa$ and such that for $p_k := [t_0 a_k]$ we have

$$|t_0 a_k - p_k| \leqslant \alpha \quad \text{for all } k \in \sigma_0. \tag{4.11}$$

Let $t := t_0 + 3\alpha$. Recall that $1 \leqslant a_k \leqslant K$ for all $k \in \{1, \dots, n\}$. By (4.11), we have for all $k \in \sigma_0$:

$$\begin{aligned} ta_k &= t_0 a_k + 3\alpha \cdot a_k \geqslant p_k - \alpha + 3\alpha > p_k + \alpha; \\ ta_k &\leqslant p_k + \alpha + 3\alpha \cdot a_k \leqslant p_k + \alpha + 1/2 < p_k + 1 - \alpha. \end{aligned} \tag{4.12}$$

In the last inequality, we used the assumption $\alpha < 1/6K \leqslant 1/6$. It follows that $\operatorname{dist}(ta_k, \mathbb{Z}) > \alpha$ for all $k \in \sigma_0$. Thus $t \notin I_{\alpha,\kappa}(a)$. Part 1 is proved.

Part 2. Since $t_1 \in I_{\alpha,\kappa}(a)$, there exists a set $\sigma_1 \subseteq \{1, \dots, n\}$ of cardinality $|\sigma_1| \geqslant n - \kappa$ and such that for $q_k := [t_1 a_k]$ we have

$$|t_1 a_k - q_k| \leqslant \alpha \quad \text{for all } k \in \sigma_1. \tag{4.13}$$

Set $\sigma := \sigma_0 \cap \sigma_1$. Then $|\sigma| \geqslant n - 2\kappa$. Moreover, (4.11) and (4.13) yield

$$\big|(t_1 - t_0)a_k - (q_k - p_k)\big| \leqslant 2\alpha \quad \text{for all } k \in \sigma.$$

Since $t_1 > t$, (4.12) implies that

$$t_1 a_k > t a_k > p_k + \alpha \quad \text{for all } k \in \sigma. \tag{4.14}$$

Hence, by (4.13) and (4.14), $q_k - p_k > 0$ for all $k \in \sigma$. By the definition of the essential LCD, this means that

$$t_1 - t_0 \geqslant D_{2\alpha,2\kappa}(a).$$

This completes the proof. □

We can use Lemma 4.6 to bound the density of the recurrence set via the reciprocal of the essential LCD.

Lemma 4.7 *(Recurrence set via essential LCD). Under the assumptions of Theorem* 4.1*, we have for every* $y > 0$

$$\operatorname{dens}\big(I_{\alpha,\kappa}(a), y\big) \leqslant 3\alpha\left(\frac{1}{2y} + \frac{2}{D_{2\alpha,2\kappa}(a)}\right). \tag{4.15}$$

Remark. The contribution of the first term in (4.15) comes from the $O(\alpha)$-neighborhood of zero, which is contained in the recurrence set. This is the initial time when all of the moving particles are still close to 0.

Proof. Denote $I := I_{\alpha,\kappa}(a) \cap [-y, y]$. This set is closed and nonempty (it contains 0). Set $t_0 := \min\{t\colon\ t \in I\}$. If $I \subseteq [t_0, t_0 + 3\alpha]$, then

$$\operatorname{dens}(I, y) = \frac{|I|}{2y} \leqslant \frac{3\alpha}{2y}, \tag{4.16}$$

which completes the proof in this case.

Assume then that $I \not\subseteq [t_0, t_0 + 3\alpha]$. Then we can define inductively the maximal sequence of points $t_1, t_2, \dots, t_L \in I$ by

$$t_l := \min\{t\colon\ t \in I;\ t > t_{l-1} + 3\alpha\}.$$

Note that by Lemma 4.6, $t_{l-1} + 3\alpha \notin I$. Thus the strict inequality in the definition of t_l can be replaced by the non-strict inequality, so the minimum makes sense.

Part 1 of Lemma 4.6 yields

$$I \subseteq \bigcup_{l=0}^{L} [t_l, t_l + 3\alpha),$$

while part 2 implies

$$t_L - t_0 \geqslant \sum_{l=1}^{L} (t_l - t_{l-1}) \geqslant L \cdot D_{2\alpha, 2\kappa}(a).$$

On the other hand, since $t_0, t_L \in I \subseteq [-y, y]$, we have $t_L - t_0 \leqslant 2y$. We conclude that

$$\operatorname{dens}(I, y) \leqslant \frac{|\bigcup_{l=0}^{L} [t_l, t_l + 3\alpha)|}{t_L - t_0} \leqslant \frac{(L+1) \cdot 3\alpha}{L \cdot D_{2\alpha, 2\kappa}(a)} \leqslant \frac{6\alpha}{D_{2\alpha, 2\kappa}(a)}.$$

This completes the proof. □

By (4.10) and Lemma 4.7, we conclude that

$$p_\varepsilon(a) \leqslant \frac{C_4 B}{\sqrt{\kappa}} \left(\varepsilon + \frac{1}{D_{2\alpha, 2\kappa}(a)} \right) + C\pi e^{-c_3 \alpha^2 \kappa / B^2}$$

for all $\varepsilon < \pi/4$ (which was our assumption (4.2)).

This completes the proof of Theorem 4.1.

4.5. Small ball probability for general coefficients

In view of the applications, we will state Theorem 1.5 for a general coefficient vector a, not necessarily with well comparable coefficients as in (1.11). This is easy to do by restricting a onto its spread part, which we define as follows:

Definition 4.8 *(Spread part).* Let $0 < K_1 < K_2$ be fixed. For a vector $x \in \mathbb{R}^n$, we consider the subset $\sigma(x) \subseteq \{1, \dots, n\}$ defined as

$$k \in \sigma(x) \quad \text{if } K_1 \leqslant \left| n^{1/2} x_k \right| \leqslant K_2,$$

and, if $\sigma(x) \neq \emptyset$, we define the *spread part of* x as

$$\hat{x} := \left(n^{1/2} x_k \right)_{k \in \sigma(x)}.$$

If $\sigma(x) = \emptyset$, the spread part of x is not defined.

As an immediate consequence of Restriction Lemma 2.1 and Theorem 1.5, we obtain:

Corollary 4.9 *(Small ball probability for general vectors). Let $\xi_1, \dots, \xi_n$ be random variables as in Theorem* 1.5. *Let $a \in \mathbb{R}^n$ be a vector of real coefficients whose spread part $\hat{a}$ is well defined (for some fixed truncation levels $K_1, K_2 > 0$). Let $\alpha \in (0,1)$ and $\beta \in (0, 1/2)$. Then for every $\varepsilon \geqslant 0$ one has*

$$p_\varepsilon(a) \leqslant \frac{C}{\sqrt{\beta}}\left(\varepsilon + \frac{1}{\sqrt{n}\, D_{\alpha,\beta n}(\hat{a})}\right) + Ce^{-c\alpha^2\beta n},$$

where $C, c > 0$ depend (polynomially) only on B, K_1, K_2.

Remark. As a convention throughout the paper, we set $D_{\alpha,\kappa}(\hat{a}) = 0$ if $\hat{a}$ is not defined.

Remark. A small ball probability bound similar to Theorem 4.1 can be proved with a weaker assumption on the coefficient vector. Namely, (4.1) can be replaced by

$$\|a\|_1 \geqslant n, \qquad \|a\|_2 \leqslant K\sqrt{n}.$$

5. Invertibility of random matrices via small ball probability

We return here to the invertibility problem for random matrices that we began to study in Section 3, and we improve the Weak Invertibility Theorem 3.1 by reducing the polynomial term $n^{1/2}$ to an exponentially small order c^n.

Theorem 5.1 *(Strong invertibility). Let $\xi_1, \dots, \xi_n$ be independent centered random variables with variances at least* 1 *and fourth moments at most B. Let A be an $n \times n$ matrix whose rows are independent copies of the random vector $(\xi_1, \dots, \xi_n)$. Let $K \geqslant 1$. Then for every $\varepsilon \geqslant 0$ one has*

$$\mathbb{P}\big(s_n(A) \leqslant \varepsilon n^{-1/2}\big) \leqslant C\varepsilon + c^n + \mathbb{P}\big(\|A\| > Kn^{1/2}\big), \tag{5.1}$$

where $C > 0$ and $c \in (0,1)$ depend (polynomially) only on B and K.

This result implies the Subgaussian Invertibility Theorem 1.2: indeed, the last term in (5.1) is exponentially small by Lemma 2.4.

The imprecise term $n^{-1/2}$ in the Weak Invertibility Theorem 3.1 came from the Weak Distance Bound, Lemma 3.8, which estimated the distance between a random vector and a random hyperplane. Thus, in order to complete the proof of the Strong Invertibility Theorem 5.1, it suffices to improve the bound in Weak Distance Bound (Lemma 3.8) as follows:

Theorem 5.2 *(Strong Distance Bound). Let A be a random matrix as in Theorem* 5.1. *Let $X_1, \dots, X_n$ denote its column vectors, and consider the subspace $H_n = \operatorname{span}(X_1, \dots, X_{n-1})$. Let $K \geqslant 1$. Then for every $\varepsilon \geqslant 0$, one has*

$$\mathbb{P}\big(\operatorname{dist}(X_n, H_n) < \varepsilon \text{ and } \|A\| \leqslant Kn^{1/2}\big) \leqslant C_7\big(\varepsilon + c^n\big),$$

where C_7 and $c \in (0,1)$ depend only on B and K.

Remark. For random vectors with independent ± 1 coordinates, a weaker bound $\mathbb{P}(\operatorname{dist}(X_n, H_n) < \frac{1}{4n}) \leqslant C \log^{-1/2} n$ was proved by Tao and Vu [29].

5.1. Essential LCD of the random normal

As in Section 3.4, we shall estimate the distance by using the *random normal* X^*, a unit normal of the subspace H_n. The inequality (3.8) reduces the problem to a lower bound on $|\langle X^*, X_n\rangle|$.

The random normal X^* is convenient to control via the random matrix A', the $(n-1) \times n$ matrix with rows $X_1, \dots, X_{n-1}$. Thus A' is the submatrix of A^T obtained by removing the last row. By the definition of the random normal,

$$A'X^* = 0.$$

We will use this observation as follows:

$$\text{If } \|A'x\|_2 > 0 \text{ for all vectors } x \text{ in some set } S, \text{ then } X^* \notin S. \tag{5.2}$$

Thus, a weak (qualitative) invertibility of the random matrix A' on S will help us to "navigate" the random normal X^* away from undesired subsets S of the unit sphere.

We shall use this approach to prove that the essential LCD of the random normal is exponentially large, with probability exponentially close to 1. This will allow us to use the full strength of the Small Ball Probability Theorem 1.5 in order to bound $|\langle X^*, X_n\rangle|$ from below.

Recall that $\hat{x}$ denotes the spread part of a vector x with some fixed truncation levels K_1, K_2, see Definition 4.8.

Theorem 5.3 *(Random normal). Let $X_1, \dots, X_{n-1}$ be random vectors as in Theorem* 5.2. *Consider a unit vector X^* orthogonal to all these vectors. Let $K \geqslant 1$. Then there exist constants $K_1, K_2, \alpha, \beta, c, c' > 0$ that depend only on B and K, and such that*

$$\mathbb{P}\big(D_{\alpha,\beta n}\big(\widehat{X^*}\big) < e^{cn} \text{ and } \|A\| \leqslant Kn^{1/2}\big) \leqslant e^{-c'n}.$$

Intuitively, the components of a random vector should be arithmetically incomparable to the extent that their essential LCD is exponential in n. In the case of the random normal X^*, its components are not independent, and it requires some work to confirm this intuition.

We shall prove that the random matrix A' is likely to be invertible on the subsets S_D of the unit sphere where the essential LCD is of order D, for each D below an exponential order. Then, by observation (5.2), the random normal X^* will not lie in such S_D. Therefore, the essential LCD of X^* will be at least of exponential order.

5.2. The level sets of the essential LCD

Fix $K \geqslant 1$ for the rest of the proof. We shall first choose the truncation levels $K_1 = K_1(B, K)$, $K_2 = K_2(B, K)$ in the definition of the spread part $\widehat{X^*}$ of the random normal.

Our soft invertibility argument in Section 3.1 was based on considering separately compressible and incompressible vectors, forming the sets $\mathit{Comp} = \mathit{Comp}(\delta, \rho)$ and $\mathit{Incomp} = \mathit{Incomp}(\delta, \rho)$ respectively, see Definition 3.2. The parameters $\delta, \rho > 0$ in the definition of these vectors were chosen in Lemma 3.3 depending only on B and K.

For every incompressible vector x, its spread part is proportionally large. Indeed, by Lemma 3.4, there exist $K_1, K_2, c_0 > 0$ that depend only on B and K, and such that for the truncation levels K_1 and K_2 one has $\operatorname{supp}(\hat{x}) \geqslant c_0 n$. For the future convenience, we consider the even integer $n_0 := 2\lfloor c_0 n/2 \rfloor$. Thus we have

$$\text{Every } x \in \mathit{Incomp} \text{ satisfies } \big|\operatorname{supp}(\hat{x})\big| \geqslant n_0 \geqslant \tfrac{c_0}{2} n. \tag{5.3}$$

We shall choose the value $\alpha \in (0, 1/2)$ later. By the definition of the essential LCD and of the spread part,

$$D_{\alpha, n_0/2}(\hat{x}) \geqslant (1-\alpha)/K_2 > 1/2K_2 =: D_0.$$

Definition 5.4 *(Level sets of LCD).* Let $D \geqslant D_0$. We define the level set $S_D \subseteq S^{n-1}$ as

$$S_D := \big\{ x \in \mathit{Incomp}\colon\ D \leqslant D_{\alpha, n_0/2}(\hat{x}) < 2D \big\}.$$

We want to show the invertibility of the random matrix A' on the level sets S_D for all D up to an exponential order. This will be done by a covering argument. We will first show the invertibility on a single vector $x \in S_D$. Next, we will find a small (α/D)-net in S_D. Then, by a union bound, the invertibility will hold for each point in this net. By approximation, we will extend the invertibility to the whole S_D.

The invertibility on a single vector $x \in S_D$ will easily follow from our general small ball probability estimates and the Tensorization Lemma 2.2.

Lemma 5.5 *(Invertibility on a single vector). There exist $c, C_8 > 0$ that depend only on B and K, and such that the following holds. Let $\alpha \in (0,1)$ and $D_0 \leqslant D < \frac{1}{\sqrt{n}} e^{c\alpha^2 n}$. Then for every vector $x \in S_D$ and for every $t \geqslant 0$, one has*

$$\mathbb{P}\big(\|A'x\|_2 < t n^{1/2}\big) \leqslant \left(C_8 t + \frac{C_8}{\sqrt{n}\, D} \right)^{n-1}.$$

Proof. Let $\xi_{k1}, \dots, \xi_{kn}$ denote the kth row of A'. The kth component of $A'x$ is then $(A'x)_k = \sum_{j=1}^n x_j \xi_{kj} =: \zeta_k$. By Corollary 4.9 and by our assumption on D, for every k we have for all $\alpha \in (0,1)$:

$$\mathbb{P}\big(|\zeta_k| < t\big) \leqslant C\left(t + \frac{1}{\sqrt{n} D_{\alpha, n_0/2}(\hat{x})} \right) + C e^{-c\alpha^2 n_0/2} \leqslant C'\left(t + \frac{1}{\sqrt{n}\, D} \right),$$

where C, c, C' depend only on B and K.

Since $\zeta_1, \dots, \zeta_{n-1}$ are independent random variables and $\|A'x\|_2^2 = \sum_{k=1}^{n-1} \zeta_k^2$, Tensorization Lemma 2.2 with $\varepsilon_0 = \frac{1}{\sqrt{n}\, D}$ completes the proof. □

Remark. This proof only used the lower bound $D_{\alpha, n_0/2}(\hat{x}) \geqslant D$ in the definition of the level set S_D.

Lemma 5.6 *(Nets of the level sets). There exist $\alpha_0 \in (0,1)$, $C_9 > 0$ and $c_9 \in (0,1)$ that depend only on B and K, and such that the following holds. Let $0 < \alpha < \alpha_0$ and $D \geqslant D_0$. Then there exists a $(4\alpha/D)$-net in S_D in the Euclidean metric, of cardinality at most*

$$\left(\frac{C_9 D}{\alpha^{1-c_9}}\right)^n.$$

Remark. By a simple volumetric estimate (see e.g. [19]), the sphere S^{n-1} has an θ-net of cardinality $(3/\theta)^n$ for every $\theta > 0$. This implies Lemma 5.6 with $c_9 = 0$. The fact that the level sets have somewhat smaller cardinality, namely with $c_9 > 0$, will be crucial in our argument.

Proof. We start by constructing a $(2\alpha/D)$-net for S_D of the desired cardinality, whose elements do not necessarily belong to S_D.

Let $x \in S_D$. Recall that $\operatorname{supp}(\hat{x}) \geqslant n_0$ by (5.3). By the definition of $D(\hat{x}) = D_{\alpha,n_0/2}(\hat{x})$, there exist $q \in \mathbb{R}^{\operatorname{supp}(\hat{x})}$ with $n_0/2$ integer coefficients and such that

$$\left\| D(\hat{x})\hat{x} - q \right\|_\infty \leqslant \alpha.$$

We can extend q to a vector in $\mathbb{R}^n$ by quantizing its non-integer coefficients uniformly with step α. Thus there exists $p \in \mathbb{R}^n$ whose $n_0/2$ coefficients are in $\mathbb{Z}$ and whose other coefficients are in $\alpha\mathbb{Z}$, and such that

$$\left\| \sqrt{n}\, D(\hat{x})x - p \right\|_\infty \leqslant \alpha. \tag{5.4}$$

(Recall that $\hat{x}$ is a restriction of a vector $\sqrt{n}\,x$.) We thus have $p \in \mathcal{P}$, where

$$\mathcal{P} := \bigcup_{|\sigma|=n_0/2} \mathbb{Z}^\sigma \oplus \alpha\mathbb{Z}^{\sigma^c}, \tag{5.5}$$

the union being over all $(n_0/2)$-element subsets σ of $\{1,\dots,n\}$.

It follows from (5.4) and Hölder's inequality that

$$\left\| \sqrt{n}\, D(\hat{x})x - p \right\|_2 \leqslant \alpha\sqrt{n}. \tag{5.6}$$

Using $x \in S_D$, we obtain

$$\left\| x - \frac{p}{\sqrt{n}\, D(\hat{x})} \right\|_2 \leqslant \frac{\alpha}{D(\hat{x})} \leqslant \frac{\alpha}{D} \leqslant \frac{\alpha_0}{D_0} \leqslant \frac{1}{4},$$

if we choose $\alpha_0 := \min(1, D_0/4)$.

Now we use the following elementary implication, which holds for every pair of vectors y and z in a Hilbert space: if $\|y\| = 1$ and $\|y - z\| \leqslant \delta \leqslant 1/4$ then $\|y - \frac{z}{\|z\|}\| \leqslant 2\delta$. This implies

$$\left\| x - \frac{p}{\|p\|_2} \right\|_2 \leqslant 2\alpha/D. \tag{5.7}$$

On the other hand, since x is a unit vector, (5.6) implies

$$\|p\|_2 \leqslant \big(D(\hat{x}) + \alpha\big)\sqrt{n} \leqslant 3\sqrt{n}\, D$$

where we used that $\alpha \leqslant \alpha_0 \leqslant D_0 \leqslant D(\hat{x})$. We have thus shown that the set

$$\mathcal{N} = \left\{ \frac{p}{\|p\|_2} \colon\ p \in \mathcal{P} \cap 3\sqrt{n}\, D \cdot B_2^n \right\} \subset \mathbb{R}^n$$

is a $(2\alpha/D)$-net for S_D.

Let us estimate the cardinality of $\mathcal{N}$. There are $\binom{n}{n_0/2} \leqslant 2^n$ ways to choose the subset σ in (5.5). Then

$$\begin{aligned} |\mathcal{N}| &\leqslant \big|\mathcal{P} \cap 3\sqrt{n}\, D \cdot B_2^n\big| \\ &\leqslant 2^n \cdot \big|\mathbb{Z}^{n_0/2} \cap 3\sqrt{n}\, D \cdot B_2^{n_0/2}\big| \cdot \big|\alpha\mathbb{Z}^{n-n_0/2} \cap 3\sqrt{n}\, D \cdot B_2^{n-n_0/2}\big|. \end{aligned}$$

The Euclidean ball in $\mathbb{R}^d$ of radius $R\sqrt{d}$ and centered at the origin contains at most $(CR)^d$ integer points, where C is an absolute constant. Then, using that $n_0 \geqslant c_0 n/2$, we conclude that

$$|\mathcal{N}| \leqslant 2^n \cdot (C \cdot 3D)^{n_0/2} \cdot (C \cdot 3D/\alpha)^{n-n_0/2} \leqslant \left(\frac{C_9 D}{\alpha^{1-c_9}}\right)^n.$$

Thus, $\mathcal{N} \subset \mathbb{R}^n$ is a $(2\alpha/D)$-net for S_D of the required cardinality. To complete the proof, note that we can make $\mathcal{N}$ a subset of S_D using the following standard observation. □

Lemma 5.7. *Let T be a metric space and let $E \subset T$. Let $\mathcal{N} \subset T$ be a θ-net of the set E. Then there exists a (2θ)-net $\mathcal{N}'$ of E whose cardinality does not exceed that of $\mathcal{N}$, and such that $\mathcal{N}' \subset E$.*

Remark. As we see from (5.5), we were able to construct a small net because of the coarse quantization of a coordinate subspace $\mathbb{R}^\sigma$ of proportional dimension, which we could afford due to the control of the essential LCD. The finer quantization of the complement $\mathbb{R}^{\sigma^c}$, i.e. $\alpha\mathbb{Z}^{\sigma^c}$, can be replaced with an arbitrary α-net of that subspace. The particular form of the net there does not matter.

Lemma 5.8 *(Invertibility on a level set). There exist $\alpha, c, c_{10} > 0$ that depend only on B and K, and such that the following holds. Let $D_0 \leqslant D < e^{cn}$. Then*

$$\mathbb{P}\left(\inf_{x \in S_D} \|A'x\|_2 < \frac{c_{10}}{D} n^{1/2} \text{ and } \|A\| \leqslant K n^{1/2} \right) \leqslant e^{-n}.$$

Proof. Recall that we can assume that n is sufficiently large. We shall therefore choose a value of α from the nonempty interval $(\frac{1}{\sqrt{n}}, \alpha_0)$. Assume that $D_0 \leqslant D < \frac{1}{\sqrt{n}} e^{c\alpha^2 n}$ as in Lemma 5.5.

We apply Lemma 5.5 with $t = 5K\alpha/D$; thus the term $C_8 t$ will dominate over the term $C_8/\sqrt{n}\, D$. We therefore obtain for each $x_0 \in S_D$:

$$\mathbb{P}\left(\|A'x_0\|_2 < \frac{5K\alpha}{D} n^{1/2} \right) \leqslant \left(\frac{C_8'\alpha}{D}\right)^{n-1}.$$

Let $\mathcal{N}$ be a $(4\alpha/D)$-net of S_D constructed in Lemma 5.6. Then taking the union bound, we obtain

$$\mathbb{P}\left(\inf_{x_0 \in \mathcal{N}} \|A' x_0\|_2 < \frac{5K\alpha}{D} n^{1/2} \right) \leqslant \left(\frac{C_9 D}{\alpha^{1-c_9}} \right)^n \left(\frac{C_8' \alpha}{D} \right)^{n-1} \leqslant C_9 D \big(C_8'' \alpha \big)^{c_9 n - 1}.$$

Using the assumption $D < e^{cn}$, we conclude that

$$\mathbb{P}\left(\inf_{x_0 \in \mathcal{N}} \|A' x_0\|_2 < \frac{5K\alpha}{D} n^{1/2} \right) \leqslant \big(C_9' \alpha \big)^{c_9 n - 1} \leqslant e^{-n}, \tag{5.8}$$

provided that we choose $\alpha \geqslant$ appropriately small in the interval $(\frac{1}{\sqrt{n}}, \alpha_0)$, depending only on C_9' and c_9, which in turn depend only on B and K.

We are now ready to bound the event V that $\|A\| \leqslant Kn^{1/2}$ and for some $x \in S_D$, $\|A'x\|_2 < \frac{c_{10}}{D} n^{1/2}$. Assume that V occurs, and choose $x_0 \in \mathcal{N}$ so that $\|x - x_0\|_2 \leqslant 4\alpha/D$. Since $\|A'\| \leqslant \|A\| \leqslant Kn^{1/2}$, we have

$$\|A' x_0\|_2 \leqslant \|A' x\|_2 + \|A'\| \|x - x_0\|_2 < \frac{c_{10}}{D} n^{1/2} + Kn^{1/2} \cdot \frac{4\alpha}{D} \leqslant \frac{5K\alpha}{D} n^{1/2},$$

if we choose $c_0 := K\alpha$ (which thus depends only on B and K). By (5.8), this completes the proof. □

5.3. Proof of the random normal theorem

Now we prove Theorem 5.3. Let α and c be as in Lemma 5.8.

If $x \in S^{n-1}$ is such that $D(\hat{x}) < e^{cn}$ then, by the definition of the level sets S_D, either x is compressible or $x \in S_D$ for some $D \in \mathcal{D}$, where

$$\mathcal{D} = \big\{ D\colon\ D_0/2 \leqslant D < e^{cn},\ D = 2^k,\ k \in \mathbb{Z} \big\}.$$

Therefore, denoting the event that $\|A\| \leqslant Kn^{1/2}$ by U_K, we have

$$\mathbb{P}\big(D\big(\widehat{X}^*\big) < e^{cn} \text{ and } U_K \big) \leqslant \mathbb{P}\big(X^* \in \mathit{Comp} \text{ and } U_K \big) + \sum_{D \in \mathcal{D}} \mathbb{P}\big(X^* \in S_D \text{ and } U_K \big).$$

By Lemma 3.6, $\mathbb{P}(X^* \in \mathit{Comp} \text{ and } U_K) \leqslant e^{-c_4 n}$. By (5.2) and Lemma 5.8, for every $D \in \mathcal{D}$ we have

$$\mathbb{P}\big(X^* \in S_D \text{ and } U_K \big) \leqslant \mathbb{P}\Big(\inf_{x \in S_D} \|A' x\|_2 = 0 \text{ and } U_K \Big) \leqslant e^{-n}.$$

Since $|\mathcal{D}| \leqslant C' n$, we conclude that

$$\mathbb{P}\big(D\big(\widehat{X}^*\big) < e^{cn} \text{ and } U_K \big) \leqslant e^{-c_4 n} + C' n \cdot e^{-n} \leqslant e^{-c' n}.$$

This completes the proof of Theorem 5.3.

5.4. Proof of the strong distance bound and the strong invertibility theorem

Now we deduce Theorem 5.2 from our small ball probability bound (Corollary 4.9) and the Random Normal Theorem 5.3.

We proceed with a conditioning argument similar to those used to prove the Weak Distance Bound, Lemma 3.8. We condition upon a realization of the random vectors $X_1, \dots, X_{n-1}$. This fixes realizations of the subspace H_n and the random normal X^*. Recall that X_n is independent of X^*. We denote the probability with respect to X_n by $\mathbb{P}_n$, and the expectation with respect to $X_1, \dots, X_{n-1}$ by $\mathbb{E}_{1,\dots,n-1}$. Then

$$\begin{aligned}\mathbb{P}\big(\big|\langle X^*, X_n\rangle\big| < \varepsilon \text{ and } \|A\| \leqslant Kn^{1/2}\big) &\leqslant \mathbb{E}_{1,\dots,n-1}\mathbb{P}_n\big(\big|\langle X^*, X_n\rangle\big| < \varepsilon \text{ and } D_{\alpha,\beta n}\big(\widehat{X}^*\big) \geqslant e^{cn}\big) \\ &\quad + \mathbb{P}\big(D_{\alpha,\beta n}\big(\widehat{X}^*\big) < e^{cn} \text{ and } \|A\| \leqslant Kn^{1/2}\big).\end{aligned}$$

By the Random Normal Theorem 5.3, the last term in the right-hand side is bounded by $e^{-c'n}$. Furthermore, by Corollary 4.9, for any fixed realization of $X_1, \dots, X_n$ such that $D_{\alpha,\beta n}(\widehat{X}^*) \geqslant e^{cn}$ we have

$$\mathbb{P}_n\big(\big|\langle X^*, X_n\rangle\big| < \varepsilon\big) \leqslant C''\varepsilon + C''e^{-c''n}.$$

It follows that

$$\mathbb{P}\big(\big|\langle X^*, X_n\rangle\big| < \varepsilon \text{ and } \|A\| \leqslant Kn^{1/2}\big) \leqslant C''\varepsilon + C''e^{-c''n} + e^{-c'n}.$$

By (3.8), the proof of Theorem 5.2 is complete.

Combining Lemma 3.5 and Theorem 5.2, we deduce a strong invertibility bound for a random matrix on the set of incompressible vectors. This improves a polynomial term in Lemma 3.9 to an exponential term:

Lemma 5.9 *(Strong invertibility for incompressible vectors). Let A be a random matrix as in Theorem* 5.1. *Let $K \geqslant 1$ and $\delta, \rho \in (0, 1)$. Then for every $\varepsilon \geqslant 0$, one has*

$$\mathbb{P}\Big(\inf_{x \in Incomp(\delta,\rho)} \|Ax\|_2 \leqslant \varepsilon\rho n^{-1/2}\Big) \leqslant \frac{C_{11}}{\delta}\big(\varepsilon + c^n\big) + \mathbb{P}\big(\|A\| > Kn^{1/2}\big),$$

where $C_{11} > 0$ and $c \in (0, 1)$ depend only on B and K.

The Strong Invertibility Theorem 5.1 now follows from the decomposition of the sphere (3.2) into compressible and incompressible vectors, and from the invertibility on each of the two parts established in Lemma 3.3 (see the remark below it) and Lemma 5.9 (used for δ, ρ as in Lemma 3.3 and for ε/ρ rather than ε).

Acknowledgment

The authors are grateful to the referee for the careful reading of the manuscript and valuable suggestions.

References

[1] Z.D. Bai, J. Silverstein, Y.Q. Yin, A note on the largest eigenvalue of a large-dimensional sample covariance matrix, J. Multivariate Anal. 26 (1988) 166–168.
[2] B. Bollobás, Combinatorics. Set Systems, Hypergraphs, Families of Vectors and Combinatorial Probability, Cambridge Univ. Press, Cambridge, 1986.
[3] E.J. Candes, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies, IEEE Trans. Inform. Theory 52 (2004) 5406–5425.
[4] K. Davidson, S.J. Szarek, Local operator theory, random matrices and Banach spaces, in: Handbook of the Geometry of Banach Spaces, vol. I, North-Holland, Amsterdam, 2001, pp. 317–366.
[5] D.L. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (2006) 1289–1306.
[6] A. Edelman, Eigenvalues and condition numbers of random matrices, SIAM J. Matrix Anal. Appl. 9 (1988) 543–560.
[7] P. Erdös, On a lemma of Littlewood and Offord, Bull. Amer. Math. Soc. 51 (1945) 898–902.
[8] P. Erdös, Extremal problems in number theory, in: Proc. Sympos. Pure Math., vol. VIII, Amer. Math. Soc., Providence, RI, 1965, pp. 181–189.
[9] C.G. Esseen, On the Kolmogorov–Rogozin inequality for the concentration function, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 5 (1966) 210–216.
[10] P. Frankl, Z. Füredi, Solution of the Littlewood–Offord problem in high dimensions, Ann. of Math. (2) 128 (2) (1988) 259–270.
[11] G. Halász, On the distribution of additive arithmetic functions, Acta Arith. 27 (1975) 143–152.
[12] G. Halász, Estimates for the concentration function of combinatorial number theory and probability, Period. Math. Hungar. 8 (1977) 197–211.
[13] J. Kahn, J. Komlós, E. Szemerédi, On the probability that a random ± 1-matrix is singular, J. Amer. Math. Soc. 8 (1) (1995) 223–240.
[14] J. Komlós, On the determinant of (0, 1) matrices, Studia Sci. Math. Hungar. 2 (1967) 7–21.
[15] R. Latala, Some estimates of norms of random matrices, Proc. Amer. Math. Soc. 133 (2005) 1273–1282.
[16] M. Ledoux, M. Talagrand, Probability in Banach Spaces. Isoperimetry and Processes, Ergeb. Math. Grenzgeb. (3), vol. 23, Springer-Verlag, Berlin, 1991.
[17] W.V. Li, Q.-M. Shao, Gaussian processes: Inequalities, small ball probabilities and applications, in: Stochastic Processes: Theory and Methods, in: Handbook of Statist., vol. 19, North-Holland, Amsterdam, 2001, pp. 533–597.
[18] A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, Adv. Math. 195 (2005) 491–523.
[19] V.D. Milman, G. Schechtman, Asymptotic Theory of Finite-Dimensional Normed Spaces, with an appendix by M. Gromov, Lecture Notes in Math., vol. 1200, Springer-Verlag, Berlin, 1986.
[20] A.M. Odlyzko, On subspaces spanned by random selections of ± 1 vectors, J. Combin. Theory Ser. A 47 (1988) 124–133.
[21] G. Pan, W. Zhou, Circular law, extreme singular values and potential theory, preprint.
[22] M. Rudelson, Invertibility of random matrices: Norm of the inverse, Ann. of Math., in press.
[23] A. Sárközy, E. Szemerédi, Über ein Problem von Erdös und Moser, Acta Arith. 11 (1965) 205–208.
[24] S. Smale, On the efficiency of algorithms of analysis, Bull. Amer. Math. Soc. (N.S.) 13 (1985) 87–121.
[25] A. Soshnikov, A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices, J. Stat. Phys. 108 (2002) 1033–1056.
[26] D. Spielman, S.-H. Teng, Smoothed analysis of algorithms, in: Proceedings of the International Congress of Mathematicians, vol. I, Beijing, 2002, Higher Ed. Press, Beijing, 2002, pp. 597–606.
[27] D.W. Stroock, Probability Theory, an Analytic View, Cambridge Univ. Press, Cambridge, 1993.
[28] S. Szarek, Condition numbers of random matrices, J. Complexity 7 (2) (1991) 131–149.
[29] T. Tao, V. Vu, On random ± 1 matrices: Singularity and determinant, Random Structures Algorithms 28 (2006) 1–23.
[30] T. Tao, V. Vu, On the singularity probability of random Bernoulli matrices, J. Amer. Math. Soc. 20 (2007) 603–628.
[31] T. Tao, V. Vu, Inverse Littlewood–Offord theorems and the condition number of random discrete matrices, Ann. of Math., in press.
[32] J. von Neumann, Collected Works, vol. V: Design of Computers, Theory of Automata and Numerical Analysis, general editor: A.H. Taub, Pergamon Press Book, The Macmillan Co., New York, 1963.
[33] Y.Q. Yin, Z.D. Bai, P.R. Krishnaiah, On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix, Probab. Theory Related Fields 78 (1988) 509–521.