Measure Concentration and Non-asymptotic Singular Values Distributions of Random Matrices

by

Feng Wei

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mathematics)
in The University of Michigan
2018

Doctoral Committee:

Professor Mark Rudelson, Chair
Professor Jinho Baik
Professor Anna Gilbert
Associate Professor Raj Rao Nadakuditi

Feng Wei

weifeng@umich.edu

ORCID iD: 0000-0002-5692-3571

To my mom and dad

ACKNOWLEDGEMENTS

First and foremost, I am greatly thankful to my advisor, Professor Mark Rudelson. A series of great courses and reading groups Mark organized helped not just me but a group of students pursuing mathematics. Mark guided me with his wisdom and experience in the process of choosing research topics, forming proofs and presenting results. The financial support from Mark provided me with undisturbed time devoted to research for several semesters. I found myself greatly indebted to him when I decide to go to industry feeling that all his effort and enthusiasm on me might be a waste. But Mark has been very supportive when I made this decision and helped me finding opportunities. There are so many other things I am grateful about that could be listed here. It is a fortune to be student of Mark, without whom this thesis would never be possible.

I want to give my second thanks to one of my friend, Yitong Sun who is also my roommate and officemate. We were good friends since we were our first year in Ann Arbor. I learned a lot from him both academically and personally. I am grateful to Yitong for all the good and bad times we shared together in the past more than six years.

I would also like to thank Professors Jinho Baik, Anna Gilbert, Raj Rao Nadakuditi, for serving on my dissertation committee. In particular, I want to thank Professor Jinho Baik for reveiwing this thesis and writing a report for it especially when the time budget is so limited.

I am grateful to Professors Roman Vershynin and Jinho Baik for several interesting analysis classes they gave. I also want to express my gratitude to all the other professors and post-docs who have been my instructors during my study at University of Michigan. I want to thank Professors Peter Miller and Joel Smoller for their nice courses and their support on my mathematics pursuit. I want to thank Professors Gabor Orosz, Galip Ulsoy and Jonah Blasiak who provided me great research opportunities during my master's study at UM.

I want to thank our awesome math department for providing me the financial support and a great community. I appreciate all the mathematical discussions with fellow graduate students at University of Michigan, including but not restricted to Qingtang Su, Elizaveta Rebrova, Yun Wei, Hao Yuan, Hao Zheng. Especially, I want to thank Han Huang and Aniket Anand Deshmukh for discussing research problems and sharing their ideas.

I would like to express my thanks to many of my friends who have been encouraging me during difficult times, sharing happiness and giving me advice in the past years, including Qingtang Su, Xiaohui Wang, Zhe Wei, Jiaqi Li, Xin Zhang, Yilun Sun, Yang Jiao, Jingchen Wu, Hao Yuan, Hao Zheng, Yingchuan Wang, Yiwei Zhang, Yebin Tao, Bowei Wu, Yun Wei and many others.

Finally, I want to finish this with saying thank you to my mom and dad. Their unconditional love made me who I am today. Their encouragement since I was a kid supported me throughout the way of my mathematics pursuit. To my parents I dedicate this thesis.

TABLE OF CONTENTS

ABSTRACT

This thesis is devoted to the non-asymptotic random matrix theory and measure concentration phenomenon. We focus on using concentration inequalities together with other probabilistic and geometric methods to study singular values distributions of several types of random matrices.

In Chapter II, we apply concentration inequalities to a convex geometry problem, namely upper bound for the Dvoretzky dimension in Milman-Schechtman theorem. Our approach combines properties of random projections and geometric observation.

In Chapter III, we study the non-asymptotic distributions of all singular values for i.i.d. sub-gaussian matrices. We prove a non-asymptotic upper bound for all singular values of i.i.d. sub-gaussian matrices under some weak condition. It is the first tight non-asymptotic upper bound for all singular values other than Gaussian matrices. The upper bound provides a two-side bound together with known lower bound.

In Chapter IV, we study the smallest singular values distributions of symmetric sparse matrices. We show that an n-dimensional sparse symmetric random matrix A is invertible with high probability under some condition on its sparsity level.

CHAPTER I

Introduction

1.1 Measure concentration in probability theory

Concentration of measure (e.g. about a mean) is a general principle that is applied in measure theory, functional analysis, probability, combinatorics other. The idea was put forward in the early seventies by V. Milman in the asymptotic theory of Banach spaces. It was further developed in the works of V. Milman and M. Gromov, B. Maurey, G. Pisier, G. Schechtman, M. Talagrand, M. Ledoux, and others. For an overview of the history and some standard results, see [31].

Measure concentration usually occurs in high dimensional measure space geometrically, or applies to a large number of random variables when there is sufficient independence among them. In the probabilistic setting, measure concentration principle states that that a good (e.g. Lipschitz) function $f : X \to \mathbb{R}$ defined on a large probability space X almost always takes values very close to the average value of f on X. To see what "close to" or "concentrated" means, let's consider following example.

Let $X_1, X_2, \cdots, X_n$ be independent symmetric Bernoulli random variables, and $a = (a_1, \cdots, a_n) \in \mathbb{R}^n$. Then, for any $t \geq 0$, by Hoeffding's inequality, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\|a\|_2^2}\right).$$

Results of this kind are known as concentration inequalities.

Choosing $a = (\frac{1}{n}, \cdots, \frac{1}{n})$, we have the average value of n independent Bernoulli random variables are within a magnitude of t with probability greater than $1 - 2\exp(-nt^2/2)$. It's worth noticing that comparing to asymptotic results such as law of large numbers, the concentration inequalities tells us more about the distribution when the number of variables is a fixed large number. On the other side, it is a common phenomenon that the dimension, or number of random variables, also appears in the probability bounds in the concentration inequalities. Thus, in many cases, the concentration inequalities may give us better bounds as dimension getting higher, which turns out to be crucial.

In Chapter II, we will discuss several more concentration inequalities and their applications. But in the rest of the thesis, we will mainly be interested in the role of concentration inequalities in non-asymptotic random matrix theory and others. Limited by space, we won't discuss much further about general measure concentration phenomenon. For more concentration inequalities and their applications, one may look at [31, 53, 74]

1.2 Non-asymptotic theory of random matrices

1.2.1 Non-asymptotic vs. Asymptotic

Random matrix theory studies properties of $N \times n$ matrices A chosen from some probability distribution on the set of all matrices. Since the beginning of the area, the classical random matrix theory has been mostly focused on asymptotic spectral properties of random matrices as their dimensions tend to infinity. Among them, there is the foundational Wigner semicircle law for the empirical measures of eigenvalues of random symmetric matrices [77, 41, 78], Marchenko-Pastur law, which is the limit of empirical measures of sample covariance matrices (or limit of singular value distributions) [69, 75, 5], Circular law, which is the limit of empirical measures of i.i.d. matrices [4, 18, 19, 62, 64], and TracyWidom distribution describing the limit of the extreme singular values of a sequence of random matrices [7, 8, 26, 27, 56, 38, 9], etc. For an introduction of classical problems and results of random matrix theory and its fascinating connections, see [14, 3, 11, 33, 66, 13, 23].

These limiting distributions are of paramount importance. The asymptotic regime that the dimensions tend to infinity is well suited for many different purposes, for example in physics when random matrices serve as finite-dimensional models of infinite-dimensional operators. However, for many problems in numerical analysis, convex geometry, statistics, theoretical computer science, data science, knowing the limit behavior is of little help. In those cases one needs information about behavior of characteristics for large fixed dimension instead. And this lead to our interest in non-asymptotic random matrix theory.

One reason we are interested in characteristics of random matrices for a fixed finite dimension comes from classical asymptotic random matrix theory. One example of those is the Stieltjes transform of measures which plays an important role in deriving limit laws. To derive the convergence of Stieltjes transforms, one often need to provide bounds on the smallest singular value of a random matrix of a fixed dimension which holds with high probability [64, 62].

In other mathematical areas, one sometimes needs to understand what happens for a fixed dimension rather than the limit. For instance, in numerical analysis, we we have a system with a small random perturbation, stability of the system of linear equations $Ax = b$ under the perturbation depends on the condition number of the random matrix A. In this case one needs to understand the spectrum of random matrices in finite large dimensions rather than in limit. There are many such examples in geometric functional analysis as well, for example, constructing a random section of an high dimensional convex body by taking the kernel or the range of a certain random matrix where one needs some probability guarantee with a fixed large dimension.

In many other areas outside mathematics such as statistics and theoretical computer science, the limiting regime may not be very useful, for example, constructing an embedding of a high dimensional subspace into another one by taking the range of some tall random matrix [49], constructing a random matrix as linear transform to reduce the dimension of a point sets while preserve distances [53], estimating error bounds of sample covariance matrices [72], constructing matrices with restricted isometry properties [49] and etc. In such applications, the dimension of the large space remains fixed, and one seeks explicit estimates of probabilities in terms of the dimension. It's worth mentioning that, non-asymptotic random matrices theory also play an important role in theoretical machine learning and data science which are rapidly growing sub-fields of statistics and computer science [74].

The difference between non-asymptotic results and asymptotic results in random matrix theory is similar to our earlier comparison between law of large numbers and Hoeffding's inequality. We now wrap up our discussion between asymptotic results and non-asymptotic ones by giving an explicit example.

Consider an $N \times n$ random matrix with i.i.d. Gaussian entries. Then in the asymptotic regime, the Bai-Yins law [3] states that as the dimensions $N \geq n$ increase to infinity while the aspect ratio n/N converges to a constant in $[0, 1]$ is fixed, we have

$$s_{\min}(A) = \sqrt{N} - \sqrt{n} + o(\sqrt{n}), \;\; s_{\max}(A) = \sqrt{N} + \sqrt{n} + o(\sqrt{n}) \;\; \text{almost surely.}$$

However, Bai-Yins law won't directly tell us anything if our matrix A has a fixed large dimension. As we pointed out earlier, we want to find non-asymptotic versions of Bai-Yins law. In the Gaussian matrix case, an exact non-asymptotic result is known as following: Let A be an $N \times n$ matrix whose entries are independent standard normal random variables. Then for every $t \geq 0$, with probability at least $1 - 2\exp(t^2/2)$ one has

$$\sqrt{N} - \sqrt{n} - t \leq s_{\min}(A) \leq \ s_{\max}(A) \leq \sqrt{N} + \sqrt{n} + t.$$

Comparing with Bai-Yins law, the non-asymptotic results provide us concentration inequalities with probability bound rather than a limit behavior. It's common that the magnitude of deviation also play a role in the probability bounds and this give us the freedom to balance between how tight the inequalities are and how strong the probability guarantee is. Choosing t in the order of $\sqrt{n}$, we also have

$$\sqrt{N} - C\sqrt{n} \leq s_{\min}(A) \leq \ s_{\max}(A) \leq \sqrt{N} + C\sqrt{n}$$

with probability at least $1 - \exp(-cn)$. Although the non-asymptotic inequalities maybe not as precise as the asymptotic limiting distributions, the probability is usually overwhelmingly large which is essential in many applications.

1.2.2 A short overview: extreme singular values and others

Singular values and eigenvalues are both important characteristics of random matrices and their magnitude agrees on symmetric matrices. Non-asymptotic random matrix theory studies spectral properties of random matrices, that is to provides probabilistic bounds for singular values, eigenvalues, etc., for random matrices of a large fixed size. In the non-asymptotic viewpoint, study of singular values are more motivated due to geometric problems in high dimensional Euclidean spaces. Here we give a short overview of non-asymptotic singular values distributions study, but due to our motivation, we only focus on the extreme singular values of real matrices with sufficient independent entries. Many of following results could be generalized to more general setting.

Recall that for an $N \times n$ real matrix A with $N \geq n$. The singular values $s_k(A)$ of A, where $k = 1, 2, \cdots, n$, are the eigenvalues of $\sqrt{A^T A}$ arranged in non-increasing order. Among all the singulars, the two extreme ones are of the most importance. When we view matrix A as a linear operator $\mathbb{R}^n \to \mathbb{R}^N$, we may want control its behavior by finding or giving useful upper and lower bounds on A. Such bounds are provided by the smallest and largest singular values of A denoted as $s_{\min}(A)$ and $s_{\max}(A)$. The extreme singular values are also referred as the operator norms of the linear operators A and A^{-1} between Euclidean spaces, that is to say $s_{\min}(A) = 1/\|A^{-1}\|$ and $s_{\max}(A) = \|A\|$.

Due to the geometric interpretation, understanding the behavior of extreme singular values of random matrices are important in many applications. For instance, in computer science and numerical linear algebra, the condition number $s_{\max}(A)/s_{\min}(A)$ is widely used to measure stability or efficiency of algorithms as the example we give in early section. In geometric functional analysis, probabilistic construction of linear operators using random matrices often depend on good bounds on the norms of these operators and their inverses [53]. In statistics, applications of extreme singular values can be found from the analysis of sample covariance matrices $A^T A$ [72].

It is widely believed that phenomena typically observed in classical random matrix theory are universal, that is independent of the particular distribution of the entries of random matrices [64]. For instance, the Circular law not only hold for i.i.d. Gaussian matrices, but also for general i.i.d. matrices with mean zero variance 1 entries [62, 64]. This principle is also adapted in the study of non-asymptotic random matrix theory, that is to find non-asymptotic characteristics of special random matrices (e.g.

Gaussian matrix or Bernoulli matrix), then prove the same properties for general matrices.

It is worth mentioning that many non-asymptotic results are known under a somewhat stronger sub-gaussian moment assumption on the entries of A, which requires their distribution to decay as fast as the normal random variable:

Definition 1.2.1. (sub-gaussian random variables). A random variable X is sub-gaussian if there exists $K > 0$ called the sub-gaussian moment of X such that

$$P(|X| > t) \leq 2\exp(-t^2/K^2) \text{ for } t > 0.$$

The are several equivalent characteritics of sub-gaussian variables

Lemma 1.2.2. (Equivalence of sub-gaussian properties [72]). *Let X be a random variable. Then the following properties are equivalent with parameters $K_i > 0$ differing from each other by at most an absolute constant factor.*

1. *Tails: $P(|X| > t) \leq \exp(1 - t^2/K_1^2)$ for all $t \geq 0$;*
2. *Moments: $(E|X|^p)^{1/p} \leq K_2\sqrt{p}$ for all $p \geq 1$;*
3. *Super-exponential moment: $\mathbb{E}\exp(X^2/K_3^2) \leq e$.*

Moreover, if $\mathbb{E}X = 0$ then properties 1-3 are also equivalent to the following one:

4. *Moment generating function: $\mathbb{E}\exp(tX) \leq \exp(t^2K_4^2)$ for all $t \in \mathbb{R}$.*

Many classical random variables are actually sub-gaussian, such as Gaussian random variables, Bernoulli random variables, Bounded random variables, etc..

As we mentioned, the concentration inequalities are powerful tools in studying high-dimensional probability. One advantage of sub-gaussian random variables is that many useful concentration inequalities are proved for sub-gaussian random variables. For example, the Hoeffding's inequality we discussed can be generalized to sub-gaussian random variables.

Theorem 1.2.3. (Hoeffding's inequality) *Let $X_1, \cdots, X_n$ be independent centered sub-gaussian random variables. Then for any $a_1, \cdots, a_n \in \mathbb{R}$ and $a = (a_1, \cdots, a_n)$*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} a_i X_i\right| \geq t\right) \leq 2\exp\left(-c\frac{t^2}{2\|a\|_2^2}\right).$$

We will introduce more concentration inequalities for sub-guassian random variables in Chapter III, further more properties of sub-guassian random variables can be found in [74, 72].

Based on the sub-gaussian properties one can prove the following non-asymptotic version of Bai-Yin's law for largest singular value on sub-gaussian matrices [49]:

Theorem 1.2.4. (Largest singular value of subgaussian matrices). *Let A be an $N \times n$ random matrix whose entries are independent mean zero sub-gaussian random variables whose sub-gaussian moments are bounded by 1. Then*

$$\mathbb{P}\left(s_{\max}(A) > C(\sqrt{N} + \sqrt{n} + t)\right) \leq 2e^{-ct^2}.$$

for $t \geq 0$. Here C, c are absolute constants.

The proof uses a simple net argument which will be discussed in Chapter III and IV.

By integration, one can easily deduce from above the correct expectation bound $\mathbb{E}s_{\max}(A) \leq C(\sqrt{N} + \sqrt{n})$. This bound later was proved to hold under much weaker moment assumptions by R. Latala [30]:

Theorem 1.2.5. (Largest singular value: fourth moment, non-i.i.d. entries). *Let A be a random matrix whose entries a_{ij} are independent mean zero random variables with finite fourth moment. Then*

$$\mathbb{E}s_{\max}(A) \leq C\left[\max_i(\sum_j \mathbb{E}a_{ij}^2)^{1/2} + \max_j(\sum_j \mathbb{E}a_{ij}^2)^{1/2} + (\sum_{i,j} \mathbb{E}a_{ij}^4)^{1/4}\right].$$

If the variance and the fourth moments of the entries are uniformly bounded, then result of Latala result yields $s_{\max}(A) = O(\sqrt{N} + \sqrt{n})$ which is not optimal but still satisfactory for many applications.

If the matrix is i.i.d. Gaussian matrix, we have the following much sharper result due to Gordon [15, 16, 17]:

Theorem 1.2.6. (Exteme singular values of Gaussian matrices). *Let A be an $N \times n$ matrix whose entries are independent standard normal random variables. Then*

$$\sqrt{N} - \sqrt{n} \leq \mathbb{E}s_{\min}(A) \leq \mathbb{E}s_{\max}(A) \leq \sqrt{N} + \sqrt{n}.$$

This above theorem is a consequence of some sharp comparison inequalities for Gaussian processes due to Slepian and Gordon [15, 16, 17]. Using measure concentration techniques, one can deduce from above results large deviation inequalities for the extreme singular values. More precise probability bound for i.i.d. sub-gaussian matrices can be found in [38]. Further results of largest singular value on non-i.i.d. Gaussian matrices proved by A. Bandeira and R. van Handel can be found in [1, 21].

So far, we mainly focused on the "soft edge" - the largest singular value distribution, and smallest singular value for tall matrices (that is $N/n > 1 + c$). When N and n are close, the smallest singular value, the "hard edge" of the spectrum is generally more difficult to analysis by classical methods of random matrix theory. This difficulty is especially significant for square matrices, $N = n$ or almost square matrices, $N - n \ll n$. For instance, our earlier bound of the type $s_{\min}(A) \sim \sqrt{N} - \sqrt{n}$ all becomes useless for square matrices.

Another reason that the smallest singular problem is important is that it tells whether a random matrix is invertible or not. One well studied important example is provided by $n \times n$ random Bernoulli matrices A, with independent ± 1 entries. Even in this problem, estimating the probability that A is invertible is a quite nontrivial problem. Komlos showed that A is invertible asymptotically almost surely, that is $\mathbb{P}(s_{\min}(A) = 0)$ tends to 0 as $n \to \infty$ [28, 29]. The bound on $\mathbb{P}(s_{\min}(A) = 0)$ was later improved in the works of Kahn, Komlos, Szemeredi, Tao, Vu, Bourgain and Wood [25, 60, 61, 24].

However, previous progress is only concerned with whether the $s_{\min}(A)$ is zero or not for a very specific matrix. Nothing was said about the quantitative invertibility problem which is more about size of $s_{\min}(A)$. The history of the quantitative invertibility problem goes back to von Neumann when he discovered that the accuracy of the matrix algorithms and their running time could depend on the condition number $\sigma(A) = s_{\max}(A)/s_{\min}(A)$. Based on heuristic and experimental evidence, von Neumann and Goldstine conjectured that with high probability $s_{\min}(a) \sim n^{-1/2}$ and $s_{\max}(a) \sim n^{1/2}$ [49]. The upper bound on largest singular value was established earlier but the lower bound of smallest singular value remained open for decades. Progress has been made by Smale, Edelman and Szarek in the Gaussian matrices case [55, 12, 58]. However, their approaches do not work for matrices other than Gaussian as they depend on explicit formula for the joint density of the singular values.

The first polynomial bound of quantitative invertibility was obtained in [36] by M. Rudelson, where it was proved that the smallest singular value of a square i.i.d. subgaussian matrix is bounded below by $n^{-3/2}$ with high probability. Later an almost sharp bound was proved by M. Rudelson and R. Vershynin in [47] up to a constant factor for general random matrices.

Theorem 1.2.7. (Smallest singular value of square random matrices). *Let A be an $n \times n$ random matrix whose entries are independent and identically distributed* *sub-gaussian random variables with zero mean and unit variance. Then*

$$\mathbb{P}\left(s_{\min}(A) \le \varepsilon n^{-1/2}\right) \le C\varepsilon + c^n,\ \varepsilon \ge 0$$

where $C > 0, c \in (0,1)$ depend only on the sub-gaussian moment of the entries.

It is worth noticing that this theorem is both qualitative and quantitative in terms of the invertibility problem. Picking $\varepsilon = 0$, this theorem implies A is invertible with probability at least $1-c^n$ which generalizes the result of Kahn, Komlos, and Szemeredi from Bernoulli to all sub-gaussian matrices. On the other hand, quantitatively, it states that $s_{\min}(A) \ge n^{1/2}$ with high probability for general random matrices. Together with a corresponding non-asymptotic upper bound was later achieved in [46], we have $s_{\min}(A) \sim n^{1/2}$ as in von Neumann-Goldstines conjecture. Weaker form of both upper and lower bounds are also proved to hold with high probability under the weaker fourth moment assumption [47, 46].

The theory and result was later extended to rectangular random matrices of arbitrary dimensions in [48].

Theorem 1.2.8. (Smallest singular value of rectangular random matrices). *Let G be an $N \times n$ random matrix, $N \ge n$, whose elements are independent copies of a centered sub-gaussian random variable with unit variance. Then for every $\varepsilon > 0$, we have*

$$\mathbb{P}\left(s_n(G) \le \varepsilon\left(\sqrt{N} - \sqrt{n-1}\right)\right) \le (C\varepsilon)^{N-n+1} + e^{-C'N}$$

where $C, C' > 0$ depend (polynomially) only on the sub-gaussian moment K.

Note that above theorem bridges all quantitative bounds of smallest singular values together. For tall matrices, it agrees with the known bounds $s_{\min}(A) \sim \sqrt{N} - \sqrt{n}$. For square matrices, $\sqrt{N} - \sqrt{n-1} \sim n^{-1/2}$. Finally, for matrices which are close to square, it gives the new optimal estimate $s_{\min}(A) \ge c(\sqrt{N} - \sqrt{n})$.

Theorem 1.2.7 and Theorem 1.2.8 comes as a consequence of an 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 $\sum_k a_k X_k$ lies near some number v [47, 48]. To solve the Littlewood-Offord problem, M. Rudelson and R. Vershynin developed theory used the quantitative arithmetic structure of the vector $(a_1, \cdots, a_n)$ to estimate the small ball probability. We will further discuss Littlewood-Offord problem, small ball probability estimate and arithmetic structure in next section and Chapter IV. One can find discussion and recent development of these topics in [47, 67, 49, 73, 45]. The theory developed in Theorem 1.2.7 and Theorem 1.2.8 turns out to be universal and motivated further research in the area.

Effort was made following the invertibility problem progress. Tao and Vu also proved a version of universality for smallest singular values [68] using a different approach. Revisiting of the Littlewood-Offord problem motivated by quantitative invertibility problem are done in [67, 63, 65] by Tao and Vu, [47, 48] by Rudelson and Vershynin, [39] by Friedland and Sodin, etc.. For symmetric matrices, quantitative version of invertibility problem was proved by Vershynin in [73]. For heavy-tailed matrices, the invertibility problem was visited by E. Rebrova and K. Tikhomirov [43]. For discrete matrices, the invertibility problem was discussed by Tao and Vu in [67]. In sparse matrices case, the first quantitative version of invertibility problem was proved by A. Basak and M. Rudelson in [6].

Above overview is made to give some necessary background for our work but it is by no means a complete survey of the area of non-asymptotic random matrix theory. Here, we completely omit several important directions. For example, work on random matrices with deterministic shift (see [62]), work on understanding gaps between eigenvalues or singular values (see [20]), work on understanding singular values close to extrem ones (see [68]), work on random matrices with less independence or more structure (see [72, 42, 71]), and etc..

1.3 Outline, Motivation and Notations

Now let us give an outline of the thesis and motivation of our work. Notations of the thesis will be explained at the end of this section.

In Chapter II, we apply concentration inequalities to study upper bound for the Dvoretzky dimension in Milman-Schechtman theorem. Informally, Milman-Schechtman theorem states that let K be a symmetric convex body in $\mathbb{R}^n$, define $k(K)$ to be the largest dimension k such that

$$\nu_{n,k}\left(F \in G_{n,k} : \forall x \in S^{n-1} \cap F\,,\ \|x\|_K \sim \text{constant}\ \right) > \frac{1}{2}.$$

Then $k(K) \sim n/b^2$ where b is the minimum width of the convex body and $\nu_{n,k}$ is the Haar measure on the Grassmannian manifold. Here $k(K) \sim n/b^2$ includes both upper and lower bound. The lower bound is guaranteed by Milman-Dvoretzky theorem but the proof of upper bound given by Milman and Schechtman would fail for a class of convex bodies. We give a proof for the upper bound in Chapter II for all symmetric convex bodies. Our proof is probabilistic and constructive based on properties of random matrices although the problem itself is geometric. This is a common phenomenon in geometric functional analysis and other areas where concentration inequalities and non-asymptotic random matrix theory are applicable. It also tells the importance of geometric interpretation for non-asymptotic random matrices theory together with other results we will discuss in Chapters III and IV. Chapter II is based on joint work with H. Huang.

In Chapters III, we move on to study the non-asymptotic distributions of singular values for i.i.d. sub-gaussian matrices. Our work is motivated by two facts.

One fact is that very few about optimal upper bound on all singular values are known other than Gaussian matrices. Prior to our result, the only progress in this direction was made by Szarek [59] who established an optimal upper bound for the Gaussian matrix. Szarek proved that for a standard Gaussian i.i.d. matrix G, $\frac{Cl}{\sqrt{n}} \leq s_{n+1-l}(G) \leq \frac{Cl}{\sqrt{n}}$ with probability at least $1-\exp(-Cl^2)$. This result suggests that lth smallest singular value of an i.i.d. sub-gaussian matrix is concentrated around $\frac{l}{\sqrt{n}}$. On the other hand, the distributions of intermediate singular values are found applicable in eigenvector delocalization problems [50].

Our main theorem in Chapter III rougly states that: for an $n \times n$ random sub-gaussian matrix A that satisfies some weak assumption. For all l between 1 and n,

$$\mathbb{P}\left(s_{n+1-l}(A) \leq C_1 \frac{tl}{\sqrt{n}}\right) \geq 1-\exp(-C_2 tl).$$

Together with a known lower bound we show that $s_{n+1-l}(A) \sim \frac{l}{\sqrt{n}}$ for all rectangular i.i.d. sub-gaussian matrices with high probability.

Note that the weak assumption which we will characterize in Chapter III depends on the following definition of Levy's concentration function.

Definition 1.3.1. Let Z be a random vector that takes values in $\mathbb{C}^n$. The concentration function of Z is defined as

$$\mathcal{L}(Z,t) = \sup_{u \in \mathbb{C}^n} \mathbb{P}\{\|Z-u\|_2 \leq t\}, t \geq 0.$$

The Levy's concentration function is very useful in characterizing small ball probability. Many classical concentration inequalities, for exmaple the Hoeffding's inequality are devoted to characterize the large deviations. These large deviation inequalities tell us probability that a random sum $\sum a_i X_i$ (or in general a random variable given by $f(X_1, \cdots, X_n)$) far away from its expectation is small. On ther other hand, the small ball probability which estimate $\mathbb{P}(\|f(X_1, \cdots X_n) - u\|_2 \leq t)$-the probability that the random vector $f(X_1, \cdots X_n)$ enters a small ball in the space, also has significant important in many applications. Our proof in Chapter III uses several recently developed concentration inequalities and small ball probability estimates, which includes the lower bound Theorem 1.2.8.

In Chapter IV, we study the quantitative smallest singular values distributions of symmetric sparse matrices. We prove a quantitative version of invertibility for sparse symmetric matrices that are not close to the critical sparsity level. This work is motivated by a recent progress of A. Basak and M. Rudelson [6].

Our proof adapts the framework developed in [47] for invertibility problems. In [47], Rudelson and Vershynin addressed that the quantitative invertibilty problem for random matrices can be divided into two parts: lower bound of $\|Ax\|_2$ over vectors x that are very close to sparse and lower bound of $\|Ax\|_2$ over vectors x which coordinates are well-spread. These two different classes of vectors needed to be handled with different approaches. The vectors which are close to sparse are called compressible vectors, the vectors which coordinates are well-spread are named as incompressible vectors. The compressible vectors are usually easier to deal for non-sparse matrices. One contribution of [6] is a combinatorial approach to address the sparsity in estimating the norm of Ax for a sparse matrix A and sparse vector x. The combinatorial lemma is generalizable in symmetric matrices case which makes it possible to prove quantitative invertibility for symmetric sparse matrix together with a decoupling method in [73].

On the other hand, the infimum of $\|Ax\|_2$ over incompressible vectors is more technical. A core component of the method in [47] is based on estimating the small ball probability of a weighted sum of independent variables which is the famous

Littlewood-Offord problem. The Littlewood-Offord problem asks to estimate the small ball probability

$$p_\varepsilon(a) = \sup_{v\in\mathbb{R}} \mathbb{P}(|\langle a, X\rangle - v| \le \varepsilon)$$

where X is a random vector with i.i.d. entries and a is arbitrary real coefficient vector [40]. In [47], Rudelson and Vershynin proposed a new and essentially sharp estimate in the Littlewood-Offord problem based on a quantitative characterization of arithmetic structure of vector a.

To see why the structure of a matters, assume X is an i.i.d. random vector with symmetric ± 1 entries. Then for $a = (1/\sqrt{2}, 1/\sqrt{2}, 0, \cdots, 0)$, we have $\mathbb{P}(|\langle a, X\rangle - v| = 0) = 1/2$. This singular behavior is due to the fact that a is sparse. If we choose vector a to be far away from sparse ones, that is to say an incompressible vector, then the small ball probability estimate can be significantly improved. For instance, choose $a = (1/\sqrt{n}, \cdots, 1/\sqrt{n})$. Then one can compute that $\mathbb{P}\left(|\langle a, X\rangle - v| = 0\right) \sim 1/\sqrt{n}$ [44].

The additive structure of a vector $a = (a_1, \cdots, a_n)$ of real numbers can be described in terms of the shortest arithmetic progression into which it embeds [47]. This length is expressed as the least common denominator of a, defined as follows:

$$\operatorname{lcd}(a) := \inf\left\{\theta > 0 : \theta a \in \mathbb{Z}^n - \{0\}\right\}$$

However, this may not work in general if no such θ exists in above definition. Instead, we define for $L \ge 1$, the least common denominator (LCD) [47, 73, 50] of $x \in S^{n-1}$ as

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

The Littlewood-Offord problem and the application of least common denominator in proving the invertibility over incompressible vectors will be detailed discussed in

Chapter IV.

Finally, we wrap up our introduction by giving some notations and basic definitions.

Through out the thesis, $\|\cdot\|_p$ denote the l_p norm of a vector and B_p^n stands for the unit ball of this norm. We use S^{n-1} for the unit Euclidean sphere. If S is a finite set, then $|S|$ denotes the cardinality of S. The canonical basis of $\mathbb{R}^n$ is denoted $e_1, \cdots, e_n$. We use $\mathrm{dist}(\cdot,\cdot)$ to denote the Euclidean distance between points, vectors or subspaces.

Given a symmetric convex body K in $\mathbb{R}^n$, we have a corresponding norm $\|x\|_K = \inf\{r > 0\,,\, x \in rK\}$. Let ν_n denote the normalized Haar measure on the Euclidean sphere, S^{n-1}, and $\nu_{n,k}$ denote the normalized Haar measure on the Grassmannian manifold $Gr_{n,k}$. Let $M = M(K) := \int_{S^{n-1}} \|x\|_K d\nu_n$ and $b = b(K) := \sup\{\|x\|_K\,,\, x \in S^{n-1}\}$ be the mean and the maximum of the norm over the unit sphere.

The norm of an operator or a matrix will be denoted by $\|\cdot\|$. Let $N \geq n$ and let A be an $N \times n$ matrix. The Hilbert-Schmidt (Frobenius) norm of a matrix which is the l_2 norm of the matrix when view as a $N \times n$ vector, will be denoted by $\|\cdot\|_{\mathrm{HS}}$ (or $\|\cdot\|_{\mathrm{F}}$). The singular values of A are the eigenvalues of $(A^*A)^{1/2}$ arranged in the decreasing order: $s_1(A) \geq s_2(A) \geq \cdots \geq s_n(A)$. So equivalently, $s_{\min}(A) = s_n(A)$ and $s_{\max}(A) = s_1(A)$. And condition number of the matrix A is $(A) = s_{\max}(A)/s_{\min}(A)$.

By $\mathbb{E}_X$, $\mathbb{P}_X$ we denote the conditional expectation and probability with respect to a random variable X, conditioned on all other variables. By $\mathbb{E}_X(\cdot|\mathrm{E})$, $\mathbb{P}_X(\cdot|\mathrm{E})$ we denote the conditional expectation and probability with respect to the event E.

Let E be a subspace of $\mathbb{R}^n$. The unit sphere on E is denoted as S_E. The orthogonal projection onto a subspace E of $\mathbb{R}^n$ is denoted P_E. The matrix A as a linear map restricted on E is denoted as $A|_E$.

Let (T, d) be a metric space and $K \subset T$. A set $\mathcal{N} \subset T$ is called an ε-net for K if for any $x \in K$, there exists $y \in \mathcal{N}$ such that $d(x, y) < \varepsilon$. We will also use the following volumetric estimate many times throughout the thesis: For any $\varepsilon < 1$ there exists an ε-net $\mathcal{N} \subset S^{n-1}$ such that $|\mathcal{N}| \leq (3/\varepsilon)^n$.

As our problem settings are different in different chapters, further notations will be introduced at the beginning of each chapter.

CHAPTER II

Upper bound for the Dvoretzky dimension in Milman-Schechtman theorem

For a symmetric convex body $K \subset \mathbb{R}^n$, the Dvoretzky dimension $k(K)$ is the largest dimension for which a random central section of K is almost spherical. A Dvoretzky-type theorem proved by V. D. Milman in 1971 provides a lower bound for $k(K)$ in terms of the average $M(K)$ and the maximum $b(K)$ of the norm generated by K over the Euclidean unit sphere. Later, V. D. Milman and G. Schechtman obtained a matching upper bound for $k(K)$ in the case when $\frac{M(K)}{b(K)} > c(\frac{\log(n)}{n})^{\frac{1}{2}}$. In this chapter, we will give an elementary proof of the upper bound in Milman-Schechtman theorem which does not require any restriction on $M(K)$ and $b(K)$. This chapter is based on a joint work with H. Huang, see [22].

Outline of the chapter.

- In Section 2.1, we first review Milman-Dvoretzky theorem and Milman-Schechtman theorem. Following that, we present our new upper bound theorem and discuss the improvement.
- In Section 2.2, we introduce the concentration inequality on sphere and prove our upper bound using a construction based on random projection.
- In Section 2.3, we discuss several characteristics of our new result.

2.1 Upper bound in Milman-Schechtman Theorem

In 1971, V. D. Milman proved the following Dvoretzky-type theorem [34]:

Theorem 2.1.1. *Let K be a symmetric convex body in $\mathbb{R}^n$. Assume that $\|x\|_K \le b|x|$ for all $x \in \mathbb{R}^n$. For any $\epsilon \in (0,1)$, there is $k \ge C_\epsilon (M/b)^2 n$ such that*

$$\nu_{n,k}\{F \in G_{n,k} : (1-\varepsilon)M < \|\cdot\|_{K\cap F} < (1+\varepsilon)M\} > 1 - \exp(-\tilde{c}k)$$

where $\tilde{c} > 0$ is a universal constant, $C_\epsilon > 0$ is a constant only depending on ϵ.

The quantity C_ϵ was of the order $\epsilon^2 \log^{-1}(\frac{1}{\epsilon})$ in the original proof of V. D. Milman. It was improved to the order of ϵ^2 by Y. Gordon [16] and later, with a simpler argument, by G. Schechtman [54].

In 1997, V. D. Milman and G. Schechtman [35] found that the bound on k appearing in Theorem 2.1.1 is essentially optimal. More precisely, they proved the following theorem.

Theorem 2.1.2. (Milman-Schechtman, see e.g., section 5.3 in [53]). *Let K be a symmetric convex body in $\mathbb{R}^n$. For $\epsilon \in (0,1)$, define $k(K)$ to be the largest dimension k such that*

$$\nu_{n,k}\left(\{F \in G_{n,k} : \forall x \in S^{n-1} \cap F\,,\ (1-\varepsilon)M < \|x\|_K < (1+\varepsilon)M\}\right) > p_{n,k} = \frac{n}{n+k}.$$

Then,

$$\tilde{C}_\epsilon n(M/b)^2 \ge k(K) \ge \bar{C}_\epsilon n(M/b)^2$$

when $\frac{M}{b} > c(\frac{\log(n)}{n})^{\frac{1}{2}}$ for some universal constant c, where $\|\cdot\|_F$ denotes the norm corresponding to the convex body $K \cap F$ in F, and $\tilde{C}_\epsilon, \bar{C}_\epsilon > 0$ are constants depending only on ϵ.

The two-sided inequality in Milman-Schechtman Theorem shows a general phenomenon in geometric functional analysis. Recall $1/b$ is the inradius of of the convex body and $M(K)$ is the mean width of the polar body. Milman Schechtman theorem connects these geometric parameters of a convex body to the Dvoretzky dimension, which is a probabilistic quantity. The fact that there is a tight connection is very rare and is remarkable. It was therefore important to show that this connection does not require any assumptions on the body.

Because, the Dvoretzky-Milman theorem cannot guarantee the lower bound with small $\frac{M}{b}$ for $p_{n,k} = \frac{n}{n+k}$, the original proof required an assumption that $\frac{M}{b} > c(\frac{\log(n)}{n})^{\frac{1}{2}}$ for some c. In [53, p. 197], S. Artstein-Avidan, A. A. Giannopoulos, and V. D. Milman addressed it as an open question whether one can prove the same result when $p_{n,k}$ is a constant, such as $\frac{1}{2}$. When $p_{n,k} = \frac{1}{2}$, the lower estimate on $k(K)$ is a direct result of Dvoretzky-Milman theorem [34], but the upper bound was unknown. In this chapter, we are going to give upper bound estimate with $p_{n,k} = \frac{1}{2}$, our main result is the following theorem:

Theorem 2.1.3. *Let K be a symmetric convex body in $\mathbb{R}^n$. Fix a constant $\epsilon \in (0,1)$, let $k(K)$ be the largest dimension such that*

$$\nu_{n,k}\{F \in G_{n,k} : (1-\varepsilon)M < \|\cdot\|_{K\cap F} < (1+\varepsilon)M\} > \frac{1}{2}.$$

Then,

$$Cn(M/b)^2 \geq k(X) \geq \bar{C}_\epsilon n(M/b)^2$$

where $C > 0$ is a universal constant and $\bar{C}_\epsilon > 0$ is a constant depending only on ϵ.

In the next section, we will provide a proof of theorem 2.1.3 with no restriction on $\frac{M}{b}$. In fact, from the proof, one can see that $\frac{1}{2}$ can be replaced by any $c \in (0,1)$ or $1 - \exp(-\tilde{c}k)$, which is the probability appearing in Milman-Dvoretzky theorem.

2.2 Proof of theorem 2.1.3

Let P_k be the orthogonal projection from S^{n-1} to some fixed k-dimensional subspace. The upper estimate is related to the distribution of $\|P_k(x)\|_2$, where x is uniformly distributed on S^{n-1} .

Recall the concentration inequality for Lipschitz functions on the sphere (see, e.g., [70]):

Theorem 2.2.1 (Measure Concentration on S^{n-1})**.** *Let $f : S^{n-1} \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant b. Then, for every $t > 0$,*

$$\nu_n(\{x \in S^{n-1} : |f(x) - \mathbb{E}(f)| \geq bt\}) \leq 4\exp(-c_0 t^2 n)$$

where $c_0 > 0$ is a universal constant.

Theorem 2.2.1 implies the following elementary lemma.

Lemma 2.2.2. *Fix any $c_1 > 0$, let P_k be an orthogonal projection from $\mathbb{R}^n$ to some subspace $\mathbb{R}^k$. If $t > \frac{c_1}{\sqrt{n}}$ and $\nu_n(\{x \in S^{n-1} : \|P_k(x)\|_2 < t\}) > \frac{1}{2}$, then $k < c_2 t^2 n$, where $c_2 > 0$ is a constant depending only on c_1.*

Proof. $\|P_k(x)\|_2$ is a 1-Lipschitz function on S^{n-1} with $\mathbb{E}\|P_k(x)\|_2$ about $\sqrt{\frac{k}{n}}$. If we want the measure of $\{x : \|P_k(x)\|_2 < t\}$ to be greater than 1/2, then measure concentration will force $\mathbb{E}|P_k|$ to be bounded by the size of t, which means $k < c_2 t^2 n$ for some universal constant c_2. Since $t^2 n > c_1^2$, we may and shall assume k is greater than some absolute constant in our proof, then adjust c_2.

To make it precise, we will first give a lower bound on $\mathbb{E}\|P_k(x)\|_2$. By Theorem

2.2.1,

$$\nu_n\left(|\|P_k(x)\|_2 - \mathbb{E}\|P_k(x)\|_2|^2 > t\right) \le 4\exp(-c_0 tn).$$

Thus,

$$\begin{aligned}
\mathbb{E}\|P_k(x)\|_2^2 - (\mathbb{E}\|P_k(x)\|_2)^2 &= \mathbb{E}\left(\|P_k(x)\|_2 - \mathbb{E}\|P_k(x)\|_2\right)^2 \\
&< \int_0^\infty \nu_n\left(|\|P_k(x)\|_2 - \mathbb{E}\|P_k(x)\|_2|^2 > t\right) dt \\
&\le \int_0^\infty 4\exp(-c_0 tn)dt = \frac{4}{c_0 n}.
\end{aligned}$$

With $\mathbb{E}\|P_k(x)\|_2^2 = \mathbb{E}\sum_{i=1}^k |x_i|^2 = \frac{k}{n}$, we get $\mathbb{E}(\|P_k(x)\|_2) > \sqrt{\frac{k}{n} - \frac{4}{c_0 n}}$. If we assume that $k > \frac{24}{c_0}$, then we have

$$\mathbb{E}(\|P_k(x)\|_2) > \sqrt{\frac{1}{2}\frac{k}{n}}.$$

Assuming $k > 8t^2 n$, we have

$$\mathbb{E}(\|P_k(x)\|_2) - t > \sqrt{\frac{1}{2}\frac{k}{n}} - t \ge \frac{1}{2}\sqrt{\frac{1}{2}\frac{k}{n}} > 0.$$

Applying Theorem 2.2.1 again, we obtain

$$\begin{aligned}
\nu_n(\|P_k(x)\|_2 < t) &< \nu_n\left(|\|P_k(x)\|_2 - \mathbb{E}\|P_k(x)\|_2| > \mathbb{E}(\|P_k(x)\|_2) - t\right) \\
&\le 4\exp(-c_0(\mathbb{E}(\|P_k(x)\|_2) - t)^2 n) \\
&\le 4\exp(-c_0(\frac{1}{2}\sqrt{\frac{1}{2}\frac{k}{n}})^2 n) \le 4\exp(-\frac{c_0}{8}k) \le 4\exp(-3) < \frac{1}{2},
\end{aligned}$$

which proves our result by contradiction. □

Theorem 2.2.3. *Let K be a symmetric convex body with inradius $\frac{1}{b}$. For $\epsilon \in (0,1)$, let k be the largest integer such that*

$$\nu_{n,k}\{F \in G_{n,k} : (1-\varepsilon)M < \|\cdot\|_{K\cap F} < (1+\varepsilon)M\} > \frac{1}{2}.$$

Then $k < Cn(\frac{M}{b})^2$ where C is an absolute constant.

Proof. We may assume $\|e_1\|_K = b$, then $K \subset S = \{x \in \mathbb{R}^n : |x_1| < \frac{1}{b}\}$, thus $\|x\|_K \geq \|x\|_S = b|\langle x, e_1\rangle|$. This implies

$$\begin{aligned} & \{V \in G_{n,k} : \forall x \in V \cap S^{n-1}, (1-\epsilon)M < \|x\|_K < (1+\epsilon)M\} \\ \subset\; & \{V \in G_{n,k} : \forall x \in V \cap S^{n-1}, \|x\|_S < (1+\epsilon)M\} \\ =\; & \{V \in G_{n,k} : \sup_{x \in V \cap S^{n-1}} \langle x, e_1\rangle < (1+\epsilon)\tfrac{M}{b}\} \\ =\; & \{V \in G_{n,k} : \|P_V(e_1)\|_2 < (1+\epsilon)\tfrac{M}{b}\} \end{aligned} \tag{2.1}$$

where P_V is the orthogonal projection from $\mathbb{R}^n$ to V. If V is uniformly distributed on $G_{n,k}$ and x is uniformly distributed on S^{n-1}, then $\|P_{V_0}(x)\|_2$ and $\|P_V(e_1)\|_2$ are equi-distributed for any fixed k-dimensional subspace V_0. Therefore,

$$\nu_{n,k}(\{V \in G_{n,k} : \|P_V(e_1)\|_2 < (1+\epsilon)\frac{M}{b}\}) = \nu_n(\{x \in S^{n-1} : \|P_{V_0}(x)\|_2 < (1+\epsilon)\frac{M}{b}\}).$$

As shown in the Remark 5.2.2(iii) of [53, p. 164], the ratio $\frac{M}{b}$ has a lower bound $\frac{c'}{\sqrt{n}}$. Setting $c_1 = c'$ and $t = (1+\epsilon)\frac{M}{b}$, it is easy to see that if

$$\nu_{n,k}\{F \in G_{n,k} : (1-\varepsilon)M < \|\cdot\|_{K\cap F} < (1+\varepsilon)M\} > \frac{1}{2},$$

then $k \leq c_1(1+\epsilon)^2 \left(\frac{M}{b}\right)^2 n < Cn(\frac{M}{b})^2$ by Lemma 2.2.2 and (2.1).

□

Now we can prove theorem 2.1.3 as a corollary of Theorem 2.2.3 and Theorem 2.1.1:

Proof of theorem 2.1.3. Theorem 2.1.1 shows that if $C_\epsilon(M/b)^2 n > \frac{\log(2)}{\tilde{c}}$, then there is $k \geq C_\epsilon(M/b)^2 n$ such that

$$\nu_{n,k}\{F \in G_{n,k} : (1-\varepsilon)M < \|\cdot\|_F < (1+\varepsilon)M\} > 1 - \exp(-\tilde{c}k) > \frac{1}{2}.$$

Otherwise, $(M/b)^2 n < \frac{\log(2)}{\tilde{c}C_\epsilon}$. Therefore, $k(K) \geq \min\{\frac{\tilde{c}C_\epsilon}{\log(2)}, C_\epsilon\}(M/b)^2 n$. Combining it with Theorem 2.2.3, we get

$$C(\frac{M}{b})^2 n \geq k(K) \geq \min\{\frac{\tilde{c}C_\epsilon}{\log(2)}, C_\epsilon\}(M/b)^2 n.$$

$\square$

2.3 Discussion

First, it is worth noticing that the number $\frac{1}{2}$ plays no special role in our proof of Theorem 2.1.3. Thus, if we define the Dvoretzky dimension to be the largest dimension such that

$$\nu_{n,k}\{F \in G_{n,k} : (1-\varepsilon)M < \|\cdot\|_{K\cap F} < (1+\varepsilon)M\} > c$$

for some $c \in (0,1)$, then exactly the same proof will work. We will still have $k(K) \sim (\frac{M}{b})^2 n$. Similarly, if we fix ϵ and replace $\frac{1}{2}$ by $1-\exp(-\tilde{c}k)$, then the lower bound of $k(K)$ is the one from Theorem 2.1.1. For k greater than some absolute constant, we have $1-\exp(-\tilde{c}k) > \frac{1}{2}$. Thus, the upper bound is still of order $\left(\frac{M}{b}\right)^2 n$. Therefore, we can replace $\frac{1}{2}$ by $1-\exp(-\tilde{c}k)$ in theorem 2.1.2. With this probability choice, it also shows Theorem 2.1.1 provides an optimal k depending on M, b.

Secondly, usually we are only interested in $\epsilon \in (0,1)$. In the lower bound, $\bar{C}_\epsilon = o_\epsilon(1)$. It is a natural question to ask if we could improve the upper bound from a universal constant C to $o_\epsilon(1)$. Unfortunately, it is not possible due to the following observation. Let $K = \operatorname{conv}(B_2^n, Re_1)^\circ$. By passing from the intersection on K to the projection of K°, one can show that $k(K)$ does not exceed the maximum dimension k such that $\nu_n(\|P_k(Rx)\|_2 < 1+\epsilon) > \frac{1}{2}$. Choosing $R = \sqrt{\frac{n}{l}}$, we get $n(\frac{M}{b})^2 \sim l$ and $k(X) \sim l$ by Theorem 2.2.1 and a similar argument to that of Lemma 2.2.2. This example shows that no matter what $\frac{M}{b}$ is, one can not improve the upper bound in theorem 2.1.2 from an absolute constant C to $o_\epsilon(1)$.

Finally, we want to address that the proof we presented uses random projections but we have freedom to use other random matrices. For example, we can prove the same result suing by i.i.d. Gaussian matrices.

CHAPTER III

Upper bound for intermediate singular values of random matrices

In this chapter, we prove that an $n \times n$ sub-gaussian matrix A with independent centered sub-gaussian entries satisfies

$$s_{n+1-l}(A) \leq C_1 t \frac{l}{\sqrt{n}}$$

with probability at least $1 - \exp(-C_2 t l)$ under some weak condition. This yields $s_{n+1-l}(A) \sim \frac{cl}{\sqrt{n}}$ in combination with a known lower bound. These results can be generalized to the rectangular matrix case. This chapter is based on one of my phd research publications, see [76] .

Outline of the chapter.

- In Section 3.1, we give a short review of results related to upper bound of singular values of random i.i.d. sub-gaussian matrices. Then we present our main theorem following necessary definitions and assumptions.
- In Section 3.2, we present some preliminary results needed in the proof of Theorem 3.1.6 which include definition of biorthogonal system, some concentration inequalities and small ball probability estimates.
- In Section 3.3, we provide proof of our main Theorem. The proof is both probabilistic and geometric which will be divided in three steps. Intuition and

outline of the proof will be explained at the beginning of Section 3.3.

- Finally, we prove Theorems 3.1.10 and 3.1.11 In Section 3.4 which generalize our results to rectangular case.

3.1 Introduction

The non-asymptotic singular value distribution of random i.i.d. sub-gaussian matrix is an important and interesting subfield in random matrix theory. The first result in this direction was obtained in [36], where it was proved that the smallest singular value of a square i.i.d. sub-gaussian matrix is bounded below by $n^{-3/2}$ with high probability. This result was later extended and improved in a number of papers, including [67, 68, 47, 6, 43]. The above mentioned results pertain to square matrices. A probabilistic lower bound for the smallest singular value of a rectangular matrix was obtained by M. Rudelson and R. Vershynin [48]. They proved that an $n \times (n-l)$ matrix has smallest singular value lower bounded by $\frac{\varepsilon l}{\sqrt{n}}$ with probability at least $1-(C\varepsilon)^l - \exp(-Cn)$. Using this result, one can show that for a square i.i.d. sub-gaussian matrix A, $s_{n+1-l}(A) > c\frac{l}{\sqrt{n}}$ with high probability where $1 \leq l \leq n$.

However, the optimal upper bound of the singular values for general sub-gaussian matrices is unknown. Prior to our result, the only progress in this direction was made by Szarek [59] who established an optimal upper bound for the Gaussian matrix. Szarek proved that for a standard Gaussian i.i.d. matrix G, $\frac{Cl}{\sqrt{n}} \leq s_{n+1-l}(G) \leq \frac{Cl}{\sqrt{n}}$ with probability at least $1-\exp(-Cl^2)$. This result suggests that lth smallest singular value of an i.i.d. sub-gaussian matrix is concentrated around $\frac{l}{\sqrt{n}}$.

Although the optimal upper bound is not proved for general matrices, some results can be deduced. T. Tao and V. Vu have established the universal behavior of small singular values in [68] (see Theorem 6.2 [68]). Combined with Szarek's Theorem 1.3

in [59], their approach allows us to deduce some non-asymptotic bounds for random i.i.d. square matrix under a moment condition. However, their bound only works for $l \leq n^c$ where c is a small constant. Tao and Vu's approach [68] is based on the Berry-Esseen Theorem for the frames and does not allow one to obtain exponential bounds for the probability as we do in our Theorem 3.1.6. Also, C. Cacciapuoti, A. Maltsev, B. Schlein estimated the rate of convergence of the empirical measure of singular values to the limit distribution near the hard edge (see [10] Theorem 3). Theorem 3 in [10] can be used to derive an upper bound of the form $\frac{cl^C}{\sqrt{n}}$ [10]. Better understood is the upper bound for the smallest singular value. M. Rudelson and R. Vershynin were the first to prove the smallest singular value of the i.i.d. sub-gaussian matrix is also bounded from above by $\frac{c}{\sqrt{n}}$ with high probability (see [46]). A different proof with an exponential tail probability can be found in a very recent paper by H. Nguyen and V. Vu [37].

In this chapter, we prove the upper bound on the singular values under two assumptions: that the entries of the matrix are non-degenerate; and that they have a fast tail decay. The first assumption is quantified in terms of the Levy concentration function and the second is quantified in terms of the ψ_θ-norm. Next we provide definitions.

Definition 3.1.1. Let Z be a random vector that takes values in $\mathbb{C}^n$. The concentration function of Z is defined as

$$\mathcal{L}(Z,t) = \sup_{u \in \mathbb{C}^n} \mathbb{P}\{\|Z-u\|_2 \leq t\}, t \geq 0.$$

Definition 3.1.2. Let $\theta > 0$. Let Z be a random variable on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Then the ψ_θ-norm of Z is defined as

$$\|Z\|_{\psi_\theta} := \inf\left\{\lambda > 0 : \mathbb{E}\exp\left(\frac{|Z|}{\lambda}\right)^\theta \leq 2\right\}$$

If $\|X\|_{\psi_\theta} < \infty$, then X is called a ψ_θ random variable. This condition is satisfied for broad classes of random variables. In particular, a bounded random variable is ψ_θ for any $\theta > 0$, a normal random variable is ψ_2, and a Poisson variable is ψ_1.

In this chapter, we prove that for all l, $s_{n+1-l}(A) \geq \frac{Ctl}{\sqrt{n}}$ with an exponentially small probability, where A is a random matrix under the following assumption:

Assumption 3.1.3. *Let $p > 0$. Let A be an $n \times m$ random matrix with i.i.d. entries that have mean 0, variance 1 and ψ_2-norm K. Assume also that there exists $0 < s \leq s_0(p, K)$ such that*

$$\mathcal{L}(A_{i,j}, s) \leq ps.$$

Here, $s_0(p, K)$ is a given function depending only on p and K.

A concrete value of $s_0(p, K)$ can be detemined by tracing the proof of Theorem 3.1.6.

Remark 3.1.4. The condition on the Levy concentration function is automatically satisfied if the density of the entries is bounded by p. However, our result holds in a much more general setting because we require this condition to hold only for one fixed s and not for all $s > 0$. This assumption can be viewed as a discrete analog of the bounded density condition.

Remark 3.1.5. The analysis of the proof for Theorem 3.1.6 shows that it is enough to take $s_0(p, K) = c(K) \min\{p^{-1}, 1\}$, where $c(K)$ is a small constant that depends only on K.

We prove the following main theorem for a random matrix A satisfying Assumption 3.1.3:

Theorem 3.1.6. (Upper bound for singular values of an i.i.d. sub-gaussian square matrix) *Let A be an $n \times n$ random matrix that satisfies Assumption 3.1.3 with some*

$s_0(p, K)$ that depends only on p, K. Then there exist constants $C_1, C_2 > 0$ such that for all $t > 1$, for all l between 1 and n,

$$\mathbb{P}\left(s_{n+1-l}(A) \leq C_1 \frac{tl}{\sqrt{n}}\right) \geq 1 - \exp(-C_2 tl)$$

where C_1, C_2 are constants that depend only on K, p.

Remark 3.1.7. In [37], Nguyen and Vu obtained a sharp bound for the smallest singular value of i.i.d. sub-gaussian matrices with exponential tail probability. Theorem 3.1.6 recovers the result obtained by Nguyen and Vu under Assumption 3.1.3 and generalizes that result to all l.

Remark 3.1.8. Unlike the probability bound deduced from [68], our probability tail bound is exponential type. Szarek's probability estimate [59] suggests that the optimal probability bound for $\mathbb{P}\left(s_{n+1-l}(A) \leq \frac{cl}{\sqrt{n}}\right)$ is $1 - \exp(-Cl^2)$.

Remark 3.1.9. Possible applications of Theorem 3.1.6 include eigenvector l_∞ delocalization of random matrices [50]. For that, one has to consider $A - zI$ instead of the matrix A. Further effort would be needed to generalize our result to the case with a shift.

Also, Theorem 3.1.6 can be extended to rectangular matrices easily (see Section 4). And more precisely, we have the following corollary:

Corollary 3.1.10. (Upper bound for singular values of an i.i.d. sub-gaussian matrix) *Let A be an $n \times (n-k)$ random matrix that satisfies Assumption 3.1.3 with some $s_0(p, K)$ that depends only on p, K. Then there exist constants $C_1, C_2 > 0$ such that for all $t > 1$ and l between 1 and n,*

$$\mathbb{P}\left(s_{n+1-l}(A) \leq C_1 \frac{tl}{\sqrt{n}}\right) \geq 1 - \exp(-C_2 tl)$$

where C_1, C_2 are constants that depend only on K, p.

A direct application of Theorem 3.1.6 and Theorem 3.2.6 leads us to a generalization of Theorem 1.3 in [59].

Corollary 3.1.11. (Non-asymptotic singular values distribution of i.i.d. sub-gaussian square matrix) *Let A be an $n \times n$ random matrix that satisfies Assumption 3.1.3 with some $s_0(p, K)$ that depends only on p, K. Then there exist $0 < C_1 < C_2$ and $C_3 > 0$, such that for all l between 1 and n,*

$$\mathbb{P}\left(\frac{C_1 l}{\sqrt{n}} \leq s_{n+1-l}(A) \leq \frac{C_2 l}{\sqrt{n}}\right) \geq 1 - \exp(-C_3 l)$$

where C_is are constants that depend only on K, p.

A similar proof can lead to an analog in the rectangular case:

Corollary 3.1.12. (Non-asymptotic distribution of singular values in the i.i.d. sub-gaussian rectangular matrix) *Let A be an $n \times (n-k)$ random matrix that satisfies Assumption 3.1.3 with some $s_0(p, K)$ that depends only on p, K. Then there exist $0 < C_1 < C_2$ and $C_3 > 0$, such that for all l between k and n,*

$$\mathbb{P}\left(\frac{C_1 l}{\sqrt{n}} \leq s_{n+1-l}(A) \leq \frac{C_2 l}{\sqrt{n}}\right) \geq 1 - \exp(-C_3 l)$$

where C_is are constants that depend only on K, p.

3.2 Notation and Preliminaries

Throughout this chapter, c denotes absolute constants, C denotes constants that may depend only on the parameters K, p. Note that these constants may vary from line to line. S^{n-1} denotes the n dimensional sphere, i.e., the sphere in $\mathbb{R}^n$ which itself is an $(n-1)$-dimensional manifold. S_E denotes the sphere of a subspace E, i.e., $S_E = S^{n-1} \cap E$.

For a $n \times n$ random matrix A, as in Theorem 3.1.6, we denote by A_l the $n \times l$ matrices of the first l columns of A. A_{n-l} denotes the $n \times (n-l)$ matrix of the last

$n-l$ columns of A. Without loss of generality, we can assume A is a.s. invertible. If not we prove the theorem for $A+\varepsilon G$, where G is an independent Gaussian matrix. Then the result holds for A up to an abssolute constant by sending ε to zero. X_k will denote the kth column of matrix A, and we use the following notations

- $H_l :=$span$(X_j)_{j>l}$.
- $H_{l,k} :=$span$(X_j)_{j>l, j\neq k}$.
- $P_l, P_{l,k}$ are the orthogonal projections onto $H_l, H_{l,k}$, respectively.
- $P_l^{\perp}, P_{l,k}^{\perp}$ are the orthogonal projections onto $H_l^{\perp}, H_{l,k}^{\perp}$, respectively.
- $X_k^* := (A^{-1})^* e_k$, i.e., the k-th column of $(A^{-1})^*$.
- $Y_k^* := P_l X_k^*$, $k = l+1, l+2, \cdots, n$.

3.2.1 Biorthogonal system

Consider vectors $(v_k)_{k=1}^n$ and $(v_k^*)_{k=1}^n$ that belong to an $n-$dimensional Hilbert space H. Recall that the system $(v_k, v_k^*)_{k=1}^n$ is called a biorthogonal system in H if $\langle v_j, v_k^*\rangle = \delta_{j,k}$ for all $j,k = 1,2,\cdots,n$. The system is called complete if span$(v_k) = H$. The following theorem summarizes some elementary known properties of biorthogonal systems.

Theorem 3.2.1. *1. Let D be an $n\times n$ invertible matrix with columns $v_k = De_k$, $k = 1,2,\cdots,n$. Define $v_k^* = (D^{-1})^* e_k$. Then $(v_k, v_k^*)_{k=1}^n$ is a complete biorthogonal system in $\mathbb{R}^n$.*

2. Let $(v_k)_{k=1}^n$ be a linearly independent system in an $n-$dimensional Hilbert space H. Then there exist unique vectors $(v_k^*)_{k=1}^n$ such that $(v_k, v_k^*)_{k=1}^n$ is a biorthogonal system in H. This system is complete.

3. Let $(v_k, v_k^*)_{k=1}^n$ *be a complete biorthogonal system in a Hilbert space* H*. Then* $\|v_k^*\|_2 = 1/\mathrm{dist}(v_k, \mathrm{span}(v_j)_{j\neq k})$ *for* $k = 1, 2, \cdots, n$*.*

Proof. (1) follows directly from

$$\langle v_j, v_k^* \rangle = \langle De_j, (D^{-1})^* e_k \rangle = \langle D^{-1} De_j, e_k \rangle = \langle e_j, e_k \rangle = \delta_{j,k}.$$

To prove (2), we use the fact that any basis on a finite dimensional vector space has a unique dual basis. Since H is Hilbert space, the dual basis also belongs to H which forms a biorthogonal system together with the original basis. The completeness follows from the dimension argument.

Since $(v_i, v_i^*)_{i=1}^n$ is a complete biorthogonal system on the Hilbert space H, for any $k = 1, 2, \cdots, n$, we have $v_k^* \perp \mathrm{span}\{v_i : i \neq k\}$ and $\{v_i : i = 1, \cdots, n, i \neq k\} \cup \{v_k^*\}$ form a basis on H. Thus we have the decomposition

$$v_k = \sum_{i\neq k} a_i v_i + \mathrm{dist}(v_k, \mathrm{span}(v_j)_{j\neq k}) \frac{v_k^*}{\|v_k^*\|_2}.$$

Take inner product with v_k^* at both sides and we have

$$1 = \mathrm{dist}(v_k, \mathrm{span}(v_j)_{j\neq k}) \|v_k^*\|_2,$$

which proves (3).

□

The next Lemma tells us the relation between Y_k^* and X_k^* for $k \geq l + 1$.

Lemma 3.2.2. $(X_k, Y_k^*)_{k=l+1}^n$ *is a complete biorthogonal system in* H_l*.*

Proof. By definition, for all $k \geq l + 1$

$$Y_k^* - X_k^* \in \ker(P_l) = H_l^\perp = \mathrm{span}(X_j^*)_{j\leq l}.$$

So we have, for all $k \geq l+1$

$$Y_k^* = X_k^* + \sum_{j=1}^{l} a_{kj} X_j^*, \text{ for some } a_{jk} \in \mathbb{R}, j = 1, 2, \cdots, l.$$

By the orthogonality, we have for all $k, i \geq l+1$

$$\langle Y_k^*, X_i \rangle = \langle X_k^*, X_i \rangle + \sum_{j=1}^{l} a_{kj} \langle X_j^*, X_i \rangle = \langle X_k^*, X_i \rangle = \delta_{k,i}.$$

Thus the biorthogonality is proved. The competeness follows since $\dim(H_l) = n - l$. □

In view of the uniqueness of Part 2 of Theorem 3.2.1, Lemma 3.2.2 has the following crucial consequence.

Corollary 3.2.3. *The system of vectors $(Y_k^*)_{k=l+1}^n$ is uniquely determined by the system $(X_k)_{k=l+1}^n$. In particular, the random vector system $(Y_k^*)_{k=l+1}^n$ is independent with random vector system $(X_k)_{k=1}^l$.*

3.2.2 Concentration thereoms

The major tools of our proof come from measure concentration theory. Here we list the concentration theorems that will be used in the proof.

The first theorem is a concentration property of sub-gaussian random vectors.

Theorem 3.2.4. *Let D be a fixed $m \times n$ matrix. Consider a random vector Z with independent entries that have mean 0, variance greater than 1, and uniformly bounded by K in ψ_2 norm. Then, for any $t \geq 0$, we have*

$$\mathbb{P}(|\|DZ\|_2 - M| > t) \leq 2 \exp\left(-\frac{ct^2}{\|D\|^2}\right)$$

where $M = (\mathbb{E}\|DZ\|_2^2)^{1/2}$ which satisfies $\|D\|_{\mathrm{HS}} \leq M \leq K\|D\|_{\mathrm{HS}}$, and $c = c(K)$ is polynomial in K.

This result can be deduced from the Hanson-Wright inequality. A modern proof of the Hanson-Wright inequality and a deduction of the above Theorem 3.2.4 are discussed in [51].

Sub-gaussian concentration paired with a standard covering argument yields the following result on norms of random matrices, see [51].

Theorem 3.2.5. (Products of random and deterministic matrices). *Let D be a fixed $m \times N$ matrix, and let G be an $N \times k$ random matrix with independent entries that satisfy $\mathbb{E}G_{ij} = 0, \mathbb{E}G_{ij}^2 \geq 1$ and $\|G_{ij}\|_{\psi_2} \leq K$. Then for any $s, t \geq 1$ we have*

$$\mathbb{P}\{\|DG\| > C(s\|D\|_{\mathrm{HS}} + t\sqrt{k}\|D\|)\} \leq 2\exp(-s^2 r - t^2 k)$$

Here $r = \|D\|_{\mathrm{HS}}^2 / \|D\|_2^2$ is the stable rank of D, and $C = C(K)$ is a polynomial in K.

The following result gives the lower bound on the smallest singular value of a rectangular i.i.d. sub-gaussian matrix. This will be used in our proof of Theorem 3.1.6; it can also yield the lower bound in Theorem 3.1.11 directly. The proof and extensions of the theorem are discussed in [36, 47, 48, 49].

Theorem 3.2.6. *Let G be an $N \times n$ random matrix, $N \geq n$, whose elements are independent copies of a centered sub-gaussian random variable with unit variance. Then for every $\varepsilon > 0$, we have*

$$\mathbb{P}\left(s_n(G) \leq \varepsilon\left(\sqrt{N} - \sqrt{n-1}\right)\right) \leq (C\varepsilon)^{N-n+1} + e^{-C'N}$$

where $C, C' > 0$ depend (polynomially) only on the sub-gaussian moment K.

As one step towards the above least singular value bound, the following distance to a random subspace theorem was proved by M. Rudelson and R. Vershynin [48]:

Theorem 3.2.7. (Distance to a random subspace). *Let Z be a vector in $\mathbb{R}^N$ whose coordinates are independent and identically distributed centered sub-gaussian random* *variables with unit variance. Let H be a random subspace in $\mathbb{R}^N$ spanned by $N-m$ vectors, $0 < m < \tilde{c}N$, whose coordinates are independent and identically distributed centered sub-gaussian random variables with unit variance, independent of Z. Then, for every $v \in \mathbb{R}^N$ and every $\varepsilon > 0$, we have*

$$\mathbb{P}(\operatorname{dist}(Z, H+v) < \varepsilon\sqrt{m}) \leq (C\varepsilon)^m + e^{-cN},$$

where $C, c, \tilde{c}$ depend only on the sub-gaussian moments.

M. Rudelson and R. Vershynin have recently proved the following results for small ball probability of a linear image of high dimensional distributions [52] (see also [32]).

Theorem 3.2.8. (Concentration function of projections.) *Consider a random vector $Z = (Z_1, \cdots, Z_n)$ where Z_i are real-valued independent random variables. Let $t, p \geq 0$ be such that*

$$\mathcal{L}(Z_i, t) \leq p \text{ for all } i = 1, \cdots, n$$

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

$$\mathcal{L}(PZ, t\sqrt{d}) \leq (cp)^d.$$

where c is an absolute constant.

In the same paper, Rudelson and Vershynin generalized Theorem 3.2.8 to general matrices:

Theorem 3.2.9. (Concentration functions of anisotropic distributions.) *Consider a random vector Z where Z_i are real-valued independent random variables. Let $t, p \geq 0$ be such that*

$$\mathcal{L}(Z_i, t) \leq p \text{ for all } i = 1, \cdots, n$$

Let D be an $m \times n$ matrix and $\varepsilon \in (0,1)$. Then

$$\mathcal{L}(DZ, t\|D\|_{\mathrm{HS}}) \leq (c_\varepsilon p)^{(1-\varepsilon)r(D)}$$

where $r(D) = \|D\|_{\mathrm{HS}}^2 / \|D\|_2^2$ *and* c_ε *depend only on* ε*.*

As a special case of Theorem 3.2.8, the following corollary usefully controls the concentration function of sums:

Theorem 3.2.10. (Concentration function of sums.) *Consider a random vector* $Z = (Z_1, \cdots, Z_n)$ *where* Z_i *are real-valued independent random variables. Let* $t, p \geq 0$ *be such that*

$$\mathcal{L}(Z_i, t) \leq p \text{ for all } i = 1, \cdots, n$$

Let $a_1, \cdots, a_n$ *be real numbers with* $\sum_{j=1}^n a_j^2 = 1$*. Then*

$$\mathcal{L}\left(\sum_{i=1}^n a_i Z_i, t\right) \leq cp.$$

where c *is an absolute constant.*

3.3 Proof of Theorem 3.1.6

Before proving the theorem, let us explain our strategy. We prove a lower bound for $s_l(A^{-1})$, rather than proving an upper bound for $s_{n+1-l}(A)$. To do this, we show that there exists an l dimensional subspace, such that the smallest singular value of the operator restricted on this subspace is bounded from below. Our target subspace will be $H_l^\perp$.

The proof uses a net argument for a specially constructed net. In Step 1, we obtain a small ball probability estimate for a random vector. A generic vector in $H_l^\perp$ can be represented as $A^{-1}P_l^\perp A_l y$ for some $y \in \mathbb{R}^l$. We will show that $\frac{\|A^{-1}P_l^\perp A_l y\|_2}{\|P_l^\perp A_l y\|_2}$ is bounded from below by $\frac{C\sqrt{n}}{l}$ for any $y \in S^{l-1}$. In steps 2 and 3, we provide a union bound argument.

There are three essential features of our proof. First, as $H_l^\perp$ is a random subspace, we cannot consider a net on $S_{H_l^\perp}$ directly. So, we consider a net $\mathcal{N}_\varepsilon$ on S^{l-1} instead, which will induce a net on $H_l^\perp$. Second, to complete the argument we need to show the union bound probability of the form $|\mathcal{N}_\varepsilon|\exp(-Cl)$ is small, where C is a small constant. Since $|\mathcal{N}_\varepsilon| \sim \left(\frac{3}{\varepsilon}\right)^l$, this bound in general can be large. To control the probability, we work not on $y \in S^{l-1}$ but on $y \in S^{l'-1}$, $l' = \lfloor \kappa l \rfloor$ for some $\kappa \in (0,1)$ instead. With this dimension reduction argument, we end up proving that $s_{l'}(A^{-1}) \geq \frac{C\sqrt{n}}{l}$, and then we rephrase it. Third, representing a vector from $H_l^\perp$ as $A^{-1}P_l^\perp A_l y$ is advantageous because

$$\|A^{-1}P_l^\perp A_l y\|_2^2 = \|BA_l y\|_2^2 + 1$$

where B is a random matrix that is independent of A_l. This allows us to analyze the property of B first and then apply tools like Theorem 3.2.9 and Theorem 3.2.5. Note that this construction was generalized from the one dimensional case presented by M. Rudelson and R. Vershynin [46].

In the proof, we will use the well-known estimate that there exists an ε-net on S^{l-1} with cardinality less than $\left(\frac{3}{\varepsilon}\right)^l$, see, e.g., Lemma 4.3 in [44].

Proof of Theorem 3.1.6. To prove Theorem 3.1.6, we only need to prove the following claim:

Claim. There exist C_1 and C_2 that only depend on K, p such that for every l between 1 and n,

$$\mathbb{P}\left(s_{n+1-l}(A) \geq C_1 \frac{l}{\sqrt{n}}\right) \leq \exp(-C_2 l).$$

To start, we derive Theorem 3.1.6 from the claim. Let $t \geq 1$, and let k be any integer between 1 and n. Set $l = \lfloor tk \rfloor$ and assume for a moment that $l < n$. Then

$$\begin{aligned} \mathbb{P}\left(s_{n+1-k}(A) \geq C_1 \frac{2tk}{\sqrt{n}}\right) &\leq \mathbb{P}\left(s_{n+1-k}(A) \geq \tfrac{C_1 l}{\sqrt{n}}\right) \\ &\leq \exp(-C_2 l) \leq \exp\left(-\tfrac{C_2 tk}{2}\right). \end{aligned} \tag{3.1}$$

In the case $l \geq n$, the sub-gaussian tail estimate for the norm of a random matrix (one may also consider this as a special case of Theorem 3.2.5) yields

$$\begin{aligned}\mathbb{P}\left(s_{n+1-k}(A) \geq C_3 \frac{2tk}{\sqrt{n}}\right) &\leq \mathbb{P}\left(s_1(A) \geq C_3 \tfrac{2tk}{\sqrt{n}}\right) \\ &\leq \exp\left(-C \tfrac{C_3^2 t^2 k^2}{n}\right) \leq \exp\left(-C_4 tk\right),\end{aligned} \tag{3.2}$$

and therefore for all k between 1 and n,

$$\mathbb{P}\left(s_{n+1-k}(A) \geq C_5 \frac{tk}{\sqrt{n}}\right) \leq \exp(-C_6 tk)$$

with constants C_5, C_6 depending on p, K only. So Theorem 3.1.6 is implied by the claim. □

Now, we prove the above claim.

Proof of the claim. In the proof of the claim, we first assume $l \leq \frac{\tilde{c}n}{2}$, where $\tilde{c}$ is the same as the $\tilde{c}$ which appeared in Theorem 3.2.7. If $l > Cn$, then the required bound follows from the estimate for $s_1(A)$. This is a standard estimate of the operator norm that can be found in many places, for example, in Theorem 2.4 of [48]. Let $\alpha > 1, \delta, \kappa < 1, \beta < \alpha^{-1} < 1$ be parameters to be chosen later. Also, assume that

$$\mathcal{L}(A_{i,j}, \beta) \leq p\beta \tag{3.3}$$

i.e. Assumption 3.1.3 is true with $s = s_0(p, K) = \beta$.

Step 1. Concentration for a random vector. Consider $y \in S^{l-1}$, define

$$U(y) := X(y) - P_l X(y) := A_l y - P_l A_l y. \tag{3.4}$$

then $X(y) := A_l y$ is still a mean 0, variance 1 sub-gaussian random vector. According to the Hoeffding inequality, the sub-gaussian moment of entries of $X(y)$ is bounded above by CK (see Theorem 3.3 in [44]). Without ambiguity, we use the notation U, X instead of $U(y), X(y)$.

In step 1, we show with high probability that

$$\|U\|_2 \lesssim \sqrt{l}, \ \|A^{-1}U\|_2 \gtrsim \frac{\sqrt{n}}{\sqrt{l}}.$$

First, we give an upper bound for $\|P_l^\perp A_l\|$. This leads to an uniform upper bound of $\|U(y)\|_2$ for all $y \in S^{l-1}$.

Step 1.1. Concentration of $\|P_l^\perp A_l\|$.

First, notice that $I - P_l = P_l^\perp$, which is an orthogonal projection onto $H_l^\perp$, so it does not depend on A_l only on A_{n-l}. Thus, $P_l^\perp$ can be treated as a fixed matrix. We apply Theorem 3.2.5 with $B = P_l^\perp$ and $G = A_l$, then we have

$$\mathbb{P}(\|P_l^\perp A_l\| > \alpha\sqrt{l}) \leq 2\exp(-C\alpha^2 l).$$

In particular, for a single vector we have

$$\mathbb{P}(\|U\|_2 > \alpha\sqrt{l}) = \mathbb{P}(\|P_l^\perp A_l y\|_2 > \alpha\sqrt{l}) \leq 2\exp(-C\alpha^2 l).$$

Step 1.2. Concentration of $\|A^{-1}U\|_2$.

Now consider

$$A^{-1}U = A^{-1}A_l y - A^{-1}P_l A_l y = y - A^{-1}P_l A_l y.$$

Notice that $A^{-1}P_l A_l y$ is supported in span $\{e_{l+1}, \cdots, e_n\}$ because $P_l A_l y \in H_l$. So we have

$$\begin{aligned}
\|A^{-1}U\|_2^2 &= \|y\|_2^2 + \|A^{-1}P_lA_ly\|_2^2 > \|A^{-1}P_lA_ly\|_2^2 \\
&= \sum_{k=1}^{n}\langle A^{-1}P_lA_ly, e_k\rangle^2 = \sum_{k=1}^{n}\langle P_lX, (A^{-1})^Te_k\rangle^2 \\
&= \sum_{k=1}^{n}\langle P_lX, X_k^*\rangle^2 = \sum_{k=l+1}^{n}\langle P_lX, X_k^*\rangle^2 \\
&= \sum_{k=l+1}^{n}\langle X, P_lX_k^*\rangle^2 = \sum_{k=l+1}^{n}\langle X, Y_k^*\rangle^2
\end{aligned} \tag{3.5}$$

where in the third line we used the fact that $(X_k, X_k^*)_{k=1}^n$ forms a complete biorthogonal system on $\mathbb{R}^n$ from Lemma 3.2.1. Thus, $X_k^* \perp H_l, k \leq l$.

Using the above property, let B be the $(n-l)\times n$ matrix whose rows are $(Y_k^*)^T, k = l+1, l+2, \cdots, n$. Then we have

$$\|A^{-1}U\|_2^2 \geq \sum_{k=l+1}^{n}\langle X, Y_k^*\rangle^2 = \|BX\|_2^2.$$

Our goal is to get a small ball probability estimate of $\|BX\|$. As B is independent of X, we would like to apply Theorem 3.2.9. Thus, we first need an estimate for $\|B\|$ and $\|B\|_{\mathrm{HS}}$.

Step 1.2.1. Lower bound of $\|B\|_{\mathrm{HS}}$.

According to Theorems 3.2.3, 3.2.2 and 3.2.1, we have

$$\|B\|_{\mathrm{HS}}^2 = \sum_{k=l+1}^{n}\|Y_k^*\|_2^2 = \sum_{k=l+1}^{n}\mathrm{dist}(X_k, H_{l,k})^{-2} = \sum_{k=l+1}^{n}\|P_{l,k}^{\perp}X_k\|_2^{-2}.$$

Denote $V_j = \mathrm{dist}^2(X_j, H_{l,j})$. Then

(3.6)
$$\begin{aligned}\mathbb{P}\left(\|B\|_{\mathrm{HS}} < \alpha^{-1}\sqrt{\frac{n-l}{l}}\right) &= \mathbb{P}\left(\left(\frac{1}{n-l}\sum_{j=l+1}^{n} V_j^{-1}\right)^{-1} > \alpha^2 l\right)\\ &\leq \mathbb{P}\left(\frac{1}{n-l}\sum_{j=l+1}^{n} V_j > \alpha^2 l\right)\\ &= \mathbb{P}\left(\frac{1}{n-l}\sum_{j=l+1}^{n} (V_j - 4(l+1)) > \alpha^2 l - 4(l+1)\right)\\ &\leq \mathbb{P}\left(\frac{1}{n-l}\sum_{j=l+1}^{n} (V_j - 4(l+1))_+ > \frac{\alpha^2}{2} l\right).\end{aligned}$$

where the first inequality follows from the inequality between harmonic mean and arithmetic mean and the second inequality is trivial if we provide $\alpha^2 > 10$. Consider

$$\begin{aligned}\mathbb{P}\left((V_j - 4(l+1))_+ > 4t\right) &\leq \mathbb{P}\left(\sqrt{V_j} > \sqrt{l+1} + \sqrt{t}\right)\\ &= \mathbb{P}\left(\|P_{l,j}^{\perp} X_j\|_2 - \sqrt{l+1} > \sqrt{t}\right)\end{aligned} \tag{3.7}$$

Then applying Theorem 3.2.4 with $A = P_{l,k}^{\perp}$ we have $M = \sqrt{l+1}$. Thus $(V_j - 4(l+1))_+$ is a sub-exponential random variable with $\|(V_j - 4(l+1))_+\|_{\psi_1} \leq C$. By the triangle inequality,

$$\left\|\frac{1}{n-l}\sum_{j=l+1}^{n} (V_j - 4(l+1))_+\right\|_{\psi_1} \leq C.$$

Recalling that $l \leq \frac{n}{2}$, we have

$$\mathbb{P}\left(\|B\|_{\mathrm{HS}} < \alpha^{-1}\sqrt{\frac{n}{l}}\right) \leq \exp(-C\alpha^2 l).$$

Step 1.2.2. Upper bound of $\|B\|$.

First, we have

$$\|B\|^2 = \sup_{x\in S^{n-1}\setminus\{0\}} \|Bx\|_2^2 = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Bx\|_2^2}{\|x\|_2^2} = \sup_{\|Bx\|_2=1} \frac{1}{\|x\|_2^2}$$
$$(3.8) \qquad = \sup_{\substack{x\in\mathbb{R}^n\setminus\{0\} \\ \sum_{k=l+1}^n \langle x,Y_k^*\rangle^2=1}} \frac{1}{\|x\|_2^2} = \sup_{\substack{x\in H_l\setminus\{0\} \\ \sum_{k=l+1}^n \langle x,Y_k^*\rangle^2=1}} \frac{1}{\|x\|_2^2}$$

where the last equality can be justified by considering the decomposition $x = x' + x'', x' \in H_l, x'' \in H_l^{\perp}$ with $\|Bx\|_2 = 1$. Since Bx', Bx have the same L_2 norm and x' has a smaller L_2 norm, the supremum must be achieved on H_l. Consider $x = A_{n-l}z, z \in S^{n-l-1}$, then

$$\langle x, Y_k^*\rangle^2 = \langle A_{n-l}z, P_l X_k^*\rangle^2 = \langle P_l A_{n-l}z, X_k^*\rangle^2 = \left\langle \sum_{k=l+1}^n z_k X_k, X_k^* \right\rangle^2 = z_k^2.$$

Thus, we have

$$\|B\|^2 = \sup_{z\in S^{n-l-1}} \frac{1}{\|A_{n-l}z\|_2^2} = s_{n+1-l}(A_{n-l})^{-2}.$$

By Theorem 3.2.6, we have

$$\mathbb{P}\left(\|B\| > \alpha\frac{\sqrt{n}}{l}\right) = \mathbb{P}\left(s_{n+1-l}(A_{n-l}) < \alpha^{-1}\frac{l}{\sqrt{n}}\right) \le (C\alpha^{-1})^l + \exp(-Cn).$$

Step 1.2.3. Concentration of $\|BX\|$.

By Lemma 3.2.3, B is independent to X. So we may condition on B such that $\|B\|_{\mathrm{HS}} > \alpha^{-1}\sqrt{\frac{n}{l}}$ and $\|B\| < \alpha\frac{\sqrt{n}}{l}$. By equation (3.3) and Theorem 3.2.10,

$$(3.9) \qquad \mathcal{L}(X_i, \beta) \le cp\beta, \text{ for all } i \in [n].$$

So, we may apply Theorem 3.2.9 with $\varepsilon = \frac{1}{2}$ and have

$$(3.10) \qquad \mathbb{P}\left(\|BX\|_2 \le \beta\alpha^{-1}\sqrt{\frac{n}{l}}\right) \le \mathbb{P}\left(\|BX\|_2 \le \beta\|B\|_{\mathrm{HS}}\right) \le (C\beta)^{cr(B)} \le (C\beta)^{\frac{l}{2\alpha^4}}.$$

Conclusion of step 1. Consider the events,

$$\begin{aligned} \mathcal{E}_1 &:= \left\{ A : \|P_l^{\perp} A_l\| > \alpha\sqrt{l} \right\} \\ \mathcal{E}_2 &:= \left\{ A : \|B\|_{\mathrm{HS}} < \alpha^{-1}\sqrt{\tfrac{n}{l}},\ \text{or } \|B\| > \alpha\tfrac{\sqrt{n}}{l} \right\}. \end{aligned} \tag{3.11}$$

We have shown

$$\begin{aligned} \mathbb{P}(\mathcal{E}_1) &\leq 2\exp(-C\alpha^2 l) \\ \mathbb{P}(\mathcal{E}_2) &\leq \exp(-C\alpha^2 l) + (C\alpha^{-1})^l + \exp(-Cn) \end{aligned} \tag{3.12}$$

By conditioning on $\mathcal{E}_2^c$ for all $y_i \in S^{l-1}$ and a vector U defined in (3.4), we have

$$\mathbb{P}\left(\|A^{-1}U\|_2 < \beta\alpha^{-1}\sqrt{\frac{n}{l}} \middle| \mathcal{E}_2^c \right) < (C\beta)^{\frac{l}{2\alpha^4}}.$$

Step 2: Preparation for the union bound argument.

Now let E_1(or in fact $\mathbb{R}^{l'}$) be an $l' := \lfloor \kappa l \rfloor$ dimensional coordinate subspace that is spanned by $e_1, \cdots, e_{l'}$. We consider an $\varepsilon-$net $\mathcal{N}_\varepsilon$ on $S^{l'-1}$(i.e. S_{E_1}), then $|\mathcal{N}_\varepsilon| \leq (3\varepsilon^{-1})^{l'}$. And for all $y_i \in \mathcal{N}_\varepsilon$, define

$$U_i = U(y_i) := X(y_i) - P_l X(y_i) := A_l y_i - P_l A_l y_i.$$

Step 2.1. $(A_l - P_l A_l)\mathcal{N}_\varepsilon$ **is a net on some ellipsoid.**

Let

$$E_2 := (A_l - P_l A_l)\mathbb{R}^{l'}, \quad S_2 := (A_l - P_l A_l)S^{l'-1}.$$

By step 1.1, with probability $1 - \exp(-C\alpha^2 l)$, $\|P_l^{\perp} A_l\| \leq \alpha\sqrt{l}$, i.e., $S_2 \subset \alpha\sqrt{l}B_2^{l'}$.

Also, consider any cap on S_2 of radius δ. Then if $P_l^{\perp} A_l \mathcal{N}_\varepsilon$ is not a $\delta\sqrt{l}$-net on S_2, there exists some $\delta\sqrt{l}$ cap that does not intersect $P_l^{\perp} A_l \mathcal{N}_\varepsilon$. This means that the pre-image of the cap does not intersect $\mathcal{N}_\varepsilon$. However, if $\|P_l^\perp A_l\| \le \alpha\sqrt{l}$, then the pre-image contains a cap of radius at least $\frac{\delta}{\alpha}$. Thus, for $\varepsilon = \frac{\delta}{\alpha}$, with probability $1 - \exp(-C\alpha^2 l)$, $P_l^\perp A_l \mathcal{N}_\varepsilon$ is a $\delta\sqrt{l}$-net on $S_2 \subset \alpha\sqrt{l}B_2^{E_2}$. We denote this $\delta\sqrt{l}$-net by $\mathcal{N}_\delta := P_l^\perp A_l \mathcal{N}_\varepsilon$

Step 2.2. Reduction of our objective.

Now, we want to show that for some small choice of κ, $\|A^{-1}U\|_2 \gtrsim \sqrt{\frac{n}{l}}$, for all $U \in S_2$ with high probability. If we can prove this, then together with step 2.1., we have $s_{l'}(A^{-1}) \gtrsim \frac{\sqrt{n}}{l}$ with high probability.

On the event that $P_l^\perp A_l \mathcal{N}_\varepsilon$ forms a $\delta\sqrt{l}$ net on S_2, we have for all $U \in S_2$, there exists some $U_i \in \mathcal{N}_\delta$ such that $\|U - U_i\|_2 \le \delta\sqrt{l}$, and

$$\|A^{-1}U\|_2 \ge \|A^{-1}U_i\|_2 - \|A^{-1}(U_i - U)\|_2.$$

For the first term, since we have

$$|\mathcal{N}_\delta| = |\mathcal{N}_\varepsilon| \le (3\varepsilon^{-1})^{l'} = \left(\frac{3\alpha}{\delta}\right)^{l'}$$

we obtain $\|A^{-1}U_i\|_2 \gtrsim \sqrt{\frac{n}{l}}$, for all U_i with high probability by choosing κ small.

To bound $\|A^{-1}(U_i - U)\|_2$ from above for $\|U_i - U\|_2 \lesssim \sqrt{l}$, we only have to prove

$$\left\|A^{-1}|_{E_2}\right\| \lesssim \frac{\sqrt{n}}{l}$$

with high probability.

Step 2.3. Upper bound for $\|A^{-1}|_{E_2}\|$.

Notice that

$$\left\|A^{-1}|_{E_2}\right\| \le \left\|A^{-1}P_l^\perp A_{l'}\right\| \cdot \left\|(P_l^\perp A_{l'})^{-1} : E_2 \to \mathbb{R}^{l'}\right\| = \frac{\left\|A^{-1}P_l^\perp A_{l'}\right\|}{s_{\min}(P_l^\perp A_{l'})}.$$

We only need to prove for κ small enough:

1. $s_{\min}(P_l^{\perp} A_{l'}) \gtrsim \sqrt{l}$ with high probability.
2. $\left\|A^{-1} P_l^{\perp} A_{l'}\right\| \lesssim \sqrt{\frac{n}{l}}$ with high probability.

Step 2.3.1. Lower bound of $s_{\min}(P_l^{\perp} A_{l'})$.

First, by Theorem 3.2.5,

$$\mathbb{P}\left(\|P_l^{\perp} A_{l'}\| \geq \alpha\sqrt{l}\right) \leq 2\exp(-C\alpha^2 l).$$

Next, consider a $\frac{1}{2\alpha^2}$-net $\mathcal{N}$ on $S^{l'-1}$, then $|\mathcal{N}| \leq (6\alpha^2)^{l'}$. And for all $y_i \in \mathcal{N}$, consider $A_{l'} y_i$ as a random vector. We use an elementary inequality $\mathcal{L}(Z, mt) \leq m\mathcal{L}(Z, t)$ which holds for any $m \in \mathbb{N}$. Consider $\beta < \alpha^{-1}$, then by equation (3.3) and Theorem 3.2.10,

$$\mathcal{L}((A_{l'} y_i)_j, \alpha^{-1}) \leq cp\beta \left\lfloor \frac{\alpha^{-1}}{\beta} + 1 \right\rfloor \leq 2cp\alpha^{-1}, \text{ for all } j \in [n]. \tag{3.13}$$

$P_l^{\perp}$ is decided by A_{n-l}, which is independent with $A_{l'} y_i$. So we may consider $P_l^{\perp}$ as a fixed matrix and apply Theorem 3.2.8 to obtain

$$\mathbb{P}\left(\|P_l^{\perp} A_{l'} y_i\|_2 \leq \alpha^{-1}\sqrt{l}\right) \leq (C\alpha^{-1})^l.$$

Let $y \in S^{l'-1}$ and choose $y_i \in \mathcal{N}$ with $\|y - y'\|_2 < \frac{1}{2\alpha^2}$. Conditioning on A such that $\|P_l^{\perp} A_{l'}\| \leq \alpha\sqrt{l}$, then we have

$$\begin{aligned} \|P_l^{\perp} A_{l'} y\|_2 &\geq \|P_l^{\perp} A_{l'} y_i\|_2 - \|P_l^{\perp} A_{l'}\| \|y - y_i\|_2 \\ &\geq \alpha^{-1}\sqrt{l} - \tfrac{1}{2\alpha^2}\alpha\sqrt{l} = \tfrac{1}{2\alpha}\sqrt{l} \end{aligned} \tag{3.14}$$

with probability $1 - (C\alpha^{-1})^l$.

Thus, with a standard union bound argument, we have

$$\mathbb{P}\left(s_{\min}\left(P_l^{\perp} A_{l'}\right) \geq \frac{1}{2\alpha}\sqrt{l}\right) \geq 1 - 2\exp(-C\alpha^2 l) - \left(6\alpha^2\right)^{l'} (C\alpha^{-1})^l.$$

Step 2.3.2. Upper bound of $\left\|A^{-1}P_l^{\perp}A_{l'}\right\|$.

Recall for $y \in S^{l'-1}$

$$\|A^{-1}P_l^{\perp}A_{l'}y\|_2^2 = \|y\|_2^2 + \|A^{-1}P_lA_ly\|_2^2 = 1 + \|BA_{l'}y\|_2^2$$

where

$$\|B\|_{\mathrm{HS}}^2 = \sum_{k=l+1}^{n} \mathrm{dist}(X_k, H_{l,k})^{-2}.$$

Thus, we only need to show $\|BA_{l'}\| \lesssim \sqrt{\frac{n}{l}}$. We will prove this using Theorem 3.2.5. To apply Theorem 3.2.5, we employ an argument that is presented in [45, Section 5.4.1. and Section 13.2.]. This argument provides an upper estimate of $\|B\|_{\mathrm{HS}}$.

Recall that the weak L^p norm of a random variable Z is defined as

$$\|Z\|_{p,\infty} = \sup_{t>0} t \cdot \left(\mathbb{P}\left\{|Z| > t\right\}\right)^{1/p}.$$

Although it is not a norm, it is equivalent to a norm if $p > 1$. In particular, the weak triangle inequality holds:

$$\left\|\sum_i Z_i\right\|_{p,\infty} \leq C(p) \sum_i \|Z_i\|_{p,\infty}$$

where $C(p)$ is bounded above by an absolute constant for $p \geq 2$, see [57], Theorem 3.21.

Now by Theorem 3.2.7, for any $t > 0$,

$$\mathbb{P}\left\{\mathrm{dist}(X_k, H_{l,k}) \leq t\sqrt{l}\right\} \leq (Ct)^l + \exp(-Cn).$$

Define

$$W_k := \min\left(\mathrm{dist}(X_k, H_{l,k})^{-2}, (t_0\sqrt{l})^{-2}\right)$$

where $t_0 = \frac{C_0 l}{n}$ and C_0 is a small constant depending only on K. Then we have

$$\begin{aligned}
\|W_k\|_{l/2,\infty} &= \sup_{t>0} t \cdot \left(\mathbb{P}\left\{W_k > t\right\}\right)^{2/l} \\
&= \sup_{t>0} t^{-2} l^{-1} \cdot \left(\mathbb{P}\left\{W_k^{-\frac{1}{2}} < t\sqrt{l}\right\}\right)^{2/l} \\
&= \sup_{t>t_0} t^{-2} l^{-1} \cdot \left(\mathbb{P}\left\{\operatorname{dist}(X_k, H_{l,k}) < t\sqrt{l}\right\}\right)^{2/l} \\
&\le \tfrac{C}{l} + \tfrac{1}{t_0^2 l} \exp\left(-C\tfrac{n}{l}\right) \\
&\le \tfrac{C}{l} + \tfrac{1}{l}\left(\tfrac{n^2}{C_0^2 l^2} \exp\left(-C\tfrac{n}{l}\right)\right) \le \tfrac{C}{l}.
\end{aligned} \tag{3.15}$$

this implies

$$\left\|\sum_{k=l+1}^{n} W_k\right\|_{l/2,\infty} \le \frac{C(n-l)}{l} \le \frac{Cn}{l}.$$

Thus, we have

$$\mathbb{P}\left\{\sum_{k=l+1}^{n} W_k > t^2 \frac{n}{l}\right\} \le (Ct^{-1})^l.$$

On the other hand,

$$\begin{aligned}
&\mathbb{P}\left(\text{there exists } k,\ W_k \neq \operatorname{dist}(X_k, H_{l,k})^{-2}\right) \\
\le\ & \sum_{k=l+1}^{n} \mathbb{P}\left\{\operatorname{dist}(X_k, H_{l,k}) \le t_0\sqrt{l}\right\} \\
\le\ & (n-l)\left(\left(\tfrac{CC_0 l}{n}\right)^l + \exp(-Cn)\right) \\
\le\ & \exp(-Cl).
\end{aligned} \tag{3.16}$$

So we have

$$\begin{aligned}
&\mathbb{P}\left\{\|B\|_{\mathrm{HS}} > t\sqrt{\frac{n}{l}}\right\} \\
\le\ & \mathbb{P}\left\{\sum_{k=l+1}^{n} W_k > t^2 \frac{n}{l}\right\} + \mathbb{P}\left(\text{there exists } k,\ W_k \neq \operatorname{dist}(X_k, H_{l,k})^{-2}\right) \\
\le\ & (Ct^{-1})^l + \exp(-Cl).
\end{aligned} \tag{3.17}$$

Now, denote

$$\mathcal{E}_2' := \mathcal{E}_2 \cup \left\{A : \|B\|_{\mathrm{HS}} > \alpha\sqrt{\frac{n}{l}}\right\}.$$

Applying Theorem 3.2.5 with $D = B, G = A_{l'}, Cs = \frac{1}{2}\alpha, Ct = \frac{1}{2}\alpha\sqrt{\frac{l}{l'}}$, we have for α large enough,

$$\begin{aligned}
& \mathbb{P}\left\{\|BA_{l'}\| > \alpha^2\sqrt{\frac{n}{l}}\middle|(\mathcal{E}_2')^c\right\} \\
\leq\ & \mathbb{P}\left\{\|BA_{l'}\| > \frac{1}{2}\alpha\|B\|_{\mathrm{HS}} + \left(\frac{1}{2}\alpha\sqrt{\frac{l}{l'}}\right)\sqrt{l'}\|B\|\middle|(\mathcal{E}_2')^c\right\} \\
\leq\ & 2\exp\left(-C\alpha^2 l\left(\alpha^{-4}+1\right)\right) \leq 2\exp\left(-C\alpha^2 l\right).
\end{aligned} \tag{3.18}$$

So we have

$$\mathbb{P}\left(\|A^{-1}P_l^{\perp}A_{l'}\| \geq \alpha^2\sqrt{\frac{n}{l}}\middle|(\mathcal{E}_2')^c\right) \leq 2\exp\left(-C\alpha^2 l\right).$$

Conclusion of Step 2.

Denote

$$\begin{aligned}
\mathcal{E}_3 &:= \left\{A : \|A^{-1}|_{E_2}\| \geq 2\alpha^3\tfrac{\sqrt{n}}{l}\right\} \cup \mathcal{E}_1 \cup \mathcal{E}_2', \\
\mathcal{E}_4 &:= \left\{A : \text{there exists } y_i \in \mathcal{N}_\varepsilon \text{ such that } \|A^{-1}U(y_i)\|_2 \leq \beta\alpha^{-1}\sqrt{\frac{n}{l}}\right\}.
\end{aligned} \tag{3.19}$$

Then we have

$$\mathbb{P}\left(\mathcal{E}_3\right) \leq \left(6\alpha^2\right)^{l'}\left(C\alpha^{-1}\right)^l + 4\exp(-C\alpha^2 l) + \exp(-Cl) + \mathbb{P}\left(\mathcal{E}_2'\right) + \mathbb{P}\left(\mathcal{E}_1\right).$$

Since $|N_\varepsilon| \leq \left(\frac{\alpha\sqrt{l}}{\delta}\right)^{l'-1}$ as we discussed in step 2.1,

$$\mathbb{P}\left(\mathcal{E}_4 \mid \mathcal{E}_2^c\right) \leq \left(\frac{3\alpha\sqrt{l}}{\delta}\right)^{l'}(C\beta)^{\frac{l}{2\alpha^4}}$$

Step 3. The union bound argument.

Denote

$$\begin{aligned}
\mathcal{E} := \ & \Big\{A : \text{there exists } y \in S^{l'-1}, \text{ such that } \|A^{-1}U(y)\|_2 \leq \tfrac{\beta}{2\alpha}\sqrt{\tfrac{n}{l}}, \\
& \text{or } \|U(y)\|_2 \geq \alpha\sqrt{l}\Big\}.
\end{aligned} \tag{3.20}$$

Choose δ such that $2\alpha^4\delta = \frac{\beta}{2}$. Let $y \in S^{l'-1}$ and choose $y_i \in \mathcal{N}_\varepsilon$ with $\|y - y'\|_2 < \delta$. If $A \notin \mathcal{E}_3 \cup \mathcal{E}_4$, then $\|U(y)\| \leq \alpha\sqrt{l}$ and

$$
\begin{aligned}
\|A^{-1}U(y)\|_2 &\geq \|A^{-1}U(y)\|_2 - \|A^{-1}(U(y_i) - U(y))\|_2 \\
&\geq \beta\alpha^{-1}\sqrt{\tfrac{n}{l}} - 2\alpha^3\tfrac{\sqrt{n}}{l} \cdot \delta \\
&\geq \tfrac{\beta}{2\alpha}\sqrt{\tfrac{n}{l}}.
\end{aligned} \tag{3.21}
$$

Thus, we have $\mathcal{E} \subset \mathcal{E}_3 \cup \mathcal{E}_4$. On the other hand,

$$
\begin{aligned}
A \in \mathcal{E}^c &\Rightarrow s_{l'}(A^{-1}) \geq \tfrac{\beta}{2\alpha^2}\tfrac{\sqrt{n}}{l} \Rightarrow s_{n+1-l'}(A) \leq \tfrac{2\beta}{\alpha^2}\tfrac{l}{\sqrt{n}} \\
&\Rightarrow s_{n+1-l'}(A) \leq \tfrac{4\beta}{\kappa\alpha^2}\tfrac{l'}{\sqrt{n}}, \text{ for all } l' < \tfrac{\kappa\tilde{c}n}{2}.
\end{aligned} \tag{3.22}
$$

The $\frac{\tilde{c}}{2}$ in $l' < \frac{\tilde{c}\kappa n}{2}$ comes from the requirement that $l \leq \frac{\tilde{c}n}{2}$.

Now let $\beta = \exp(-\alpha^5)$, then $\delta = \frac{1}{4}\alpha^{-4}\exp(-\alpha^5)$. Choose α to be a big enough constant, then

$$
\begin{aligned}
&\mathbb{P}(\mathcal{E}) \\
\leq\ &\mathbb{P}(\mathcal{E}_3) + \mathbb{P}(\mathcal{E}_4) \\
\leq\ &\left(\tfrac{3\alpha}{\delta}\right)^{l'}(C\beta)^{\frac{l}{2\alpha^4}} + 7\exp(-C\alpha^2 l) + 2(C\alpha^{-1})^l + \exp(-Cl) + \exp(-Cn) \\
&+ (6\alpha^2)^{l'}(C\alpha^{-1})^l \\
\leq\ &\left((12\alpha^5\exp(\alpha^5))^\kappa C\exp(-\tfrac{1}{2}\alpha)\right)^l + (C\alpha^{2\kappa-1})^l + 9(C\alpha^{-1})^l + \exp(-Cl).
\end{aligned} \tag{3.23}
$$

Replace l', l by $l, \kappa^{-1}l$, then for a sufficiently small κ depending on α, and $l < \frac{\kappa\tilde{c}n}{2}$, we have

$$
\begin{aligned}
&\mathbb{P}\left(s_{n+1-l}(A) \geq \tfrac{4\beta}{\kappa\alpha^2}\tfrac{l}{\sqrt{n}}\right) \\
\leq\ &\left(\exp(-\tfrac{1}{4}\alpha)\right)^{\kappa^{-1}l} + \left(C\alpha^{-\frac{1}{2}}\right)^{\kappa^{-1}l} + 9(C\alpha^{-1})^{\kappa^{-1}l} + \exp(-Cl) \\
\leq\ &(C\alpha^{-\frac{1}{2}})^{\kappa^{-1}l} + \exp(-Cl).
\end{aligned} \tag{3.24}
$$

Choosing a sufficiently large α, we show that there exist C_1, C_2, C_3 depending only on K, p, such that

$$
\mathbb{P}\left(s_{n+1-l}(A) \geq C_1\frac{l}{\sqrt{n}}\right) \leq \exp(-C_2 l)
$$

for all $l \leq C_3 n$. For $l > C_3 n$, the above bound follows from the estimate for $s_1(A)$. So we have for all $1 \leq l \leq n$,

$$\mathbb{P}\left(s_{n+1-l}(A) \geq C_1 \frac{l}{\sqrt{n}}\right) \leq \exp(-C_2 l).$$

□

Remark 3.3.1. As the proof demonstrates, Assumption 3.1.3 is satisfied with $s = \beta$ which only depends on p and K.

Remark 3.3.2. The only place we used the non-degeneracy condition is in the application of 3.2.9. We expect that the same result holds without the concentration function condition. To remove that condition, the application of Theorem 3.2.9 on BX must be replaced by showing matrix B does not have a good arithmetic structure with high probability (for arithmetic structure of random matrices and its application, see [36, 44, 47, 49, 45]).

3.4 Deduction of Corollary 3.1.10 and 3.1.11

Both Theorem 3.1.10 and 3.1.11 are direct corollaries of Theorem 3.1.6.

Proof of Theorem 3.1.10. Construct an $n \times n$ random matrix J such that its first $n-k$ columns are matrix A and the rest are i.i.d. entries with the same distribution as $A_{i,j}$. Then, by Theorem 3.1.6, for all $t > 0$ and k between l and n, $s_{n+1-l}(J) \leq \frac{C_1 tl}{\sqrt{n}}$, with probability $1 - \exp(-C_2 tl)$, where C_1, C_2 are constants that depend only on K, p. This implies, with the same probability, there exists an l-dimensional subspace E such that $\|Jy\|_2 \leq \frac{C_1 tl}{\sqrt{n}}$ for all $y \in S_E$.

Let $F := \operatorname{span}\{e_1, \cdots, e_{n-k}\}$, then $\|Jy\|_2 \leq \frac{C_1 tl}{\sqrt{n}}$, for all $y \in S_{E\cap F}$ with probability $1 - \exp(-C_2 tl)$. This implies

$$\mathbb{P}_J \left\{ s_{n+1-l}(A) \geq \frac{C_1 tl}{\sqrt{n}} \right\} \leq \exp(-C_2 tl)$$

Then we only need to notice that the above event is independent of the last k rows of J; thus the probability is also with respect to A. □

To prove Theorem 3.1.11, in addition to applying Theorem 3.1.6, we only need Theorem 3.2.6 to give a lower bound.

Proof of Theorem 3.1.11. For the lower bound, denote J as the matrix of the first $n-l$ rows of A. Then we have by Theorem 3.2.6

(3.25)
$$\begin{aligned}\mathbb{P}\left\{s_{n+1-l}(A) < \tfrac{C_1 l}{\sqrt{n}}\right\} &\leq \mathbb{P}\left\{s_{n+1-l}(J) < \tfrac{C_1 l}{\sqrt{n}}\right\} \\ &\leq (CC_1)^l + \exp(-Cn) \leq \tfrac{1}{2}\exp(-\tfrac{1}{2}C_3 l) + \exp(-Cn) \\ &\leq \tfrac{1}{2}\exp(-C_3 l)\end{aligned}$$

with some small constant C_3. The upper bound follows directly with a large t in Theorem 3.1.6.

□

Note that Theorem 3.1.11 is a generalization of Theorem 1.3 in [59]. Theorem 3.1.12 can be proved in the same way as Theorem 3.1.11.

CHAPTER IV

Investigate invertibility of sparse symmetric matrix

In this chapter, we investigate the invertibility of sparse symmetric matrices. We will show that an $n \times n$ sparse symmetric random matrix A with $A_{ij} = \delta_{ij}\xi_{ij}$ is invertible with high probability. Here, δ_{ij}s, $i \geq j$ are i.i.d. Bernoulli random variables with $\mathbb{P}(\delta_{ij} = 1) = p \geq n^{-c}$, $\xi_{ij}, i \geq j$ are i.i.d. random variables with mean 0, variance 1 and finite fourth moment M_4, and c is constant depending on M_4. More precisely,

$$s_{\min}(A) > \varepsilon\sqrt{\frac{p}{n}}.$$

with high probability.

Outline of the chapter.

- In Section 4.1, we introduce setup of the problem and present our main results.
- In Section 4.2, we recall the necessary concepts and some technical lemmas. We also recall the method of separating compressible and incompressible vectors (see [47]) in Section 4.2.
- In Section 4.3, we bound $\|Ax\|_2$ over compressible vectors. The method we used to bound the infimum over compressible vectors for sparse matrix is invented in Section 3 in [6].
- In Section 4.4, 4.5 and 4.6 we bound $\|Ax\|_2$ over incompressible vectors. In Sec

4.4 we recall the definition of LCD and regularized LCD and reduce the infimum to a distance problem which can be written as a quadratic form, see [73]. In Section 4.5, we prove the structure theorem for large LCD vectors which is an analog of Theorem 7.1 in [36]. In Section 4.6, we estimate the distance problem using the decoupling technique in [36].

- In Section 4.7, we combine the estimate for compressible and incompressible part to prove our main theorem.
- In Section 4.8, we prove an upper bound of the spectral norm for sparse symmetric sub-gaussian matrix which is an analog of Theorem 1.7 in [6].

4.1 Introduction

The quantitative smallest singular value distribution of random matrix is an important and interesting topic in non-asymptotic random matrix theory, and as we discussed, it was intensively studied in the past decade [36, 67, 68, 47, 6, 43, 48, 73]. However, very few are know about sparse matrices until very recently, Basak and Rudelson proved that for a non-Hermitian i.i.d. sparse matrix [6],

$$\begin{aligned} &\mathbb{P}\left\{ s_{\min}(A) \geq C\varepsilon \exp\left(-c\frac{\log(1/p_n)}{\log(np_n)}\right)\sqrt{\frac{p_n}{n}} \bigcap \|A\| \leq C\sqrt{pn} \right\} \\ \leq\ & \varepsilon + \exp(-cnp_n) \end{aligned} \tag{4.1}$$

where $\mathbb{P}(a_{ij} \neq 0) = p_n$. One may notice that for $p_n \geq n^{-c}$, where $0 < c < 1$, the above result of Basak and Rudelson implies an upper bound on condition number. That is to say

$$\sigma(A_n) := \frac{s_{\max}(A)}{s_{\min}(A)} \leq n$$

with high probability. This generalized the optimal upper bound on condition number for non-sparse random matrices. So it is a nature question to ask, whether one can use the similar technique to develop the invertibility for sparse symmetric matrices.

This work is motivated by the above result of non-Hermitian sparse matrices of A. Basak and M. Rudelson and the paper of R. Vershynin for non-sparse symmetric matrices, see [73]. Without special notice, we always assume the following for our random matrix A_n:

Assumption 4.1.1. *$A_n = \{a_{i,j}\}_{i,j=1}^n$ is an $n \times n$ symmetric random matrix with i.i.d entries on the upper triangular part, and $a_{i,j} = \xi_{ij}\delta_{ij}$. Here δ_{ij}s are i.i.d. Bernoulli random variables with $\mathbb{P}(\delta_{ij} = 1) = p_n$. ξ_{ij}s are i.i.d. random variables with mean zero, variance 1 and fourth moment bounded by M_4^4.*

Remark 4.1.2. The dependence of c_p on M_4 is tracked in the the proof.

Remark 4.1.3. Through out the chapter, we are going to call p_n the sparsity level of A.

Remark 4.1.4. For the ease of writing, hereafter, we will often drop the sub-script in A_n, p_n, write A, p instead. But please have it in mind that the sparsity level will depend on n.

Our proof will also use an upper bound for operator norm. For convenience, throughout the proof, we denote $\mathcal{E}_{op}$ as the event that $\|A\| \leq C_{op}\sqrt{pn}$.

Our main theorem is the following:

Theorem 4.1.5. (Smallest singular value for sparse symmetric matrices.) *For A satisfies Assumption 4.1.1 and $p \geq n^{-c_p}$, where c_p is a constant depending only on M_4, C_{op}, one has*

$$\mathbb{P}\left(s_n(A) \leq \varepsilon\sqrt{\frac{p}{n}} \cap \mathcal{E}_{op}\right) \leq C_{4.1.5}\varepsilon^{1/9} + e^{-n^{c_{4.1.5}}}.$$

Here $C_{4.1.5}, c_{4.1.5} > 0$ depend only on M_4 and C_{op}.

Remark 4.1.6. Theorem 4.1.5 can be also generalized to the case A is replaced by $A+D$ where D is a diagonal matrix and $\|D\| = O(\sqrt{pn})$. For simplicity, we do not include the proof in this chapter, see [6] for more details.

Recall that for a random variable Z on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The sub-gaussin norm or ψ_2-norm of Z is defined as

$$\|Z\|_{\psi_2} := \inf\left\{\lambda > 0 : \mathbb{E}\exp\left(\frac{|Z|}{\lambda}\right)^2 \leq 2\right\}.$$

A random variable is called sub-gaussian if it has finite sub-gaussian norm. For properties of sub-gaussian random variables, see [44]. For sparse symmetric matrix with ξ_{ij}s are sub-gaussian, we have the following result about spectral norm.

Theorem 4.1.7. *There exists $C'_{4.1.7} \geq 1$ such that the following holds. Let $n \in \mathcal{N}$ and $p \in (0,1]$ be such that $p \geq C'_{4.1.7}\frac{\log n}{n}$. Let A_n be a random matrix as in Assumption 4.1.1. Moreover, we require ξ_{ij} to be sub-gaussian random variables in the assumption. Then there exist positive constants $C_{4.1.7}, c_{4.1.7}$ depending on the sub-gaussian norm of ξ_{ij}, such that*

$$\mathbb{P}\left(\|A_n\| \geq C_{4.1.7}\sqrt{np_n}\right) \leq \exp(-c_{4.1.7}np).$$

Theorem 4.1.5 and 4.1.7 together give us the following result:

Corollary 4.1.8. (Smallest singular value for sparse symmetric sub-gaussian matrices.) *For A as in Theorem 4.1.5 and moreover ξ_{ij}s are sub-gaussian random variables, one has*

$$\mathbb{P}\left(s_n(A) \leq \varepsilon\sqrt{\frac{p}{n}}\right) \leq C_{4.1.8}\varepsilon^{1/9} + e^{-n^{c_{4.1.8}}} + \exp(-c'_{4.1.8}np).$$

Here $C_{4.1.8}, c_{4.1.8}, c'_{4.1.8} > 0$ depend only on the sub-gaussian norm.

Remark 4.1.9. For sparse sub-gaussian matrix, above theorems directly yield a bound on the condition number that $n \gtrsim \sigma(A) := \frac{s_{max}(A)}{s_{min}(A)}$ with high probability.

4.2 Notations and Preliminaries

We first explain our notations in this chapter. Through out the chapter c, C, c_0, c_1, $c', \cdots$ denote absolute constants or constants that are going to be used only locally. These constants are different in proofs of different lemmas or theorems. Constants with double indices, triple indices or letter indices are global constants. These constants are uniform through out the chapter and we will keep track of these constants through out the chapter, for example $c_{4.1}, c'_{4.3.2}, c_p$.

First, recall that

$$s_{\min}(A_n) = \inf_{x \in S^{n-1}} \|A_n x\| .$$

Thus, to prove Theorem 4.1.5, we need to find a lower bound on the infimum. For dense matrices, this can be done via decomposing the unit sphere into compressible and incompressible vectors, and obtaining necessary bound on the infimum on both of these parts, see [47, 48]. To carry out the argument for sparse matrices, Basak and Rudelson introduced another class of vectors which they called dominated vectors, see [6].

Below, we state necessary concepts, starting with the definition of compressible and incompressible vectors, see [47].

Definition 4.2.1. Fix $m < n$. The set of $m-$sparse vectors is given by

$$\text{Sparse}(m) := \{x \in \mathbb{R}^n || \text{supp}(x)| \leq m\}$$

where $|S|$ denotes the cardinality of a set S. Furthermore, for any $\delta > 0$, the vectors which are δ-close to m-sparse vectors in Euclidean norm, are called $(m, \delta)-$compressible vectors. The set of all such vectors, will be denoted by $\text{Comp}(m, \delta)$. Thus

$$\text{Comp}(m, \delta) := \left\{x \in S^{n-1} | \exists y \in \text{Sparse}(m) \text{ such that } \|x - y\|_2 \leq \delta\right\} .$$

The vectors in S^{n-1} which are not compressible, are defined to be incompressible, and the set of all incompressible vectors is denoted as $\mathrm{Incomp}(m, \delta)$.

The dominated vectors are also close to sparse vectors, but in a different sense, see [6].

Definition 4.2.2. For any $x \in S^{n-1}$, let $\pi_x : [n] \to [n]$ be a permutation which arranges the absolute values of the coordinates of x in non-increasing order. For $1 \leq m \leq m' \leq n$, denote by $x_{[m:m']} \in \mathbb{R}^n$ the vector with coordinates

$$x_{[m:m']}(j) = x(j)1_{[m:m']}(\pi_x(j)).$$

In other words, we include in $x_{[m:m']}$ the coordinates of x which take places from m to m' in the non-increasing rearrangement. For $\alpha < 1$ and $m \leq n$ define the set of vectors with dominated tail as follows:

$$\mathrm{Dom}(m, \alpha) := \left\{ x \in S^{n-1} | \left\| x_{[m+1:n]} \right\|_2 \leq \alpha \sqrt{m} \| x_{[m+1:n]} \|_\infty \right\}.$$

One may notice that for $m-$sparse vectors $x_{[m+1:n]} = 0$, thus we have $\mathrm{Sparse}(m) \cap S^{n-1} \subset \mathrm{Dom}(m, \alpha)$.

Theorem 4.1.5 will be proved by first bounding the infimum over compressible and dominated vectors, and then the same for the incompressible vectors. As in [6], the first step is to control the infimum for sparse vectors. To this end, we need some estimates on the small ball probability. For the estimates, recall the definition of Levy concentration function.

Definition 4.2.3. Let Z be random variable in $\mathbb{R}^n$. For every $\varepsilon > 0$, the Levy's concentration function of Z is defined as

$$\mathcal{L}(Z, \varepsilon) := \sup_{u \in \mathbb{R}^n} \mathbb{P}\left(\|Z - u\|_2 \leq \varepsilon \right),$$

where $\| \cdot \|_2$ denotes the Euclidean norm.

The following Paley-Zygmund inequality is useful on estimating Levy's concerntration function:

Lemma 4.2.4. *If ξ is a random variable with finite variance and $0 \le \theta \le 1$, then*

$$\mathbb{P}(\xi > \theta \mathbb{E}\xi) \ge \frac{(\mathbb{E}\xi - \theta \mathbb{E}\xi)^2}{\mathbb{E}\xi^2}.$$

Remark 4.2.5. We note that there exist $\delta_0, \varepsilon_0' \in (0,1)$, such that for any $\varepsilon < \varepsilon_0'$, $\mathcal{L}(\xi\delta, \varepsilon) \le 1 - \delta_0 p$, where ξ is a random variable with unit variance and finite fourth moment, and δ is a Ber(p) random variable, independent of each other (for more details see [[73], Lemma 3.3]).

For application of Levy's concerntration function, the following tensorization lemma can be very useful to transfer bounds for the Levy concentration function from random variables to random vectors.

Lemma 4.2.6. (Tensorization, Lemma 3.4 in [73]). *Let $X = (X_1, \cdots, X_n)$ be a random vector in $\mathbb{R}^n$ with independent coordinates X_k.*

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

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

Then

$$\mathcal{L}(X, \varepsilon\sqrt{n}) \le (CL\varepsilon)^n \text{ for all } \varepsilon \ge \varepsilon_0,$$

where C is an absolute constant .

2. *Suppose there exists number $\varepsilon > 0$ and $q \in (0,1)$ such that*

$$\mathcal{L}(X_k, \varepsilon) \le q \text{ and all } k.$$

There exists numbers $\varepsilon_1 = \varepsilon_1(\varepsilon, q) > 0$ and $q_1 = q_1(\varepsilon, q) \in (0,1)$ such that

$$\mathcal{L}(X, \varepsilon_1\sqrt{n}) \le q_1^n.$$

Remark 4.2.7. A useful equivalent form of Lemma 4.2.6 (part 1) is the following. Suppose there exist numbers $a, b \geq 0$ such that

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

Then

$$\mathcal{L}(X, \varepsilon\sqrt{n}) \leq (C(a\varepsilon + b))^n \text{ for all } \varepsilon \geq 0,$$

Where C is an absolute constant

4.3 Invertibility over compressible vectors

The main theorem in this section is the following:

Theorem 4.3.1. *Consider A satisfies 4.1.1 and $p \geq (1/4)n^{-1/3}$, then there exist $c_{4.3.1}$, $c'_{4.3.1}$, $c''_{4.3.1}$, $c'''_{4.3.1}$, $C_{4.3.1} > 0$ depending only on C_{op}, M_4, such that for any $p^{-1} \leq M \leq c'''_{4.3.1}n$, we have for any $u \in \mathbb{R}^n$*

$$\mathbb{P}\Bigg(\exists x \in \mathrm{Dom}(M, C_{4.3.1}^{-1}) \cup \mathrm{Comp}(M, c'_{4.3.1}) \\ \|Ax - u\|_2 \leq c''_{4.3.1}\sqrt{np} \text{ and } \|A\| \leq C_{op}\sqrt{pn} \Bigg) \leq \exp(-c_{4.3.1}pn). \tag{4.2}$$

Remark 4.3.2. Although for the purpose our our proof we do not need to bound the dominated vectors close to moderately sparse, we still work on it due to it's own interest for future work.

Remark 4.3.3. Theorem 4.3.1 can be extended to the sparsity level of n^{-1+c} for arbitrary c following our framework. The reason we can't not reach n^{-1+c} in Theorem 4.1.5 is due to incompressible part.

A direct proof following the paper of Vershynin [73] won't work due to the sparsity phenomenon found in the sparse paper of Basak and Rudelson, see [6]. So we need to adapt the technique for sparse matrix and deal with the symmetricity at the same time. The proof splits into two steps as in [6]. First, we consider vectors which are close to $(1/8p)$-sparse. The small ball probability estimate is not strong enough for such vectors. This forces us to use the method designed for sparse matrices in [6]. We prove Lemma 4.3.4 which a generalized version of Lemma 3.2 in [6] for symmetric matrix. Lemma 4.3.4 allows us to control $\|Ax\|_2$ for very sparse vectors without cancellation and ε-net argument. For more intuition of the technique for vectors close to very sparse, see Section 3.1 in [6]. Later, one needs to improve these estimates for vectors which are close to M-sparse. For such moderately sparse vectors, a better control of the Levy concentration function is available. After we obtain such estimates for sparse vectors, we extend them to compressible vectors using the standard ε-net and the union bound argument.

4.3.1 Vectors close to very sparse

Now we state a combinatorial lemma similar to Lemma 3.2 in [6] but designed for symmetric matrices. The proof is a variant of Lemma 3.2 in [6] to deal with the symetricity.

Lemma 4.3.4. *Consider A_n be an $n \times n$ random matrix with $a_{ij} = \delta_{ij}\xi_{ij}$ for $i \leq j$ and $a_{ji} = \pm a_{ij}$ for $i > j$. Here δ_{ij} are i.i.d. Bernoulli random variables with $\mathbb{P}(\delta_{ij} = 1) = p$, where $p \geq C \log n/n$. And ξ_{ij} are independent mean zero random variables with* $\min\{\mathbb{P}(\xi_{i,j} \geq c_1), \mathbb{P}(\xi_{i,j} \leq -c_1)\} \geq c_0$. *For $\kappa \in \mathbb{N}$, $s \in \{-1, 1\}^\kappa$ and for $J, J' \subset [n]$, let $\mathcal{A}_c^{J,J',s}$ denote the event that satisfies the following conditions:*

(i) There are at least $c\kappa pn$ rows of the matrix have non-zero entry in the columns corresponding to J, and all zero entries in the columns corresponding to J'.

(ii) Denote $I^{J,J'}$ be the indices of those $c\kappa pn$ rows. Then $I^{J,J'} \cap (J \cup J') = \emptyset$.

(iii) Suppose $i \in I^{J,J'}$ and $j_i \in J$ is the non-zero entry as in (i), then $|a_{ij_i}| \geq c_1$

and $\operatorname{sign}(a_{ij_i}) = s_{j_i}$.

Denote

$$m = m(\kappa) := \kappa\sqrt{pn} \wedge \frac{1}{8p}.$$

Then, there exist absolute constants $0 < c_{4.3.4}, c'_{4.3.4} < \infty$ *depending only on* c_0, c_1, *such that*

$$\mathbb{P}\left(\bigcap_{\kappa \le (8p\sqrt{pn})^{-1} \vee 1} \bigcap_{s \in \{-1,1\}^\kappa} \bigcap_{J \in \binom{[n]}{\kappa}} \bigcap_{J' \in \binom{[n]}{m}, J \cap J' = \emptyset} \mathcal{A}^{J,J',s}_{c'_{4.3.4}} \right) \ge 1 - \exp(-c_{4.3.4} pn).$$

Proof. The proof is done by bounding the complement event. It is similar to Lemma 3.2 in [6] but we need to take care of the sign and symmetricity.

Without loss of generality, we assume $c_1 = 1$ and only need to consider $s = (1, \cdots, 1)$. For different choice of signs, the argument is identical. Fix $\kappa \le (8p\sqrt{pn})^{-1} \vee 1$ and $J \in \binom{[n]}{\kappa}$, $J' \in \binom{[n]}{m}$. Let

$$I^1(J, J') := \Big\{ i \in [n] \backslash (J \cup J') : a_{ij_i} \ge 1 \text{ for some } j_i \in J, \text{ and } a_{ij_i} = 0 \text{ for all } j \in J \backslash j_i \Big\}. \tag{4.3}$$

Similarly, we define

$$I^0(J, J') := \{ i \in [n] \backslash (J \cup J') : a_{ij} = 0 \text{ for all } j \in J' \}.$$

Here we require $(I^1 \cup I^0) \cap (J \cup J') = \emptyset$ so that we can get rid of symmetricity and achieve independence. On the other hand, since $m, \kappa \ll n$, this won't harm our probability bound.

To prove our desired result, we need to show the cardinality of $I^1(J, J')$ is at least $c\kappa pn$ with high probability for some constant c firstly. Then we can apply Chernoff's inequality to prove that $|I'(J, J') \cap I^0(J, J')|$ is large with high probability. Finally, we take union bound over all different choices of J, J', s.

We start with obtaining a lower bound on $\mathbb{P}(i \in I^1(J, J'))$ for every $i \in [n]$. By our assumption on a_{ij}, we have for any $i \notin J \cup J'$,

$$\mathbb{P}(i \in I^1(J, J')) \geq c_0|J|p(1-p)^{|J|-1} \geq c_0\kappa p(1-\kappa p) \geq \frac{c_0\kappa p}{2}.$$

Therefore, by Chernoff's inequality and the fact that $\kappa, m \ll n$, we have

$$\mathbb{P}(|I^1(J, J')| \leq \frac{c_0\kappa pn}{4}) \leq \exp(-c_1 pn).$$

Next, we fix a set $J' \in \binom{[n]}{m}$, for any $i \in [n]\backslash(J \cup J')$, we have that

$$\mathbb{P}(i \in I^0(J, J')) = (1-p)^{|J'|} \geq 1 - p|J'| = 1 - pm \geq \frac{3}{4}.$$

Thus, for any given $I \subset [n]$, the random variable $I\backslash I^0(J, J')$ can be represented as the sum of independent Bernoulli variables taking value 1 with probability less than pm. Also, note that

$$\mathbb{E}|I\backslash I^0(J, J')| \leq pm|I| \leq \frac{|I|}{4}$$

by the assumption on κ and m. Now, use Chernoff's inequality again, we have

$$\mathbb{P}\left(|I\backslash I^0(J, J')| \geq \frac{|I|}{2}\right) \leq \exp\left(-\frac{|I|}{16}\log\left(\frac{1}{4pm}\right)\right).$$

So for any $I \subset [n]$ such that $|I| \geq \frac{c_0\kappa np}{4}$, we can deduce that for any $J \in \binom{[n]}{\kappa}$,

$$\begin{aligned}
& \mathbb{P}\left(\exists J' \in \binom{[n]}{m} \text{ such that } |I^0(J, J') \cap I| \leq \frac{c_0\kappa pn}{8}\right) \\
\leq \quad & \sum_{J' \in \binom{[n]}{m}} \mathbb{P}(|I\backslash I^0(J, J')| \geq \frac{1}{2}|I|) \\
\leq \quad & \binom{n}{m}\exp\left(-\frac{|I|}{16}\log\left(\frac{1}{4pm}\right)\right) \leq \exp(-\kappa pnU),
\end{aligned} \tag{4.4}$$

where

$$U := \frac{c_0}{64}\log\left(\frac{1}{4pm}\right) - \frac{m}{\kappa pn}\log\left(\frac{en}{m}\right).$$

Here we have $U \geq \frac{c_0}{100}$ (lower bound of U is a direct computation which was done in proof of Lemma 3.2 in [6] so we omit details here). Now for any $J \in \binom{[n]}{\kappa}$, define

$$\begin{aligned} p_J \quad := \mathbb{P}\Big(& J' \in \tbinom{[n]}{m} \text{ such that } J' \cap J = \emptyset, \\ & |I'(J,J') \cap I^0(J,J')| < \tfrac{c_0 \kappa p n}{8} \Big). \end{aligned} \tag{4.5}$$

As J, J' are disjoint, we have independence between random subsets $I^1(J,J')$ and $I^0(J,J')$. Thus

$$\begin{aligned} p_J \quad \leq & \sum_{I \subset [n],\ |I| \leq \frac{c_0}{4}\kappa p n} \mathbb{P}(I^1(J,J') = I) \\ & + \sum_{I \subset [n],\ |I| > \frac{c_0}{4}\kappa p n} \mathbb{P}(I^1(J,J') = I)\mathbb{P}\Bigg(\exists J' \in \binom{[n]}{m} \\ & \text{such that } |I^0(J,J') \cap I| \leq \tfrac{c_0}{8}\kappa p n \Bigg) \\ & \leq \exp(-c_1 \kappa p n) + \exp(-c_2 \kappa p n) \leq \exp(-c_3 \kappa p n). \end{aligned} \tag{4.6}$$

To finish the proof, we only need to take union bound over all different choices of J, s and κ. Set $c'_{4.3.4} = c_0/8$. We have

$$\mathbb{P}\left(\bigcup_{s \in \{-1,1\}^\kappa} \bigcup_{J \in \binom{[n]}{\kappa}} \bigcup_{J' \in \binom{[n]}{m}, J \cap J' = \emptyset} \left(\mathcal{A}^{J,J',s}_{c'_{4.3.4}} \right)^c \right) \leq 2^\kappa \binom{n}{\kappa} \exp(-c_3 \kappa p n).$$

Notice that the probability bound $\exp(-c_3\kappa p n)$ dominate $2^\kappa \binom{n}{\kappa}$ for C large enough in $p \geq \frac{C \log n}{n}$, we have the above probability is bound by $\exp(-c_3 \kappa p n/2)$. Finally take another union bound over κ with finish our proof.

□

Notice that to apply Lemma 4.3.4, we need a two side tail probability estimate of a random variable with mean zero, variance 1 and bounded fourth moment. The following lemma although simple may have its own interest in some applications.

Lemma 4.3.5. *Let ξ be a random variable with mean zero, unit variance, and finite fourth moment M_4^4. Then there exist constant $c_{4.3.5}, c'_{4.3.5} > 0$ depending only on M_4*

such that,

$$\min(\mathbb{P}(\xi \leq -c_{4.3.5}), \mathbb{P}(\xi \geq c_{4.3.5})) \geq c'_{4.3.5}.$$

Proof. This lemma is a two-sided version of lemma 3.2 in [48]. We derive a lower bound for second moment of positive and negative part separately and then use Paley-Zymmund inequality.

Let $\xi^+(t) = 1_{t>0}(t)\xi(t),\ \xi^-(t) = 1_{t<0}(t)\xi(t)$ be the positive and negative part of ξ. Suppose $\mathbb{E}(\xi^+)^2 = a$. Then by Cauchy-Schwartz inequality, we have $\mathbb{E}\xi^+ \leq a^{1/2}$. By $\mathbb{E}\xi = 0$ and $\mathbb{E}\xi^2 = 1$, we have

$$\mathbb{E}|\xi^-| = \mathbb{E}\xi^+ \leq a^{1/2}, \mathbb{E}(\xi^-)^2 = 1 - a.$$

Apply Hölder's inequality and $\mathbb{E}|\xi|^4 = M_4^4$, we have

$$1 - a = \mathbb{E}(\xi^-)^2 = \mathbb{E}|\xi^-|^{2/3}|\xi^-|^{4/3} \leq (\mathbb{E}|\xi^-|)^{2/3}(\mathbb{E}|\xi^-|^4)^{1/3} \leq a^{1/3}M_4^{4/3}.$$

Thus a is lower bounded by some constants c depending only on M_4. Apply Paley-Zygmund inequality we have

$$\mathbb{P}\left(\xi^+ \geq \sqrt{\frac{c}{2}}\right) = \mathbb{P}\left(|\xi^+|^2 \geq \frac{c}{2}\right) \geq \frac{(\mathbb{E}|\xi^+|^2 - c/2)^2}{M_4^4} \geq \frac{c^2}{4M_4^4}.$$

The Lemma is proved by repeating the same argument for positive part. □

We now use the above Lemma 4.3.4 to establish a uniform small ball probability bound for the set of dominated vectors. Without loss of generality, we many assume that $1/(8p) > 1$. For $p \geq 1/8$, we only need to apply result on dense matrix (see [73]) to prove our main theorem.

Lemma 4.3.6. *Consider A satisfies 4.1.1 and $p \geq (1/4)n^{-1/3}$. For any $u \in \mathbb{R}^n$,* *there exist $c_{4.3.6}, c'_{4.3.6}, c''_{4.3.6}$ depending only on C_{op}, M_4, such that*

$$\begin{aligned}
&\mathbb{P}\Big(\exists x \in \mathrm{Dom}((8p)^{-1}, c'_{4.3.6}) \text{ such that} \\
&\|Ax-u\|_2 \le c''_{4.3.6}\sqrt{np} \text{ and } \|A\| \le C_{op}\sqrt{pn}\Big) \qquad (4.7)\\
\le\ &\exp(-c_{4.3.6}pn).
\end{aligned}$$

Proof of Lemma 4.3.6. Our proof is similar to Lemma 3.3 in [6]. The major difference is how to deal with the symmetricity. We start with proving the result for $\mathrm{Sparse}((8p)^{-1})$ vectors of unit length. Then we can prove that these estimates can be easily extended to the dominated vectors. The proof strategy for sparse vectors may depends on p (see Lemma 3.3 in [6]), but for our purpose, we only need to prove it for $p \ge (1/4)n^{-1/3}$.

Since $p \ge (1/4)n^{-1/3}$, we apply the combinatorial Lemma 4.3.4 with $\kappa = 1$ and $m = \frac{1}{8p}$. Assuming that the event described in this lemma occurs, we split the vector into blocks with disjoint support. One of these blocks has a large l_2-norm. By Lemma 4.3.4, a large number of rows of the matrix has only one non-zero entry in the columns corresponding to the support of this block. This will be sufficient for us to conclude that $\|Ax-u\|_2$ is bounded from below for $x \in \mathrm{Sparse}((8p)^{-1})$. Note that to get the small ball probability estimate, we also need $\min(\mathbb{P}(\xi \le -c), \mathbb{P}(\xi \ge c)) \ge c'$. This is guaranteed by Lemma 4.3.5.

With out loss of generality, we only need to work on $\mathrm{sign}(u) = \{-1\}_{i=1}^n$. For general cases, we only need to work on $A' = -\mathrm{diag}(\mathrm{sign}(u))A$ and $u' = -\mathrm{diag}(\mathrm{sign}(u))u$ where A' still satisfies condition of Lemma 4.3.4. For $k \in [n]$, set $J_k = \{k\}$ and $J'_l = \mathrm{supp}(x)\backslash J_k$. Let $\mathcal{A}$ be the event that for each $k \in [n]$, $v \in \{-1, 1\}$ there exists a set $I_k \subset [n]$ of rows such that $|I_k| = c_{4.3.4'}pn$, and for any $i \in I_k$, $a_{ik}v \ge c_{4.3.5}$ and $a_{ij} = 0$ for $j \in \mathrm{supp}(x)\backslash k$, and $\mathrm{supp}(x)$ is non-intersect with I_k. The definition of the sets I_k immediately implies that $I_k \cap I_{k'} = \emptyset$ for $k \ne k' \in \mathrm{supp}(x)$. By Lemma

4.3.4 and Lemma 4.3.5, $\mathbb{P}(\mathcal{A}) \geq 1 - \exp(-c_{4.3.4}pn)$ where $c_{4.3.4}$ depend only on M_4. This shows that condition on this large probability event $\mathcal{A}$, we have

$$\|Ax-u\|_2^2 \geq \|Ax\|_2^2 \geq \sum_{k \in \mathrm{supp}(x)} \sum_{i \in I_k} |(Ax)_i|^2 \geq \sum_{k \in \mathrm{supp}(x)} c'_{4.3.4} pn c_{4.3.5}^2 |x(k)|^2.$$

Thus $\|Ax-u\|_2 \geq c_1\sqrt{pn}$ where c_1 depend only on c_{op}, M_4. So we get the result proved for sparse vectors. This estimate can be automatically extended to the set $\mathrm{Dom}\,((8p)^{-1}, c'_{4.3.6})$, provided that the constant $c'_{4.3.6}$ is small enough. Indeed, assume that

$$\|Ax-u\|_2 < \tfrac{c_1}{2}\sqrt{pn} \tag{4.8}$$

for some $x \in \mathrm{Dom}\,((8p)^{-1}, c'_{4.3.6})$. Set $m = (8p)^{-1}$, it is easy to notice that $\|x_{[m+1:n]}\|_\infty \leq m^{-1/2}$. Hence,

$$\|x_{[m+1:n]}\|_2 \leq c'_{4.3.6}\sqrt{m}\|x_{[m+1:n]}\|_\infty \leq c'_{4.3.6},$$

and therefore

$$\begin{aligned} \|Ax_{[1:m]}\|_2 &\leq \|Ax\|_2 + \|A\|\|x_{[m+1:n]}\|_2 \\ &\leq \frac{1}{2}\sqrt{c_1pn} + C_{op}\sqrt{pn}c'_{4.3.6} \leq \frac{3}{4}\sqrt{c_1pn} \end{aligned} \tag{4.9}$$

provide $c'_{4.3.6}$ small enough. Furthermore,

$$\begin{aligned} \Big| \big\|A(x_{[1:m]}/\|x_{[1:m]}\|_2)\big\|_2 - \|Ax_{[1:m]}\| \Big| &\leq C_{op}\left|1 - \|x_{[1:m]}\|_2\right| \\ &\leq \tfrac{1}{4}\sqrt{c_1pn}. \end{aligned} \tag{4.10}$$

Since $x_{[1:m]}/\|x_{[1:m]}\|_2 \in \mathrm{Sparse}((8p)^{-1}) \cap S^{n-1}$, combining the above steps we note equality (4.8) holds only in $\mathcal{A}^c$. Therefore, we proved the lemma with $c_{4.3.6} = c_{4.3.4}$ and $c''_{4.3.6} = c_1$.

$\square$

Similar to dominated vectors, we can extend the result of Lemma 4.3.6 to compressible vectors. This step is simply an approximation. Recall that $\mathrm{Sparse}((8p)^{-1}) \cap S^{n-1} \subset \mathrm{Dom}((8p)^{-1}, c)$ for any c.

Lemma 4.3.7. *Consider A satisfies 4.1.1 and $p \geq (1/4)n^{-1/3}$. For any $u \in \mathbb{R}^n$, there exist $c_{4.3.7}, c'_{4.3.7}, c''_{4.3.7}$ depending only on C_{op}, M_4, such that*

$$\begin{aligned} & \mathbb{P}\Big(\exists x \in \mathrm{Comp}((8p)^{-1}, c'_{4.3.7}) \text{ such that} \\ & \|Ax - u\|_2 \leq c''_{4.3.7}\sqrt{np} \text{ and } \|A\| \leq C_{op}\sqrt{pn}\Big) \\ \leq \;& \exp(-c_{4.3.7}pn). \end{aligned} \tag{4.11}$$

Proof. We first denote following set

$$\begin{aligned} \Omega := \; & \Big\{\forall x \in \mathrm{Sparse}(1/(8p)) \cap S^{n-1},\ \|Ax - u\|_2 \geq c_{4.3.6}\sqrt{pn} \\ & \text{and } \|A\| \leq C_{op}\sqrt{pn}\Big\}. \end{aligned} \tag{4.12}$$

Then on Ω, for any $\bar{x} \in \mathrm{Comp}((8p)^{-1}, c'_{4.3.7})$, we can find $x \in \mathrm{Sparse}(1/(8p))$ such that

$$\|Ax/\|x\|_2 - u\|_2 \geq c_{4.3.6}\sqrt{pn} \text{ and } \|x - \bar{x}\|_2 \leq c'_{4.3.7}.$$

This also implies $|1 - \|x\|_2| \leq c'_{4.3.7}$. Therefore

$$\begin{aligned} \|A\bar{x} - u\|_2 & \geq \|Ax/\|x\|_2 - u\|_2 - \|A\| \left\| x - \frac{x}{\|x\|_2} \right\| - \|A\|\|x - \bar{x}\|_2 \\ & \geq c''_{4.3.7}\sqrt{pn} \end{aligned} \tag{4.13}$$

by choosing $c'_{4.3.7}$ small enough. Since by Lemma 4.3.6, $\mathbb{P}(\Omega) \geq 1 - \exp(-c_{4.3.6}pn)$, the result follows. □

4.3.2 Vectors very close to moderately sparse

Lemma 4.3.6 provided uniform lower bound on $\|Ax\|$ for vectors which are close to very sparse vectors. To prove Theorem 4.3.1, we need to uplift these estimates for vectors which are less sparse (see Section 3.2 in [6]). These vectors are well spread ones which allows us to obtain a strong small ball probability estimate so that we can use the standard net argument. The argument is a modification of proof of Lemma 3.8 in [6].

As a direct application of Corollary 3.7 in [6], we have the following corollary.

Corollary 4.3.8. *Let A_n be an $n \times m$ matrix with i.i.d. entries of the form $a_{ij} = \xi_{ij}\delta_i j$, where ξ_{ij}, δ_{ij} are the same as in Assumption 4.1.1. Then for any $\alpha > 1$, there exist $\beta, \gamma > 0$, depending on α and the fourth moment of ξ_{ij}, such that for any $x \in \mathbb{R}^m$, satisfying $\|x\|_\infty / \|x\|_2 \le \alpha\sqrt{p}$, we have*

$$\mathcal{L}\left(A_n x, \beta\sqrt{pn}\|x\|_2 \le \exp(-\gamma n)\right).$$

Applying these results on Levy concentration we now prove uniform lower bound on $\|Ax\|_2$ for vectors in $\mathrm{Dom}(M, c)$. Note that proof of following lemma is a direct modification of first part of Lemma 3.8 in [6]. The only variation is we need to restrict on a block of A to get the small ball probability estimate.

Lemma 4.3.9. *Consider A satisfies 4.1.1 and $p \ge (1/4)n^{-1/3}$. For any $u \in \mathbb{R}^n$ and $p^{-1} \le M \le c'''_{4.3.9}n$, there exist $c_{4.3.9}, c'_{4.3.9}, c''_{4.3.9}$ depending only on C_{op}, M_4 such that, for any $u \in \mathbb{R}^n$,*

$$\begin{aligned} &\mathbb{P}\Big(\exists x \in \mathrm{Dom}\,(M, c'_{4.3.9}) \text{ such that} \\ &\qquad \|Ax - u\|_2 \le c''_{4.3.9}\sqrt{np} \text{ and } \|A\| \le C_{op}\sqrt{pn}\Big) \qquad (4.14)\\ \le\ & \exp(-c_{4.3.9}pn). \end{aligned}$$

Proof. For convenience, denote $m = (8p)^{-1}$, so we have $m < M/2$. Due to Lemma 4.3.6 and 4.3.7, it is enough to obtain a uniform lower bound for all vectors from the set

$$W := \mathrm{Dom}\,(M, c'_{4.3.9}) \setminus \left(\mathrm{Comp}((8p)^{-1}, c'_{4.3.7}) \cup \mathrm{Dom}((8p)^{-1}, c'_{4.3.6})\right).$$

We start with a set with only M-sparse vectors

$$V := \mathrm{Sparse}(M) \setminus \left(\mathrm{Comp}((8p)^{-1}, c'_{4.3.7}) \cup \mathrm{Dom}((8p)^{-1}, c'_{4.3.6})\right).$$

Since $p \ge (1/4)n^{-1/3}$, the proof is based on the straightforward ε-net argument as in Lemma 3.8 in [6]. Since for any $x \in V, x \notin (\mathrm{Comp}((8p)^{-1}, c'_{4.3.7}) \cup \mathrm{Dom}((8p)^{-1}, c'_{4.3.6}))$, we have that

$$\frac{\|x_{[m+1:M]}\|_\infty}{\|x_{[m+1:M]}\|_2} \le (c'_{4.3.6})^{-1}\sqrt{8p}.$$

Now for this given x, define A^x to be the sub-matrix restricted on the columns corresponding to $\mathrm{supp}(x)$ and rows corresponding to $[n]\backslash\mathrm{supp}(x)$. Then A^x is an $(n-M)\times M$ submatrix with i.i.d. entries. By Corollary 4.3.8 and properties of Levy's concentration function, we have

$$\begin{aligned} & \mathcal{L}\left(Ax, c_1\sqrt{pn}\|x_{m+1:M}\|_2\right) \\ \le\ & \mathcal{L}\left(A^x x, c_1\sqrt{pn}\|x_{m+1:M}\|_2\right) \\ \le\ & \exp(-c_2 n) \end{aligned} \tag{4.15}$$

where c_1, c_2 depending only on C_{op}, M_4.

Now, we will use this estimate of the Levy concentration function to control the infimum over V. Since $V \subset \mathrm{Sparse}(M)$, note that the set V is contained in S^{n-1} intersected with the union of coordinate subspaces of dimension M. Thus, for $\varepsilon < c'_{4.3.7}c'_{4.3.9}$, there exists an $\varepsilon-$ net $\mathcal{N} \subset V$ of cardinality less than

$$\binom{n}{M}\left(\frac{3}{\varepsilon}\right)^M \le \exp\left(c'''_{4.3.9} n \log\left(\frac{3e}{c'''_{4.3.9}\varepsilon}\right)\right).$$

We used the assumption $M \le c'''_{4.3.9}n$ in above estimate. Moreover, we can choose the constant $c'''_{4.3.9}$ sufficiently small (depending on ε) so that $|\mathcal{N}| \le \exp(c_2 n/2)$. Using the union bound argument, we have

$$\mathbb{P}\left(\exists x \in \mathcal{N}, u \in \mathbb{R}^n | \|Ax - u\|_2 \le c_1\sqrt{pn}\|x_{[m+1:M]}\|_2\right) \le \exp(-c_2 n/2).$$

Now we can approximate any point of W by a point of $\mathcal{N}$. Assume that for any $x \in \mathcal{N}$,

$$\|Ax - u\|_2 \ge c_1\sqrt{pn}\|x_{[m+1:M]}\|_2.$$

Let $x' \in W$, then we can find $x \in \mathcal{N}$ such that

$$\|x'_{[1:M]}/\|x'_{[1:M]}\|_2 - x\|_2 \le \varepsilon.$$

Now, we show that x and x' are close. Since $m \le M/2$ and all coordinates of $x'_{[M+1:n]}$ have smaller absolute value than those of $x'_{[1:M]}$, we have

$$\sqrt{M}\|X'_{[M+1:n]}\|_\infty \le \sqrt{2}\|x'_{[m+1:M]}\|_2.$$

Now recall that $x' \in \mathrm{Dom}(M, c'_{4.3.9})$, so we have

$$\|x'_{[M+1:n]}\|_2 \le c'_{4.3.9}\sqrt{M}\|x'_{[M+1:n]}\|_\infty \le \sqrt{2}c'_{4.3.9}\|x'_{[m+1:M]}\|_2.$$

Now, we can use the fact that $\|x'_{[1:M]}/\|x'_{[1:M]}\|_2 - x\|_2 \le \varepsilon$ together with triangle inequality. Therefore, we have

$$\|x'_{[m+1:M]}\|_2 \le \|x'_{[1:M]}\|_2(\|x_{[m+1:M]}\|_2 + \varepsilon) \le \|x_{[m+1:M]}\|_2 + \varepsilon.$$

Now, for any $x \in \mathcal{N}$, $x \notin \mathrm{Comp}(m, c'_{4.3.7})$, $\|x_{[m+1:M]}\|_2 \ge c'_{4.3.7} \ge \varepsilon$. Applying previous two inequalities, we also have

$$\begin{aligned} \|x'_{[M+1:n]}\|_2 &\le \sqrt{2}c'_{4.3.9}\|x'_{[m+1:M]}\|_2 \\ &\le 2c'_{4.3.9}(\|x_{[m+1:M]}\|_2 + \varepsilon) \le 4c'_{4.3.9}\|x_{[m+1:M]}\|_2 \end{aligned} \tag{4.16}$$

and

$$\begin{aligned} \|x - x'\|_2 &\le \left\| x - x'_{[1:M]}/\|x'_{[1:M]}\|_2 \right\|_2 + \left| 1 - \|x'_{[1:M]}\|_2 \right| + \|x'_{[M+1:n]}\|_2 \\ &\le \varepsilon + 2\|x'_{[M+1:n]}\|_2 \le \varepsilon + 8c'_{4.3.9}\|x_{[m+1:M]}\|_2 \\ &\le 9c'_{4.3.9}\|x_{[m+1:M]}\|_2. \end{aligned} \tag{4.17}$$

Finally, by choosing $c'_{4.3.9}$ sufficiently small, by the triangle inequality,

$$\begin{aligned} \|Ax' - u\|_2 &\ge \|Ax - u\| - \|A\|\|x - x'\|_2 \\ &\ge (c_1 - 9c'_{4.3.9}C_{op})\sqrt{pn}\|x_{[m+1:M]}\|_2 \ge c''_{4.3.9}\sqrt{pn}. \end{aligned} \tag{4.18}$$

$\square$

Now, we can conclude Theorem 4.3.1.

Proof. Theorem 4.3.1 follow directly from a similar argument of Lemma 4.3.7.

□

Now, by Theorem 4.3.1, we have following small probability estimate similar to Proposition 4.2 in [73].

Theorem 4.3.10. (Small ball probability for compressible vectors). *Consider A satisfies 4.1.1 and $p \geq (1/4)n^{-1/3}$. For every $u \in \mathbb{R}^n$, one has*

$$\mathbb{P}\left(\inf_{\frac{x}{\|x\|_2}\in \mathrm{Comp}(c_s n, c_d)} \|Ax-u\|_2/\|x\|_2 \leq c'_{4.3.10}\sqrt{pn} \wedge \mathcal{E}_{op}\right) \leq 2\exp(-c_{4.3.10}pn)$$

where $c_s, c_d, c_{4.3.10}, c'_{4.3.10}$ depending only on M_4, C_{op}.

Proof. Let $\mathcal{E}$ be the event in the left hand side whose probability need to be estimated. We start with some fixed small positive numbers of c_s, c_d and $c'_{4.3.10}$ which specific choice will be decided later. Conditioning on $\mathcal{E}$, we have that there exist vectors $u_0 := u/\|x\|_2 \in \mathrm{span}(u)$ and $x_0 := x/\|x\|_2 \in \mathrm{Comp}(c_s n, c_d)$ such that

$$\|Ax_0 - u_0\|_2 \leq c'_{4.3.10}\sqrt{pn}.$$

By definition of event $\mathcal{E}_{op}$, we have

$$\|u_0\|_2 \leq \|Ax_0\| + c'_{4.3.10}\sqrt{pn} \leq C_{op}\sqrt{pn} + c'_{4.3.10}\sqrt{pn} \leq 2C_{op}\sqrt{pn}$$

Therefore

$$u_0 \in \mathrm{span}(u) \cap 2C_{op}\sqrt{pn}B_2^n =: E$$

Let $\mathcal{M}$ be a $(c_1\sqrt{pn})$-net of the interval E such that

$$|\mathcal{M}| \leq \frac{2C_{op}\sqrt{pn}}{c_1\sqrt{pn}} = \frac{2C_{op}}{c_1}$$

and choose $v_0 \in |\mathcal{M}|$ such that $\|v_0 - u_0\|_2 \leq c_1\sqrt{pn}$. Then

$$\|Ax_0 - v_0\|_2 \leq c'_{4.3.10}\sqrt{pn} + c_1\sqrt{pn}.$$

Now we may choose $c'_{4.3.10}, c_1 \in (0,1)$ such that $c'_{4.3.10} + c_1 \leq c''_{4.3.1}$. So the event $\mathcal{E}$ implies the existence of vector $x \in \mathrm{Comp}(c_s n, c_d), v_0 \in \mathcal{M}$ such that $\|Ax_0 - v_0\|_2 \leq c''_{4.3.1}\sqrt{pn}$. Taking the union bound over $\mathcal{M}$, we have

$$\mathbb{P}(\mathcal{E}) \leq |\mathcal{M}| \max_{v_0 \in \mathcal{M}} \mathbb{P}\left\{\exists x \in \mathrm{Comp}(c_s n, c_d) \text{ such that } \|Ax - v_0\|_2 \leq c''_{4.3.1}\sqrt{np}\right\}.$$

Now we may apply Theorem 4.3.1 together with the net cardinalities estimates and we get

$$\mathbb{P}(\mathcal{E}) \leq \frac{2C_{op}}{c_1}\exp(-c_{4.3.10}np).$$

Use the condition on p then we are done. The c_s, c_d in this theorem can be chosen as $c'_{4.3.1}$ and $c'''_{4.3.1}$ in Theorem 4.3.1.

□

Remark 4.3.11. Note that the constants c_s, c_d can be chosen depending only on C_{op}, M_4. These to constants are fixed in the later part of the proof. An immediate consequence of Theorem 4.3.10 is

$$\mathbb{P}\left\{\inf_{x \in \mathrm{Comp}(c_s n, c_d)} \|Ax\|_2 \leq \varepsilon\sqrt{\frac{p}{n}} \wedge \mathcal{E}_{op}\right\} \leq 2\exp(-c_{4.3.10}pn) \tag{4.19}$$

4.4 Invertibility over incompressible vectors

Our goal in the following sections is to show, with high probability

$$\min_{x \in \mathrm{Incomp}(c_s n, c_d)} \|Ax\|_2 \gtrsim \sqrt{\frac{p}{n}}.$$

4.4.1 Incompressible vectors are spread

Note that in Theorem 4.3.10 and from now on, we will adapt the methodology of Vershynin in [73] in order to decouple the symmetric matrix. Although some proofs are very similar to those in [73], we still need to went through several proofs in much detail under our setting. This need to be done to ensure the methodology works as well in sparsity setting. And what is more important is to catch the affect of sparsity especially how it affect the probability bounds. For convenience of reader and to show the connection in methodology, we will try to use similar notation and structure as proofs in [73].

First, we want to note that although the incompressible vectors have many non-negligible coordinate but they have different advantage. Incompressible vectors x have many coordinates that are well spread, that is to say a set of coordinates of size of order n whose magnitudes are all of the order $n^{-1/2}$. More precisely, we have the following lemma, see Lemma 3.4 in [47]:

Lemma 4.4.1. (Incompressible vectors are spread). *For every $x \in \mathrm{Incomp}(c_0 n, c_1)$, one has*

$$\frac{c_1}{\sqrt{2n}} \leq |x_k| \leq \frac{1}{\sqrt{c_0 n}} \tag{4.20}$$

for at least $\frac{1}{2}c_0c_1^2 n$ coordinates x_k of x.

We fix some constant c_{oo} such that as in [47]

$$\frac{1}{4}c_s c_d^2 \leq c_{oo} \leq \frac{1}{4}.$$

Here note that the value of c_{oo} depend only on c_s and c_d, which depend only on the parameters C_{op} and M_4. We may assign a subset called $\mathrm{spread}(x) \subset [n]$ for every vector $x \in \mathrm{Incomp}(c_s n, c_d)$ such that

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

and the property in Lemma 4.4.1 hold for any $k \in \mathrm{spread}(x)$. The point here is that not all of the coordinates x_k satisfying Lemma 4.4.1 will be good, the set spread(x) will allow us to only focus on the good coordinates. At this point, we may consider an arbitrary valid assignment of spread(x) to x, the particular choice will be decide later in the proof.

4.4.2 Distance problem via small ball probabilities for quadratic forms

To derive incompressible part of the invertibility problem, we need the following Lemma, see Lemma 2.4 in [6].

Lemma 4.4.2. (Invertibility via distance). *For $j \in [n]$, let A_j denote the $j-th$ column of A_n, and let H_j be the subspace of $\mathbb{R}^n$ spanned by $A_i, i \in [n]\backslash j$. Then for any $\varepsilon, \rho > 0$, and $M < n$, we have*

$$\mathbb{P}\left(\inf_{x \in \mathrm{Incomp}(M,\rho)} \|Ax\|_2 \leq \varepsilon\sqrt{\frac{p}{n}}\right) \leq \frac{1}{M}\sum_{j=1}^{n} \mathbb{P}\left(\mathrm{dist}(A_j, H_j) \leq \sqrt{p}\varepsilon\right)$$

So we may reduces the invertibility problem to the distance problem, namely an upper bound on the probability

$$\mathbb{P}\left(\mathrm{dist}(A_1, H_1) \leq c_1\sqrt{p}\varepsilon\right)$$

where A_1 is the first column of A and H_1 is the span of the other column. (By a permutation of the indices in $[n]$, the same bound would hold for all dist(A_k, H_k) as required in Lemma 4.4.2).

But we have a symmetric matrix, to do the decoupling we need tools to evaluate the distance problem. To this end, the following proposition in [73] reduces the distance problem to the small ball probability for quadratic forms of random variables:

Proposition 4.4.3. (Distance problems via quadratic forms). *Let $A = (a_{ij})$ be an arbitrary $n \times n$ matrix. Let A_1 denote the first column of A and H_1 denote the span of the other columns. Furthermore, let B denote the $(n-1) \times (n-1)$ minor of A obtained by removing the first row and the first column from A, and let $X \in \mathbb{R}^{n-1}$ denote the first column of A with the first entry removed. Then*

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

Remark 4.4.4. We may apply Proposition 4.4.3 to the $n \times n$ random matrix A which we studied. Consider $a_{1,1}$ as an arbitrary fixed number and bound our probability uniformly for all $a_{1,1}$, the problem reduces to estimating the small ball probability for the quadratic form $\langle B^{-1}X, X\rangle$. The random matrix B has the same structure as A except for the dimension is $n-1$. Thus it will be convenient to develop the theory in dimension n for the quadratic forms $\langle A^{-1}X, X\rangle$, where X is an independent random vector (see Remark 5.2 in [73]).

4.4.3 Small ball probabilities for quadratic forms via additive structure

It is a popular and powerful to estimate small ball probabilities using the additive structure of vectors. For completion of our argument, let us first review the the Littlewood-Offord theory and its extension to quadratic forms by decoupling, see [73].

Linear Littlewood-Offord theory concerns the small ball probabilities for the sums of the form $S = \sum x_k \xi_k$ where ξ_k are identically distributed independent random variables, and $x = (x_1, \cdots, x_n) \in S^{n-1}$ is a given coefficient vector. The additive structure of $x \in \mathbb{R}^n$ is characterized by the least common denominator (LCD) of x.

If the coordinates $x_k = p_k/q_k$ are rational numbers, one can measure the additive structure in x using the least denominator $D(x)$ of these ratios, which is the common multiple of the integers q_k. In the other words, $D(x)$ is the smallest number $\theta > 0$ such that $\theta x \in \mathbb{Z}^n$. An extension of this concept for general vectors with real coefficients was developed in [47, 48, 73] which give us following definition of LCD.

Definition 4.4.5. (Least Common Denominator). Let $L \geq 1$. We defined the least common denominator (LCD) of $x \in S^{n-1}$ as

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

Remark 4.4.6. If the vector x is considered in $\mathbb{R}^I$ for some subset $I \subset [n]$, then in this definition we replace $\mathbb{Z}^n$ by $\mathbb{Z}^I$.

It can be easily seen that we always have $D_L(x) > L$. We may also notice that the parameter L is up to our choice. Recall by Remark 4.2.5 that there exists $\delta_0, \varepsilon_0' \in (0,1)$, such that for any $\varepsilon < \varepsilon_0'$, $\mathcal{L}(\xi_{ij}\delta_{ij}, \varepsilon) \leq 1 - \delta_0 p$. Due to the sparsity, we will often use the parametrization $L = (\delta p)^{-1/2}$ in our proofs (also see Section 4 of [6]).

Remark 4.4.7. We may refer $D_L(x)$ as $D(x)$ for convenience.

Another useful bound is the following, see Lemma 6.2 in [73].

Lemma 4.4.8. *For every $x \in S^{n-1}$ and every $L \geq 1$, one has*

$$D_L(x) \geq \frac{1}{\|x\|_\infty}$$

Now we can try to express the small ball probabilities of sums $\mathcal{L}(S, \varepsilon)$ in terms of $D_L(x)$. This was done in the following theorem, see Theorem 6.3 in [73].

Theorem 4.4.9. (Small ball probabilities via LCD). *Let $\xi_1, \cdots, \xi_n$ be independent and identically distributed random variables. Assume that there exist numbers*

$\varepsilon_0, p_0, M_1 > 0$ such that $\mathcal{L}(\xi_k, \varepsilon_0) \leq 1 - p_0$ and $\mathbb{E}|\xi_k| \leq M_1$ for all k. Then there exists $C_{6,3}$ which depends only on ε_0, p_0 and M_1, and such that the following holds. Let $x \in S^{n-1}$ and consider the sum $S = \sum_{k=1}^n x_k \xi_k$. Then for every $L \geq p_0^{-1/2}$ and $\varepsilon \geq 0$ one has

$$\mathcal{L}(S, \varepsilon) \leq C_{4.4.9} L \left(\varepsilon + \frac{1}{D_L(x)} \right)$$

for some constant $C_{4.4.9}$ depending only on second and fourth moments of ξ.

Applying the above theorem to the sparse vector, one may get following theorem for sparse vector, see Proposition 4.2 in [6].

Theorem 4.4.10. (Small ball probabilities via LCD). *Let $S \in \mathbb{R}^n$ be a random vector with i.i.d. coordinates of the form $S_j = \delta_j \xi_j$, where $\mathbb{P}(\delta_j = 1) = p$, and ξ_js are random variables with unit variance, and finite fourth moment, which are independent of δ_j. Then for any $v \in S^{n-1}$, $L = (\delta p)^{-1/2}$ and $\delta < \delta_0$*

$$\mathcal{L}\Big(\sum_{j=1}^n S_j v_j, \sqrt{p}\varepsilon \Big) \leq C_{4.4.10} \left(\varepsilon + \frac{1}{\sqrt{p} D_L(v)} \right)$$

for some constant $C_{4.4.10}, \delta_0$ depending only on fourth moments of ξ_j.

4.4.4 Regularized LCD

As we discussed, the distance problem reduces to a quadratic Littlewood-Offord problem. Similar to [73], we want to the use the same technique to reduce the quadratic problem to a linear one by decoupling and conditioning arguments. This process requires a more robust version of the concept of the LCD, which R. Vershynin developed in [73].

Definition 4.4.11. (Regularized LCD). Let $\lambda \in (0, c_{oo})$ and $L \geq 1$. We define the regularized LCD of a vector $x \in \text{Incomp}(c_s n, c_d)$ as

$$\hat{D}_L(x, \lambda) = \max \left\{ D_L(x_I / \|x_I\|_2) : I \subset \text{spread}(x), |I| = \lambda n \right\}.$$

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

Remark 4.4.12. Since the sets I in this definition are subsets of spread(x), inequality are subsets of spread(x), inequalities (4.20) imply that

$$c\sqrt{\lambda} \leq \|x_I\|_2 \leq C\sqrt{\lambda}$$

where $c = c_d/\sqrt{2}$ and $C = 1/\sqrt{c_s}$.

We also have the following estimate for regularized LCD, see Lemma 6.8 in [73].

Lemma 4.4.13. *For every $x \in \mathrm{Incomp}(c_s n, c_d)$ and every $\lambda \in (0, c_{oo})$ and $L \geq 1$, one has*

$$\hat{D}_L(x, \lambda) \geq c_{4.4.13}\sqrt{\lambda n}$$

where $c_{4.4.13}$ depends only on c_s and c_d.

We now state a version of Theorem 4.4.9 for regularized LCD, see Proposition 6.9 in [73].

Theorem 4.4.14. (Small ball probabilities via regularized LCD). *Let $\xi_1, \cdots, \xi_n$ be independent and identically distributed random variables. Assume that there exist numbers $\varepsilon_0, p_0, M_1 > 0$ such that $\mathcal{L}(\xi_k, \varepsilon_0) \leq 1 - p_0$ and $\mathbb{E}|\xi_k| \leq M_1$ for all k. Then there exists $C_{4.4.14}$ which depends only on ε_0, p_0 and M_1, and such that the following holds.*

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

$$\mathcal{L}(S_J, \varepsilon) \leq C_{4.4.14} L \left(\frac{\varepsilon}{\sqrt{\lambda}} + \frac{1}{\hat{D}_L(x, \lambda)} \right).$$

Similarly, we can rewrite it for sparse random sums.

Theorem 4.4.15. *Let $S \in \mathbb{R}^n$ be a random vector with i.i.d. coordinates of the form $S_j = \delta_j \xi_j$, where $\mathbb{P}(\delta_j = 1) = p$, and $\xi_j s$ are random variables with unit variance, and finite fourth moment, which are independent of δ_j. Consider a vector $x \in \mathrm{Incomp}(c_s n, c_d)$ and a subset $J \subseteq [n]$ such that $J \supseteq I(x)$. Then for every $\lambda \in (0, c_{oo})$, $v \in S^{n-1}$, $L = (\delta p)^{-1/2}$ and $\delta < \delta_0$*

$$\mathcal{L}\Big(\sum_{j=1}^{n} S_j v_j, \sqrt{p}\varepsilon\Big) \le C_{4.4.15} \left(\frac{\varepsilon}{\sqrt{\lambda}} + \frac{1}{\sqrt{p}\hat{D}_L(x, \lambda)} \right)$$

for some constant $C_{4.4.15}, \delta_0$ depending only on fourth moments of ξ_j.

By Theorem 4.2.6, one has the following proposition as a corollary, see Proposition 6.11 in [73]:

Proposition 4.4.16. (Small ball probabilities for Ax via regularized LCD.) *Let A be a random symmetric matrix with mean zero variance one and fourth moment M_4^4 i.i.d. entries above diagonal. Let $x \in \mathrm{Incomp}(c_s n, c_d)$ and $\lambda \in (0, c_{oo})$. Then for every $L \ge L_0$ and $\varepsilon \ge 0$, one has*

$$\mathcal{L}(Ax, \varepsilon\sqrt{n}) \le \left[\frac{C_{4.4.16} L \varepsilon}{\sqrt{\lambda}} + \frac{C_{4.4.16} L}{\hat{D}_L(x, \lambda)} \right]^{n - \lambda n}.$$

Here $C_{4.4.16}$ and L_0 depend only on the parameters M_4.

It can be easily derived as a corollary that for A is a sparse matrix, we have the following result:

Proposition 4.4.17. (Small ball probabilities for Ax via regularized LCD where A is sparse.) *Let A be a random matrix satisfies Assumption 4.1.1. Let $x \in$ $\mathrm{Incomp}(c_s n, c_d)$ and $\lambda \in (0, c_{oo})$. Then one has for $L = (\delta p)^{-1/2}$ and $\delta < \delta_0$*

$$\mathcal{L}(Ax, \varepsilon\sqrt{pn}) \le \left[\frac{C_{4.4.17} \varepsilon}{\sqrt{\lambda}} + \frac{C_{4.4.17}}{\sqrt{p}\hat{D}_L(x, \lambda)} \right]^{n - \lambda n}.$$

Here $C_{4.4.16}, \delta_0$ depends only on the parameters M_4.

4.5 Estimating additive structure

To estimate the small ball probability for quadratic form $\langle A^{-1}X, X\rangle$, we will first need to estimate the additive structure in the random vector $A^{-1}X$. In this section, we will show that the regularized LCD of $A^{-1}X$ is large for every fixed X which is an analog of Theorem 7.1 in [73] for sparse matrices.

Theorem 4.5.1. (Structure theorem for sparse matrix.) *Let A be a random matrix which satisfies Assumption 4.1.1 and $p \geq n^{-c_p}$. Let $u \in \mathbb{R}^n$ be an arbitrary fixed vector, and consider $x_0 := A^{-1}u/\|A^{-1}u\|_2$. Let $n^{c_{4.5.1}n/6}p^{-1/2} \geq L = (p\delta)^{-1/2} \geq (p\delta_0)^{-1/2}$, $p \geq n^{-c_p}$ and $n^{-c_{4.5.1}} \leq \lambda \leq c_{4.5.1}/4$. Consider the event*

$$\mathcal{E} = \left\{ x_0 \in \mathrm{Incomp}(c_s n, c_d) \text{ and } \hat{D}_L(x_0, \lambda) \geq L^{-2} n^{c_{4.5.1}/\lambda} \right\}$$

Then

$$\mathbb{P}(\mathcal{E}^c \cap \mathcal{E}_{op}) \leq 2e^{-c'_{4.5.1}pn}.$$

Here $c_p, c_{4.5.1}, c'_{4.5.1}, \delta_0 > 0$ depend only on the parameters C_{op} and M_4.

Remark 4.5.2. Theorem 4.5.1 is the step that $p \geq n^{-c_p}$ is needed. To improve Theorem 4.1.5, one just need to improve Theorem 4.5.1 to work for a greater range of p.

We shall first prove the easier part that $x_0 \in \mathrm{Incomp}(c_s n, c_d)$ w.h.p.. The more difficult part of the theorem is the estimate on the LCD.

Lemma 4.5.3. ($A^{-1}u$ is incompressible.) *In the setting of Theorem 4.5.1, consider the event*

$$\mathcal{E}_1 = \{x_0 \in \mathrm{Incomp}(c_s n, c_d)\}$$

Then

$$\mathbb{P}(\mathcal{E}_1^c \cap \mathcal{E}_{op}) \leq 2\exp(-c_{4.5.3}pn)$$

Here $c_{4.5.3}$ depends only on the parameters C_{op} and M_4.

Proof. Denote $x = A^{-1}u$, then $Ax = u$. Hence

$$\mathcal{E}_1^c \subseteq \left\{\exists x \in \mathbb{R}^n : \frac{x}{\|x\|_2} \in \mathrm{Comp}(c_s n, c_d) \wedge Ax = u\right\}$$

By Proposition 4.3.10, $\mathbb{P}(\mathcal{E}_1^c \cap \mathcal{E}_{op}) \leq 2\exp(-c_{4.3.10}np)$. □

Following the strategy in [73], to get the structure theorem, we also need a special entropy estimate. This is done in Proposition 7.4 of [73]. To state the result, we need the following definition first.

Definition 4.5.4. (Sublevel sets of LCD). Let us fix $\lambda \in (0, c_{oo})$. For every value $D \geq 1$, we define the set

$$S_D = \left\{x \in \mathrm{Incomp}(c_s n, c_d) : \hat{D}_L(x, \lambda) \leq D\right\}$$

Then recall following covering Lemma, see Proposition 7.4 in [73].

Lemma 4.5.5. (Covering sublevelsets of regularized LCD). *Let $\lambda \in (C_{4.5.5}/n, c_{oo})$ and $L \geq 1$. For every $D \geq 1$, the sublevel set S_D has a $\beta-$net $\mathcal{N}$ such that*

$$\beta = \frac{L\sqrt{\log D}}{\sqrt{\lambda}D},\ \ |\mathcal{N}| \leq \left[\frac{C_{4.5.5}D}{(\lambda n)^{c_{4.5.5}}}\right]^n D^{1/\lambda}$$

where $C_{4.5.5}, c_{4.5.5}$ depend only on c_s, c_d. More precisely, $c_{4.5.5} = c_{oo}/4$.

Remark 4.5.6. The dominating term in the net size is the term $(\lambda n)^c$. However, once we adapt this cardinality estimate in the sparse case, the $(\lambda n)^{-cn}$ term need to dominate p^n, this end up with a limitation of the sparsity level p in our proof.

In Proposition 4.3.10, we estimated the small ball probabilities for the random vector Ax for a fixed vector x. Now we combine it with Lemma 4.5.5 to obtain a bound that is uniform over all x with small regularized LCD.

Lemma 4.5.7. (Small ball probabilities on a sublevel set of LCD.) *There exist $\delta_0, c_{4.5.7}, c_{4.5.7}, c_p$ depend only on C_{op} and M_4, and such that the following hold. Let $n^{c_{4.5.7}n/6}p^{-1/2} \geq L = (p\delta)^{-1/2} \geq (p\delta_0)^{-1/2}$, $n^{-c_{4.5.7}} \leq \lambda \leq c_{4.5.7}/4$, $p \geq n^{-c_p}$ and $1 \leq D \leq (L)^{-2}n^{c_{4.5.7}/\lambda}$. Then*

$$\mathbb{P}\Big\{\exists x \in S_D : \|Ax-u\|_2 \leq C_{op}\beta\sqrt{pn} \wedge \mathcal{E}_{op}\Big\} \leq n^{-c'_{4.5.7}n}$$

where

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

Proof. In this proof, the sparsity would play an important role. Unlike the non-sparse case in proof of Lemma 7.9 in [73]. This proof would only work when p is relatively large. And this is the reason we have to force some assumption for our main theorem of the paper.

We start with estimating the probability for $S_D/S_{D/2}$ instead of S_D. Proposition 4.4.17 implies that for every $s \in S_D \backslash S_{D/2}$,

$$\mathbb{P}\left\{\|Ax-u\|_2 \leq \varepsilon\sqrt{pn}\right\} \leq \left[\frac{C_{4.4.17}\varepsilon}{\sqrt{\lambda}} + \frac{C_{4.4.17}}{\sqrt{p}D}\right]^{n-\lambda n}, \ \varepsilon \geq 0.$$

Now we apply this for $\varepsilon = 2C_{op}\beta$. Since $\frac{\varepsilon}{\sqrt{\lambda}}$ dominates $\frac{1}{\sqrt{p}D}$, we have

$$\mathbb{P}\left\{\|Ax-u\|_2 \leq 2C_{op}\beta\sqrt{pn}\right\} \leq \left[\frac{CL\sqrt{\log(2D)}}{\lambda D}\right]^{n-\lambda n} =: p_0$$

where C depend only on M_4, C_{op}. Now, choose a $\beta-$ net $\mathcal{N}$ of $S_D \backslash S_{D/2}$ according to Lemma 4.5.5. We have

$$\begin{aligned} &\mathbb{P}\left\{\exists x \in \mathcal{N} : \|Ax-u\|_2 \leq C_{op}\beta\sqrt{n}\right\} \leq |\mathcal{N}|p_0 \\ \leq\ &\left[\frac{C_{4.5.5}D}{(\lambda n)^{c_{4.5.5}}}\right]^n D^{1/\lambda}\left[\frac{CL\sqrt{\log(2D)}}{\lambda D}\right]^{n-\lambda n} =: p_1. \end{aligned} \tag{4.21}$$

To estimate p_1, notice that n is sufficiently large, $n^{-c} \leq \lambda \leq c_{4.5.7}/4$ and $1 \leq D \leq L^{-2} n^{c/\lambda}$. By choosing c small enough, we have

$$\begin{aligned} p_1 &\leq C^n D^{\lambda n + 1/\lambda} (\lambda n)^{-c_{4.5.5} n} L^n \lambda^{-n} (\sqrt{\log(2D)})^n \\ &\leq C^n n^{2cn + 1/\lambda^2} n^{-c_{4.5.5} n/2} L^n \lambda^{-n} (c \log n/\lambda)^n \\ &\leq n^{-c_{4.5.5} n/3} L^n. \end{aligned} \tag{4.22}$$

Choosing the constant c_p sufficient small and we obtain

$$p_1 \leq n^{-c' n}$$

where c' depend only on M_4, C_{op}. Assume event $\mathcal{E}_{op}$ hold and there exists $x \in S_D \backslash S_{D/2}$ such that $\|Ax - u\|_2 \leq C_{op} \beta \sqrt{n}$. Then there exists $x_0 \in \mathcal{N}$ such that $\|x - x_0\|_2 \leq \beta$. Therefore

$$\begin{aligned} \|Ax_0 - u\|_2 &\leq \|Ax - u\|_2 + \|A(x - x_0)\|_2 \leq \|Ax - u\|_2 + \|A\| \|x - x_0\|_2 \\ &\leq 2 C_{op} \beta \sqrt{pn}. \end{aligned} \tag{4.23}$$

The probability of the later event is bounded by $p_1 \leq n^{-c'n}$. So we have

$$\mathbb{P}\left\{\exists x \in S_D \backslash S_{D/2} : \|Ax - u\|_2 \leq C_{op} \beta \sqrt{pn} \wedge \mathcal{E}_{op}\right\} \leq n^{-c'n}.$$

To remove $S_{D/2}$ in this bound, we divide it into level sets. Since β decreases in D, the previous result can be applied for $D/2$ instead of D if $D \geq 2$. Therefore

$$\mathbb{P}\left\{\exists x \in S_{D/2} \backslash S_{D/4} : \|Ax - u\|_2 \leq C_{op} \beta \sqrt{pn} \wedge \mathcal{E}_{op}\right\} \leq n^{-c'n}.$$

We can continue defining such sets for $S_{D/4} \backslash S_{D/8}$ and so on. On the other hand, $S = \bigcup_{k=0}^{k_0} (S_{2^{-k} D})$, where k_0 is the largest integer such that $2^{-k_0} D \geq c_{4.4.13} \sqrt{\lambda n}$. By Proposition 4.4.13, S_{D_0} is empty set if $D_0 < c_{4.4.13} \sqrt{\lambda n}$. Since $c_{4.4.13} \sqrt{\lambda n} \geq 1$, we have $k_0 \leq \log_2(D)$. Therefore

$$\mathbb{P}\left\{\exists x \in S_D : \|Ax - u\|_2 \leq K \beta \sqrt{pn} \wedge \mathcal{E}_{op}\right\} \leq \log_2(D) n^{-c'n} \leq n^{c'' n}$$

if the constant c'' is chosen appropriately small.

$\square$

Proof of Theorem 4.5.1. This is a direct analog of proof of Theorem 7.1 in [73]. We now fix constants $\delta_0, c_{4.5.7}, c_{4.5.7}, c_p$ in Lemma 4.5.7. Define

$$\mathcal{E}_0 = \left\{ \hat{D}_L(x_0, \lambda) > L^{-2} n^{c_{4.5.7}/\lambda} =: D_0 \text{ or } \hat{D}_L(x_0, \lambda) \text{ is undefined} \right\}$$

and

$$\mathcal{E}_1 = \{x_0 \in \mathrm{Incomp}(c_s n, c_d)\}\,.$$

Note $\hat{D}_L(x_0, \lambda)$ is defined if $\mathcal{E}_1$ holds. Thus we may rewrite $\mathcal{E}$ as

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

Then

$$\mathcal{E}^c = \mathcal{E}_1^c \cup (\mathcal{E}_1 \cap \mathcal{E}^c) = \mathcal{E}_1^c \cup (\mathcal{E}_1 \cap \mathcal{E}_0^c).$$

So the probability we want to estimate can be rephrased as

$$\mathcal{E}^c \cap \mathcal{E}_K \subseteq (\mathcal{E}_1^c \cap \mathcal{E}_K) \cup (\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K).$$

Thus

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

By Lemma 4.5.3, the first term can be bounded to be:

$$\mathbb{P}(\mathcal{E}_1^c \cap \mathcal{E}_K) \leq 2\exp(-c_{4.5.3} pn).$$

To estimate the second term $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K)$, consider

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

Define $u_0 := Ax_0 = u/\|A^{-1}u\|_2$ and $\mathcal{E}_K$ implies

$$\|u_0\|_2 = \|Ax_0\|_2 \leq \|A\| \leq C_{op}\sqrt{pn}.$$

Thus, u_0 belongs to a one-dimensional interval. More precisely,

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

So

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

Now, choose

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

Let $\mathcal{M}$ be some fixed $(C_{op}\beta_0\sqrt{pn})-$net of the interval E with cardinality

$$|\mathcal{M}| \leq \frac{C_{op}\sqrt{pn}}{C_{op}\beta_0\sqrt{pn}} = \frac{1}{\beta_0} \leq D_0.$$

Therefore for $u_0 \in E$ we there exists $v_0 \in \mathcal{M}$ such that $\|u_0 - v_0\|_2 \leq C_{op}\beta_0\sqrt{pn}$. We also have $\|Ax_0 - v_0\|_2 \leq C_{op}\beta\sqrt{np}$ since $Ax_0 = u_0$. Therefore

$$\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_K \subseteq \{\exists x_0 \in S_{D_0}, \exists v_0 \in \mathcal{M} : \|Ax_0 - v_0\|_2 \leq C_{op}\beta_0\sqrt{pn} \wedge \mathcal{E}_{op}\}.$$

Finally, applying Lemma 4.5.7 and a union bound argument for all $v_0 \in \mathcal{M}$,

$$\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_0^c \cap \mathcal{E}_{op}) \leq |\mathcal{M}|n^{-c'_{4.5.7}n} \leq D_0 n^{-c'_{4.5.7}n} \leq n^{-c'_{4.5.7}n/2}$$

where $D_0 \leq n^{c/\lambda}$, and since we can assume that constant $c_{4.5.7} > 0$ sufficient small. Our proof is complete.

$\square$

4.6 Small ball probability for quadratic forms

Now, we use the machinery developed in [73] to estimate small ball probabilities. Recall that by Proposition 4.4.3, the distance problem reduces to estimating Levy concentration function for the self-normalized quadratic forms:

$$\mathcal{L}\left\{\frac{|\langle A^{-1}X, X\rangle|}{\sqrt{1+\|A^{-1}X\|_2^2}}, \varepsilon\sqrt{p}\right\}. \tag{4.24}$$

The goal of this section is to prove the following estimate, for the non-sparse version, see Theorem 8.1 in [73].

Theorem 4.6.1. (Small ball probabilities for quadratic forms.) *Let A be an $n \times n$ random matrix satisfies Assumption 4.1.1 and $p \geq n^{-c_p}$. Let X be a random vector in $\mathbb{R}^n$ whose entries are identically distributed, and satisfy the same assumption as those of A. There exist constants $c_p, C_{4.6.1}, c_{4.6.1}, c'_{4.6.1}$ depend only on the parameters C_{op} and M_4, and such that the following holds. For every $\varepsilon \geq 0$ and $u \in \mathbb{R}$, one has*

$$\mathbb{P}\left\{\frac{|\langle A^{-1}X, X\rangle - u|}{\sqrt{1 + \|A^{-1}X\|_2^2}} \leq \varepsilon\sqrt{p} \wedge \mathcal{E}_{op}\right\} \leq C_{4.6.1}\varepsilon^{1/9} + 2\exp(-n^{c_{4.6.1}}) + \exp(-c'_{4.6.1}pn).$$

To prove Theorem 4.6.1, we will first decouple the enumerator $\langle A^{-1}X, X\rangle$ from the denominator $\sqrt{1 + \|A^{-1}X\|_2^2}$ by showing that $\|A^{-1}X\|_2 \sim \|A^{-1}\|_{\mathrm{HS}}$ with high probability. Then we adapt argument from [73] to decouple $\langle A^{-1}X, X\rangle$. Finally, by condition on X we obtain a linear form, and we can estimate its small ball probabilities using the Littlewood-Offord theory.

The following result is an analog of Proposition 8.2 in [73], it compares the size of the denominator $\sqrt{1 + \|A^{-1}X\|_2^2}$ to $\|A^{-1}\|_{\mathrm{HS}}$.

Proposition 4.6.2. *(Size of $A^{-1}X$) Let A be an $n \times n$ random matrix satisfies Assumption 4.1.1. Let X be a random vector in $\mathbb{R}^n$ whose entries are identically distributed, and satisfy the same assumption as those of A. There exist constants $c_{4.6.2}, C_{4.6.2}, c'_{4.6.2} > 0$ that depend only on the parameter C_{op} and M_4 from the assumption, and such that the following holds. Let $n^{-c_{4.6.2}} \leq \lambda \leq c_{4.6.2}$. The random matrix A has the following property with probability at least $1 - \exp(-c_{4.6.2}np)$. If C_{op} holds, then for every $\varepsilon > 0$, one has:*

(i) with probability of at least $1 - \exp(-c'_{4.6.2}pn)$ in X, we have

$$\|A^{-1}X\|_2 \geq C_{4.6.2}^{-1};$$

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

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

(iii) with probability at least $1 - C_{4.6.2}\varepsilon/\sqrt{\lambda} - n^{c'_{4.6.2}/\lambda}$ *in* X*, we have*

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

And the same result of (iii) still hold if we replace X *by an i.i.d. random vector with* $\mathcal{L}(X_i, \varepsilon_0) \leq 1 - c_0 p$*. In this case* $C_{4.6.2}, c'_{4.6.2}$ *would also depend on* p_0, ε_0*.*

The proof of this result uses the following elementary lemma, see Lemma 8.3 in [73].

Lemma 4.6.3. (Sums of dependent random variables.) *Let* $Z_1, \cdots, Z_n$ *be arbitrary non-negative random variables (not necessarily independent), and* $p_1, \cdots, p_n$ *be non-negative numbers such that*

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

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

$$\mathbb{P}\left\{\sum_{k=1}^{n} p_k Z_k \leq \varepsilon\right\} < 2\sum_{k=1}^{n} p_k \mathbb{P}\left\{Z_k \leq 2\varepsilon\right\}.$$

Proof of Proposition 4.6.2. Denote $e_1, \cdots, e_n$ the canonical basis of $\mathbb{R}^n$, and

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

Now, apply Structure Theorem 4.5.1 together with a union bound over $k = 1, \cdots, n$. More specifically, choose $L = L_0 = (\delta_0 p)^{-1/2}$ (the choice of δ_0 see remark 4.2.5). The random matrix with probability at least $1 - n2e^{-c'_{4.5.1}pn} \geq 1 - 2e^{-c'_{4.5.1}pn/2}$ has following property: if $\mathcal{E}_{op}$ holds then

$$x_k \in \mathrm{Incomp}(c_s n, c_d), \hat{D}_L(x_k, \lambda) \geq L^{-2} n^{c_{4.5.1}/\lambda}, k = 1, 2, \cdots, n.$$

From now on, let us fix a realization of A satisfies above property. Without loss of generality, we may also assume that $\mathcal{E}_{op}$ holds.

(i) First, we have

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

By the definition of event $\mathcal{E}_{op}$, we have $\|A\| \le C_{op}\sqrt{pn}$. Moreover, Chernoff's inquality together with the Tensorization Lemma 4.2.6 implies that the random vector X satisfies $\|X\|_2 \ge c\sqrt{pn}$ with probability at least $1 - \exp(-cpn)$. Here c is a constant only depending on M_4. Then we have $\|A^{-1}X\|_2 \ge \frac{c}{C_{op}}$ with the same probability. So we proved (i).

(ii) Using the fact that A is symmetric, we have

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

Recall that we also have $X_i = \delta_i \xi_i$, where δ_is are Bernoulli with parameter p and ξ_is are random variables with mean 0 variance 1. Therefore,

$$\mathbb{E}_X \langle x_k, X\rangle^2 = \mathbb{E}_X \sum_{i=1}^{n} x_{k,i}^2 X_i^2 = p$$

So,

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

Part (ii) follows directly from an application of Markov's inequality.

(iii) Now, we fix $k \le n$. Then $\langle x_k, X\rangle$ is a sum of independent random variables: $\sum_{i=1}^{n} x_{k,i}X_i$. We can estimate this sum using Proposition 4.4.15 combined with the estimated on the regularized LCD of x_k. Therefore

$$\mathcal{L}\left(\langle x_k, X\rangle, \sqrt{2p}\varepsilon\right) \le C_{4.4.10}\left(\frac{\varepsilon}{\sqrt{\lambda}} + p^{-1/2}L^2 n^{-c_{4.5.1}/\lambda}\right). \tag{4.25}$$

Now, together with estimates for all k using (4.25), Lemma 4.6.3 with $p_k = \|A^{-1}e_k\|_2^2/\|a^{-1}\|_{\mathrm{HS}}^2$ and that $\sum p_k = 1$. We have

$$
\begin{aligned}
\mathbb{P}_X\left\{\|A^{-1}X\|_2 \le \varepsilon\sqrt{p}\|A^{-1}\|_{\mathrm{HS}}\right\} &= \mathbb{P}\left\{\sum_{k=1}^n p_k\langle x_k, X\rangle^2 \le p\varepsilon^2\right\} \\
&\le 2\sum_{k=1}^n p_k\mathbb{P}\left\{\langle x_k, X\rangle^2 \le 2p\varepsilon^2\right\} \\
&\le 2C\left(\frac{\varepsilon}{\sqrt{\lambda}} + p^{-3/2}n^{-c_{4.5.1}/\lambda}\right)
\end{aligned}
\tag{4.26}
$$

We complete the proof using the range of λ and p. To prove the same result hold for X replaced by an i.i.d. random vector with $\mathcal{L}(X_i, \varepsilon_0) \le 1 - c_0 p$. We only need to notice that to derive (4.25) from Theorem 4.4.14, above condition is sufficient. $\square$

Decoupling the quadratic form is based on the following Lemma, see Lemma 8.4 in [73].

Lemma 4.6.4. (Decoupling quadratic forms). *Let G be an arbitrary symmetric $n\times n$ matrix, and let X be a random vector in $\mathbb{R}^n$ with independent coordinates. Let X' denote an independent copy of X. Consider a subset $J \subset [n]$. Then for every $\varepsilon \ge 0$, one has*

$$
\begin{aligned}
\mathcal{L}(\langle GX, X\rangle, \varepsilon)^2 &= \sup_{u\in\mathbb{R}}\mathbb{P}\left\{|\langle GX, X\rangle - u| \le \varepsilon\right\}^2 \\
&\le \mathbb{P}_{X,X'}\left\{|\langle G(P_{J^c}(X - X')), P_J X\rangle - v| \le \varepsilon\right\}
\end{aligned}
\tag{4.27}
$$

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

Now, we are ready to prove Theorem 4.6.1. The argument is based on the decoupling lemma and Littlewood-Offord theory which stated earlier. The proof is a modification of Section 8.3 in [73]. Although the proof structure is the same as in [73], we still need to go into details to catch the effect of sparsity.

Step 1: Constructing a random subset J and assignment spread(x). We start by decomposing $[n]$ into two random sets J and J^c. To this end, we consider independent $0, 1-$-valued random variables $\gamma_1, \cdots, \gamma_n$. with $\mathbb{E}\gamma_i = c_{oo}/2$. We also define

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

Then $\mathbb{E}|J^c| = c_{oo}n/2$. By a in large deviation inequality ([2] Theorem A.1.4), the inequality

$$|J^c| \leq c_{oo}n \tag{4.28}$$

holds with high probability:

$$\mathbb{P}\{(4.28) \text{ holds}\} \geq 1 - 2\exp(-c'_{oo}pn)$$

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

Fix a realization of J that satisfies (4.28). By Lemma 4.4.1, at least $2c_{oo}n$ coordinates of a vector $x \in \mathrm{Incomp}(c_s n, c_d)$ satisfy the regularity condition. So for each vector $x \in \mathrm{Incomp}(c_s n, c_d)$ we can assign a subset

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

so that the regularity condition holds for all $k \in$spread(x). If there is more than one way to assign spread(x) to x, we only need to choose one fixed way. This results in an assignment that depends only on the choice of the random set J. We will use this specific assignment J in applications of Definition 4.4.11 for regularized LCD.

Step 2. Estimating the denominator $\sqrt{1 + \|A^{-1}\|_2^2}$ and LCD of the inverse. By Lemma 4.6.2, we may replace the denominator $\sqrt{1 + \|A^{-1}X\|_2^2}$ by $\|A^{-1}\|_{\mathrm{HS}}$ in (4.24). Let $\varepsilon_0 \in (0, 1)$ and let X' denote an independent copy of the random vector X. Then we consider following event which is determined by the random matrix A, random vectors X, X' and the random set J:

$$\sqrt{\varepsilon_0 p^{-1}}\sqrt{1+\|A^{-1}X\|_2^2} \leq \|A^{-1}\|_{\mathrm{HS}} \leq \frac{1}{\varepsilon_0\sqrt{p}}\|A^{-1}(P_{J^c}(X-X'))\|_2 \tag{4.29}$$

Denote $Y := P_{J^c}(X-X')$, then we have Y_is are i.i.d. random variables and

$$\mathcal{L}(Y_i, c_0) \leq \mathcal{L}(P_{J^c}X, c_0) \leq 1 - c_1 p.$$

where c_0, c_1 depends only on M_4, C_{op}. Here we simply used the fact that $P_{J^c}X$ is a sparse random variable with sparsity level $c_{oo}p/2$ and Remark 4.2.5. So we can apply Proposition 4.6.2 with $A^{-1}X$ and $A^{-1}Y$. We have

$$\mathbb{P}_{A,X,X',J}\{(4.29) \text{ and holds } \wedge \mathcal{E}_{op}^c\} \geq 1 - \frac{C_{4.6.2}\varepsilon_0}{\sqrt{\lambda}} - n^{-c'_{4.6.2}/\lambda} - 2e^{-c'_{4.6.2}pn}.$$

where $c'_{4.6.2}, C_{4.6.2}$ depend only on C_{op} and M_4.

Denote the random vector

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

and condition on an arbitrary realization of random vectors X, X' and on realization of J which satisfies (4.28). Fix a value of parameter λ that satisfying $n^{-c_{4.5.1}} \leq \lambda \leq \frac{c_{4.5.1}}{4}$ as needed in Theorem 4.5.1. Then consider the event

$$x_0 \in \mathrm{Incomp}(c_s n, c_d) \text{ and } \hat{D}_{L_0}(x_0, \lambda) \geq \delta_0 p n^{c_{4.5.1}/\lambda} \tag{4.30}$$

which depends on the random matrix A. By Theorem 4.5.1, we have

$$\mathbb{P}_A\left\{(4.30) \text{ holds } \vee \mathcal{E}_{op}^c | X, X', J \text{satisfies (4.28)}\right\} \geq 1 - 2e^{-c'_{4.5.1}pn}.$$

Therefore

$$\begin{aligned} & \mathbb{P}_{A,X,X',J}\left\{(4.28, 4.29, 4.30) \text{ hold } \vee \mathcal{E}_{op}^c\right\} \\ \geq\ & 1 - 2e^{-c'_{oo}n} - \frac{C_{4.6.2}\varepsilon_0}{\sqrt{\lambda}} - n^{-c'_{4.6.2}/\lambda} - 2e^{-c'_{4.6.2}pn} - 2e^{-c'_{4.5.1}pn} \qquad (4.31) \\ =:\ & 1 - p_0 \end{aligned}$$

Thus there exists a realization of J that satisfies (4.28) and

$$\mathbb{P}_{A,X,X'}\{(4.29, 4.30) \text{ hold} \vee \mathcal{E}_{op}^c\} \geq 1 - p_0.$$

Now, fix such a realization of J in the rest of the proof. Applying Fubini's theorem and we have A has the following property with probability at least $1 - \sqrt{p_0}$:

$$\mathbb{P}_{X,X'}\{(4.29, 4.30) \text{ hold } \vee \mathcal{E}_{op}^c | A\} \geq 1 - \sqrt{p_0}$$

Since event $\mathcal{E}_{op}^c$ depends on A only, the random matrix A has the following property with probability at least $1 - \sqrt{p_0}$. Either $\mathcal{E}_{op}^c$ holds, or:

$$\mathcal{E}_{op} \text{ holds and } \mathbb{P}_{X,X'}\{(4.29), (4.30) \text{ hold}|A\} \geq 1 - \sqrt{p_0} \tag{4.32}$$

Step 3: Decoupling. Recall the event we want to estimate probability is

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

So we only need to estimate

$$\mathbb{P}_{A,X}(\mathcal{E} \cap \mathcal{E}_{op}) \leq \mathbb{P}_{A,X}\{\mathcal{E} \wedge \text{ (4.32) holds}\} + \mathbb{P}_{A,X}\{\mathcal{E}_{op} \wedge \text{(4.32) fails}\}$$

The second term is bounded by $\sqrt{p_0}$. Therefore,

$$\mathbb{P}_{A,X}(\mathcal{E} \cap \mathcal{E}_{op}) \leq \sup_{A \text{ satisfies (4.32)}} \mathbb{P}_X(\mathcal{E}|A) + \sqrt{p_0}$$

Moreover, using property (4.32) in a larger probability space, we have

$$\mathbb{P}_{A,X}(\mathcal{E} \cap \mathcal{E}_{op}) \leq \sup_{A \text{ satisfies (4.32)}} \mathbb{P}_{X,X'}\{\mathcal{E} \wedge \text{ (4.32) holds}|A\} + 2\sqrt{p_0}$$

Now, we fix a realization of a random matrix A satisfying (4.32). We only need to bound the probability

$$p_1 := \mathbb{P}_{X,X'}\{\mathcal{E} \wedge \text{(4.29) holds}\}$$

By definition of $\mathcal{E}$ and property (4.29),

$$p_1 \le \mathbb{P}_{X,X'}\left\{|\langle A^{-1}X, X\rangle - u| \le \frac{p\varepsilon}{\sqrt{\varepsilon_0}}\|A^{-1}\|_{\mathrm{HS}}\right\}$$

Now we may apply decoupling Lemma 4.6.4, and therefore

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

where

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

Here v is a number that depends on A^{-1}, $P_{J^c}X, P_{J^c}X'$ only. Use property (4.32) and we have

$$p_1^2 \le \mathbb{P}_{X,X'}\{\mathcal{E}_0\} \le \mathbb{P}_{X,X'}\left\{\mathcal{E}_0 \wedge (4.29, 4.30) \text{ hold}\right\} + \sqrt{p_0}$$

Now, we may divide both sides in the inequality defining the event $\mathcal{E}_0$ by $\|A^{-1}(P_{J^c}(X - X'))\|_2$. By definition of x_0 and (4.29) and we get

$$p_1^2 \le \mathbb{P}_{x,X'}\left\{|\langle x_0, P_J X\rangle - w| \le \sqrt{p}\varepsilon_0^{-3/2}\varepsilon \wedge (4.30) \text{ holds}\right\} + \sqrt{p_0} \tag{4.33}$$

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

Step 4: The small ball probabilities of a linear form. Finally, the random vector x_0 depends only on $P_{J^c}(X - X')$, which is independent of the random vector $P_J X$. So we may fix an arbitrary realization of the random vectors $P_{J^c}X$ and $P_{J_c}X'$, this will fix vector x_0 and number w in (4.33). By (4.30) we have

$$p_1^2 \le \sup_{x_0 \text{ satisfies (4.30), w}\in\mathbb{R}} \mathbb{P}_{P_J X}\left\{|\langle x_0, P_J X\rangle - w| \le \sqrt{p}\varepsilon_0^{-3/2}\varepsilon\right\} + \sqrt{p_0}$$

So from now on, let us fix a vector $x_0 \in S^{n-1}$ such that (4.30) holds and a number $w \in \mathbb{R}$. This reduce the problem to estimating the small ball probability for the weighted sum of independent random variables

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

We now apply Proposition 4.4.15, noticing that we have $J \supseteq \text{spread}(x_0) \supseteq I(x)$ as needed in the theorem. Therefore

$$\mathbb{P}_{P_J X}\left\{|\langle x_0, P_J X\rangle - w| \le \sqrt{p}\varepsilon_0^{-3/2}\varepsilon\right\} \le \frac{C_{4.4.15}\varepsilon_0^{-3/2}\varepsilon}{\sqrt{\lambda}} + \frac{C_{4.4.15}}{\sqrt{p}\hat{D}_{L_0}(x_0,\lambda)}.$$

Using property (4.30) to bound the regularized LCD, we have

$$p_1^2 \le \frac{C_{4.4.15}\varepsilon_0^{-3/2}\varepsilon}{\sqrt{\lambda}} + C_{4.4.15}\delta_0^{-1}p^{-3/2}n^{-c_{4.5.1}/\lambda} + \sqrt{p_0}.$$

Now we set $\varepsilon_0 = \varepsilon^{1/2}/\lambda^{1/8}$ and estimate $\mathbb{P}_{A,X}(\mathcal{E}\cap\mathcal{E}_{op})$ as

(4.34)

$$\begin{aligned}
\mathbb{P}_{A,X}(\mathcal{E}\cap\mathcal{E}_{op}) \le\ & p_1 + 2\sqrt{p_0} \\
\le\ & \left(\frac{C_{4.4.15}\varepsilon_0^{-3/2}\varepsilon}{\sqrt{\lambda}}\right)^{1/2} + \left(C_{4.4.15}\delta_0^{-1}p^{-3/2}n^{-c_{4.5.1}/\lambda}\right)^{1/2} \\
& +3\left(2e^{-c'_{oo}n} + \frac{C_{4.6.2}\varepsilon_0}{\sqrt{\lambda}} + n^{-c'_{4.6.2}/\lambda} + 2e^{-c'_{4.6.2}pn} + 2e^{-c'_{4.5.1}pn}\right)^{1/4} \\
\le\ & C\left(e^{-cpn} + n^{-c/\lambda} + \frac{\varepsilon_0^{1/4}}{\lambda^{1/8}} + \frac{\varepsilon_0^{-3/4}\varepsilon^{1/2}}{\lambda^{1/4}}\right)^{1/2} \\
\le\ & n^{-c'/\lambda} + C'\frac{\varepsilon^{1/8}}{\lambda^{5/32}} + e^{-c'pn}
\end{aligned}$$

Optimizing above probability using $n^{-c_{4.5.1}} \le \lambda \le \frac{c_{4.5.1}}{4}$ (see page 49 and Fact 8.6 in [73]), we have

$$\mathbb{P}_{A,X}(\mathcal{E}\cap\mathcal{E}_{op}) \le C''\varepsilon^{1/9} + \exp(-n^{c''}) + \exp(-c'np)$$

where c', c'', C'' depend only on M_4, C_{op}.

4.7 Proof of Theorem 4.1.5

Now we can combine the incompressible and compressible part to prove Theorem 4.1.5.

Proof of Theorem 4.1.5. We consider

$$\begin{aligned}
&\mathbb{P}\left\{\min_{x\in S^{n-1}}\|Ax\|_2\le\varepsilon\sqrt{\frac{p}{n}}\wedge\mathcal{E}_{op}\right\}\\
\le\ &\mathbb{P}\left\{\inf_{x\in\mathrm{Comp}(c_sn,c_d)}\|Ax\|_2\le\varepsilon\sqrt{\frac{p}{n}}\wedge\mathcal{E}_{op}\right\}\\
&+\mathbb{P}\left\{\inf_{x\in\mathrm{Incomp}(c_sn,c_d)}\|Ax\|_2\le\varepsilon\sqrt{\frac{p}{n}}\wedge\mathcal{E}_{op}\right\}
\end{aligned}\tag{4.35}$$

The first term is bounded by $2\exp(-c_{4.3.10}pn)$ as in (4.19). The probability for the incompressible vectors is estimated via distances in Lemma 4.4.2. Finally, we only need to apply Theorem 4.6.1 and Proposition 4.4.3, and notice that $e^{-n^{c_{4.6.1}}}$ dominate the term e^{-cpn} for $p\ge n^{-c_p}$. $\square$

4.8 Estimate of the Spectral Norm

In this section, we prove Theorem 4.1.7, that is to say when ξ_{ij} is sub-gaussian, $\|A\|\le C\sqrt{np}$ w.h.p.. The proof use the same moment technique and structure as the proof of Theorem 1.7 in [6].

Proof of Theorem 4.1.7. First, let's consider ξ'_{ij}, $i,j\in[n]$ to be independent copies of ξ_{ij}, $i,j\in[n]$ and $\eta_{ij}:=\xi_{ij}-\xi'_{ij}$. Let A'_n and B_n be the matrices with entries $a'_{ij}=\delta_{ij}\xi'_{ij}$ and $b_{ij}=\delta_{ij}\eta_{ij}$. Denote $\mathbb{E}_\xi$ as the expectation with respect to ξ, conditioned on $\delta:=(\delta_{ij})_{i,j\in[n]}$. Consider $q\ge1$ to be an even integer. By Jensen's inequality, as operator norm is convex function of matrix entries, we have

$$\mathbb{E}_\xi\|A_n\|^q=\mathbb{E}_\xi\|A_n-\mathbb{E}_{\xi'}A'_n\|^q\le\mathbb{E}_\eta\|B_n\|^q.$$

Then, let g_{ij}, $i,j\in[n]$ be independent $N(0,1)$ random variables. Clearly, $\xi_{ij}-\xi'_{ij}$ is a sub-gaussian random variable, by moment condition of sub-gaussian random variable there exists a constant C_1, depending on the sub-guassian norm of ξ_{ij}, such that $\mathbb{E}|\eta_{ij}|^q\le\mathbb{E}|C_1g_{ij}|^q$ for all $q\ge1$. Let W_n be the $n\times n$ random matrix with entries $w_{ij} = \delta_{ij} g_{ij}$. Since

$$\mathbb{E}_\eta \|B_n\|^q \le \mathbb{E}_\eta \mathrm{Tr}\left((B_n B_n^*)^{q/2}\right)$$

where right hand side is a polynomial of the even moments of η_{ij} with non-negative coefficients, we have

$$\mathbb{E}_\eta \mathrm{Tr}\left((B_n B_n^*)^{q/2}\right) \le C_1^q n \mathbb{E}_g \|W_n\|^q.$$

Above inequality uses the elementary identity that $\mathrm{Tr}\left((W_n W_n^*)^{q/2}\right) = \sum_{j=1}^n \lambda_j^{q/2}(W_n W_n^*)$. Here eigenvalues $\lambda_j(W_n W_n^*)$ satisfy $|\lambda_j(W_n W_n^*)| \le \|W_n\|^2$ for all j.

Now we are ready to estimate $\mathbb{E}\|W\|^2$. Here we need to apply the following result due to Bandeira and van Handel [1].

Lemma 4.8.1. *Let X be the $n \times n$ symmetric matrix with $X_{ij} = g_{ij} b_{ij}$, where $\{g_{ij} : i \ge j\}$ are i.i.d.$\sim N(0,1)$ and $\{b_{ij} : i \ge j\}$ are given scalars. Let*

$$\sigma := \max_i \sqrt{\sum_j b_{ij}^2}, \quad \sigma_* := \max_{ij} |b_{ij}|$$

Then

$$\mathbb{E}\|X\| \le (1+\varepsilon)\left\{2\sigma + \frac{6}{\sqrt{\log(1+\varepsilon)}}\sigma_* \sqrt{\log n}\right\}$$

for any $\varepsilon \in (0, 1/2)$.

Let Ω be the event for all $i \in [n]$, $\sum_{j=1}^n \delta_{ij} \le \bar{C} pn$, for some $\bar{C} \ge 2$. Since $p \ge C_0 \frac{\log n}{n}$, applying Chernoff's inequality and union bound argument, we can choose the C_0 large enough, such that $\mathbb{P}(\Omega^c) \le e^{-cpn}$ for some $c > 0$. And c depends only on C_0. Now, we can use the above Lemma 4.8.1 and assume that $\delta \in \Omega$. Conditionally on δ, we have

$$\mathbb{E}\left(\|W_n\| | \delta\right) \le \sqrt{\bar{C} pn} + C^* \sqrt{\log n} \le \sqrt{C' pn}.$$

Here C^* is some absolute constnat, and $C' = 2(C^*)^2 \bar{C}$. Conditioning on δ, $\|W_n\|$ can be viewed as a $\sqrt{2}$-Lipschitz function on $\mathbb{R}^{n(n+1)/2}$ with the standard Gaussian measure. Applying standard Gaussian concentration inequality [31], we have

$$\mathbb{P}\left(\|W_n\| \geq \mathbb{E}[\|W_n\||\delta] + t\right) \leq \tilde{C}\exp(-c't^2)$$

for some absolute constants $\tilde{C}, c' > 0$, and any $t > 0$. Therefore,

$$\begin{aligned} \mathbb{E}_g\|W_n\|^q &\leq (C'pn)^{q/2} + \int_{\sqrt{C'pn}}^{\infty} qs^{q-1}\mathbb{P}\left[\|W_n\| \geq s|\delta]\right] ds \\ &\leq (C'pn)^{q/2} + (C''q)^{q/2}, \end{aligned} \tag{4.36}$$

for some absolute constant C''. Now choose $q = pn$. This inequality in combination with previous inequalities yields

$$\mathbb{E}_\xi\|A_n\|^{pn} \leq n(C_2pn)^{pn/2} \leq (C_2^2pn)^{pn/2}.$$

where C_2 is a positive constant depending on C_0 and the sub-gaussian norm of ξ_{ij}. Here we used the condition $p \geq C_0\frac{\log n}{n}$ to absorb the factor n. Finally, choosing $C_{op} > C_2^2$, we have for any $\delta \in \Omega$, there exists a small positive constant c_{op}, depending on C_{op}, such that

$$\mathbb{P}\left(\|A_n\| \geq C_{op}\sqrt{pn}|\delta\right) \leq \exp(-c_{op}pn)$$

by applying Markov inequality. Now picking c_{op} small enough, we have

$$\mathbb{P}\left(\|A_n\| \geq C_{op}\sqrt{pn}\right) \leq \max_{\delta\in\Omega}\mathbb{P}\left(\|A_n\| \geq C_{op}\sqrt{pn}|\delta\right) + \mathbb{P}(\Omega^c) \leq \exp(-c_{op}pn).$$

$\square$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] R. Handel A. Bandeira. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Annals of Probability*, 44(2479-2506), 2016.

[2] N. Alon and J. Spencer. *The probability method*. Wiley-Interscience, New York, 2000.

[3] Z. Bai and J. Silverstein. *Spectral analysis of large dimensional random matrices. Second edition.* New York: Springer, 2010.

[4] Z. D. Bai. Circular law. *Ann. Probab.*, 25(494-529), 1997.

[5] Z. D. Bai. Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, 9(611-677), 1999.

[6] A. Basak and M. Rudelson. Invertibility of sparse non-hermitian matrices. *arXiv:1507.03525*.

[7] H. Widom C. Tracy. Level-spacing distributions and the airy kernel. *Comm. Math. Phys.*, 159(151-174), 1994.

[8] H. Widom C. Tracy. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, no. 3(727-754), 1996.

[9] H. Widom C. Tracy. Distribution functions for largest eigenvalues and their applications. *Proceedings of the International Congress of Mathematicians*, Vol. I (Beijing, 2002)(587-596), Beijing, 2002.

[10] C. Cacciapuoti, A. Maltsev, and B. Schlein. Local Marchenko-Pastur law at the hard edge of sample covariance matrices. *Journal of Mathematical Physics*, 54(43302), 2012.

[11] P. Deift and D. Gioev. *Random matrix theory: invariant ensembles and universality.* American Mathematical Society, 2009.

[12] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(543-560), 1988.

[13] P. Di Francesco G. Akemann, J. Baik. *Handbook of Random Matrix Theory.* Oxford University Press, 2011.

[14] O. Zeitouni G. W. Anderson, A. Guionnet. *An introduction to random matrices.* Cambridge University Press, 2010.

[15] Y. Gordon. On dvoretzkys theorem and extensions of slepians lemma. *Israel seminar on geometrical aspects of functional analysis (1983/84)*, 1984.

[16] Y. Gordon. Some inequalities for gaussian processes and applications. *Israel J. Math.*, 50(265-289), 1985.

[17] Y. Gordon. Majorization of gaussian processes and geometric applications. *Probab. Theory Related Fields*, 91(251-267), 1992.

[18] F. Gotze and A. Tikhomirov. On the circular law. *arXiv:math/0702386.*

[19] F. Gotze and A. Tikhomirov. The circular law for random matrices. *Ann. Prob.*, 38(1444-1491), 2010.

[20] V. Vu H. Nguyen, T. Tao. Random matrices: tail bounds for gaps between eigenvalues. *Probability Theory and Related Fields*, 167(777-816), 2017.

[21] R. Handel. On the spectral norm of gaussian random matrices. *Trans. Amer. Math. Soc.*, 369(8161-8178), 2017.

[22] H. Huang and F. Wei. Upper bound for the dvoretzky dimension in milman-schechtman theorem. *Geometric aspects of functional analysis, 2014-2016*, (181-186), 2017.

[23] T. Suidan J. Baik, P. Deift. *Combinatorics and Random Matrix Theory.* American Mathematical Society, 2016.

[24] V. Vu J. Bourgain, P. Wood. On the singularity probability of random discrete matrices. *Journal of Functional Analysis*, 258(559-603), 2010.

[25] E. Szemeredi J. Kahn, J. Komlos. On the probability that a random ± 1-matrix is singular. *J. Amer. Math. Soc.*, 8(223-240), 1995.

[26] K. Johansson. Shape fluctuations and random matrices. *Comm. Math. Phys.*, 209(437-476), 2000.

[27] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(295-327), 2001.

[28] J. Komlos. On the determinant of $(0, 1)$ matrices. *Studia Sci. Math. Hungar.*, 2(7-22), 1967.

[29] J. Komlos. On the determinant of random matrices. *Studia Sci. Math. Hungar.*, 3(387-399), 1968.

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

[31] M. Ledoux. *The concentration of measure phenomenon.* American Mathematical Soc., 2005.

[32] G. Livshyts, G. Paouris, and P. Pivovarov. On sharp bounds for marginal densities of product measures. *arXiv:1507.07949.*

[33] M. L. Mehta. *Random matrices. Pure and Applied Mathematics (Amsterdam).* Amsterdam: Elsevier/Academic Press, 2004.

[34] V. D. Milman. New proof of the theorem of a. dvoretzky on sections of convex bodies. *Funct. Anal. Appl.*, 5(4)(288-295), 1971.

[35] V. D. Milman and G. Schechtman. Global versus local asymptotic theories of finite dimensional normed spaces. *Duke Math. Journal*, 90(73-93), 1997.

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

[37] Hoi H. Nguyen and Van H. Vu. Normal vector of a random hyperplane. *arXiv:1604.04897.*

[38] S. Sodin O. Feldheim. A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geometric and Functional Analysis*, 20(88-123), 2010.

[39] S. Sodin O. Friedland. Bounds on the concentration function in terms of the diophantine approximation. *C. R. Math. Acad. Sci. Paris*, 345(513-518), 2007.

[40] Z. Fredi P. Frankl. Solution of the littlewood-offord problem in high dimensions. *Ann. of Math.*, 128(259-270), 1988.

[41] L. A Pastur. On the spectrum of random matrices. *Teoret. Mat. Fiz.*, 10(102-112), 1973.

[42] A. Pajor N. Tomczak-Jaegermann R. Adamczak, A. Litvak. Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc.*, 23(535-561), 2010.

[43] E. Rebrova and K. Tikhomirov. Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries. *arXiv:1508.06690.*

[44] M. Rudelson. Lecture notes on non-asymptotic random matrix theory. *notes from the AMS Short Course on Random Matrices*, 2013.

[45] M. Rudelson and R. Vershynin. No-gaps delocalization for general random matrices. arXiv:1506.04012.

[46] M. Rudelson and R. Vershynin. The least singular value of a random square matrix is $o(n^{-1/2})$. *Comptes rendus de l'Académie des sciences - Mathématique*, 346(893-896), 2008.

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

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

[49] M. Rudelson and R. Vershynin. Non-asymptotic theory of random matrices: extreme singular values. *Proceedings of the International Congress of Mathematicians*, Volume III(1576-1602), 2010.

[50] M. Rudelson and R. Vershynin. Delocalization of eigenvectors of random matrices with independent entries. *Duke Math. Journal*, 164(2507-2538), 2013.

[51] M. Rudelson and R. Vershynin. Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18(1-9), 2013.

[52] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high dimensional distributions. *International Mathematics Research Notices*, 2015(9594-9617), 2014.

[53] A. A. Giannopoulos S. Artstein-Avidan and V. D. Milman. *Asymptotic geometric analysis. Part I.* 2015.

[54] G. Schechtman. A remark concerning the dependence on ϵ in dvoretzky's theorem. *Geometric aspects of functional analysis (1987-88), Lecture Notes in Math.*, (274-277), 1989.

[55] S. Smale. On the efficiency of algorithms of analysis. *Bull. Amer. Math. Soc.*, 13(87-121), 1985.

[56] A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(1033-1056), 2002.

[57] E. M. Stein and G. Weiss. *Introduction to Fourier analysis on Euclidean spaces. Princeton Mathematical Series, No. 32.* 1971.

[58] S. Szarek. Condition numbers of random matrices. *J. Complexity*, 7(131-149), 1991.

[59] S. J. Szarek. Spaces with large distance to l_∞^n and random matrices. *American Journal of Mathematics*, 112(899-942), 1990.

[60] V. Vu T. Tao. On random ± 1 matrices: singularity and determinant. *Random Structures and Algorithms*, 28(1-23), 2006.

[61] V. Vu T. Tao. On the singularity probability of random bernoulli matrices. *J. Amer. Math. Soc.*, 20(603-628), 2007.

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

[63] V. Vu T. Tao. From the littlewood-offord problem to the circular law: universality of the spectral distribution of random matrices. *Bull. Amer. Math. Soc.*, 46(377-396), 2009.

[64] V. Vu T. Tao. Random matrices: Universality of esds and the circular law. *Ann. Prob.*, 38(2023-2065), 2010.

[65] V. Vu T. Tao. A sharp inverse littlewood-offord theorem. *Random Structures and Algorithms*, 37(525-539), 2010.

[66] T. Tao. *Topics in random matrix theory.* American Mathematical Society, 2012.

[67] T. Tao and V. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, 169(595-632), 2009.

[68] T. Tao and V. Vu. Random matrices: The distribution of the smallest singular values. *Geometric and Functional Analysis*, 20(260-297), 2010.

[69] L. A. Pastur V. A. Marchenko. The distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.*, 72(507-536), 1967.

[70] G. Schechtman V. D. Milman. *Asymptotic theory of finite-dimensional normed spaces.* 1986.

[71] R. Vershynin. How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, 25(655-686), 2012.

[72] R. Vershynin. *Introduction to the non-asymptotic analysis of random matrices, Compressed sensing, 210-268.* Cambridge University Press, 2012.

[73] R. Vershynin. Invertibility of symmetric random matrices. *Random Structures and Algorithms*, 44, 2014.

[74] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge University Press, to appear.

[75] K. W. Wachter. Strong limits of random matrix spectra for sample covariance matrices of independent elements. *Annals of Probability*, 6(1-18), 1978.

[76] F. Wei. Upper bound for intermediate singular values of random matrices. *Journal of Mathematical Analysis and Applications*, (1530-1547), 2017.

[77] P. Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(325-327), 1958.

[78] J. Silverstein Z. D. Bai. *Spectral analysis of large dimensional random matrices, 2nd ed., Springer Series in Statistics.* 2010.