The smallest singular value of deformed random rectangular matrices

Fan Yang *[1]

[1]Department of Mathematics, University of Wisconsin-Madison

February 8, 2017

Abstract

We prove an estimate on the smallest singular value of a type of multiplicatively and additively deformed random rectangular matrices. Suppose $n \leqslant N \leqslant M \leqslant CN$ for some positive constant C. Let X be a random $M \times n$ matrix with independent entries, which have zero mean, unit variance and arbitrarily high moments. Let T be an $N \times M$ deterministic matrix with comparable singular values $c \leqslant s_N(T) \leqslant s_1(T) \leqslant C$, and let A be an $N \times n$ deterministic matrix with $\|A\| = O(\sqrt{N})$. Then we prove that with high probability, the smallest singular value of $TX - A$ is at least of the order $N^{-\epsilon}(\sqrt{N} - \sqrt{n-1})$ for any $\epsilon > 0$. If we assume further the entries of X have subgaussian decay, then the smallest singular value of $TX - A$ is at least of the order $\sqrt{N} - \sqrt{n-1}$ with high probability, which is an essentially optimal bound.

1 Introduction

1.1 Smallest singular values of random matrices

In this paper, we prove a lower bound for the smallest singular values of deformed random rectangular matrices of the form $TX - A$, where T, A are fixed matrices and X is a random matrix with independent centered entries.

Consider an $N \times n$ real or complex matrix A with $N \geqslant n$. The singular values $s_i(A)$ of A are the eigenvalues of $(A^*A)^{1/2}$ arranged in the non-increasing order

$$s_1(A) \geqslant s_2(A) \geqslant \ldots \geqslant s_n(A).$$

Of particular importance are the largest singular value $s_1(A)$, which controls the norm $\|A\|$, and the smallest singular value $s_n(A)$, which measures the invertibility of A^*A.

A natural random matrix model is given by a rectangular matrix X whose entries are independent centered random variables with unit variance and certain moment assumptions. In this paper, we focus on random variables with *arbitrarily high moments* (see (1.8)). This includes all the *subgaussian* and *subexponential* variables, and some other random variables with heavy tails. The asymptotic behavior of the extreme singular values of X has been well-studied. Let X be an $N \times n$ random

*E-mail: fyang75@math.wisc.edu.

matrix. Let $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ be the eigenvalues of $N^{-1}X^*X$ and define the empirical spectral distribution as $\mu_N := n^{-1}\sum_{i=1}^n \delta_{\lambda_i}$. If $n/N \to \lambda \in (0,1)$ as $N \to \infty$, then μ_N converges (in distribution) to the famous Marchĕnko-Pastur (MP) law almost surely [13]. Moreover, the MP distribution has a density with positive support on $[(1-\sqrt{\lambda})^2, (1+\sqrt{\lambda})^2]$, which suggests that asymptotically,

$$s_1(X) \to \sqrt{N}(1+\sqrt{\lambda}) = \sqrt{N}+\sqrt{n}, \quad \text{and} \quad s_n(X) \to \sqrt{N}(1-\sqrt{\lambda}) = \sqrt{N}-\sqrt{n}. \tag{1.1}$$

The almost sure convergence of the largest singular value was proved in [7] for random matrices whose entries have arbitrarily high moments. The almost sure convergence of the smallest singular value was proved in [20] for Gaussian random matrices (i.e. the Wishart matrix). These results were later generalized to random matrices with i.i.d entries with finite fourth moment in [24] and [3].

A considerably harder problem is to establish non-asymptotic versions of (1.1), which would hold for any fixed dimensions N and n. Most often needed are upper bounds for the largest singular value $s_1(X)$ and lower bounds for the smallest singular value $s_n(X)$. With a standard ϵ-net argument, it is not hard to show that the operator norm of X is at most of the optimal order $\sqrt{N}$ for all dimensions, see e.g. [6, 11, 17]. On the other hand, the lower bound for the smallest singular value is much harder to obtain. There has been much progress in this direction during last decade.

Tall matrices. It was proved in [12] that for arbitrary aspect ratios $\lambda < 1-c/\log N$ and random matrices with independent subgaussian entries, one has

$$\mathbb{P}\left(s_n(X) \leqslant c_\lambda \sqrt{N}\right) \leqslant e^{-cN}, \tag{1.2}$$

where $c_\lambda > 0$ depends only on λ and the maximal subgaussian moment of the entries.

Square matrices. For square random matrices (i.e. when $N = n$), the estimate on the smallest singular value was first obtained in [15], where it was proved that for subgaussian matrix X, $s_n(X) \geqslant \epsilon N^{-3/2}$ with high probability. This result was later improved to [16]

$$\mathbb{P}\left(s_N(X) \leqslant \epsilon N^{-1/2}\right) \leqslant C\epsilon + e^{-cN}, \tag{1.3}$$

and essentially optimal result for subgaussian matrices. Subsequently, the lower bound for $s_N(X)$ was proved under weakened moments assumptions [21, 14, 9]. In [22], Tao and Vu proved $s_N(X) \geqslant N^{-O(1)}$ with probability $1-o(1)$ assuming only unit variance of matrix entries.

Almost square matrices. The gap $1-c/\log N \leqslant \lambda < 1$ was filled in [17]. It was shown that for subgaussian random rectangular matrices,

$$\mathbb{P}\left(s_n(X) \leqslant \epsilon(\sqrt{N}-\sqrt{n-1})\right) \leqslant (C\epsilon)^{N-n+1} + e^{-cN}, \tag{1.4}$$

for all fixed dimensions $N \geqslant n$. This bound is essentially optimal for subgaussian matrices with all aspect ratios. It is easy to see that (1.2) and (1.3) are now the special cases of (1.4).

There is an alternative way to study the extreme singular values, that is, through a local estimate (often referred to as *local law*) on the Stieltjes transform of the empirical spectral distribution μ_N. Following this approach, one can prove the following finer results for X whose entries have arbitrarily high moments [5]: for any (small) $\epsilon > 0$ and (large) $D > 0$, there exists $n(\epsilon, D)$ such that

$$\mathbb{P}\left(\left|s_1(X) - \left(\sqrt{N}+\sqrt{n}\right)\right| \geqslant (Nn)^{1/4} n^{-2/3+\epsilon}\right) \leqslant n^{-D}, \quad \text{for all } n \geqslant n(\epsilon, D). \tag{1.5}$$

If in addition $|\lambda - 1| \geqslant c$ for some constant $c > 0$ (so that the left edge $(1-\sqrt{\lambda})^2$ of the MP law is strictly away from zero), then a similar result holds for the smallest singular value:

$$\mathbb{P}\left(\left|s_n(X) - \left(\sqrt{N}-\sqrt{n}\right)\right| \geqslant (Nn)^{1/4} n^{-2/3+\epsilon}\right) \leqslant N^{-D}, \quad \text{for all } n \geqslant n(\epsilon, D). \tag{1.6}$$

Here the assumption $|\lambda - 1| \geqslant c$ is crucial. The bound (1.6) does not hold for $\lambda \to 1$ case.

In this paper, we are interested in the extreme eigenvalues of a multiplicatively and additively deformed random rectangular matrix. Given an $M \times n$ random matrix X with independent entries, we consider the matrix $TX - A$, where T and A are deterministic $N \times M$ and $N \times n$ matrices, respectively. It is easy to control the largest singular value through $\|TX - A\| \leqslant \|T\|\|X\| + \|A\|$. On the other hand, we expect that if $n \leqslant N \leqslant M$ and the singular values of T satisfy $c \leqslant s_N(T) \leqslant s_1(T) \leqslant C$, then the bound $s_n(TX-A) \leqslant \epsilon(\sqrt{N}-\sqrt{n-1})$ would hold for subgaussian X. Intuitively, we can regard T as a nice "isomorphic projection" of $\mathbb{R}^M$ onto an N-dimensional subspace. On the other hand, by (1.4), X is a nice "isomorphic embedding" of $\mathbb{R}^n$ into $\mathbb{R}^M$ with high probability. Hence TX should be a nice embedding of $\mathbb{R}^n$ into an N-dimensional space. Moreover, adding a fixed matrix A can make the matrix become singular with very small probability (actually with 0 probability if the entries of X have continuous distributions). Thus $TX - A$ should have a suitable lower bound for the smallest singular value. In this paper, we will make this argument rigorous.

One of our motivations is the potential application in statistical science. Consider sample covariance matrices of the form $Q = n^{-1}BB^*$, where B is an $N \times n$ matrix. The columns of B represent n independent observations of some random N-dimensional vector $\mathbf{b}$. For the sample vector $\mathbf{b}$, we take a linear model $\mathbf{b} = T\mathbf{x}$, where T is a deterministic $N \times M$ matrix and $\mathbf{x}$ is a random M-dimensional vector with independent entries. Then we can write $\mathbf{b} = T\hat{\mathbf{x}}+\mathbf{a}$, where $\hat{\mathbf{x}}$ is a centered random vector and $\mathbf{a} = T\mathbb{E}\mathbf{x}$. In addition, we may without loss of generality assume that the entries of $\hat{\mathbf{x}}$ have unit variance by absorbing the variance of $\hat{x}_i$ into T. Hence we can write B into the form $B = TX - A$, and our result provides a good a priori estimate on the smallest singular values of B. It is worth mentioning that in [10], the authors proved a local law for sample covariance matrices of the form $Q = n^{-1}TXX^*T^*$. As a corollary, they showed that there exist constants $0 < a_1(\lambda, T) < a_2(\lambda, T)$, depending only on λ and the singular values of T, such that (1.5) and (1.6) hold if one replace $\sqrt{N} + \sqrt{n}$ with $\sqrt{N}a_2(\lambda, T)$ and $\sqrt{N} - \sqrt{n}$ with $\sqrt{N}a_1(\lambda, T)$. However, on the one hand, the local law for $n^{-1}(TX - A)(TX - A)^*$ is still unknown; on the other hand, (1.6) fails when $a_1(\lambda, T) \to 0$ (which may happen, for example, when $\lambda \to 1$).

Another application of our result is the *circular law* for square random matrices. Let X be an $N \times N$ random matrix with i.i.d. entries with mean zero and variance one. It is well known that the spectral measure of eigenvalues of $N^{-1/2}X$ converges to the circular law, i.e. the uniform distribution on the unit disk [8, 1]. An important input of the proof is the upper and lower bounds for the extreme singular values of $N^{-1/2}X - z$ for any fixed $z \in \mathbb{C}$ [9, 14, 21, 22]. In [23], we obtained a generalized *local circular law* for square random matrices of the form $N^{-1/2}TX$. In order to get a lower bound for the smallest singular value of $N^{-1/2}TX - z$, we originally assumed that the entries of X have continuous distributions. This assumption rules out some important models such as the Bernoulli random matrices. Now with the result of this paper, we can relax this assumption greatly to include all random variables with sufficiently high moments.

1.2 Main results and reduction to subgaussian matrices

Let $\xi_1, \ldots, \xi_n$ be independent random variables such that for all $1 \leqslant i \leqslant n$,

$$\mathbb{E}\xi_i = 0, \quad \mathbb{E}|\xi_i|^2 = 1, \tag{1.7}$$

and for any $p \in \mathbb{N}$, there is an N-independent constant C_p such that

$$\mathbb{E}|\xi_i|^p \leqslant C_p. \tag{1.8}$$

We assume that X is an $M \times n$ random matrix, whose rows are independent copies of the random vector $(\xi_1, \ldots, \xi_n)$. In this paper, we want to prove a lower bound for the smallest singular value of

$TX - B$, where T and B are fixed $N \times M$ and $N \times n$ matrices, respectively. We assume that

$$n \leqslant N \leqslant M \leqslant CN, \quad \|B\| \leqslant C\sqrt{N} \tag{1.9}$$

for some constant $C > 0$. Moreover, we assume the eigenvalues of TT^* satisfy that

$$C \geqslant \sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_N \geqslant C^{-1}. \tag{1.10}$$

For definiteness, in this paper we focus on *real matrices*. However, our results and proofs also hold, after minor changes, in the complex case if we assume in addition that X_{ij} have independent real and imaginary parts, such that

$$\mathbb{E}\,(\mathrm{Re}\, X_{ij}) = 0, \quad \mathbb{E}\,(\mathrm{Re}\, X_{ij})^2 = \frac{1}{2},$$

and similarly for $\mathrm{Im}\, X_{ij}$. The main result of this paper is the following theorem.

Theorem 1.1. *Suppose the assumptions (1.7), (1.8), (1.9) and (1.10) hold. Then for every* $\tau > 0$, $\epsilon \geqslant 0$ *and* $D > 0$, *there exists a constant* $C > 0$ *such that*

$$\mathbb{P}\left(s_n(TX - B) \leqslant N^{-\tau}\epsilon(\sqrt{N} - \sqrt{n-1})\right) \leqslant (C\epsilon)^{N-n+1} + N^{-D} \tag{1.11}$$

for large enough N.

To prove this theorem, we first truncate the entries of X at level N^{ω} for some small $\omega > 0$. Combining condition (1.8) with Markov's inequality, we get that for any (small) $\omega > 0$ and (large) $D > 0$, there exists $N(\omega, D)$ such that

$$\mathbb{P}\,(|\xi_i| > N^{\omega}/2) \leqslant N^{-D-2}$$

for all $N \geqslant N(\epsilon, D)$. Hence with a loss of probability $O(N^{-D})$, it is sufficient to control the smallest singular values of the random matrix $T\tilde{X} - B$, where

$$\tilde{X} = 1_{\Omega} X, \quad \Omega := \{|X_{ij}| \leqslant N^{\omega}/2 \text{ for all } 1 \leqslant i \leqslant M, 1 \leqslant j \leqslant n\}.$$

By (1.8) and integration by parts, we can verify that

$$\mathbb{E}\left(\xi_i \mathbf{1}_{\{|\xi_i| \leqslant N^{\omega}/2\}}\right) = O\left(N^{-D-2+\omega}\right), \quad \mathrm{Var}\left(\xi_i \mathbf{1}_{\{|\xi_i| \leqslant N^{\omega}/2\}}\right) = 1 + O\left(N^{-D-2+2\omega}\right), \quad 1 \leqslant i \leqslant n. \tag{1.12}$$

Let D_1 be an $n \times n$ diagonal matrix such that $(D_1)_{ii} = \mathrm{Var}\left(\xi_i \mathbf{1}_{\{|\xi_i| \leqslant N^{\omega}/2\}}\right)^{1/2}$. Let $T = U\tilde{D}V$ be a singular decomposition, where U is an $N \times N$ unitary matrix, V is an $M \times M$ unitary matrix and $\tilde{D} = (D, 0)$ such that $D = \mathrm{diag}(d_1, d_2, \ldots, d_N)$ with $|d_i|^2 = \sigma_i$ (so D is nonsingular by (1.10)). We denote $V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$, where V_1 has size $N \times M$ and V_2 has size $(M - N) \times M$. Then we have

$$T\tilde{X} - B = UDV_1(\tilde{X} - \mathbb{E}\tilde{X}) - (B - T\mathbb{E}\tilde{X}) = UD\left[V_1(\tilde{X} - \mathbb{E}\tilde{X})D_1^{-1} - \left(D^{-1}U^{-1}B - V_1\mathbb{E}\tilde{X}\right)D_1^{-1}\right]D_1.$$

Due to (1.10) and (1.12), we only need to control $s_n(V_1 Y - A)$, where $Y := (\tilde{X} - \mathbb{E}\tilde{X})D_1^{-1}$ and $A := (D^{-1}U^{-1}B - V_1\mathbb{E}\tilde{X})D_1^{-1}$. Now by (1.9), (1.10), (1.12) and the definition of Ω, it is easy to see that Y is a random matrix with independent centered entries satisfying

$$\mathbb{E}Y_{ij} = 0, \quad \mathrm{Var}(Y_{ij}) = 1, \quad |Y_{ij}| \leqslant N^{\omega}, \tag{1.13}$$

and A is a deterministic matrix such that

$$\|A\| \leqslant C\left(\|B\|+\|\mathbb{E}\tilde{X}\|\right) \leqslant C\left(\sqrt{N}+N^{-D-1+\omega}\right) \leqslant C\sqrt{N}. \tag{1.14}$$

Moreover, by Theorem 2.10 of [5], there exists a constant $C>0$ such that

$$\mathbb{P}\left(\|X^*X\| \leqslant CN\right) \geqslant N^{-D}$$

for large enough N. Then using $\|\tilde{X}\| \leqslant \|X\|$, we get

$$\mathbb{P}(\|Y\| \leqslant C\sqrt{N}) \geqslant 1-N^{-D}. \tag{1.15}$$

Now it is easy to see that Theorem 1.1 is a simple corollary of the following theorem.

Theorem 1.2. *Let $\xi_1,\dots,\xi_n$ be independent centered random variables with variance 1, finite fourth moment and subgaussian moment bounded by K for some $K \equiv K(N) \leqslant N^\omega$. Let Y be an $M\times n$ random matrix such that its rows are independent copies of the random vector $(\xi_1,\dots,\xi_n)$, and suppose that $\|Y\| \leqslant C\sqrt{N}$. Let P be an $N\times M$ matrix with $PP^T=1$, and A be a fixed $N\times n$ matrix with $\|A\| \leqslant C\sqrt{N}$. Then for every $\epsilon \geqslant 0$, there exists constants $c, C>0$ such that*

$$\mathbb{P}\left(s_n(PY-A) \leqslant \epsilon\left(\sqrt{N}-\sqrt{n-1}\right)\right) \leqslant \left(CK^{12}\epsilon\right)^{N-n+1}+e^{-cN/K^4} \tag{1.16}$$

for large enough N.

Recall that a random variable ξ is called subgaussian if there exists $K>0$ such that

$$\mathbb{P}\left(|\xi|>t\right) \leqslant 2\exp(-t^2/K^2) \quad \text{for all } t>0. \tag{1.17}$$

The infimum of such K is called the subgaussian moment of ξ or the ψ_2-norm $\|\xi\|_{\psi_2}$. By (1.13), Y_{ij} are uniformly bounded by N^ω, so they are subgaussian with $\|Y_{ij}\|_{\psi_2} \leqslant N^\omega$. Combining with (1.15), we see that Y satisfies the assumptions in Theorem 1.2 with a loss of probability N^{-D}. Note that the condition (1.17) can be equivalently formulated as the moment condition

$$\left(\mathbb{E}\,|\xi|^p\right)^{1/p} \leqslant CK\sqrt{p}, \quad \text{for all } p \geqslant 1, \tag{1.18}$$

which is stronger than our condition (1.8).

Remark 1.3. *Suppose X_{ij} are subgaussian random variables with subgaussian moment bounded by a constant $K>0$. Then*

$$\mathbb{P}(\|X\| \geqslant tN^{1/2}) \leqslant e^{-c_0t^2N} \quad \textit{for } t \geqslant C_0,$$

where $c_0, C_0>0$ depend only on the subgaussian moment K (see [17, Proposition 2.4]). Combining with Theorem 1.2, we obtain the following optimal bound

$$\mathbb{P}\left(s_n(TX-B) \leqslant \epsilon\left(\sqrt{N}-\sqrt{n-1}\right)\right) \leqslant (C\epsilon)^{N-n+1}+e^{-cN} \tag{1.19}$$

for subgaussian random matrices, which corresponds to the bound in (1.4).

Remark 1.4. *The K^4 and K^{12} factors in Theorem 1.2 can be improved during the proof. However, we do not attempt to get the optimal factors in this paper.*

The bulk of this paper is devoted to the proof of Theorem 1.2. In the preliminary Section 2, we introduce some notations and tools that will be used in the proof. In Section 3, we first reduce the problem into bounding below $\|(PY - A)x\|_2$ for *compressible unit vectors* $x \in S^{n-1}$, whose l^2-norm is concentrated in a small number of coordinates, and for *incompressible unit vectors* comprising the rest of the sphere S^{n-1}. Then we prove a lower bound for compressible vectors using a large deviation estimate (Lemma 2.7) and a standard ϵ-net argument. The incompressible vectors are dealt with in Sections 4 and 5. In Section 4, we consider the case $1 \leqslant n < \lambda_0 N$ for some constant $\lambda_0 \in (0,1)$, i.e. when $PY - A$ is a tall matrix. The proof can be finished with a small ball probability result (Lemma 2.6) and the ϵ-net argument. The almost square case with $\lambda_0 N \leqslant n \leqslant N$ is considered in Section 5. We first reduce the problem into bounding the distance between a random vector and a random subspace, and then complete the proof with the random distance Lemma 5.4, whose proof will be given in Section 6.

Acknowledgements. I would like to thank Roman Vershynin for valuable suggestions and for pointing out several useful references. I am also grateful to my advisor Jun Yin for his direction and helpful comments on this paper. Finally, many thanks to my friend Haokai Xi for useful discussions.

2 Basic notations and tools

In this paper, we use C to denote a generic large positive constant, which may depend on fixed parameters and whose value may change from one line to the next. Similarly, we use c, ϵ or ω to denote a generic small positive constant. If a constant depends on a quantity a, we use $C(a)$ or C_a to indicate this dependence.

The canonical inner product on $\mathbb{R}^n$ is denoted $\langle \cdot, \cdot \rangle$, and the Euclidean norm is denoted $\| \cdot \|_2$. The distance from a point x to a set D in $\mathbb{R}^n$ is denoted $\mathrm{dist}(x, D)$. The unit sphere centered at the origin in $\mathbb{R}^n$ is denoted S^{n-1}. The orthogonal projection in $\mathbb{R}^n$ onto a subspace E is denoted P_E. For a subset of coordinates $J \subseteq \{1, \ldots, n\}$, we often write P_J for $P_{\mathbb{R}^J}$. The unit sphere of E is denoted $S(E) := S^{n-1} \cap E$.

For any (complex) matrix A, we use A^* to denote its conjugate transpose, A^T the transpose, $\|A\|$ the operator norm and $\|A\|_{HS}$ the Hilbert-Schmidt norm. We usually write an identity matrix as 1 without causing any confusions.

The following tensorization lemma is Lemma 2.2 of [16]

Lemma 2.1 (Tensorization). *Let $\zeta_1, \ldots, \zeta_n$ be independent non-negative random variables, and let $B, \epsilon_0 \geqslant 0$.*

(1) Assume that for each k,

$$\mathbb{P}(\zeta_k < \epsilon) \leqslant B\epsilon \quad \text{for all } \epsilon \geqslant \epsilon_0.$$

Then

$$\mathbb{P}\left(\sum_{k=1}^{n} \zeta_k^2 < \epsilon^2 n \right) \leqslant (CB\epsilon)^n \quad \text{for all } \epsilon \geqslant \epsilon_0,$$

where C is a universal 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 exists $\lambda_1 > 0$ and $\mu_1 \in (0,1)$ that depend on λ and μ only and such that

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

Consider a subset $D \subset \mathbb{R}^n$ and $\epsilon > 0$. An ϵ-net of D is a subset $\mathcal{N} \subset D$ such that for every $x \in D$ one has $\operatorname{dist}(x, \mathcal{N}) \leqslant \epsilon$. The following lemma about the cardinality of ϵ-nets is stated as Propositions 2.1 and 2.2 in [17].

Lemma 2.2 (Nets)**.** *Fix any $\epsilon > 0$.*

(1) There exists an ϵ-net of S^{n-1} of cardinality at most $2n\left(1 + 2\epsilon^{-1}\right)^{n-1}$.

(2) Let S be a subset of S^{n-1}. There exists an ϵ-net of S of cardinality at most $2n\left(1 + 4\epsilon^{-1}\right)^{n-1}$.

Next we define the small ball probability for a random vector.

Definition 2.3. *The Lévy concentration function of a random vector $S \in \mathbb{R}^m$ is defined for $\epsilon > 0$ as*

$$\mathcal{L}(S, \epsilon) = \sup_{v \in \mathbb{R}^m} \mathbb{P}(\|S - v\|_2 \leqslant \epsilon),$$

which measures the small ball probabilities.

With Definition 2.3, it is easy to prove the following lemma. It will allow us to select a nice subset of the coefficients a_k when computing the small ball probability.

Lemma 2.4. *For any $\sigma \subseteq \{1, \ldots, n\}$, any $a \in \mathbb{R}^n$ and any $\epsilon \geqslant 0$, we have*

$$\mathcal{L}\left(\sum_{k=1}^{n} a_k \xi_k, \epsilon\right) \leqslant \mathcal{L}\left(\sum_{k \in \sigma} a_k \xi_k, \epsilon\right).$$

The following three lemmas give some useful small ball probability bounds. They correspond to [17, Lemma 3.2], [16, Corollary 2.9] and [18, Corollary 2.4] respectively.

Lemma 2.5. *Let ξ be a random variable with mean zero, unit variance, and finite fourth moment. Then for every $\epsilon \in (0,1)$, there exists a $p \in (0,1)$ which depends only on ϵ and on the fourth moment, and such that*

$$\mathcal{L}(\xi, \epsilon) \leqslant p.$$

Lemma 2.6. *Let $\xi_1, \ldots, \xi_n$ 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 $\epsilon \geqslant 0$, one has*

$$\mathcal{L}\left(\sum_{k=1}^{n} a_k \xi_k, \epsilon\right) \leqslant \sqrt{\frac{2}{\pi}} \frac{\epsilon}{\|a\|_2} + C_1 B \left(\frac{\|a\|_3}{\|a\|_2}\right)^3,$$

where C_1 is an absolute constant.

Lemma 2.7. *Let A be a fixed $N \times M$ matrix. Consider a random vector $\zeta = (\zeta_1, \ldots, \zeta_M)$ where ζ_i are independent random variables satisfying $\mathbb{E}\zeta_i = 0$, $\mathbb{E}\zeta_i^2 = 1$ and $\|\zeta_i\|_{\psi_2} \leqslant K$. Then for every $y \in \mathbb{R}^N$, we have*

$$\mathbb{P}\left\{\|A\zeta - y\|_2 \leqslant \frac{1}{2}\|A\|_{HS}\right\} \leqslant 2\exp\left(-\frac{c\|A\|_{HS}^2}{K^4\|A\|^2}\right).$$

3 Decomposition of the sphere

The rest of this paper is devoted to prove Theorem 1.2. We will make use of a partition of the unit sphere into two sets of compressible and incompressible vectors. They are first defined in [16].

Definition 3.1. *Let $\delta \equiv \delta(N), \rho \in (0,1)$. A vector $x \in \mathbb{R}^n$ is called sparse if $|supp(x)| \leqslant \delta n$. 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 is called incompressible if it is not compressible. The sets of sparse, compressible and incompressible vectors in $\mathbb{R}^n$ will be denoted by $Sparse_n(\delta)$, $Comp_n(\delta,\rho)$ and $Incomp_n(\delta,\rho)$. We sometimes neglect the subindex n when the dimension is clear.*

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

$$\mathbb{P}\left(s_n(PY-A) \leqslant \epsilon(\sqrt{N}-\sqrt{n-1})\right) \leqslant \mathbb{P}\left(\inf_{x\in Comp_n(\delta,\rho)} \|(PY-A)x\|_2 \leqslant \epsilon(\sqrt{N}-\sqrt{n-1})\right) \tag{3.1}$$

$$+\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|(PY-A)x\|_2 \leqslant \epsilon(\sqrt{N}-\sqrt{n-1})\right). \tag{3.2}$$

The bound for compressible vectors follows from the following lemma, which is a variant of Lemma 3.3 from [16].

Lemma 3.2 (Invertibility for compressible vectors)**.** *There exits N-independent constants $\rho, c_0, c_1 > 0$ such that for $\delta \leqslant \min\left\{c_1 N/\left(nK^4 \log K\right), 1\right\}$, we have*

$$\mathbb{P}\left(\inf_{x\in Comp_n(\delta,\rho)} \|(PY-A)x\|_2 \leqslant c_0\sqrt{N}\right) \leqslant e^{-c_0 N/K^4}.$$

Proof. We first prove a weaker result. For any fixed $x \in S^{n-1}$, we define the random vector $\zeta := Yx \in \mathbb{R}^N$. It is easy to verify that $\mathbb{E}\zeta_i = 0$, $\mathbb{E}\zeta_i^2 = 1$ and $\|\zeta_i\|_{\psi_2} \leqslant CK$. Then with $\|P\| = 1$ and $\|P\|_{HS}^2 = N$, we conclude from the Lemma 2.7 that

$$\mathbb{P}\left\{\|(PY-A)x\|_2 \leqslant \frac{1}{2}\sqrt{N}\right\} \leqslant 2\exp\left(-\frac{cN}{K^4}\right). \tag{3.3}$$

Let $S_1 := \{x \in S^{n-1} : x_k = 0, k > \delta n\}$. There exists an ϵ-net $\mathcal{N}$ in S_1 of cardinality $|\mathcal{N}| \leqslant (3/\epsilon)^{\delta n}$. Then by (3.3) and the union bound, we get

$$\mathbb{P}\left(\exists x \in \mathcal{N} : \|(PY-A)x\|_2 \leqslant \frac{1}{2}\sqrt{N}\right) \leqslant 2e^{-cN/K^4}\left(3\epsilon^{-1}\right)^{\delta n}. \tag{3.4}$$

Let V be the event that $\|(PY-A)y\| \leqslant \sqrt{N}/4$ for some $y \in S_1$. Recall that by the assumptions of Theorem 1.2, $\|PY-A\| \leqslant C_1\sqrt{N}$ for some constant C_1. Assume that V occurs and choose a point $x \in \mathcal{N}$ such that $\|y-x\| \leqslant \epsilon$. Then

$$\|(PY-A)x\|_2 \leqslant \|(PY-A)y\|_2 + \|PY-A\|\|x-y\|_2 \leqslant \frac{1}{4}\sqrt{N} + C_1\epsilon\sqrt{N} \leqslant \frac{1}{2}\sqrt{N}$$

if we choose $\epsilon \leqslant 1/(4C_1)$. Fix one such ϵ, using (3.4) we obtain that

$$\mathbb{P}\left(\inf_{x\in S_1}\|(PY-A)x\|_2 \leqslant \frac{1}{4}\sqrt{N}\right) = \mathbb{P}(V) \leqslant 2e^{-cN/K^4}\left(3\epsilon^{-1}\right)^{\delta n} \leqslant e^{-c_2 N/K^4},$$

if we choose c_1 (and hence δ) to be sufficiently small. We use this result and take the union bound over all $\lceil \delta n \rceil$-element subsets σ of $\{1, \ldots, n\}$:

$$\begin{aligned}
&\mathbb{P}\left(\inf_{x \in Sparse(\delta) \cap S^{n-1}} \|(PY - A)x\|_2 \leqslant \frac{1}{4}\sqrt{N}\right) \\
=&\mathbb{P}\left(\exists \sigma, |\sigma| = \lceil \delta n \rceil : \inf_{x \in \mathbb{R}^\sigma \cap S^{n-1}} \|(PY - A)x\|_2 \leqslant \frac{1}{4}\sqrt{N}\right) \\
\leqslant &\binom{n}{\lceil \delta n \rceil} e^{-c_2 N / K^4} \leqslant \exp\left(4e\delta \log\left(\frac{e}{\delta}\right) n - \frac{c_2 N}{K^4}\right) \leqslant \exp\left(-\frac{c_2 N}{2K^4}\right),
\end{aligned} \tag{3.5}$$

with an appropriate choice of c_1, which depends only on c_2.

Finally 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 control the event W that $\|(PY - A)x\|_2 \leqslant c_3\sqrt{N}$ for some vector $x \in Comp(\delta, \rho)$. Assume W occurs, then every such vector x can be written as a sum $x = y + z$ with $y \in Sparse(\delta)$ and $\|z\|_2 \leqslant \rho$. Thus $\|y\|_2 \geqslant 1 - \rho \geqslant 1/2$, and

$$\|(PY - A)y\|_2 \leqslant \|(PY - A)x\|_2 + \|(PY - A)\|_2 \|z\|_2 \leqslant c_3\sqrt{N} + \rho C_1 \sqrt{N}.$$

We choose $c_3 = 1/16$ and $\rho = 1/(16C_1)$, so that $\|(PY - A)y\|_2 \leqslant \sqrt{N}/8$. Since $\|y\|_2 \geqslant 1/2$, we can find a unit vector $u = y/\|y\|_2 \in Sparse(\delta)$ such that $\|(PX - A)u\|_2 \leqslant \sqrt{N}/4$. This shows that event W implies the event in (3.5), so we have $\mathbb{P}(W) \leqslant e^{-c_2 N/(2K^4)}$. This concludes the proof. □

Remark 3.3. *If $n < c_1 N/(K^4 \log K)$, then all the vectors in S^{n-1} are compressible by Definition 3.1. Hence throughout the rest of this paper, it suffices to assume*

$$n \geqslant c_1 N/(K^4 \log K). \tag{3.6}$$

In particular, n is assumed to be an arbitrarily large number in the following proof.

It only remains to prove the bound for incompressible vectors in (3.2). We will divide the proof into two cases: the case where the aspect ratio $\lambda = n/N \leqslant \lambda_0$ for some constant $0 < \lambda_0 < 1$, and the case where $\lambda_0 < \lambda \leqslant 1$. For simplicity we choose $\lambda_0 = 1/2$ in the following, although λ_0 can be any constant between 0 and 1. We record here an important property of the incompressible vectors, which will be used in the proof of both cases. It is Lemma 3.4 of [16].

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

$$\frac{\rho}{\sqrt{2n}} \leqslant |x_k| \leqslant \frac{1}{\sqrt{\delta n}} \quad \text{for all } k \in \sigma. \tag{3.7}$$

4 Tall matrices

In this section, we deal with the bound in (3.2) for the $\lambda < \lambda_0 = 1/2$ case, i.e. $PY - A$ is a *tall matrix.* Then it is equivalent to control the probability

$$\mathbb{P}\left(\inf_{x \in Incomp_n(\delta, \rho)} \|(PY - A)x\|_2 \leqslant t\sqrt{N}\right).$$

Our proof is a modification of the one in [12].

Let x be a vector in $Incomp_n(\delta,\rho)$ for δ and ρ in Lemma 3.2. Then we take the set σ given by Lemma 3.4. Note that every entry of Yx is of the form $(Yx)_i = \sum_{k=1}^n Y_{ik}x_k$, $1 \leqslant i \leqslant M$, where Y_{ik} are independent centered random variables with unit variance and bounded fourth moment. Hence we can use Lemma 2.6 and Lemma 2.4 to get that

$$\mathcal{L}\left((Yx)_i, t\right) \leqslant \sqrt{\frac{2}{\pi}}\frac{t}{\|P_\sigma x\|_2} + C\left(\frac{\|P_\sigma x\|_3}{\|P_\sigma x\|_2}\right)^3 \leqslant C_2(\rho)\left(\frac{t}{\sqrt{\delta}} + \frac{1}{\delta\sqrt{n}}\right), \tag{4.1}$$

for some constant $C_2(\rho) > 0$ depending only on ρ and the fourth moment. Here in the second step, we used the bound

$$\|P_\sigma x\|_2 \geqslant \frac{1}{2}\rho^2\sqrt{\delta}, \quad \left(\frac{\|P_\sigma x\|_3}{\|P_\sigma x\|_2}\right)^3 \leqslant \frac{2}{\rho^2\delta\sqrt{n}},$$

deduced from Lemma 3.4. With (4.1) as the input, the next lemma provides a small ball probability bound for PYx. One can refer to [19, Corollary 1.4] for the proof.

Lemma 4.1. *Consider a random vector $X = (\xi_1, \ldots, \xi_M)$ where ξ_i are real-valued independent random variables. Let $t, p \geqslant 0$ such that*

$$\mathcal{L}(\xi_i, t) \leqslant p \quad \text{for all } i = 1, \ldots, M.$$

Let P be an orthogonal projection in $\mathbb{R}^M$ onto an N-dimensional subspace. Then

$$\mathcal{L}\left(PX, t\sqrt{N}\right) \leqslant (Cp)^N$$

for some absolute constant $C > 0$.

Apply the above lemma to random vector Yx, we obtain that

$$\mathbb{P}\left(\|(PY-A)x\|_2 \leqslant t\sqrt{N}\right) \leqslant \mathcal{L}\left(PYx, t\sqrt{N}\right) \leqslant \left[C_3(\rho)\left(\frac{t}{\sqrt{\delta}} + \frac{1}{\delta\sqrt{n}}\right)\right]^N \tag{4.2}$$

for some constant $C_3(\rho) > 0$ depending only on ρ and the fourth moment. Then, as in the proof for Lemma 3.2, we can take a union bound over all x in an ϵ-net of $Incomp_n(\delta,\rho)$ and complete the proof by approximation.

Without loss of generality, we may assume $t \geqslant 1/\sqrt{\delta n}$. Otherwise, we can use the trivial bound

$$\mathbb{P}\left(s_n(PY-A) \leqslant t\sqrt{N}\right) \leqslant \mathbb{P}\left(s_n(PY-A) \leqslant (\delta n)^{-1/2}\sqrt{N}\right), \quad \text{for } t < 1/\sqrt{\delta n},$$

and the bound from Lemma 3.2 will dominate. Under the above assumption, the $t/\sqrt{\delta}$ term in (4.2) dominates. Thus we have

$$\mathbb{P}\left(\|(PY-A)x\|_2 \leqslant t\sqrt{N}\right) \leqslant \left(2C_3 t/\sqrt{\delta}\right)^N.$$

By Lemma 2.2, there exists an ϵ-net $\mathcal{N}$ in $Incomp_n(\delta,\rho)$ of cardinality $|\mathcal{N}| \leqslant 2n(5/\epsilon)^{n-1}$. Taking the union bound, we get

$$\mathbb{P}\left(\exists x \in \mathcal{N} : \|(PY-A)x\|_2 \leqslant t\sqrt{N}\right) \leqslant 2n\left(\frac{2C_3 t}{\sqrt{\delta}}\right)^N\left(\frac{5}{\epsilon}\right)^{n-1}. \tag{4.3}$$

Let V be the event that $\|(PY-A)y\| \leqslant t\sqrt{N}/2$ for some $y \in Incomp_n(\delta,\rho)$. Recall that $\|PY-A\| \leqslant C_1\sqrt{N}$. Assume that V occurs and choose a point $x \in \mathcal{N}$ such that $\|x-y\| \leqslant \epsilon$. Then

$$\|(PY-A)x\|_2 \leqslant \|(PY-A)y\|_2 + \|PY-A\|\|x-y\|_2 \leqslant \frac{1}{2}t\sqrt{N} + C_1\epsilon\sqrt{N} \leqslant t\sqrt{N}$$

if we choose $\epsilon \leqslant t/(2C_1)$. Fix one such ϵ, from (4.3) we obtain that

$$\begin{aligned} &\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|(PY-A)x\|_2 \leqslant \frac{t}{2}\sqrt{N}\right) = \mathbb{P}(V) \\ &\leqslant 2n\left(\frac{2C_3t}{\sqrt{\delta}}\right)^N \left(\frac{10C_1}{t}\right)^{n-1} \leqslant \left(C_4\delta^{-1}t\right)^{N-n+1}, \end{aligned} \tag{4.4}$$

where in the last step we used $n/N \leqslant \lambda_0 = 1/2$. Together with Lemma 3.2, this concludes the proof of Theorem 1.2 for the $\lambda < \lambda_0$ case.

5 Almost square matrices

In this section, we deal with the bound in (3.2) for the $1/2 = \lambda_0 < \lambda \leqslant 1$ case. In particular, when $\lambda \to 1$, $PY-A$ becomes an *almost square matrix* and (4.4) cannot provide a satisfactory probability bound. For instance, for the square matrix with $N = n$, it is easy to see that the $\delta^{-N/2}$ term dominates over the t term. To handle this difficulty, we will use the mathods in [17], which reduce the problem of bounding $\|(PY-A)x\|_2$ for incompressible vectors x to a random distance problem. We denote $N = n-1+d$ for some $d \geqslant 1$. Note that $\sqrt{N}-\sqrt{n-1} \leqslant d/\sqrt{n}$. Hence to bound (3.2), it suffices to bound

$$\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|(PY-A)x\|_2 \leqslant \epsilon\frac{d}{\sqrt{n}}\right).$$

We denote

$$m := \min\left\{d, \left\lfloor \frac{1}{2}\rho^2\delta n \right\rfloor\right\}. \tag{5.1}$$

Let $Z_1 := PY_1 - A_1, \ldots, Z_n := PY_n - A_n$ be the columns of the matrix $Z := PY - A$. Given a subset $J \subset \{1,\ldots,n\}$ of cardinality m, we define the subspaces

$$H_{J^c} := \operatorname{span}(Z_k)_{k\in J^c} \subset \mathbb{R}^N. \tag{5.2}$$

For levels $K_1 = \rho\sqrt{\delta/2}$ and $K_2 = K_1^{-1}$, we define the set of totally spread vectors

$$S^J := \left\{y \in S(\mathbb{R}^J) : \frac{K_1}{\sqrt{m}} \leqslant |y_k| \leqslant \frac{K_2}{\sqrt{m}} \text{ for all } k \in J\right\}. \tag{5.3}$$

In the following lemma, we let J be a random subset uniformly chosen over all subsets of $\{1,\ldots,n\}$ of cardinality m. We shall write P_J for $P_{\mathbb{R}^J}$, the orthogonal projection onto the subspace $\mathbb{R}^J$, and denote the probability and expectation over the random subset J by $\mathbb{P}_J$ and $\mathbb{E}_J$.

Lemma 5.1. *There exists constant $c_2 > 0$ depending only on ρ such that for every $x \in Incomp_n(\delta,\rho)$, the event*

$$\mathcal{E}(x) := \left\{\frac{P_Jx}{\|P_Jx\|_2} \in S^J \text{ and } \frac{\rho\sqrt{m}}{\sqrt{2n}} \leqslant \|P_Jx\|_2 \leqslant \frac{\sqrt{m}}{\sqrt{\delta n}}\right\}$$

satisfies $\mathbb{P}_J(\mathcal{E}(x)) \geqslant (c_2\delta)^m$.

Proof. Let $\sigma \subset \{1, \ldots, n\}$ be the subset from Lemma 3.4. Then we have

$$\mathbb{P}_J (J \subset \sigma) = \binom{|\sigma|}{m} / \binom{n}{m}.$$

Using Stirling's approximation, for $d \leqslant \frac{1}{4}\rho^2 \delta n$, we have

$$\mathbb{P}_J (J \subset \sigma) \geqslant \left(\frac{c|\sigma|}{n}\right)^m \geqslant (c_2\delta)^m,$$

and for $d > \frac{1}{4}\rho^2 \delta n$, we have

$$\mathbb{P}_J (J \subset \sigma) \geqslant \binom{n}{m}^{-1} \geqslant \frac{m!}{n^m} \geqslant \left(\frac{cm}{n}\right)^m \geqslant (c_2\delta)^m.$$

If $J \subset \sigma$, then summing (3.7) over $k \in J$, we obtain the required two-sided bound for $\|P_J x\|_2$. This and (3.7) yield $P_J x/\|P_J x\|_2 \in S^J$. Hence $\mathcal{E}(x)$ holds. □

Lemma 5.1 implies the following lemma, whose proof is a minor modification of the one for [17, Lemma 6.2].

Lemma 5.2. *Let J denote the m-element subsets of $\{1, \ldots, n\}$. Then for every $\epsilon > 0$,*

$$\mathbb{P}\left(\inf_{x \in Incomp_n(\delta,\rho)} \|Zx\|_2 < \epsilon\rho\sqrt{\frac{m}{2n}}\right) \leqslant (c_2\delta)^{-m} \max_J \mathbb{P}\left(\inf_{x\in S^J} \operatorname{dist}(Zx, H_{J^c}) < \epsilon\right). \tag{5.4}$$

It remains to bound $\mathbb{P}\left(\inf_{x\in S^J} \operatorname{dist}(Zx, H_{J^c}) < \epsilon\right)$ for any m-element subset J.

Theorem 5.3 (Uniform distance bound)**.** *Let Y be a random matrix satisfying the assumptions in Theorem 1.2. Then for any m-element subset J and any $t > 0$, there exist $C, c > 0$ independent of J such that*

$$\mathbb{P}\left(\inf_{x\in S^J} \operatorname{dist}(Zx, H_{J^c}) < t\sqrt{d}\right) \leqslant (CtK^5 \log K)^d + e^{-cN}. \tag{5.5}$$

By the definition of m in (5.1), we have

$$(c_2\delta)^{-m} \leqslant \left[(c_2\delta)^{-\rho^2\delta/2}\right]^n \leqslant e^{-cN/2}, \tag{5.6}$$

where we used that $\delta = c_1/(K^4 \log K)$ for some sufficiently small $c_1 > 0$ (see Lemma 3.2). Then we conclude from Theorem 5.3 and Lemma 5.2 that

$$\begin{aligned}\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|(PY - A)x\|_2 < \epsilon\rho\frac{\sqrt{md}}{\sqrt{n}}\right) &\leqslant (c_2\delta)^{-m}(CK^5 \log K\epsilon)^d + (c_2\delta)^{-m}e^{-cN} \\ &\leqslant \left(C'K^9(\log K)^2\epsilon\right)^d + e^{-cN/2},\end{aligned} \tag{5.7}$$

where in the last step we used (5.6), $m \leqslant d$ and $d/m \leqslant CK^4 \log K$. Hence we obtain from (5.7) that

$$\mathbb{P}\left(\inf_{x\in Incomp_n(\delta,\rho)} \|(PY - A)x\|_2 < \epsilon\rho\frac{d}{\sqrt{n}}\right) \leqslant \left(C'K^{11}(\log K)^{5/2}\epsilon\right)^d + e^{-cN/2},$$

which, together with Lemma 3.2, concludes the proof of Theorem 1.2.

To prove Theorem 5.3, we need the following lemma, which gives a lower bound for the distance between a single random vector and a random subspace. We will prove it in Section 6. Note that by choosing small enough ρ and δ in (5.1), we can assume $m + d - 1 \leqslant 3N/4$.

Lemma 5.4 (Distance to a random subspace)**.** *Let J be any m-element subset of $\{1, \ldots, n\}$ and let H_{J^c} be the random subspace of $\mathbb{R}^N$ as defined in (5.2). Let X be a random vector in $\mathbb{R}^M$ whose coordinates are i.i.d. centered random variables with unit variance and finite fourth moment, independent of H_{J^c}. Let $l = m + d - 1$ such that $l \leqslant 3N/4$. Then for every $\epsilon > 0$, we have*

$$\mathbb{P}\left(\sup_{v\in\mathbb{R}^N} \mathbb{P}\left(\operatorname{dist}(PX - v, H_{J^c}) < \epsilon\sqrt{l} \,\middle|\, H_{J^c}\right) > (C\epsilon)^l + e^{-cN}\right) \leqslant e^{-cN}, \tag{5.8}$$

where $C, c > 0$ depend only on the fourth moment of the entries of X. Moreover, (5.8) implies the following weaker result:

$$\sup_{v\in\mathbb{R}^N} \mathbb{P}\left(\operatorname{dist}(PX - v, H_{J^c}) < \epsilon\sqrt{l}\right) \leqslant (C\epsilon)^l + e^{-cN}. \tag{5.9}$$

Note that for any fixed $x \in S^J$, we have $Zx = PYx - Ax$, where Yx is a random vector that satisfies the assumption for X in Lemma 5.4. So (5.9) gives a useful probability bound for a single $x \in S^J$. We might then take a union bound over all z in an ϵ-net of S^J and complete by approximation. Before proving Theorem 5.3, we make several reductions. First, we can assume that the entries of Y have absolute continuous distributions. In fact we can add to each entry an independent Gaussian random variable with small variance σ, and later let $\sigma \to 0$ (all the estimates below do not depend on σ). The main purpose of this reduction is to have the following equality

$$\dim(H_{J^c}) = n - m \quad \text{a.s.} \tag{5.10}$$

Second, without loss of generality, in the proof of (5.5) we can assume

$$t \geqslant e^{-c'N/d}, \quad \text{for some small } c' > 0. \tag{5.11}$$

Let $P_{H^\perp}$ be the orthogonal projection in $\mathbb{R}^N$ onto $H_{J^c}^\perp$, and define

$$W := P_{H^\perp} PY|_{\mathbb{R}^J}\,. \tag{5.12}$$

Then for every $x \in \mathbb{R}^n$, we have

$$\operatorname{dist}(PYx - v, H_{J^c}) = \|Wx - w\|_2\,, \quad \text{where } w = P_{H^\perp} v. \tag{5.13}$$

By (5.10), $\dim(H_{J^c}^\perp) = N - n + m = l$ almost surely. Thus W acts as an operator from an m-dimensional subspace into an l-dimensional subspace almost surely. If we have a proper operator bound for W, we can run the approximation argument on the ϵ-net of S^J and prove a uniform distance bound over all $x \in S^J$.

Lemma 5.5. *Let W be a random matrix as in (5.12) and let w be a random vector as in (5.13). For every t satisfying (5.11) and every $C_0 > 0$, we have*

$$\mathbb{P}\left(\inf_{x\in S^J} \|Wx - w\|_2 < t\sqrt{d}, \|W\| \leqslant C_0 K\sqrt{d}\right) \leqslant K^{m-1}(C_1 t)^d.$$

Proof. Fix any $x \in S^J$. It is easy to verify that Yx is a random vector that satisfies the assumption for X in Lemma 5.4, which implies

$$\mathbb{P}\left(\|Wx - w\|_2 < t\sqrt{d}\right) \leqslant \mathbb{P}\left(\operatorname{dist}(PX - v, H_{J^c}) < t\sqrt{l}\right) \leqslant (Ct)^{m+d-1}, \tag{5.14}$$

for t satisfying (5.11).

Let $\epsilon = t/(C_0 K)$. By Lemma 2.2, there exists an ϵ-net $\mathcal{N}$ of $S^J \subset S(\mathbb{R}^J)$ of cardinality

$$|\mathcal{N}| \leqslant 2m\left(1+\frac{2}{\epsilon}\right)^{m-1} \leqslant 2m\left(\frac{3C_0K}{t}\right)^{m-1}.$$

Consider the event

$$\mathcal{E} := \left\{\inf_{x\in\mathcal{N}} \|Wx-w\|_2 < 2t\sqrt{d}\right\}.$$

Taking union bound, we get

$$\mathbb{P}(\mathcal{E}) \leqslant |\mathcal{N}|\mathbb{P}(\|Wx-w\|_2 \leqslant 2t\sqrt{d}) \leqslant 2m\left(\frac{3C_0K}{t}\right)^{m-1}(Ct)^{m+d-1} \leqslant K^{m-1}(C_1t)^{d-1}. \tag{5.15}$$

Suppose there exists $y \in S^J$ such that

$$\|Wy-w\|_2 < t\sqrt{d}, \quad \|W\| \leqslant C_0K\sqrt{d}.$$

Then we choose $x \in \mathcal{N}$ such that $\|x-y\|_2 \leqslant \epsilon$, and by triangle inequality we obtain that

$$\|Wx-w\|_2 \leqslant \|Wy-w\|_2 + \|W\|\|x-y\|_2 < t\sqrt{d} + C_0K\sqrt{d}\epsilon \leqslant 2t\sqrt{d},$$

i.e. the event $\mathcal{E}$ holds. Then the bound (5.15) concludes the proof. □

To bound the norm of W, we have the following proposition.

Proposition 5.6. *Let W be a random matrix as in (5.12). Then*

$$\mathbb{P}\left(\|W\| \geqslant sK\sqrt{d} \,\middle|\, H_{J^c}\right) \leqslant e^{-\tilde{c}s^2d}, \quad \textit{for } s \geqslant \tilde{C},$$

where $\tilde{C}, \tilde{c} > 0$ are absolute constants.

Proof. For simplicity of notations, we fix an H_{J^c} during the proof and omit the conditioning on it from the expressions below. Let $\mathcal{N}$ be an $(1/2)$-net of $S(\mathbb{R}^J)$ and $\mathcal{M}$ be an $(1/2)$-net of $S(H_{J^c}^{\perp})$. By Lemma 2.2, we can choose $\mathcal{N}$ and $\mathcal{M}$ such that

$$|\mathcal{N}| \leqslant 2m \cdot 5^{m-1}, \quad |\mathcal{M}| \leqslant 2l \cdot 5^{l-1}.$$

It is easy to prove that

$$\|W\| \leqslant 4 \sup_{x\in\mathcal{N}, y\in\mathcal{M}} |\langle Wx, y\rangle|. \tag{5.16}$$

For every $x \in \mathcal{N}$ and $y \in \mathcal{M}$, $\langle Wx, y\rangle = \langle PYx, y\rangle = \langle Yx, P^Ty\rangle$ is a random variable with subgaussian moment bounded by CK for some absolute constant $C > 0$. Hence we have

$$\mathbb{P}\left(|\langle Wx, y\rangle| \geqslant sK\sqrt{d}\right) \leqslant 2e^{-cs^2d}.$$

Using (5.16) and taking the union bound, we get

$$\mathbb{P}\left(\|W\| \geqslant sK\sqrt{d}\right) \leqslant 32ml \cdot 5^{m-1} \cdot 5^{l-1} \cdot e^{-cs^2d} \leqslant e^{-\tilde{c}s^2d},$$

where in the last step we used $m \leqslant l \leqslant 2d$ and $s \geqslant \tilde{C}$ for some sufficiently large $\tilde{C} > 0$. □

With Lemma 5.5 and Proposition 5.6, we have

$$\mathbb{P}\left(\inf_{x\in S^J}\|Wx-w\|_2<t\sqrt{d}\right)\leqslant K^{m-1}(C_1t)^d+e^{-\tilde{c}C_0^2d}.$$

Unfortunately, the bound $e^{-\tilde{c}C_0^2d}$ is too weak for small d. Following the idea in [17], we refine Lemma 5.5 by decoupling the information about $\|Wx-w\|_2$ from the information about $\|W\|$.

Lemma 5.7 (Decoupling)**.** *Let W be an $N\times m$ matrix whose columns are independent random vectors, and let w be an arbitrary fixed vector in $\mathbb{R}^N$. Let $z\in S^J$, which satisfies $z_k\geqslant K_1/\sqrt{m}$ for all $1\leqslant k\leqslant m$ (see (5.3)). Then for every $0<a<b$, we have*

$$\mathbb{P}\left(\|Wz-w\|_2<a,\|W\|>b\right)\leqslant 2\sup_{x\in S^{m-1},v\in\mathbb{R}^N}\mathbb{P}\left(\|Wx-v\|_2<\frac{\sqrt{2}a}{K_1}\right)\mathbb{P}\left(\|W\|>\frac{b}{\sqrt{2}}\right).$$

Proof. If $m=1$, then $\|W\|=\|Wz\|_2$, so the probability on the left hand side is zero. Hence in the following we only consider $m\geqslant 2$. Then we can decompose the index set $\{1,\dots,m\}$ into two disjoint subsets I and H with $|I|=\lceil m/2\rceil$. We write $W=W_I+W_H$, where W_I and W_H are the submatrices of W with columns in I and H, respectively. Similarly, we write $z=z_I+z_H$ for $z\in S^J$. Using $\|W\|^2\leqslant\|W_I\|^2+\|W_H\|^2$, we have

$$\mathbb{P}\left(\|Wz-w\|_2<a,\|W\|>b\right)\leqslant p_I+p_H,$$

where

$$\begin{aligned}p_I&=\mathbb{P}\left(\|Wz-w\|_2<a,\|W_H\|>b/\sqrt{2}\right)\\&=\mathbb{P}\left(\|Wz-w\|_2<a\,\middle|\,\|W_H\|>b/\sqrt{2}\right)\mathbb{P}\left(\|W_H\|>b/\sqrt{2}\right),\end{aligned}$$

and similarly for p_H. We shall bound p_I in the following, while the arguments for p_H are similar.

Writing $Wz=W_Iz_I+W_Hz_H$ and using the independence of W_I and W_H, we conclude that

$$\begin{aligned}p_I&\leqslant\sup_{u\in\mathbb{R}^N}\mathbb{P}\left(\|W_Iz_I-u\|_2<a\right)\mathbb{P}\left(\|W_H\|>\frac{b}{\sqrt{2}}\right)\\&\leqslant\sup_{u\in\mathbb{R}^N}\mathbb{P}\left(\|Wz_I-u\|_2<a\right)\mathbb{P}\left(\|W\|>\frac{b}{\sqrt{2}}\right).\end{aligned}\tag{5.17}$$

By the assumption on z and using $|I|\geqslant m/2$, we have $\|z_I\|_2\geqslant K_1/\sqrt{2}$. Hence for $x=z_I/\|z_I\|_2$ and $v=u/\|z_I\|_2$, we obtain

$$\mathbb{P}\left(\|Wz_I-u\|_2<a\right)\leqslant\mathbb{P}\left(\|Wx-v\|_2<\sqrt{2}a/K_1\right).$$

Together with (5.17), this concludes the proof. □

With this decoupling lemma, we can prove the following refinement of Lemma 5.5.

Lemma 5.8. *Let W be a random matrix as in (5.12) and let w be a random vector as in (5.13). For every $s>1$, every $C_0>0$ and every t satisfying (5.11), we have*

$$\mathbb{P}\left(\inf_{x\in S^J}\|Wx-w\|_2<t\sqrt{d}\text{ and }sC_0K\sqrt{d}<\|W\|\leqslant 2sC_0K\sqrt{d}\right)\leqslant K^{m-1}K_1^{-(m+d-1)}\left(C_2te^{-c_2s^2}\right)^d.$$

Proof. Let $\epsilon = t/(2sC_0K)$. By Lemma 2.2, there exists an ϵ-net $\mathcal{N}$ of $S^J \subset S(\mathbb{R}^J)$ of cardinality

$$|\mathcal{N}| \leqslant 2m\left(1+\frac{4}{\epsilon}\right)^{m-1} \leqslant 2m\left(\frac{9sC_0K}{t}\right)^{m-1}.$$

Consider the event

$$\mathcal{E} := \left\{\inf_{x\in\mathcal{N}} \|Wx-w\|_2 < 2t\sqrt{d} \text{ and } \|W\| > sC_0K\sqrt{d}\right\}.$$

Note that conditioning on H_{J^c}, W and w satisfy the assumptions in Lemma 5.7. Using Lemma 5.7 and taking union bound, we get

$$\mathbb{P}\left(\mathcal{E}\middle|\, H_{J^c}\right) \leqslant |\mathcal{N}|\cdot 2 \sup_{x\in S^{d-1}, v\in\mathbb{R}^N} \mathbb{P}\left(\|Wx-v\|_2 < \frac{\sqrt{2}}{K_1}\cdot 2t\sqrt{d}\,\middle|\, H_{J^c}\right)\mathbb{P}\left(\|W\| \geqslant \frac{sC_0K\sqrt{d}}{\sqrt{2}}\,\middle|\, H_{J^c}\right)$$

Taking expectation over H_{J^c} and using Proposition 5.6, we obtain that

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\leqslant 4m\left(\frac{9sC_0K}{t}\right)^{m-1} e^{-c's^2d}\,\mathbb{E}\left[\sup_{x\in S^{d-1}, v\in\mathbb{R}^N} \mathbb{P}\left(\|Wx-v\|_2 < \frac{\sqrt{2}}{K_1}\cdot 2t\sqrt{d}\,\middle|\, H_{J^c}\right)\middle|\, H_{J^c}\right]\\ &\leqslant 4m\left(\frac{9sC_0K}{t}\right)^{m-1} e^{-c's^2d}\left(\frac{C't}{K_1}\right)^{m+d-1} \leqslant K^{m-1}K_1^{-(m+d-1)}\left(C_2te^{-c_2s^2}\right)^d, \qquad (5.18)\end{aligned}$$

where in the second step we used the representation in (5.13), (5.8) and (5.11), and in the last step we used $s \geqslant 1$ and $1 \leqslant m \leqslant d$.

Suppose there exists $y \in S^J$ such that

$$\|Wy-w\|_2 < t\sqrt{d} \quad\text{and}\quad sC_0K\sqrt{d} < \|W\| \leqslant 2sC_0K\sqrt{d}.$$

Then we choose $x \in \mathcal{N}$ such that $\|x-y\|_2 \leqslant \epsilon$, and by triangle inequality we obtain that

$$\|Wx-w\|_2 \leqslant \|Wy-w\|_2 + \|W\|\|x-y\|_2 < t\sqrt{d} + 2sC_0K\sqrt{d}\epsilon \leqslant 2t\sqrt{d},$$

i.e. the event $\mathcal{E}$ holds. The bound (5.18) concludes the proof. □

Proof of Theorem 5.3. Recall that we assume (5.11) holds. Summing the probability bounds in Lemma 5.5 and Lemma 5.8 for $s = 2^k$, $k \in \mathbb{Z}_+$, we conclude that

$$\begin{aligned}\mathbb{P}\left(\inf_{x\in S^J}\|Wx-w\|_2 < t\sqrt{d}\right) &\leqslant K^{m-1}(C_1t)^d + \sum_{s=2^k, k\in\mathbb{Z}_+} K^{m-1}K_1^{-(m+d-1)}\left(C_2te^{-c_2s^2}\right)^d\\ &\leqslant \left(C_3KK_1^{-2}t\right)^d.\end{aligned}$$

Recall that $K_1 = \rho\sqrt{\delta/2}$ (see (5.3)) and $\delta = c_1N/(nK^4\log K)$ (see Lemma 3.2). Then we get that

$$\mathbb{P}\left(\inf_{x\in S^J}\|Wx-w\|_2 < t\sqrt{d}\right) \leqslant \left(C_3tK^5\log K\right)^d.$$

This concludes the proof. □

6 Proof of Lemma 5.4

We will first prove a general inequality that holds for any fixed subspace H. This probability bound depends on the arithmetic structure of the subspace, which can be expressed using the *least common denominator* (LCD). Then we will prove a bound for the LCD of the random subspace H_{J^c}.

For $\alpha > 0$ and $\gamma \in (0,1)$, the least common denominator of a vector $a \in \mathbb{R}^M$ is defined as

$$\mathrm{LCD}_{\alpha,\gamma}(a) := \inf\left\{\theta > 0 : \mathrm{dist}(\theta a, \mathbb{Z}^M) < \min(\gamma\|\theta a\|_2, \alpha)\right\}.$$

The above definition can be generalized to higher dimensions. Let $a = (a_1, \ldots, a_M)$ be a sequence of vectors $a_k \in \mathbb{R}^l$. We define the product of the multi-vector a and a vector $\theta \in \mathbb{R}^l$ as

$$\theta \cdot a = (\langle\theta, a_1\rangle, \ldots, \langle\theta, a_M\rangle) \in \mathbb{R}^M.$$

Then we define, for $\alpha > 0$ and $\gamma \in (0,1)$,

$$\mathrm{LCD}_{\alpha,\gamma}(a) := \inf\left\{\|\theta\|_2 : \theta \in \mathbb{R}^l, \mathrm{dist}(\theta\cdot a, \mathbb{Z}^M) < \min(\gamma\|\theta\cdot a\|_2, \alpha)\right\}.$$

Finally for $\alpha > 0$ and $\gamma \in (0,1)$, the least common denominator of a *subspace* $E \subset \mathbb{R}^M$ is defined as

$$\mathrm{LCD}_{\alpha,\gamma}(E) := \inf\{\mathrm{LCD}_{\alpha,\gamma}(a) : a \in S(E)\} = \inf\left\{\|\theta\|_2 : \theta \in E, \mathrm{dist}(\theta, \mathbb{Z}^M) < \min(\gamma\|\theta\|_2, \alpha)\right\}. \quad (6.1)$$

A key to the proof is the next small ball probability theorem. It is stated as Theorem 3.3 in [17].

Theorem 6.1. *Consider a sequence $a = (a_1, \ldots, a_M)$ of vectors $a_k \in \mathbb{R}^l$, which satisfies*

$$\sum_{k=1}^{M}\langle x, a_k\rangle^2 \geqslant \|x\|_2^2 \quad \text{for every } x \in \mathbb{R}^l. \quad (6.2)$$

Let $\xi_1, \ldots, \xi_M$ be independent and identically distributed, mean zero random variables, such that $\mathcal{L}(\xi_k, 1) \leqslant 1 - b$ for some $b > 0$. Consider the random sum $S = \sum_{k=1}^{M} a_k\xi_k \in \mathbb{R}^l$. Then for every $\alpha > 0$ and $\gamma \in (0,1)$, and for

$$\epsilon \geqslant \frac{\sqrt{l}}{\mathrm{LCD}_{\alpha,\gamma}(a)},$$

we have

$$\mathcal{L}(S, \epsilon\sqrt{l}) \leqslant \left(\frac{C\epsilon}{\gamma\sqrt{b}}\right)^l + C^l e^{-2b\alpha^2}.$$

Let H be a fixed subspace in $\mathbb{R}^N$ of dimension $N - l$. We denote an orthonormal basis of $H^\perp$ by $\{n_1, \ldots, n_l\} \subset \mathbb{R}^N$, and write X in coordinates as $X = (\xi_1, \ldots, \xi_M)$. Then using $PP^T = 1$, we get

$$\begin{aligned}\mathrm{dist}(PX - v, H) &= \|P_{H^\perp}(PX - v)\|_2 = \left\|\sum_{r=1}^{r}\langle PX, n_r\rangle n_r - P_{H^\perp}v\right\|_2 = \left\|\sum_{r=1}^{r}\langle X, P^T n_r\rangle n_r - P_{H^\perp}v\right\|_2 \\ &= \left\|\sum_{r=1}^{r}\langle X, P^T n_r\rangle P^T n_r - P^T P_{H^\perp}v\right\|_2 = \|P_E X - w\|_2 = \left\|\sum_{k=1}^{M} a_k\xi_k - w\right\|_2,\end{aligned}$$

where $E \equiv E(H) := P^T H^\perp$, $a_k = P_E e_k$, $k = 1, \ldots, M$, and $w = P^T P_{H^\perp}v \in \mathbb{R}^M$. Notice that

$$\sum_{k=1}^{M}\langle x, a_k\rangle^2 = \|x\|_2^2, \quad \text{for any } x \in E.$$

Hence the subsequence of vectors $a = (a_1, \ldots, a_M)$ satisfies the assumption (6.2). So we can use Theorem 6.1 in the space E, which is identified with $\mathbb{R}^l$ by a suitable isometry. For every $\theta = (\theta_1, \ldots, \theta_M) \in E$ and every k, we have $\langle \theta, a_k \rangle = \langle \theta, e_k \rangle = \theta_k$. So $\theta \cdot a = \theta$, where the right hand side is considered as a vector in $\mathbb{R}^M$. Therefore, we have

$$\mathrm{LCD}_{\alpha,\gamma}(E) = \mathrm{LCD}_{\alpha,\gamma}(a).$$

By Lemma 2.5, we have $\mathcal{L}(\xi_k, 1/2) \leqslant 1 - b$ for some $b > 0$ that depends only on the fourth moment of ξ_k. Hence we can apply Theorem 6.1 to $S = \sum_{k=1}^M a_k \xi_k$ and conclude that for all $\epsilon > 0$,

$$\mathbb{P}\left(\mathrm{dist}(PX - v, H) < \epsilon\sqrt{l}\right) \leqslant \mathcal{L}(S, \epsilon\sqrt{l}) \leqslant \left(\frac{C\epsilon}{\gamma}\right)^l + \left(\frac{C\sqrt{l}}{\gamma \mathrm{LCD}_{\alpha,\gamma}(E)}\right)^l + C^l e^{-c\alpha^2}. \tag{6.3}$$

Applying the above inequality to the random subspaces H_{J^c} and $E = E(H_{J^c})$, we conclude Lemma 5.4 if we can prove the following theorem. Heuristically, it shows that the randomness will remove any arithmetic structure from the subspace E and make the LCD exponentially large.

Theorem 6.2. *Suppose $\xi_1, \ldots, \xi_{N-l}$ are independent centered random variables with unit variance and uniformly bounded fourth moment. Let $\tilde{Y}$ be an $M \times (N-l)$ random matrix whose rows are independent copies of the random vector $(\xi_1, \ldots, \xi_{N-l})$, and $\tilde{A}$ be a fixed $N \times (N-l)$ matrix. Assume that $\|\tilde{Y}\| + \|\tilde{A}\| \leqslant C_1\sqrt{N}$ for some constant $C_1 > 0$. Let H be the random subspace of $\mathbb{R}^N$ spanned by the column vectors of $P\tilde{Y} - \tilde{A}$, and $E \equiv E(H)$ be the random subspace $P^T H^\perp$ of $\mathbb{R}^M$. Then for $\alpha = c\sqrt{N}$, we have*

$$\mathbb{P}\left(\mathrm{LCD}_{\alpha,c}(E) < c\sqrt{N} e^{cN/l}\right) \leqslant e^{-cN},$$

where c depends only on the maximal fourth moment.

The rest of this section is devoted to proving this theorem. Note that if $a \in E(H)$, then $a = P^T x$ for some $x \in H^\perp$. On the other hand, using $x = Pa$ we obtain that

$$x \in H^\perp \Leftrightarrow \tilde{Y}^T P^T x - \tilde{A}^T x = 0 \Leftrightarrow \tilde{Y}^T a - \tilde{A}^T P a = 0.$$

We define the $(N-l) \times M$ matrix $\tilde{B} := \tilde{A}^T P$. Then for every set S in E, we have

$$\inf_{x \in S} \left\| \tilde{Y}^T x - \tilde{B} x \right\|_2 > 0 \text{ implies } S \cap E = \varnothing. \tag{6.4}$$

This helps us to navigate the random subspace E away from undesired sets S on the unit sphere.

As in Definition 3.1, we can define the compressible and incompressible vectors on S^{M-1}, which are denoted by $Comp_M(\delta, \rho)$ and $Incomp_M(\delta, \rho)$, respectively. We can prove the following result for compressible vectors as in Lemma 3.2.

Lemma 6.3 (Random subspaces are incompressible)**.** *There exist N-independent constants $\rho, c_0, c_1 > 0$ such that for $0 < \delta \leqslant c_1$, we have*

$$\mathbb{P}\left(\inf_{x \in Comp_M(\delta,\rho)} \left\| \left(\tilde{Y}^T - \tilde{B}\right) x \right\|_2 \leqslant c_0\sqrt{N}\right) \leqslant e^{-c_0 N}. \tag{6.5}$$

In particular, this implies that

$$\mathbb{P}\left(E \cap S^{M-1} \subseteq Incomp_M(\delta, \rho)\right) \geqslant 1 - e^{-c_0 N}. \tag{6.6}$$

Proof. The proof is similar to the one for Lemma 3.2. The only difference is that instead of using Lemma 2.7, we shall use that $\tilde{Y}^T$ has independent row vector $\tilde{Y}_1, \ldots, \tilde{Y}_{N-l}$. For any $x \in S^{M-1}$, it is easy to verify that $\langle \tilde{Y}_k, x \rangle$ have variance 1 and uniformly bounded fourth moment. Then by Lemma 2.5, we have for any fixed $v = (v_1, \ldots, v_{N-l}) \in \mathbb{R}^{N-l}$,

$$\mathbb{P}\left(|\langle \tilde{Y}_k, x\rangle - v_k| \leqslant 1/2\right) \leqslant p$$

for some $0 < p < 1$. Combining with the tensorization Lemma 2.1, we can find constants $\eta, \nu \in (0,1)$ depending on p such that

$$\mathbb{P}\left\{\|\tilde{Y}^T x - v\|_2 \leqslant \eta\sqrt{N-l}\right\} \leqslant \nu^{N-l}. \tag{6.7}$$

Note that $l \leqslant 3N/4$ and $M \leqslant CN$ by our assumptions. Then using estimate (6.7) instead of (3.3), we can complete the rest of the proof as in Lemma 3.2. □

Fix δ and ρ given by Lemma 6.3 for the rest of this section. Note that we can take δ to be a constant of order 1 (in contrast to $\delta \leqslant c_1 N/\left(nK^4 \log K\right)$ in Lemma 3.2). We will further decompose $Incomp_M(\delta, \rho)$ into level sets S_D according to the value D of the least common denominator. We shall prove a nontrivial lower bound on $\inf_{x \in S_D} \|(\tilde{Y}^T - \tilde{B})x\|_2$ for each level set up to D of the exponential order. By (6.4), this means that E is disjoint from every such level set. Therefore, E must have exponentially large least common denominators. First, as a consequences of Lemma 3.4, we have the following lemma, which gives a weak lower bound for the LCD. It is Lemma 3.6 of [17].

Lemma 6.4. *For every $\delta, \rho \in (0,1)$, there exists $c_2(\delta, \rho) > 0$ and $c_3(\delta) > 0$ such that the following holds. Let $a \in Incomp_M(\delta, \rho)$. Then for every $0 < c < c_2(\delta, \rho)$ and every $\alpha > 0$, one has*

$$\mathrm{LCD}_{\alpha,c}(a) > c_3(\delta)\sqrt{M}.$$

We then decompose $Incomp_M(\delta, \rho)$ according to the values of LCD.

Definition 6.5 (Level sets)**.** *Let $D \geqslant c_3(\delta)\sqrt{M}$. Define $S_D \subseteq S^{M-1}$ as*

$$S_D := \{x \in Incomp_M(\delta, \rho) : D \leqslant \mathrm{LCD}_{\alpha,c}(x) < 2D\} \cap \left(P^T \mathbb{R}^N\right).$$

To obtain a lower bound for $\|(\tilde{Y}^T - \tilde{B})x\|_2$ on S_D, we first prove a bound for a single vector x and then use an ϵ-net argument.

Lemma 6.6. *Let $x \in S_D$. Then for every $t > 0$ we have*

$$\mathbb{P}\left(\|(\tilde{Y}^T - \tilde{B})x\|_2 < t\sqrt{N}\right) \leqslant \left(Ct + \frac{C}{D} + Ce^{-c\alpha^2}\right)^{N-l}.$$

Proof. Denoting the elements of $\tilde{Y}^T$ by ξ_{jk}, we can write the j-th coordinate of $\tilde{Y}^T x$ as

$$\left(\tilde{Y}^T x\right)_j = \sum_{k=1}^{M} \xi_{jk} x_k =: \zeta_j, \quad j = 1, \ldots, N-l.$$

We denote the coordinates of $\tilde{B}x$ by v_j, $1 \leqslant j \leqslant N-l$. Applying Theorem 6.1 in dimension 1, we obtain that for every j and every $t > 0$,

$$\mathbb{P}\left(|\zeta_j - v_j| < t\right) \leqslant Ct + \frac{C}{\mathrm{LCD}_{\alpha,c}(x)} + Ce^{-c\alpha^2} \leqslant Ct + \frac{C}{D} + Ce^{-c\alpha^2}.$$

Since ζ_j are independent random variables, we can use tensorization Lemma 2.1 to conclude that for every $t > 0$,

$$\mathbb{P}\left(\sum_{j=1}^{N-l} |\xi_j - v_j|^2 \leqslant t^2(N-l)\right) \leqslant \left(C't + \frac{C'}{D} + C'e^{-c\alpha^2}\right)^{N-l}.$$

This concludes the proof since $\|(\tilde{Y}^T - \tilde{B})x\|_2^2 = \sum_{j=1}^{N-l} |\xi_j - v_j|^2$ and $l \leqslant 3N/4$. □

Lemma 6.7. *There exists a $(4\alpha/D)$-net of S_D of cardinality at most $(CD/\sqrt{N})^N$.*

The following lemma is a classical result in geometric functional analysis, proved in [4].

Lemma 6.8. *If $S \subseteq \mathbb{R}^M$ is a subspace with dimension k, then*

$$|S \cap Q_M| \leqslant (\sqrt{2})^{M-k},$$

where $Q_M = [-1/2, 1/2]^M$ is the unit cube centered at the origin.

Proof of Lemma 6.7. We can assume that $4\alpha/D \leqslant 1$, otherwise the conclusion is trivial. For $x \in S_D$, denote $D(x) := \mathrm{LCD}_{\alpha,c}(x)$. By the definition of S_D, we have $D \leqslant D(x) \leqslant 2D$. By the definition of the least common denominator, there exists $p \in \mathbb{Z}^M$ such that

$$\|D(x)x - p\|_2 < \alpha. \tag{6.8}$$

Therefore,

$$\left\| x - \frac{p}{D(x)} \right\| < \frac{\alpha}{D(x)} \leqslant \frac{1}{4}.$$

Since $\|x\|_2 = 1$, it follows that

$$\left\| x - \frac{p}{\|p\|_2} \right\|_2 \leqslant \frac{2\alpha}{D}. \tag{6.9}$$

On the other hand, by (6.8) and using that $\|x\|_2 = 1$, $D(x) \leqslant 2D$ and $4\alpha/D \leqslant 1$, we obtain

$$\|p\|_2 < D(x) + \alpha \leqslant 2D + \alpha \leqslant 3D \tag{6.10}$$

Since x is in the subspace $F := P^T\mathbb{R}^N$, we also have $D(x)x \in F$. We can chose p such that it is the closest integer point to $D(x)x$. Hence p must lie in the "cube covering" of F, defined as

$$\tilde{F} := \bigcup_{b \in F} \left(\prod_{i=1}^{M} [b_i - 1/2, b_i + 1/2] \right). \tag{6.11}$$

Then (6.9) and (6.10) show that

$$\mathcal{N} := \left\{ \frac{p}{\|p\|_2} : p \in \mathbb{Z}^N \cap B(0, 3D) \cap \tilde{F} \right\}$$

is a $(2\alpha/D)$-net of S_D. The cardinality of $\mathcal{N}$ can be bounded by the volume of $B(0, 3D) \cap \tilde{F}$. Note that $B(0, 3D) \cap \tilde{F}$ is the "cube covering" of the N-dimension ball $B(0, 3D) \cap F$. Then by Fubini's theorem, we get

$$\left| B(0, 3D) \cap \tilde{F} \right| \leqslant |B(0, 3D) \cap F| \cdot \left| F^{\perp} \cap Q_M \right|.$$

Then by the volume formula for an N-dimension ball and Lemma 6.8, we obtain that

$$|\mathcal{N}| \leqslant (CD/\sqrt{N})^N.$$

Finally, we can find a $4\alpha/D$-net of the same cardinality, which lies in S_D (see Lemma 5.7 of [16]). □

Lemma 6.9. *There exist $c_4, c_5, \mu \in (0,1)$ such that the following holds. Let $\alpha = \mu\sqrt{N} \geqslant 1$ and $D \leqslant c_4\sqrt{N}e^{c_4 N/l}$. Then*

$$\mathbb{P}\left(\inf_{x\in S_D}\left\|\left(\tilde{Y}^T - \tilde{B}\right)x\right\|_2 < c_5 N/D\right) \leqslant e^{-N}.$$

Proof. Recall that by the assumption of Theorem 6.2, we have

$$\|\tilde{Y}^T\| + \|\tilde{B}\| \leqslant \|\tilde{Y}\| + \|\tilde{A}\| \leqslant C_1\sqrt{N}$$

for some positive constant C_1. To conclude the proof, it is enough to find $\nu > 0$ such that the event

$$\mathcal{E} := \left\{\inf_{x\in S_D}\left\|\left(\tilde{Y}^T - \tilde{B}\right)x\right\|_2 < \frac{\nu N}{2D}\right\}$$

has probability $\leqslant e^{-N}$.

Let $\nu > 0$ be a small constant to be chosen later, which depends only on the fourth moment. We apply Lemma 6.6 with $t = \nu\sqrt{N}/D$. By the assumptions on α and D, the term Ct dominates in the right hand side: $t \geqslant D^{-1}$ and $t \geqslant e^{-c\mu^2 N}$. This gives for arbitrary $x \in S_D$,

$$\mathbb{P}\left(\left\|\left(\tilde{Y}^T - \tilde{B}\right)x\right\|_2 < \frac{\nu N}{D}\right) \leqslant \left(\frac{C\nu\sqrt{N}}{D}\right)^N.$$

We take the $(4\alpha/D)$-net $\mathcal{N}$ of S_D given in Lemma 6.7, and take the union bound to get

$$p := \mathbb{P}\left(\inf_{x\in\mathcal{N}}\left\|\left(Y^T - B\right)x\right\|_2 < \frac{\nu N}{D}\right) \leqslant \left(\frac{CD}{\sqrt{N}}\right)^N\left(\frac{C\nu\sqrt{N}}{D}\right)^{N-l} \leqslant \left(\frac{CD}{\sqrt{N}}\right)^l\left(C'\nu\right)^{N-l}.$$

Under the assumption on D, we can choose ν small enough such that

$$p \leqslant \left(C''\right)^l e^{c_4 N}\left(C'\nu\right)^{N-l} \leqslant e^{-N},$$

where we used $l \leqslant 3N/4$ again in the last step.

Now assume $\mathcal{E}$ holds. Fix $x \in S_D$ such that $\|(\tilde{Y}^T - \tilde{B})x\| < \nu N/(2D)$. Then we can find $y \in \mathcal{N}$ such that

$$\|x - y\| \leqslant \frac{4\alpha}{D} = \frac{4\mu\sqrt{N}}{D}.$$

Then by the triangle inequality we have

$$\begin{aligned}\left\|\left(\tilde{Y}^T - \tilde{B}\right)y\right\|_2 &\leqslant \left\|\left(\tilde{Y}^T - \tilde{B}\right)x\right\|_2 + \left(\left\|\tilde{Y}^T\right\| + \left\|\tilde{B}\right\|\right)\cdot\|x - y\|_2\\ &\leqslant \frac{\nu N}{2D} + C_1\sqrt{N}\frac{4\mu\sqrt{N}}{D} < \frac{\nu N}{D},\end{aligned}$$

where we have chosen μ to satisfy $\mu < \nu/(8C_1)$. Thus the event $\mathcal{E}$ implies the event that

$$\inf_{y\in\mathcal{N}}\left\|\left(\tilde{Y}^T - \tilde{B}\right)y\right\|_2 \leqslant \frac{\nu N}{D},$$

whose probability $\leqslant e^{-N}$. This concludes the proof. □

Proof of Theorem 6.2. Let $x \in S^{M-1} \cap E$ such that

$$\mathrm{LCD}_{\alpha,c}(x) < c_4\sqrt{N}e^{c_4 N/l}$$

where c_4 is the constant from Lemma 6.9. Then, by Definition 6.5, either x is compressible or $x \in S_D$ for some $D \in \mathcal{D}$ where

$$\mathcal{D} := \left\{ D : c_3\sqrt{N} \leqslant D < c_4\sqrt{N}e^{c_4 N/l}, D = 2^k, k \in \mathbb{N} \right\}.$$

Therefore, we can decompose the desired probability as follows:

$$p := \mathbb{P}(\mathrm{LCD}_{\alpha,c}(E) < c_4\sqrt{N}e^{c_4 N/l}) \leqslant \mathbb{P}\left(E \cap Comp_M(\delta,\rho) \neq \varnothing\right) + \sum_{D\in\mathcal{D}} \mathbb{P}(E \cap S_D \neq \varnothing).$$

The first term can be bounded by $e^{-c_0 N}$ by Lemma 6.3. For the other terms, we have

$$x \in E \cap S_D \Leftrightarrow \left(\tilde{Y}^T - \tilde{B}\right) x = 0.$$

Thus by Lemma 6.9, we get

$$\mathbb{P}(E \cap S_D \neq \varnothing) \leqslant \mathbb{P}\left(\inf_{x \in S_D} \left\| \left(\tilde{Y}^T - \tilde{B}\right) x \right\|_2 = 0 \right) \leqslant e^{-N}.$$

Since there are $|\mathcal{D}| \leqslant CN$ terms in the sum, we conclude that

$$p \leqslant e^{-c_0 N} + CNe^{-N} \leqslant e^{-c' N}.$$

This concludes the proof. □

References

[1] Z. D. Bai. Circular law. *Ann. Probab.*, 25(1):494–529, 1997.

[2] Z. D. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*, volume 2 of *Mathematics Monograph Series*. Science Press, Beijing, 2006.

[3] Z. D. Bai and Y. Q. Yin. Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *Ann. Probab.*, 21(3):1275–1294, 1993.

[4] K. Ball. *Volumes of sections of cubes and related problems*, pages 251–260. Springer Berlin Heidelberg, Berlin, Heidelberg, 1989.

[5] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Isotropic local laws for sample covariance and generalized Wigner matrices. *Electron. J. Probab.*, 19(33):1–53, 2014.

[6] K. R. Davidson and S. J. Szarek. Local operator theory, random matrices and banach spaces. volume 1 of *Handbook of the Geometry of Banach Spaces*, pages 317 – 366. North-Holland, Amsterdam, 2001.

[7] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.

[8] J. Ginibre. Statistical ensembles of complex, quaternion, and real matrices. *J. Math. Phys.*, 6(3):440–449, 1965.

[9] F. Gőtze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38(4):1444–1491, 2010.

[10] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *arXiv:1410.3516.*

[11] A. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.*, 195(2):491 – 523, 2005.

[12] A. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann. Smallest singular value of random matrices and geometry of random polytopes. *Advances in Mathematics*, 195(2):491 – 523, 2005.

[13] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1:457, 1967.

[14] G. Pan and W. Zhou. Circular law, extreme singular values and potential theory. *J. Multivar. Anal.*, 101(3):645–656, 2010.

[15] M. Rudelson. Invertibility of random matrices: norm of the inverse. *Ann. Math.*, 168(2):575–600, 2008.

[16] M. Rudelson and R. Vershynin. The littlewood-offord problem and invertibility of random matrices. *Adv. Math.*, 218:600–633, 2008.

[17] M. Rudelson and R. Vershynin. The smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.*, 62:1707–1739, 2009.

[18] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electron. Commun. Probab.*, 18:9 pp., 2013.

[19] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Notices*, 2015(19):9594–9617, 2015.

[20] J. W. Silverstein. The smallest eigenvalue of a large dimensional Wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 1985.

[21] T. Tao and V. Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008.

[22] T. Tao, V. Vu, and M. Krishnapur. Random matrices: Universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010.

[23] H. Xi, F. Yang, and J. Yin. Local circular law for the product of a deterministic matrix with a random matrix. *arXiv:1603.04066.*

[24] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. On the limit of the largest eigenvalue of the large dimensional sample covariance matrix. *Probability Theory and Related Fields*, 78(4):509–521, 1988.